Latent Space Modeling for Human Disease Network with Temporal Variations: Analysis of Medicare Data

Guojun Zhu

School of Mathematical Sciences,
University of Chinese Academy of Sciences

April 21, 2025

Outline

Outline

Background: Human Disease Network

In the HDN analysis, the concept of network analysis is adopted. One node corresponds to a disease, and two diseases are connected with a network edge if they are "interconnected".

- gene-centric HDN [Goh et al., 2007]
- phenotypic HDN [Zhou et al., 2014, 2022]
- clinical outcome HDN [Yang et al., 2022, Mei et al., 2025]
    - Two diseases are defined as interconnected if their clinical treatments and/or outcomes – such as inpatient length-of-stay (LOS), number of outpatient visits, and treatment costs – are "correlated".

Background: Latent Space Model

Our literature review suggests that one family of techniques, which has been widely adopted and shown as powerful in other contexts [Liu et al., 2024, Zhang et al., 2024a] but limitedly examined for HDNs, is latent space modeling [Hoff et al., 2002].

- Single layer network [Hoff et al., 2002, Ma et al., 2020]
- Multi-layer networks [Zhang et al., 2020]
- Time-varying networks [Sarkar and Moore, 2005, Liu et al., 2024]

Background: Latent Space Model

Our literature review suggests that one family of techniques, which has been widely adopted and shown as powerful in other contexts [Liu et al., 2024, Zhang et al., 2024a] but limitedly examined for HDNs, is latent space modeling [Hoff et al., 2002].

- Single layer network [Hoff et al., 2002, Ma et al., 2020]
- Multi-layer networks [Zhang et al., 2020]
- Time-varying networks [Sarkar and Moore, 2005, Liu et al., 2024]

Limitation: Difficult to analyze the following sensible temporal structure

- There are time intervals within which network structures remain constant – they correspond to small and slow changes in disease diagnosis, treatment.
- Between those intervals, network structures are assumed to change smoothly.

Outline

Data Exploration: Motivation

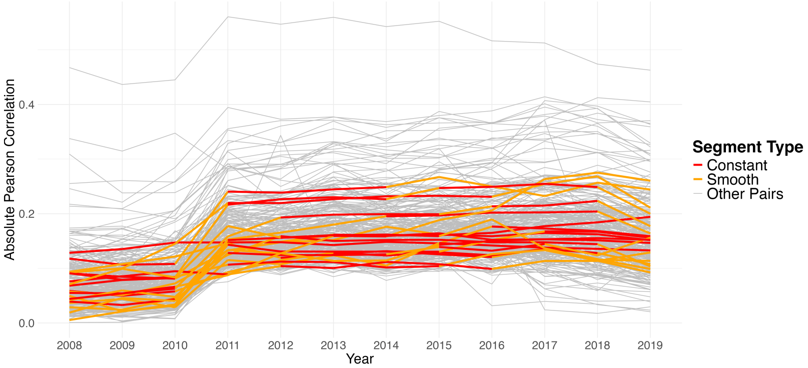


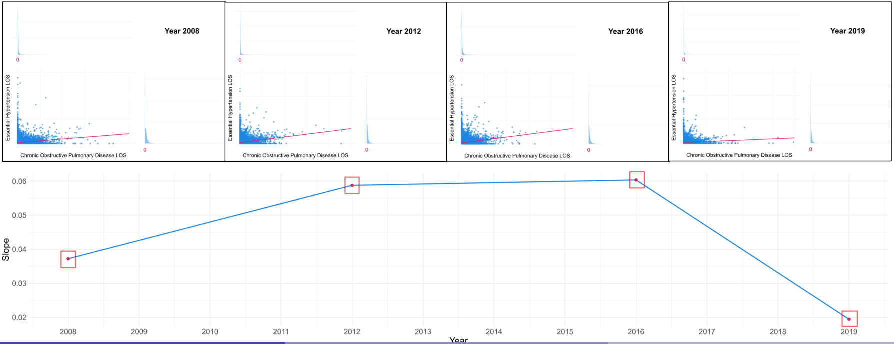
Figure: Pairwise interconnections: Pearson correlation

Outline

Data and Modeling Framework

- p: number of diseases
- T: number of time periods
- n: number of subjects
- $\{y_{ij}^{(t)}\}_{i\in[n],j\in[p],t\in[T]}$: clinical treatment measurement

As noted in the literature and can be seen from Figure 1, the marginal distributions of disease-specific LOS are highly zero-inflated.



Time-varying Latent Space Modeling

- We adopt a two-part modeling approach for the estimation of network adjacency matrices [Mei et al., 2025].
  - $\boldsymbol{A} = \left[\left(A_{jk}^{(1)} \right)_{j,k=1}^{p} ; \cdots ; \left(A_{jk}^{(T)} \right)_{j,k=1}^{p} \right] \in \{0,1\}^{T \times p \times p}$: a tensor for the time-varying adjacency matrices.
- We develop the latent space modeling based on the estimated $\boldsymbol{A}$.

