Statistically and Computationally Optimal Estimation and Inference in the Common Subspace Model

Joshua Agterberg

IMSI, 2026

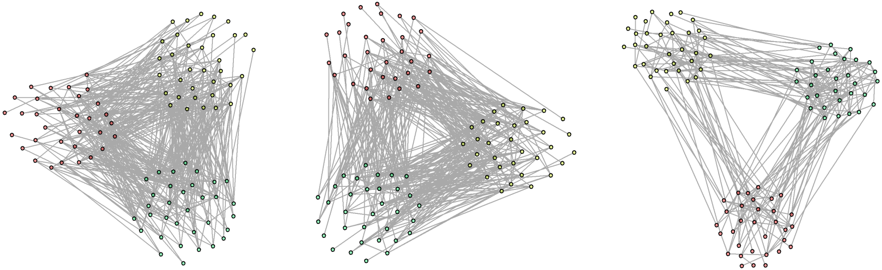
UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

Outline

Multilayer Networks

Setting

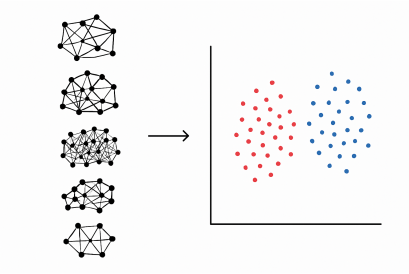
Observe L networks on same n vertices

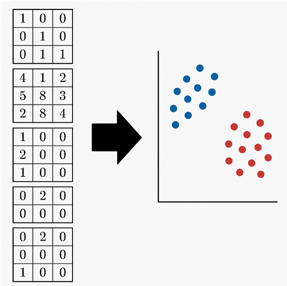


Data Integration

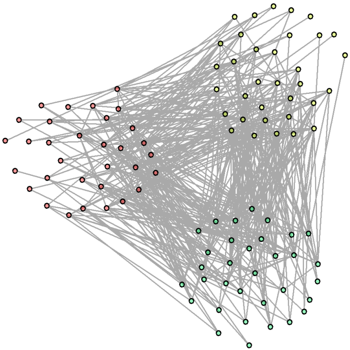
Goal

Find a common representation of all datasets

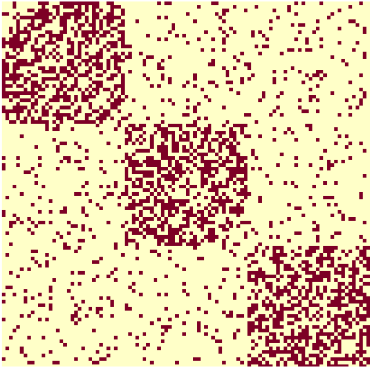




Adjacency Matrix

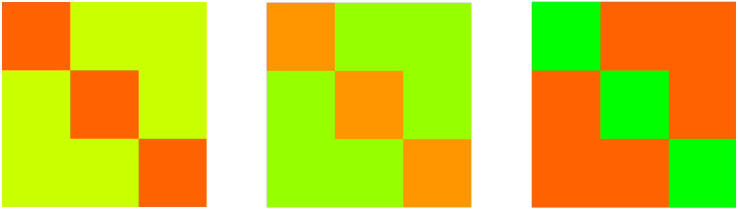


$\Longrightarrow$



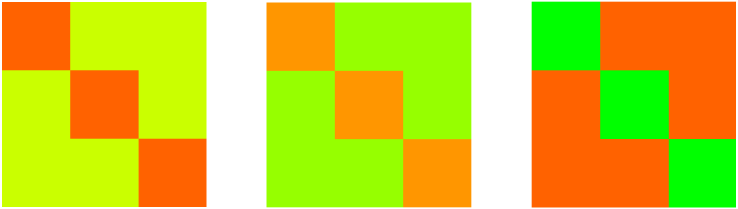
Multilayer Blockmodel

Multilayer SBM (e.g. (Lei and Lin, 2022; Lei et al., 2020)) incorporates network-level idiosyncracies:



Multilayer Blockmodel

Multilayer SBM (e.g. (Lei and Lin, 2022; Lei et al., 2020)) incorporates network-level idiosyncracies:



Problem

Edges determined solely by community memberships!

Many Generalizations

- Jing et al. (2021)
- Fan et al. (2021)
- Pensky and Wang (2021)
- Noroozi and Pensky (2022)
- Arroyo et al. (2021)
- Young et al. (2022)
- Agterberg and Zhang (2025)
- Agterberg et al. (2025)
- Hu and Wang (2023)
- Xie (2024)
- Pensky (2025)
- Agterberg and Cape (2025)
- Lyu et al. (2021)
- Lyu and Xia (2023)
- Lyu and Xia (2025)
- Wang et al. (2025)
- Yang et al. (2024)
- Chen et al. (2022)
- Chen and Hero (2017)
- Tang et al. (2025)
- Tian et al. (2025)
- Han et al. (2015)
- Han et al. (2021)
- Yan and Levin (2025)

(This list is incomplete...)