Time-varying Latent Space Modeling

- We adopt a two-part modeling approach for the estimation of network adjacency matrices [Mei et al., 2025].
  - $\boldsymbol{A} = \left[\left(A_{jk}^{(1)}\right)_{j,k=1}^{p}; \cdots ; \left(A_{jk}^{(T)}\right)_{j,k=1}^{p}\right] \in \{0,1\}^{T\times p\times p}$: a tensor for the time-varying adjacency matrices.
- We develop the latent space modeling based on the estimated $\boldsymbol{A}$.
- Overall, the model is defined as:

$$A_{jk}^{(t)} \sim \text{Bernoulli}\left(P_{jk}^{(t)}\right), \text{logit}\left(P_{jk}^{(t)}\right) := \Theta_{jk}^{(t)} = \alpha_j^{(t)} + \alpha_k^{(t)} + \boldsymbol{z}_j^{\top} \boldsymbol{\Lambda}^{(t)} \boldsymbol{z}_k,$$

where $\text{logit}(x) = \log[x/(1-x)]$, $P_{jk}^{(t)}$ represents the connection probability between diseases j and k, and $\alpha_j^{(t)}$ is disease j's heterogeneity parameter for period t.

Identifiability

Proposition

Suppose that two sets of parameters $\left(\{\boldsymbol{\alpha}^{(t)}\}_{t=1}^{T}, \{\boldsymbol{\Lambda}^{(t)}\}_{t=1}^{T}, \boldsymbol{Z}\right)$ *and* $\left(\left\{\boldsymbol{\alpha}_{\dagger}^{(t)}\right\}_{t=1}^{T}, \left\{\boldsymbol{\Lambda}_{\dagger}^{(t)}\right\}_{t=1}^{T}, \boldsymbol{Z}_{\dagger}\right)$ *satisfy the following conditions:*

1. $\boldsymbol{J}_p\boldsymbol{Z} = \boldsymbol{Z}, \boldsymbol{J}_p\boldsymbol{Z}_{\dagger} = \boldsymbol{Z}_{\dagger}$, *where* $\boldsymbol{J}_p = \boldsymbol{I}_p - \frac{1}{p}\mathbf{1}_p\mathbf{1}_p^{\top}$;
2. $\boldsymbol{Z}^{\top}\boldsymbol{Z} = p\boldsymbol{I}_r$ *and* $\boldsymbol{Z}_{\dagger}^{\top}\boldsymbol{Z}_{\dagger} = p\boldsymbol{I}_r$;
3. *At least one of* $\boldsymbol{\Lambda}^{(t)}$ *'s,* $t = 1, 2, \cdots, T$, *is full rank.*

Then,

$$\boldsymbol{\alpha}^{(t)}\mathbf{1}_p^{\top} + \mathbf{1}_p\boldsymbol{\alpha}^{(t)\top} + \boldsymbol{Z}\boldsymbol{\Lambda}^{(t)}\boldsymbol{Z}^{\top} = \boldsymbol{\alpha}_{\dagger}^{(t)}\mathbf{1}_p^{\top} + \mathbf{1}_p\boldsymbol{\alpha}_{\dagger}^{(t)\top} + \boldsymbol{Z}_{\dagger}\boldsymbol{\Lambda}_{\dagger}^{(t)}\boldsymbol{Z}_{\dagger}^{\top},$$

for $t = 1, \ldots, T$, *which implies that there exists an orthonormal matrix* $\boldsymbol{O} \in \mathbb{R}^{r\times r}$ *where* $\boldsymbol{O}^{\top}\boldsymbol{O} = \boldsymbol{O}\boldsymbol{O}^{\top} = \boldsymbol{I}_r$, *such that*

$$\boldsymbol{\alpha}_{\dagger}^{(t)} = \boldsymbol{\alpha}^{(t)}, \boldsymbol{Z}_{\dagger} = \boldsymbol{Z}\boldsymbol{O}, \boldsymbol{\Lambda}_{\dagger}^{(t)} = \boldsymbol{O}^{\top}\boldsymbol{\Lambda}^{(t)}\boldsymbol{O}$$

for $t = 1, \cdots, T$.

Estimation

The negative log-likelihood is:

$$L_{\mathsf{NLL}}\left(\boldsymbol{Z},\boldsymbol{\alpha},\boldsymbol{\Lambda}\right)=-\sum_{t=1}^{T}\sum_{j,k=1}^{p}\left\{A_{jk}^{(t)}\Theta_{jk}^{(t)}+\log\left(1-\sigma\left(\Theta_{jk}^{(t)}\right)\right)\right\},$$

As discussed above, we consider the temporal structure with "piecewise constant + smoothly varying" properties. To achieve this, we propose a penalty built on the combination of ℓ_1 and ℓ_2 norms. Following Tibshirani [2014] and other literature, we refer it to as the **M**ixed **T**rend **F**ilter **(MTF)** penalty, which has the form:

$$\begin{aligned}Q\left(\boldsymbol{\alpha},\boldsymbol{\Lambda}\right)=&\lambda_1\left(\left\|\boldsymbol{D}^{(1)}\boldsymbol{\alpha}\right\|_{\ell_1}+\sum_{t=1}^{T-1}\left\|\left(\boldsymbol{D}^{(1)}\boldsymbol{\Lambda}\right)^{(t)}\right\|_F\right)\\&+\frac{\lambda_2}{2}\left(\left\|\boldsymbol{D}^{(3)}\boldsymbol{\alpha}\right\|_F^2+\left\|\boldsymbol{D}^{(3)}\boldsymbol{\Lambda}\right\|_F^2\right),\end{aligned}$$

where λ_1,λ_2 are data-dependent tuning parameters, and $D^{(k)}\in\mathbb{R}^{(T-k)\times T}$ is the discrete difference operator of order k.

Computation

Algorithm 1 Coordinate-Proximal-Projected Gradient Descent for MTF-LSM

Require: $\boldsymbol{A} \in \mathbb{R}^{T\times p\times p}$; initial estimates: $\boldsymbol{Z}_0, \boldsymbol{\alpha}_0, \boldsymbol{\Lambda}_0$; step sizes $\eta_z, \eta_\alpha, \eta_\lambda$; tunings:$\lambda_1, \lambda_2$;

while not convergent **do**

  for t in $1{:}T$ **do**

    $\boldsymbol{\Lambda}^{(t)} \leftarrow \boldsymbol{\Lambda}^{(t)} + \eta_\lambda \boldsymbol{Z}^\top \left(\boldsymbol{A}^{(t)} - \sigma\left(\boldsymbol{\Theta}^{(t)}\right)\right)\boldsymbol{Z} - \eta_\lambda \frac{\lambda_1}{2} \nabla_{\boldsymbol{\Lambda}^{(t)}} \left\|\boldsymbol{D}^{(3)}\boldsymbol{\Lambda}\right\|_F^2$

    $\boldsymbol{\Lambda}^{(t)} \leftarrow \mathrm{Prox}(\boldsymbol{\Lambda}^{(t)}) = \arg\min_{\boldsymbol{\beta}} \frac{1}{2}\|\boldsymbol{\beta} - \boldsymbol{\Lambda}^{(t)}\|_F^2 + \frac{\lambda_1}{2}(\|\boldsymbol{\beta} - \boldsymbol{\Lambda}^{(t-1)}\|_F + \|\boldsymbol{\beta} - \boldsymbol{\Lambda}^{(t+1)}\|_F)$

  end for

  for t in $1{:}T$ **do**

    $\boldsymbol{\alpha}^{(t)} \leftarrow \boldsymbol{\alpha}^{(t)} + 2\eta_\alpha \left(\boldsymbol{A}^{(t)} - \sigma\left(\boldsymbol{\Theta}^{(t)}\right)\right)\mathbf{1}_p - \eta_\alpha \frac{\lambda_2}{2} \nabla_{\boldsymbol{\alpha}^{(t)}} \left\|\boldsymbol{D}^{(3)}\boldsymbol{\alpha}\right\|_F^2$

    $\boldsymbol{\alpha}^{(t)} \leftarrow \mathrm{Prox}(\boldsymbol{\alpha}^{(t)}) = \arg\min_{\boldsymbol{\beta}} \frac{1}{2}\|\boldsymbol{\beta} - \boldsymbol{\alpha}^{(t)}\|_2^2 + \frac{\lambda_1}{2}(\|\boldsymbol{\beta} - \boldsymbol{\alpha}^{(t-1)}\|_1 + \|\boldsymbol{\beta} - \boldsymbol{\alpha}^{(t+1)}\|_1)$

  end for

  $\boldsymbol{Z} \leftarrow \boldsymbol{Z} + 2\eta_z \sum_{t=1}^{T} \left(\boldsymbol{A}^{(t)} - \sigma\left(\boldsymbol{\Theta}^{(t)}\right)\right)\boldsymbol{Z}\boldsymbol{\Lambda}^{(t)}$