Common Subspace Model of Arroyo et al. (2021)

Assume each matrix shares a *common subspace*

$$\mathbb{E}(\boldsymbol{A}^{(l)}) = \boldsymbol{U}\boldsymbol{R}^{(l)}\boldsymbol{U}^\top \in \mathbb{R}^{n\times n}$$

$$\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_r; \qquad \boldsymbol{R}^{(l)} \in \mathbb{R}^{r\times r}$$

Common Subspace Model of Arroyo et al. (2021)

Assume each matrix shares a *common subspace*

$$\mathbb{E}(\boldsymbol{A}^{(l)}) = \boldsymbol{U}\boldsymbol{R}^{(l)}\boldsymbol{U}^\top \in \mathbb{R}^{n \times n}$$
$$\boldsymbol{U}^\top \boldsymbol{U} = \boldsymbol{I}_r; \qquad \boldsymbol{R}^{(l)} \in \mathbb{R}^{r \times r}$$

- Rows of $\boldsymbol{U} \in \mathbb{R}^{n \times r}$ are r-dimensional latent positions for each *node*
- $\boldsymbol{R}^{(l)}$ is a *layer-wise* idiosyncracy

Common Subspace Model of Arroyo et al. (2021)

Assume each matrix shares a *common subspace*

$$\mathbb{E}(\boldsymbol{A}^{(l)}) = \boldsymbol{U}\boldsymbol{R}^{(l)}\boldsymbol{U}^\top \in \mathbb{R}^{n\times n}$$

$$\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}_r; \qquad \boldsymbol{R}^{(l)} \in \mathbb{R}^{r\times r}$$

- Rows of $\boldsymbol{U} \in \mathbb{R}^{n\times r}$ are r-dimensional latent positions for each *node*
- $\boldsymbol{R}^{(l)}$ is a *layer-wise* idiosyncracy

Generalizes many other models!

A Simple Statistical Model

- Observe

$$\boldsymbol{A}^{(l)} = \underbrace{\boldsymbol{U}\boldsymbol{R}^{(l)}\boldsymbol{U}^\top}_{\text{Signal}} + \underbrace{\boldsymbol{N}^{(l)}}_{\text{Noise}}; \qquad l \in [L]$$

A Simple Statistical Model

- Observe

$$A^{(l)} = \underbrace{UR^{(l)}U^{\top}}_{\text{Signal}} + \underbrace{N^{(l)}}_{\text{Noise}}; \qquad l \in [L]$$

- $N^{(l)}$ is symmetric and satisfies

$$N_{ij}^{(l)} \sim \begin{cases} \mathcal{N}(0, \frac{\sigma^2}{2}) & i < j; \\ \mathcal{N}(0, \sigma^2) & i = j \\ N_{ji}^{(l)} & i > j. \end{cases}$$

A Simple Statistical Model

- Observe

$$\boldsymbol{A}^{(l)} = \underbrace{\boldsymbol{U}\boldsymbol{R}^{(l)}\boldsymbol{U}^\top}_{\text{Signal}} + \underbrace{\boldsymbol{N}^{(l)}}_{\text{Noise}}; \qquad l \in [L]$$

- $\boldsymbol{N}^{(l)}$ is symmetric and satisfies

$$\boldsymbol{N}^{(l)}_{ij} \sim \begin{cases} \mathcal{N}(0, \frac{\sigma^2}{2}) & i < j; \\ \mathcal{N}(0, \sigma^2) & i = j \\ \boldsymbol{N}^{(l)}_{ji} & i > j. \end{cases}$$

L observations of the classical *matrix denoising model*.

(Doesn't need to be Gaussian)

Key Parameters

Signal Strength Parameter

$$\lambda = \sigma_{\min}\left(\frac{1}{\sqrt{L}}\mathcal{R}\right) = \sigma_{\min}\left(\frac{1}{\sqrt{L}}\left[\boldsymbol{R}^{(1)}, \boldsymbol{R}^{(2)}, \cdots, \boldsymbol{R}^{(L)}\right]\right)$$

Key Parameters

Signal Strength Parameter

$$\lambda = \sigma_{\min}\left(\frac{1}{\sqrt{L}}\mathcal{R}\right) = \sigma_{\min}\left(\frac{1}{\sqrt{L}}\left[\boldsymbol{R}^{(1)}, \boldsymbol{R}^{(2)}, \cdots, \boldsymbol{R}^{(L)}\right]\right)$$

Noise and Dimensions

σ = noise level;

n = dimension of each matrix;

L = number of Layers;

r = rank of each matrix

A Simple Statistical Question

What are the fundamental limits for:

- Estimation of $\boldsymbol{U}$?
- Inference for $\boldsymbol{U}$?

A Simple Statistical Question

What are the fundamental limits for:

- Estimation of $\boldsymbol{U}$?
- Inference for $\boldsymbol{U}$?

$0 \qquad \frac{\sqrt{n}}{\sqrt{L}} \qquad \frac{\sqrt{n}}{L^{1/4}} \qquad \frac{n}{\sqrt{L}} \qquad \longrightarrow \text{SNR} = \lambda/\sigma$

Loss Function

Projection Loss

$$\mathrm{dist}_F(\boldsymbol{U}_1, \boldsymbol{U}_2) = \|\sin\Theta(\boldsymbol{U}_1, \boldsymbol{U}_2)\|_F = \frac{1}{\sqrt{2}}\|\boldsymbol{U}_1\boldsymbol{U}_1^\top - \boldsymbol{U}_2\boldsymbol{U}_2^\top\|_F$$

Loss Function

Projection Loss

$$\mathrm{dist}_F(\boldsymbol{U}_1, \boldsymbol{U}_2) = \left\| \sin\Theta(\boldsymbol{U}_1, \boldsymbol{U}_2) \right\|_F = \frac{1}{\sqrt{2}} \| \boldsymbol{U}_1 \boldsymbol{U}_1^\top - \boldsymbol{U}_2 \boldsymbol{U}_2^\top \|_F$$

- Focus on $r, \kappa = O(1)$ for simplicity.

Loss Function

Projection Loss

$$\mathrm{dist}_F(\boldsymbol{U}_1, \boldsymbol{U}_2) = \left\| \sin \Theta(\boldsymbol{U}_1, \boldsymbol{U}_2) \right\|_F = \frac{1}{\sqrt{2}} \| \boldsymbol{U}_1 \boldsymbol{U}_1^\top - \boldsymbol{U}_2 \boldsymbol{U}_2^\top \|_F$$

- Focus on $r, \kappa = O(1)$ for simplicity.
- Assume that $L \lesssim n$ also for simplicity.