  $\boldsymbol{Z} \leftarrow \boldsymbol{J}_p\boldsymbol{Z}, \boldsymbol{Z} \leftarrow \boldsymbol{Z}\boldsymbol{W}$ for $\boldsymbol{W} \in \mathbb{R}^{r\times r}$ s.t. $\boldsymbol{Z}^\top\boldsymbol{Z} = p\boldsymbol{I}_r, \boldsymbol{\Lambda}^{(t)} \leftarrow \boldsymbol{W}^{-1}\boldsymbol{\Lambda}^{(t)}\left(\boldsymbol{W}^{-1}\right)^\top$

end while

Outline

Conditions and Assumptions

Definition 1. For $p, T, r \in \mathbb{N}, \mu_1, \mu_2, \mu_3 \in \mathbb{R}_+$, the feasible parameter space is defined as:

$$\begin{aligned}\mathcal{F} =& \mathcal{F}_{p,T,r}\left(\mu_1, \mu_2, \mu_3\right) \\ =& \left\{ \mathcal{T} = \left[\boldsymbol{\Theta}^{(1)}; \boldsymbol{\Theta}^{(2)}; \cdots; \boldsymbol{\Theta}^{(T)}\right] \in \mathbb{R}^{T \times p \times p} : \right. \\ & \boldsymbol{\Theta}^{(t)} = \boldsymbol{\alpha}^{(t)} \mathbf{1}_p^\top + \mathbf{1}_p \boldsymbol{\alpha}^{(t)\top} + \boldsymbol{Z} \boldsymbol{\Lambda}^{(t)} \boldsymbol{Z}^\top; \\ & \boldsymbol{Z} \in \mathbb{R}^{p \times r}, \boldsymbol{Z}^\top \boldsymbol{Z} = p \boldsymbol{I}_r, \boldsymbol{J}_p \boldsymbol{Z} = \boldsymbol{Z}, \boldsymbol{\alpha}^{(t)} \in \mathbb{R}^p \\ & \boldsymbol{\Lambda}^{(t)} \in \mathbb{S}^{r \times r}, \left\| \boldsymbol{\Theta}^{(t)} \right\|_{\max} \leq \mu_1, t = 1, 2, \cdots, T \\ & \left. \|\boldsymbol{\alpha}\|_{\max} \leq \mu_2, \left\| \boldsymbol{\Lambda}^{(t)} \right\|_{\max} \leq \mu_3, t = 1, 2, \cdots, T \right\},\end{aligned}$$

where $\boldsymbol{J}_p = \boldsymbol{I}_p - \frac{1}{p} \mathbf{1}_p \mathbf{1}_p^\top, \mathbb{S}^{k \times k}$ includes all symmetric $k \times k$ matrices, and $\| \cdot \|_{\max}$ calculates the maximum absolute value of entries for a matrix. For the estimator $\widehat{\mathcal{T}} = \arg\min_{\mathcal{T} \in \mathcal{F}} L(\mathcal{T})$ and the true parameter $\mathcal{T}_\star \in \mathcal{F}$ associated with $\boldsymbol{\alpha}_\star$ and $\boldsymbol{Z}_\star$. Theorem 1 establishes the error bound.

Statistical Properties of $\widehat{\mathcal{T}}$

Theorem

Assume that $\sigma_{\min}\left(\boldsymbol{\Lambda}_{\star}^{(t)}\right) \geq \kappa, \quad t = 1, 2, \cdots, T$, *for some constant* $\kappa > 0$. *Further assume that* $\lambda_1 + 32\lambda_2\mu_3 r \leq \frac{\kappa^2 p}{32\mu_3 r M_{\mu_1}\sqrt{T}}$, *where* $b(x) = \log(1 + \exp(x))$, $M_{\mu_1} = \frac{2}{\min_{|v|<\mu_1} b''(v)}$. *Then, there exist constants* $c_1, c_2 > 0$, *such that with probability at least* $1 - T\exp(-c_1 p) - \exp(-c_2(2p + T))$,

$$\left\|\widehat{\mathcal{T}} - \mathcal{T}_{\star}\right\|_F^2 \leq C_1 pT + C_2(2p + T) + C_3 T$$

where positive constants C_1 *and* C_2 *depend solely on* (μ_1, μ_2, μ_3, r), *while positive constant* C_3 *additionally depends on* (λ_1, λ_2).

- The term $C_1 pT$ is induced by $\left\{\boldsymbol{\alpha}^{(t)}\right\}_{t=1}^{T}$.
- The second term $C_2(2p + T)$ is induced by $\left\{\boldsymbol{Z}\boldsymbol{\Lambda}^{(t)}\boldsymbol{Z}^{\top}\right\}_{t=1}^{T}$.
- The third term $C_3 T$ is induced by the mixed trend penalty.

Statistical Properties of $\widehat{\boldsymbol{\alpha}}$

Corollary

Assume that the conditions of Theorem 1 are satisfied and there exists a constant $\delta > 0$ *such that* $T \leq \delta p$*. Then, there exist constants* $c_1, c_2 > 0$*, such that with probability at least* $1 - T\exp(-c_1 p) - \exp(-c_2(2p+T))$*,*

$$\frac{1}{T}\left\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_\star\right\|_F^2 \leq \tilde{C}_1 + \tilde{C}_2 T^{-1}$$

where positive constant $\tilde{C}_1$ *depends solely on* (μ_1, μ_2, μ_3, r)*, while positive constant* $\tilde{C}_2$ *additionally depends on* (λ_1, λ_2)*.*

- Even in the worst-case scenario (without any constant-smooth trends), the penalty does not significantly increase the error.
- Under general conditions, the penalty may further reduce the optimization error, although this effect is not considered in the theoretical analysis.

Statistical Properties of $\widehat{\boldsymbol{Z}}$

Theorem

Assume that the conditions of Theorem 1 are satisfied and there exists a constant $\delta > 0$ *such that* $T \leq \delta p$*. Then, there exist constants* $c_1, c_2 > 0$*, such that with probability at least* $1 - T\exp(-c_1 p) - \exp(-c_2(2p+T))$*,*

$$\min_{\boldsymbol{O} \in \mathbb{S}^{r\times r}} \left\| \boldsymbol{Z}_\star \boldsymbol{O} - \hat{\boldsymbol{Z}} \right\|_F^2 \leq C_4 + C_5 T^{-1} + C_6 p^{-1}$$

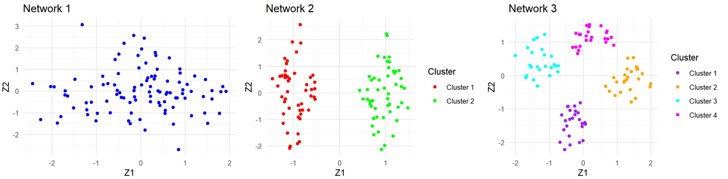
where positive constants C_4 *and* C_5 *depend solely on* (μ_1, μ_2, μ_3, r)*, while positive constant* C_6 *additionally depends on* (λ_1, λ_2)*.*

- Our approach maintains the same order of complexity as other relevant models [Ma et al., 2020, Zhang et al., 2020].

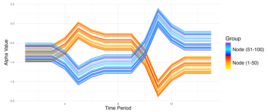
Outline

Data generation

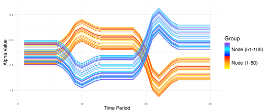
We conduct simulations with three network structures and various parameter settings: $p = 100, 200, T = 15, 30, 50$ and $r = 2, 4$.



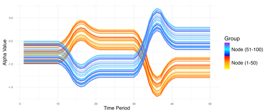
(a) Latent space $\boldsymbol{Z}$ under different network structures $(\boldsymbol{r = 2})$



(b) Temporal trends of $\boldsymbol{\alpha}$ $(\boldsymbol{T = 15})$



(c) Temporal trends of $\boldsymbol{\alpha}$ $(\boldsymbol{T = 30})$



(d) Temporal trends of $\boldsymbol{\alpha}$ $(\boldsymbol{T = 50})$

Figure: Simulated Data Generation

Alternative methods

- **LSM** [Ma et al., 2020]: $\mathbf{\Theta}^{(t)} = \boldsymbol{\alpha}^{(t)}\mathbf{1}_p^\top + \mathbf{1}_p\boldsymbol{\alpha}^{(t)\top} + \boldsymbol{Z}^{(t)}\boldsymbol{Z}^{(t)\top}$
- **TDCPD** [Zhang et al., 2024c]: $\mathbf{\Theta}^{(t)} = \boldsymbol{Z}^{(t)}\operatorname{diag}(\boldsymbol{\beta}^{(t)})\boldsymbol{Z}^{(t)\top}$
- **LCSC-LSM** [Liu et al., 2024]: $\mathbf{\Theta}^{(t)} = \boldsymbol{\alpha}^{(t)}\mathbf{1}_p^\top + \mathbf{1}_p\boldsymbol{\alpha}^{(t)\top} + \boldsymbol{Z}\boldsymbol{Z}^\top$
- **KCN-LSM** [Zhang et al., 2024b]: $\mathbf{\Theta}^{(t)} = \boldsymbol{\alpha}\mathbf{1}_p^\top + \mathbf{1}_p\boldsymbol{\alpha}^\top + \boldsymbol{Z}^{(t)}\boldsymbol{Z}^{(t)\top}$
- **PLSM** [Zhang et al., 2024a]:

$$\mathbf{\Theta}^{(t)} = \boldsymbol{\alpha}\mathbf{1}_p^\top + \mathbf{1}_p\boldsymbol{\alpha}^\top + \left(\operatorname{diag}(\boldsymbol{\beta}^{(t)})\boldsymbol{Z}\right)\left(\operatorname{diag}(\boldsymbol{\beta}^{(t)})\boldsymbol{Z}\right)^\top$$

- **FlexMn** [Zhang et al., 2020]: (Does not incorporate any penalty)

$$\mathbf{\Theta}^{(t)} = \boldsymbol{\alpha}^{(t)}\mathbf{1}_p^\top + \mathbf{1}_p\boldsymbol{\alpha}^{(t)\top} + \boldsymbol{Z}\mathbf{\Lambda}^{(t)}\boldsymbol{Z}^\top$$

- **Oracle**: Assume that the constant segments are known in advance for the proposed method.