Minimax Lower Bound

Minimax Lower Bound

Theorem (Minimax Lower Bound)

It holds that

$$\inf_{\widehat{\boldsymbol{U}}} \sup_{\mathcal{P}(\lambda)} \mathbb{E}\| \sin \Theta(\widehat{\boldsymbol{U}}, \boldsymbol{U})\|_F^2 \gtrsim \frac{\sigma^2 n}{\lambda^2 L},$$

where the infimum is over all estimators $\widehat{\boldsymbol{U}}$ of $\boldsymbol{U}$.

Minimax Lower Bound

Theorem (Minimax Lower Bound)

It holds that

$$\inf_{\widehat{\boldsymbol{U}}} \sup_{\mathcal{P}(\lambda)} \mathbb{E}\| \sin\Theta(\widehat{\boldsymbol{U}}, \boldsymbol{U})\|_F^2 \gtrsim \frac{\sigma^2 n}{\lambda^2 L},$$

where the infimum is over all estimators $\widehat{\boldsymbol{U}}$ of $\boldsymbol{U}$.

Projected Gradient Descent

Under Gaussian noise, the MLE is the minimizer of

$$L\left(\widehat{\boldsymbol{U}}, \{\widehat{\boldsymbol{R}}^{(l)}\}_{l=1}^{L}\right) = \frac{1}{2}\sum_{l=1}^{L} \left\|\boldsymbol{A}^{(l)} - \widehat{\boldsymbol{U}}\widehat{\boldsymbol{R}}^{(l)}\widehat{\boldsymbol{U}}^{\top}\right\|_F^2$$

$$\text{subject to } \widehat{\boldsymbol{U}}^{\top}\widehat{\boldsymbol{U}} = \boldsymbol{I}_r$$

Projected Gradient Descent

Under Gaussian noise, the MLE is the minimizer of

$$L\left(\widehat{\boldsymbol{U}}, \{\widehat{\boldsymbol{R}}^{(l)}\}_{l=1}^{L}\right) = \frac{1}{2}\sum_{l=1}^{L} \left\| \boldsymbol{A}^{(l)} - \widehat{\boldsymbol{U}}\widehat{\boldsymbol{R}}^{(l)}\widehat{\boldsymbol{U}}^\top \right\|_F^2$$

$$\text{subject to } \widehat{\boldsymbol{U}}^\top \widehat{\boldsymbol{U}} = \boldsymbol{I}_r$$

Can try minimizing this loss function directly!

Projected Gradient Descent

Under Gaussian noise, the MLE is the minimizer of

$$L\left(\widehat{\boldsymbol{U}}, \{\widehat{\boldsymbol{R}}^{(l)}\}_{l=1}^{L}\right) = \frac{1}{2}\sum_{l=1}^{L}\left\|\boldsymbol{A}^{(l)} - \widehat{\boldsymbol{U}}\widehat{\boldsymbol{R}}^{(l)}\widehat{\boldsymbol{U}}^{\top}\right\|_F^2$$

subject to $\widehat{\boldsymbol{U}}^{\top}\widehat{\boldsymbol{U}} = \boldsymbol{I}_r$

Can try minimizing this loss function directly!

Initialize with leading r left singular vectors of $\sum_l (\boldsymbol{A}^{(l)})^2$
(no bias because of homoskedasticity)

Projected Gradient Descent

Algorithm Projected Gradient Descent Initialized via $\widehat{\boldsymbol{U}}_0$

Input: Collection of matrices $\{\boldsymbol{A}^{(l)}\}_{l=1}^{L}$; rank r, step-size η, number of iterations T, initialization $\widehat{\boldsymbol{U}}_0$.

1. While $t \leq T$:
    1. Set $\widehat{\boldsymbol{R}}_t^{(l)} := \widehat{\boldsymbol{U}}_t^{\top} \boldsymbol{A}^{(l)} \widehat{\boldsymbol{U}}_t$.
    2. Gradient step:
    $$\widehat{\boldsymbol{U}}_{t+.5} := \widehat{\boldsymbol{U}}_t - \frac{\eta}{L} \sum_{l=1}^{L} \left(\widehat{\boldsymbol{U}}_t \widehat{\boldsymbol{R}}_t^{(l)} \widehat{\boldsymbol{U}}_t^{\top} - \boldsymbol{A}^{(l)} \right) \widehat{\boldsymbol{U}}_t \widehat{\boldsymbol{R}}_t^{(l)}. \quad (1)$$
    3. Projection step:
    $$\widehat{\boldsymbol{U}}_{t+1} := \mathrm{SVD}_r(\widehat{\boldsymbol{U}}_{t+.5}) \quad (2)$$

Output: $\widehat{\boldsymbol{U}}_T$.

Theoretical Guarantee

Theorem (Upper Bounds for Iterates)

Suppose that

$$\mathrm{SNR} = \lambda/\sigma \geq C_0 \frac{\sqrt{n}}{L^{1/4}}. \tag{3}$$

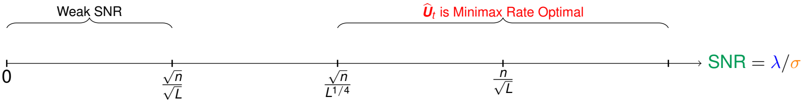
Then with high probability,

$$\|\sin\Theta(\widehat{\boldsymbol{U}}_t, \boldsymbol{U})\|_F \leq \frac{C_1 \sigma \sqrt{n}}{\lambda \sqrt{L}} + \frac{1}{2^{t+1}}.$$

Theoretical Guarantee

When SNR $= \lambda/\sigma \gtrsim \frac{\sqrt{n}}{L^{1/4}}$, after logarithmically many iterations,