Evaluation metrics

When evaluating and comparing different methods, we first consider the relative errors defined as:

$$RE_{\boldsymbol{\Theta}} := \frac{\sum_{t=1}^{T} \left\| \widehat{\boldsymbol{\Theta}}^{(t)} - \boldsymbol{\Theta}_{\star}^{(t)} \right\|_F^2}{\sum_{t=1}^{T} \left\| \boldsymbol{\Theta}_{\star}^{(t)} \right\|_F^2}, \quad RE_{\boldsymbol{\alpha}} := \frac{\sum_{t=1}^{T} \left\| \widehat{\boldsymbol{\alpha}}^{(t)} - \boldsymbol{\alpha}_{\star}^{(t)} \right\|_F^2}{\sum_{t=1}^{T} \left\| \boldsymbol{\alpha}_{\star}^{(t)} \right\|_F^2}.$$

We also consider the special relative error of $\boldsymbol{Z}$ based on the identifiability:

$$RE_{\boldsymbol{Z}} := \frac{\min_{\boldsymbol{O} \in \mathbb{S}^{r \times r}, \boldsymbol{O}\boldsymbol{O}^{\top} = \boldsymbol{I}_r} \left\| \widehat{\boldsymbol{Z}} - \boldsymbol{Z}_{\star}\boldsymbol{O} \right\|_F^2}{\left\| \boldsymbol{Z}_{\star} \right\|_F^2}.$$

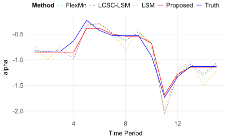
Finding the optimal $\boldsymbol{O}$ is known as the orthogonal Procrustes problem, which can be solved by singular value decomposition (SVD). In particular, if we denote the SVD of $\widehat{\boldsymbol{Z}}^{\top}\boldsymbol{Z}_{\star}$ by $\boldsymbol{S}\boldsymbol{\Sigma}\boldsymbol{V}^{\top}$, then the optimal $\boldsymbol{O}$ is given by $\boldsymbol{V}\boldsymbol{S}^{\top}$.

Results: Tables

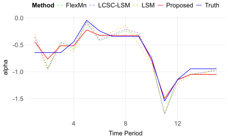
Table: Simulation results for **Network 3, $p = 200$, $T = 50$ and $r = 4$.** In each cell, mean (sd) based on 100 replicates. The bold and underlined values indicate the smallest and the second smallest, respectively.

T	**Method**	$RE_{\boldsymbol{\Theta}}$	$RE_{\boldsymbol{\alpha}}$	$RE_{\boldsymbol{Z}}$
50	LSM	0.2133(0.0161)	0.1587(0.0113)	0.1156(0.0072)
	TDCPD	0.6141(0.0167)	-	0.8716(0.0005)
	LCSC-LSM	0.0348(0.0028)	0.0627(0.0019)	0.0194(0.0040)
	KCN-LSM	0.1502(0.0128)	8.1669(0.1833)	0.0953(0.0062)
	PLSM	0.0706(0.0027)	8.1836(0.2248)	0.2781(0.0029)
	FlexMn	0.0348(0.0028)	0.0627(0.0019)	0.0018(0.0001)
	Proposed	<u>0.0102(0.0004)</u>	**0.0253(0.0009)**	<u>0.0017(0.0001)</u>
	Oracle	**0.0100(0.0004)**	<u>0.0265(0.0008)</u>	**0.0016(0.0001)**

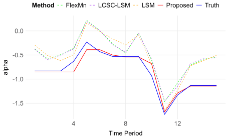
Results: Figures of temporal trends



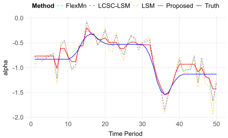
(a) Network 1, $T = 15$



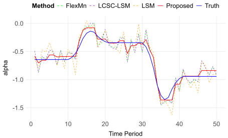
(b) Network 2, $T = 15$



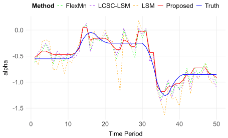
(c) Network 3, $T = 15$



(d) Network 1, $T = 50$



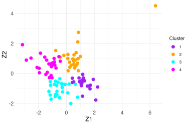
(e) Network 2, $T = 50$



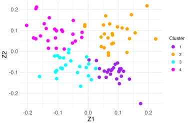
(f) Network 3, $T = 50$

Figure: Simulation results: estimation of temporal trends ($\boldsymbol{r = 2}$ and $\boldsymbol{p = 200}$).