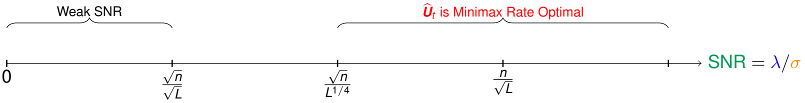
$$\|\sin\Theta(\widehat{\boldsymbol{U}}_t, \boldsymbol{U})\|_F \lesssim \frac{\sigma\sqrt{n}}{\lambda\sqrt{L}}.$$



Theoretical Guarantee

When SNR $= \lambda/\sigma \gtrsim \frac{\sqrt{n}}{L^{1/4}}$, after logarithmically many iterations,

$$\|\sin\Theta(\widehat{\boldsymbol{U}}_t, \boldsymbol{U})\|_F \lesssim \frac{\sigma\sqrt{n}}{\lambda\sqrt{L}}.$$

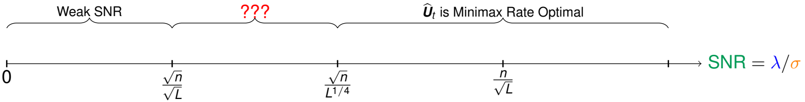


Is this assumption necessary?

Theoretical Guarantee

When SNR $= \lambda/\sigma \gtrsim \frac{\sqrt{n}}{L^{1/4}}$, after logarithmically many iterations,

$$\|\sin\Theta(\widehat{\boldsymbol{U}}_t, \boldsymbol{U})\|_F \lesssim \frac{\sigma\sqrt{n}}{\lambda\sqrt{L}}.$$



Is this assumption necessary?

Well, Is It?

Well, Is It?

Conjecture (Schramm and Wein (2022))

Let $LR^{\leq D}$ be the projection of LR onto degree at most D polynomials.

Well, Is It?

Conjecture (Schramm and Wein (2022))

Let $LR^{\leq D}$ be the projection of LR onto degree at most D polynomials. If there exists $D \geq \log(n)^{1+\varepsilon}$ for some $\varepsilon > 0$ such that $\mathbb{E}_{H_a}(LR^{\leq D})^2 = 1 + o(1)$, then no polynomial time algorithm hypothesis test can strongly distinguish H_0 and H_a.

Well, Is It?

Conjecture (Schramm and Wein (2022))

Let $LR^{\leq D}$ be the projection of LR onto degree at most D polynomials. If there exists $D \geq \log(n)^{1+\varepsilon}$ for some $\varepsilon > 0$ such that $\mathbb{E}_{H_a}(LR^{\leq D})^2 = 1 + o(1)$, then no polynomial time algorithm hypothesis test can strongly distinguish H_0 and H_a.

Theorem

If $\mathrm{SNR} = \lambda/\sigma \ll \frac{\sqrt{n}}{L^{1/4}}$*, then* $\mathbb{E}_{H_a}(LR^{\leq D})^2 = 1 + o(1)$.

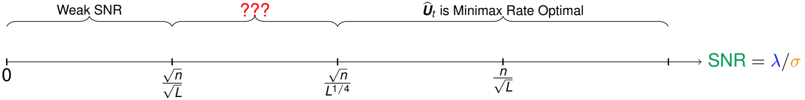
Well, Is It?

Conjecture (Schramm and Wein (2022))

Let $LR^{\leq D}$ be the projection of LR onto degree at most D polynomials. If there exists $D \geq \log(n)^{1+\varepsilon}$ for some $\varepsilon > 0$ such that $\mathbb{E}_{H_a}(LR^{\leq D})^2 = 1 + o(1)$, then no polynomial time algorithm hypothesis test can strongly distinguish H_0 and H_a.

Theorem

If $\text{SNR} = \lambda/\sigma \ll \frac{\sqrt{n}}{L^{1/4}}$, *then* $\mathbb{E}_{H_a}(LR^{\leq D})^2 = 1 + o(1)$.



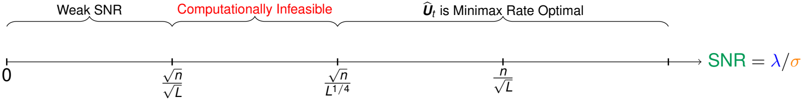
Well, Is It?

Conjecture

Let $LR^{\leq D}$ be the projection of LR onto degree at most D polynomials. If there exists $D \geq \log(n)^{1+\varepsilon}$ for some $\varepsilon > 0$ such that $\mathbb{E}_{H_a}(LR^{\leq D})^2 = 1 + o(1)$, then no polynomial time algorithm hypothesis test can strongly distinguish H_0 and H_a.

Theorem

If $\text{SNR} = \lambda/\sigma \ll \frac{\sqrt{n}}{L^{1/4}}$, *then* $\mathbb{E}_{H_a}(LR^{\leq D})^2 = 1 + o(1)$.



Confidence Intervals for the Loss

Goal

Find *rate-optimal* confidence intervals for $\| \sin \Theta(\widehat{\boldsymbol{U}}_t, \boldsymbol{U}) \|_F^2$.