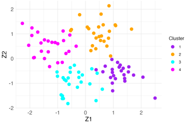
Results: Figures of latent spaces



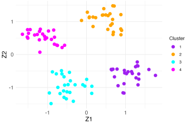
(a) LSM



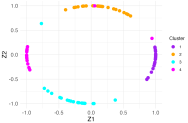
(b) TDCPD



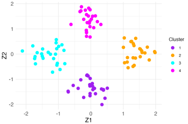
(c) KCN-LSM



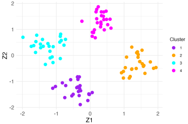
(d) LCSC-LSM



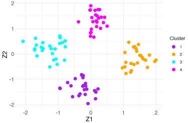
(e) PLSM



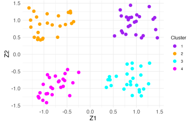
(f) FlexMn



(g) **Proposed**



(h) **Oracle**



(i) **True**

Outline

Medicare

- We first retrieve 133 million Medicare inpatient records collected during the period from January 2008 to December 2019, representing service utilization of 35 million Medicare beneficiaries.
- As in the literature, we focus on subjects aged 65 years and above.
- Following the literature [Wei et al., 2017, Jiang et al., 2018], we extract the length-of-stay (LOS) information.
- The final data for analysis is a array containing the LOS measurements for each subject, each of the 108 diseases, and each of the 12 years.

Adopting the approach developed in Mei et al. [2023] to Medicare inpatient claims data from 2008 to 2019, we first estimate the 12-year HDN adjacency matrices.

Analysis of shared latent space

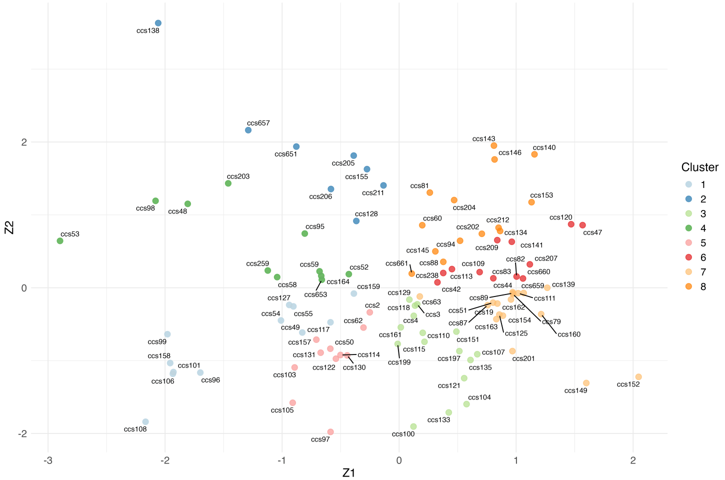
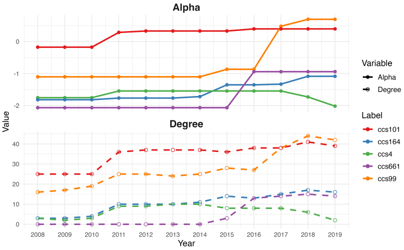
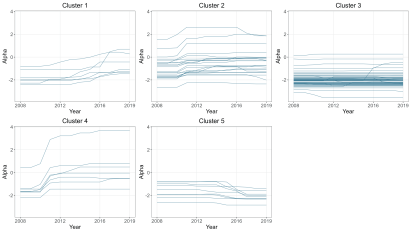


Figure: Clustering result of latent space $(r = 2)$

Analysis of temporal trends



(a) Trends of alpha v.s. degree for selected nodes



(b) Clustering result of $\{\alpha^{(t)}\}_{t=1}^{T}$ trends

Analysis using the alternative methods (e.g. LCSC-LSM)

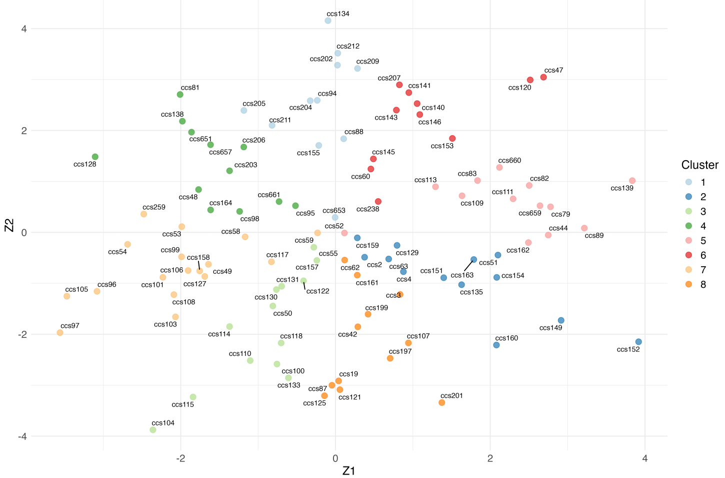
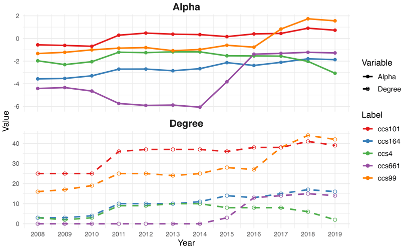
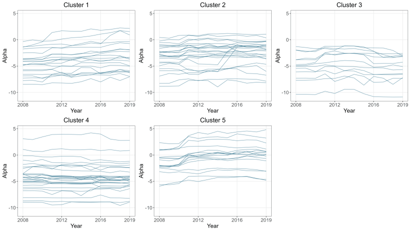