Minimax Lower Bound on Length

Theorem

Suppose $\lambda/\sigma \geq C_0\sqrt{\frac{n}{L}}$. *Then minimax length of any confidence interval* CI_α *that is valid over* $\mathcal{P}(\lambda)$ *satisfies*

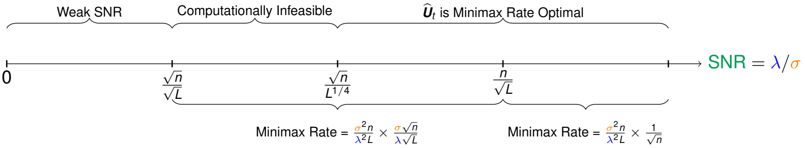
$$\inf_{\mathrm{CI}_\alpha} \sup_{\theta \in \mathcal{P}(\lambda)} \mathbb{E}_\theta \mathrm{Length}(\mathrm{CI}_\alpha) \gtrsim \frac{\sigma^2 n}{\lambda^2 L}\left(\frac{\sigma\sqrt{n}}{\lambda\sqrt{L}} + \frac{1}{\sqrt{n}}\right)$$

Minimax Lower Bound on Length

Theorem

Suppose $\lambda/\sigma \geq C_0\sqrt{\frac{n}{L}}$. *Then minimax length of any confidence interval* CI_α *that is valid over* $\mathcal{P}(\lambda)$ *satisfies*

$$\inf_{\mathsf{CI}_\alpha} \sup_{\theta\in\mathcal{P}(\lambda)} \mathbb{E}_\theta \mathrm{Length}(\mathsf{CI}_\alpha) \gtrsim \frac{\sigma^2 n}{\lambda^2 L}\left(\frac{\sigma\sqrt{n}}{\lambda\sqrt{L}} + \frac{1}{\sqrt{n}}\right)$$



A Limit Theorem

Theorem (Asymptotic Normality)

Suppose that $\lambda/\sigma \geq C_0 \frac{n}{\sqrt{L}}$.

A Limit Theorem

Theorem (Asymptotic Normality)

Suppose that $\lambda/\sigma \geq C_0 \frac{n}{\sqrt{L}}$. *Recall*

$$\mathcal{R} = \left[\boldsymbol{R}^{(1)}, \boldsymbol{R}^{(2)}, \cdots, \boldsymbol{R}^{(L)}\right].$$

Then after logarithmically many iterations,

$$\frac{\|\sin\Theta(\widehat{\boldsymbol{U}}_t, \boldsymbol{U})\|_F^2 - \frac{\sigma^2 n}{2}\mathrm{Tr}\left((\mathcal{R}\mathcal{R}^\top)^{-1}\right)}{\sigma^2\sqrt{n/2}\|(\mathcal{R}\mathcal{R}^\top)^{-1}\|_F} \xrightarrow{d} \mathcal{N}(0,1).$$

Proposed Confidence Interval

$$\widehat{\boldsymbol{R}}^{(l)} = \widehat{\boldsymbol{U}}_t^\top \boldsymbol{A}^{(l)} \widehat{\boldsymbol{U}}_t; \qquad \widehat{\mathcal{R}} := \left[\widehat{\boldsymbol{R}}^{(1)}, \widehat{\boldsymbol{R}}^{(2)}, \cdots, \widehat{\boldsymbol{R}}^{(L)}\right];$$

$$\widehat{\mathrm{CI}} := \frac{\sigma^2 n}{2}\mathrm{Tr}\left(\left(\widehat{\mathcal{R}}\widehat{\mathcal{R}}^\top\right)^{-1}\right) \pm z_{\alpha/2}\sigma^2\sqrt{\frac{n}{2}}\left\|\left(\widehat{\mathcal{R}}\widehat{\mathcal{R}}^\top\right)^{-1}\right\|_F.$$

Proposed Confidence Interval

$$\widehat{\boldsymbol{R}}^{(l)} = \widehat{\boldsymbol{U}}_t^\top \boldsymbol{A}^{(l)} \widehat{\boldsymbol{U}}_t; \qquad \widehat{\mathcal{R}} := \left[\widehat{\boldsymbol{R}}^{(1)}, \widehat{\boldsymbol{R}}^{(2)}, \cdots, \widehat{\boldsymbol{R}}^{(L)}\right];$$

$$\widehat{\mathrm{CI}} := \frac{\sigma^2 n}{2} \mathrm{Tr}\Big(\big(\widehat{\mathcal{R}}\widehat{\mathcal{R}}^\top\big)^{-1}\Big) \pm z_{\alpha/2}\sigma^2 \sqrt{\frac{n}{2}} \big\| \big(\widehat{\mathcal{R}}\widehat{\mathcal{R}}^\top\big)^{-1} \big\|_F.$$

Theorem

Suppose that $\lambda/\sigma \geq C_0 \frac{n}{\sqrt{L}}$. *Then*

$$\mathbb{P}\Big(\| \sin \Theta(\widehat{\boldsymbol{U}}_t, \boldsymbol{U}) \|_F^2 \in \widehat{\mathrm{CI}} \Big) \geq 1 - \alpha;$$

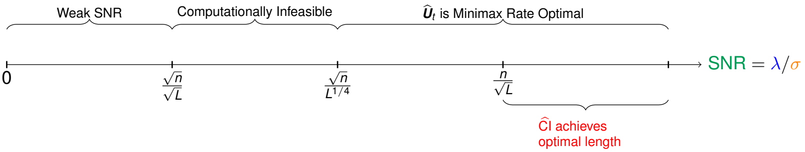
$$\mathrm{L}(\widehat{\mathrm{CI}}) \lesssim \frac{\sigma^2 n}{\lambda^2 L} \frac{1}{\sqrt{n}} = \textit{ minimax optimal length.}$$

Confidence Intervals

When SNR $\gtrsim \frac{n}{\sqrt{L}}$, $\widehat{\text{CI}}$ attains the minimax-optimal length!

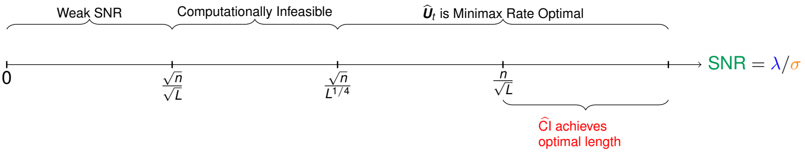
Confidence Intervals

When SNR $\gtrsim \frac{n}{\sqrt{L}}$, $\widehat{\text{CI}}$ attains the minimax-optimal length!