Figure: Clustering result of latent space ($r = 2$) by LCSC-LSM

Analysis using the alternative methods (e.g. LCSC-LSM)



(a) Trends of alpha v.s. degree for selected nodes by LCSC-LSM



(b) Clustering result of $\{\alpha^{(t)}\}_{t=1}^{T}$ trends by LCSC-LSM

- This instability in temporal patterns not only complicates trend interpretation but also limits the utility of clustering analysis.

THANKS!

Reference

Kwang-Il Goh, Michael E Cusick, David Valle, Barton Childs, Marc Vidal, and Albert-László Barabási. The human disease network. *Proceedings of the National Academy of Sciences*, 104(21):8685–8690, 2007.

Peter D Hoff, Adrian E Raftery, and Mark S Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460): 1090–1098, 2002.

Yefei Jiang, Shuangge Ma, Ben-Chang Shia, and Tian-Shyug Lee. An epidemiological human disease network derived from disease co-occurrence in taiwan. *Scientific Reports*, 8(1):4557, 2018.

Mengque Liu, Xinyan Fan, and Shuangge Ma. A quantitative linguistic analysis of a cancer online health community with a smooth latent space model. *The Annals of Applied Statistics*, 18(1):144–158, 2024.

Zhuang Ma, Zongming Ma, and Hongsong Yuan. Universal latent space model fitting for large networks with edge covariates. *Journal of Machine Learning Research*, 21(4): 1–67, 2020.

Hao Mei, Ruofan Jia, Guanzhong Qiao, Zhenqiu Lin, and Shuangge Ma. Human disease clinical treatment network for the elderly: analysis of the medicare inpatient length of stay and readmission data. *Biometrics*, 79(1):404–416, 2023.

Hao Mei, Haonan Xiao, Ben-Chang Shia, Guanzhong Qiao, and Yang Li. Interconnections of multimorbidity-related clinical outcomes: Analysis of health administrative claims data with a dynamic network approach. *arXiv preprint arXiv:2504.06540*, 2025.

Purnamrita Sarkar and Andrew W Moore. Dynamic social network analysis using latent space models. *Acm sigkdd explorations newsletter*, 7(2):31–40, 2005.

Ryan J. Tibshirani. Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics*, 42(1):285–323, 2014.

Wei-Qi Wei, Lisa A Bastarache, Robert J Carroll, Joy E Marlo, Travis J Osterman, Eric R Gamazon, Nancy J Cox, Dan M Roden, and Joshua C Denny. Evaluating phecodes, clinical classification software, and icd-9-cm codes for phenome-wide association studies in the electronic health record. *PLOS One*, 12(7):e0175508, 2017.

Ping Yang, Hang Qiu, Liya Wang, and Li Zhou. Early prediction of high-cost inpatients with ischemic heart disease using network analytics and machine learning. *Expert Systems with Applications*, 210:118541, 2022.

Maoyu Zhang, Biao Cai, Dong Li, Xiaoyue Niu, and Jingfei Zhang. Preferential latent space models for networks with textual edges. *arXiv preprint arXiv:2405.15038*, 2024a.

Xuefei Zhang, Songkai Xue, and Ji Zhu. A flexible latent space model for multilayer networks. In *International conference on machine learning*, pages 11288–11297. Proceedings of Machine Learning Research, 2020.

Yan Zhang, Rui Pan, Xuening Zhu, Kuangnan Fang, and Hansheng Wang. A latent space model for weighted keyword co-occurrence networks with applications in knowledge discovery in statistics. *Journal of Computational and Graphical Statistics*, pages 1–16, 2024b.

Yuzhao Zhang, Jingnan Zhang, Yifan Sun, and Junhui Wang. Change point detection in dynamic networks via regularized tensor decomposition. *Journal of Computational and Graphical Statistics*, 33(2):515–524, 2024c.

Dejia Zhou, Liya Wang, Shuhan Ding, Minghui Shen, and Hang Qiu. Phenotypic disease network analysis to identify comorbidity patterns in hospitalized patients with ischemic heart disease using large-scale administrative data. In *Healthcare*, volume 10, page 80. Multidisciplinary Digital Publishing Institute, 2022.

XueZhong Zhou, Jörg Menche, Albert-László Barabási, and Amitabh Sharma. Human symptoms–disease network. *Nature Communications*, 5(1):4212, 2014.