Confidence Intervals

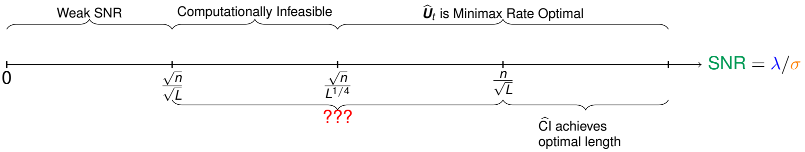
When SNR $\gtrsim \frac{n}{\sqrt{L}}$, $\widehat{\mathrm{CI}}$ attains the minimax-optimal length!



Is this assumption *necessary*?

Confidence Intervals

When SNR $\gtrsim \frac{n}{\sqrt{L}}$, $\widehat{\mathrm{CI}}$ attains the minimax-optimal length!



Is this assumption *necessary*?

Adaptivity

Adaptivity

A confidence interval is *optimally adaptive* to signal strength if it achieves the rate-optimal length without *a priori* knowledge of the signal strength.

Adaptivity

A confidence interval is *optimally adaptive* to signal strength if it achieves the rate-optimal length without *a priori* knowledge of the signal strength.

Theorem

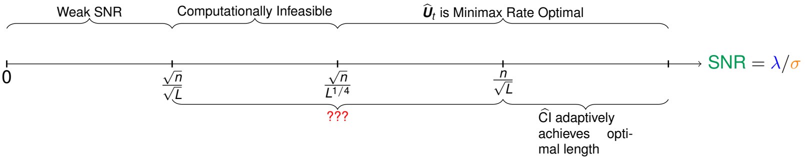
If $\lambda/\sigma \ll \frac{n}{\sqrt{L}}$*, then rate-optimal adaptive inference is impossible.*

Adaptivity

A confidence interval is *optimally adaptive* to signal strength if it achieves the rate-optimal length without *a priori* knowledge of the signal strength.

Theorem

If $\lambda/\sigma \ll \frac{n}{\sqrt{L}}$, then rate-optimal adaptive inference is impossible.

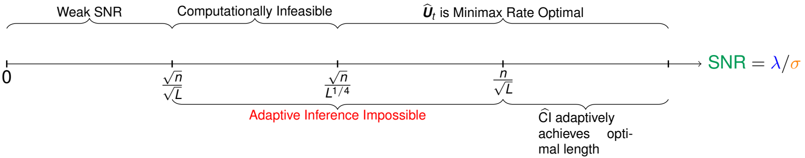


Adaptivity

A confidence interval is *optimally adaptive* to signal strength if it achieves the rate-optimal length without *a priori* knowledge of the signal strength.

Theorem

If $\lambda/\sigma \ll \frac{n}{\sqrt{L}}$, then rate-optimal adaptive inference is impossible.



Conclusion

Main Question

What are the fundamental limits for:

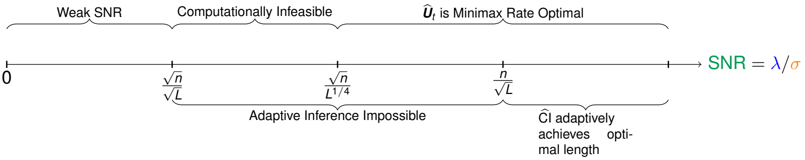
- Estimation of $\boldsymbol{U}$?
- Inference for $\boldsymbol{U}$?

Conclusion

Main Question

What are the fundamental limits for:

- Estimation of $\boldsymbol{U}$?
- Inference for $\boldsymbol{U}$?



Questions?

(ArXiv paper to appear soon...)

References I

Joshua Agterberg and Joshua Cape. An Overview of Asymptotic Normality in Stochastic Blockmodels: Cluster Analysis and Inference, January 2025. arXiv:2305.06353 [math].

Joshua Agterberg and Anru R. Zhang. Estimating Higher-Order Mixed Memberships via the l2,∞ Tensor Perturbation Bound. *Journal of the American Statistical Association*, 120(550): 1214–1224, April 2025. ISSN 0162-1459. doi: 10.1080/01621459.2024.2404265.

Joshua Agterberg, Zachary Lubberts, and Jesús Arroyo. Joint Spectral Clustering in Multilayer Degree-Corrected Stochastic Blockmodels. *Journal of the American Statistical Association*, 120(551):1607–1620, July 2025. ISSN 0162-1459. doi: 10.1080/01621459.2025.2516201.

References II

Jesús Arroyo, Avanti Athreya, Joshua Cape, Guodong Chen, Carey E. Priebe, and Joshua T. Vogelstein. Inference for Multiple Heterogeneous Networks with a Common Invariant Subspace. *Journal of Machine Learning Research*, 22(142): 1–49, 2021. ISSN 1533-7928.

Pin-Yu Chen and Alfred O. Hero. Multilayer Spectral Graph Clustering via Convex Layer Aggregation: Theory and Algorithms. *IEEE Transactions on Signal and Information Processing over Networks*, 3(3):553–567, September 2017. ISSN 2373-776X. doi: 10.1109/TSIPN.2017.2731123.

Shuxiao Chen, Sifan Liu, and Zongming Ma. Global and Individualized Community Detection in Inhomogeneous Multilayer Networks, May 2022. arXiv:2012.00933 [cs, math, stat].

References III

Xing Fan, Marianna Pensky, Feng Yu, and Teng Zhang. ALMA: Alternating Minimization Algorithm for Clustering Mixture Multilayer Network, October 2021. arXiv:2102.10226 [cs, math, stat].

Qiuyi Han, Kevin Xu, and Edoardo Airoldi. Consistent estimation of dynamic and multi-layer block models. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 1511–1520. PMLR, June 2015. ISSN: 1938-7228.

Rungang Han, Yuetian Luo, Miaoyan Wang, and Anru R. Zhang. Exact Clustering in Tensor Block Model: Statistical Optimality and Computational Limit, October 2021. arXiv:2012.09996 [math, stat].

References IV

Jiaxin Hu and Miaoyan Wang. Multiway Spherical Clustering via Degree-Corrected Tensor Block Models. *IEEE Transactions on Information Theory*, 69(6):3880–3919, June 2023. ISSN 0018-9448, 1557-9654. doi: 10.1109/TIT.2023.3239521.

Bing-Yi Jing, Ting Li, Zhongyuan Lyu, and Dong Xia. Community detection on mixture multilayer networks via regularized tensor decomposition. *The Annals of Statistics*, 49(6):3181–3205, December 2021. ISSN 0090-5364, 2168-8966. doi: 10.1214/21-AOS2079.

Jing Lei and Kevin Z. Lin. Bias-Adjusted Spectral Clustering in Multi-Layer Stochastic Block Models. *Journal of the American Statistical Association*, 0(0):1–13, March 2022. ISSN 0162-1459. doi: 10.1080/01621459.2022.2054817.

References V

Jing Lei, Kehui Chen, and Brian Lynch. Consistent community detection in multi-layer network data. *Biometrika*, 107(1): 61–73, March 2020. ISSN 0006-3444. doi: 10.1093/biomet/asz068.

Zhongyuan Lyu and Dong Xia. Optimal estimation and computational limit of low-rank Gaussian mixtures. *The Annals of Statistics*, 51(2):646–667, April 2023. ISSN 0090-5364, 2168-8966. doi: 10.1214/23-AOS2264.

Zhongyuan Lyu and Dong Xia. Optimal clustering by Lloyd's algorithm for low-rank mixture model. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkaf041, July 2025. ISSN 1369-7412. doi: 10.1093/jrsssb/qkaf041.

References VI

Zhongyuan Lyu, Dong Xia, and Yuan Zhang. Latent Space Model for Higher-order Networks and Generalized Tensor Decomposition. *arXiv:2106.16042 [cs, math, stat]*, June 2021.

Majid Noroozi and Marianna Pensky. Sparse Subspace Clustering in Diverse Multiplex Network Model, June 2022. arXiv:2206.07602 [cs, stat].

Marianna Pensky. Signed Diverse Multiplex Networks: Clustering and Inference. *IEEE Transactions on Information Theory*, 71(9):7076–7096, September 2025. ISSN 1557-9654. doi: 10.1109/TIT.2025.3588786.

Marianna Pensky and Yaxuan Wang. Clustering of Diverse Multiplex Networks. *arXiv:2110.05308 [stat]*, October 2021.

References VII

Tselil Schramm and Alexander S. Wein. Computational barriers to estimation from low-degree polynomials. *The Annals of Statistics*, 50(3):1833–1858, June 2022. ISSN 0090-5364, 2168-8966. doi: 10.1214/22-AOS2179.

Runshi Tang, Runbing Zheng, Anru R. Zhang, and Carey E. Priebe. A functional tensor model for dynamic multilayer networks with common invariant subspaces and the RKHS estimation, September 2025. arXiv:2509.05221 [stat].

Yuang Tian, Jiajin Sun, and Yinqiu He. Efficient Analysis of Latent Spaces in Heterogeneous Networks, August 2025. arXiv:2412.02151 [stat].

References VIII

Fan Wang, Wanshan Li, Oscar Hernan Madrid Padilla, Yi Yu, and Alessandro Rinaldo. Multilayer random dot product graphs: estimation and online change point detection1. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkaf051, August 2025. ISSN 1369-7412. doi: 10.1093/jrsssb/qkaf051.

Fangzheng Xie. Bias-Corrected Joint Spectral Embedding for Multilayer Networks With Invariant Subspace: Entrywise Eigenvector Perturbation and Inference. *IEEE Transactions on Information Theory*, 70(12):9036–9083, December 2024. ISSN 1557-9654. doi: 10.1109/TIT.2024.3471953.

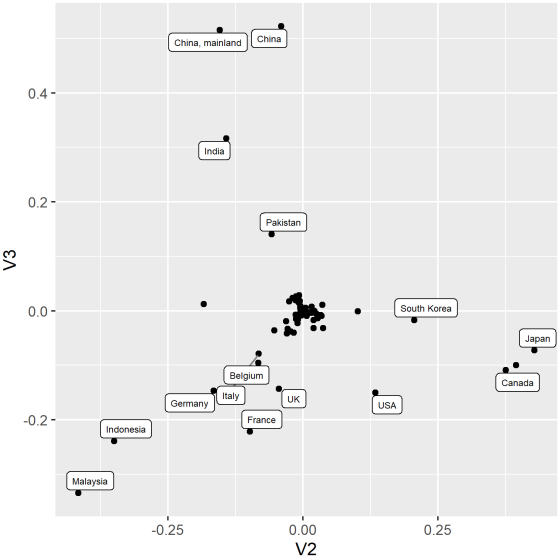
Hao Yan and Keith Levin. Estimating Multiple Weighted Networks with Node-Sparse Differences and Shared Low-Rank Structure, June 2025. arXiv:2506.15915 [math].

References IX

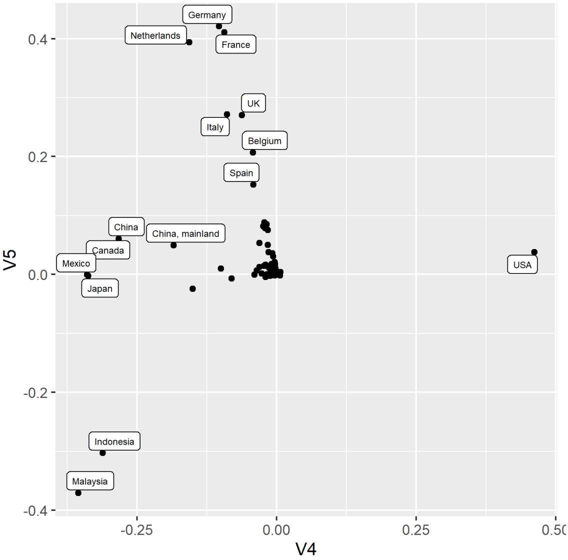
Xiaodong Yang, Buyu Lin, and Subhabrata Sen. Fundamental limits of community detection from multi-view data: multi-layer, dynamic and partially labeled block models, January 2024. arXiv:2401.08167 [math].

Jean-Gabriel Young, Alec Kirkley, and M. E. J. Newman. Clustering of heterogeneous populations of networks. *Physical Review E*, 105(1):014312, January 2022. doi: 10.1103/PhysRevE.105.014312.

Trade Data: Dimensions 2 and 3



Trade Data: Dimensions 4 and 5



Dimension 1

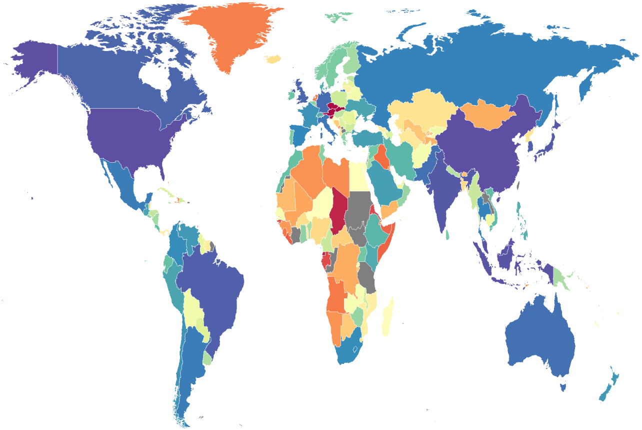


Figure: Percentile of latent dimension 1.

Dimension 2

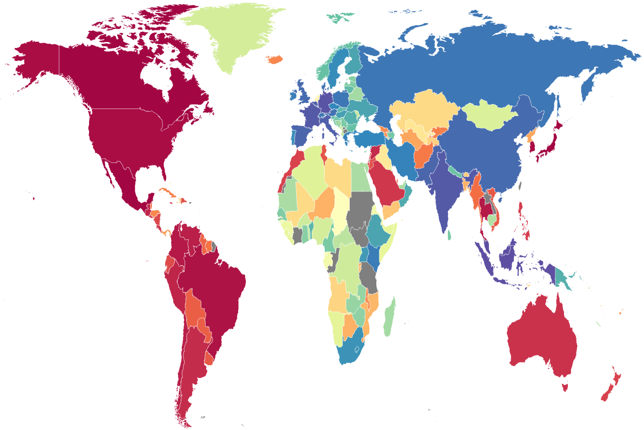


Figure: Percentile of latent dimension 2.

Dimension 3

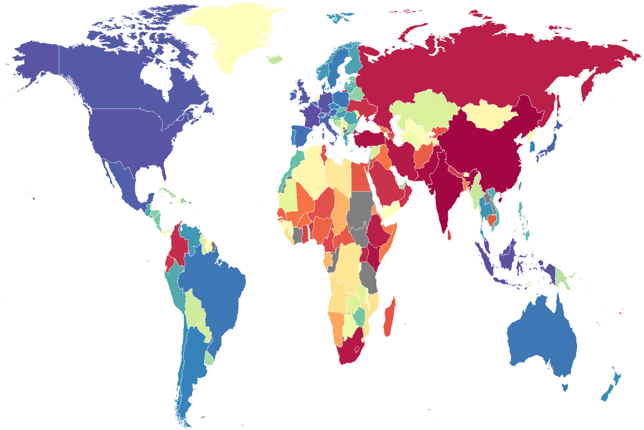


Figure: Percentile of latent dimension 3.

Dimension 4

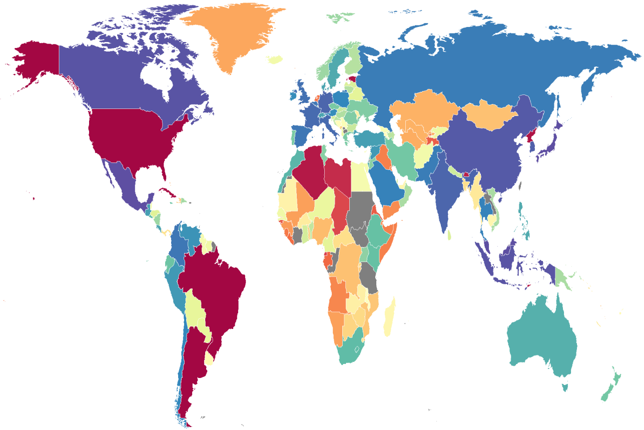


Figure: Percentile of latent dimension 4.

Dimension 5

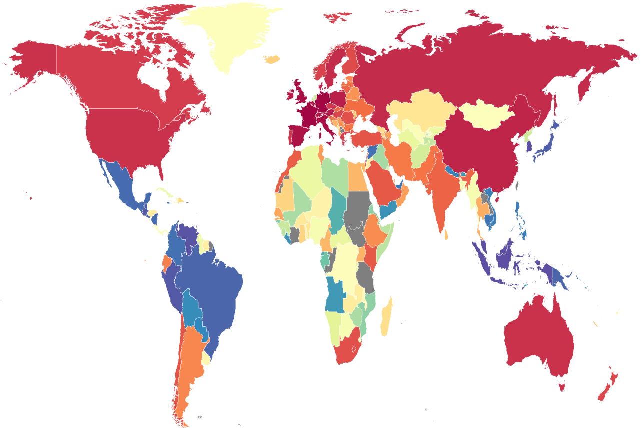


Figure: Percentile of latent dimension 5.