

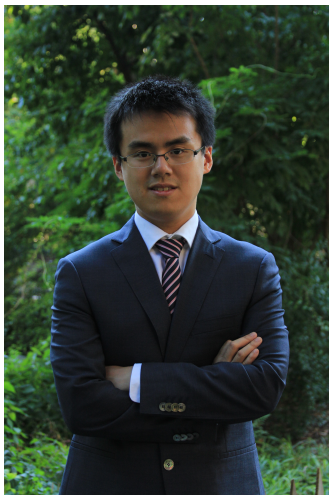
# Estimating Higher-Order Mixed Memberships via the $\ell_{2,\infty}$ Tensor Perturbation Bound

Joshua Agterberg



Department of Statistics  
University of South Carolina  
2023

# Joint Work With:



Anru Zhang (Duke)

# Outline

- 1 Motivation
- 2 Community Models
- 3 Estimation Algorithm
- 4  $\ell_{2,1}$  Tensor Perturbation
- 5 Data Analysis
- 6 Conclusion

# Outline

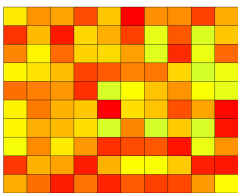
- 1 Motivation
- 2 Community Models
- 3 Estimation Algorithm
- 4  $\ell_{2,1}$  Tensor Perturbation
- 5 Data Analysis
- 6 Conclusion

# Tensor Data

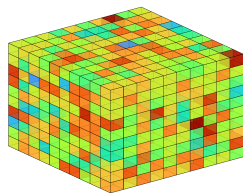
- A tensor is a multidimensional array.



Vector



Matrix



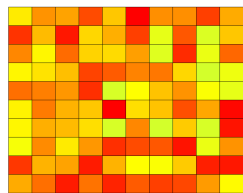
Order 3 Tensor

# Tensor Data

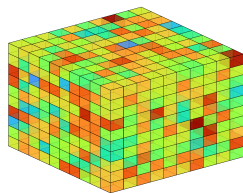
- A tensor is a multidimensional array.



Vector



Matrix



Order 3 Tensor

- Can have higher-order tensors  $\mathcal{T} \in \mathbb{R}^{p_1 \times \dots \times p_d}$ .
- This talk: focus on order 3 tensors.

# Examples of Tensor Data

- Matrix time series

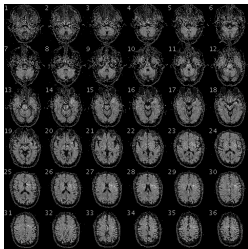
$$\mathbf{M}_t = \begin{matrix} & \text{Apple} & \text{Twitter} & \text{Tesla} & \cdots & 1 \\ \text{Revenue}_t & X_{11}^{(t)} & X_{12}^{(t)} & X_{13}^{(t)} & \cdots & \vdots \\ \text{Assets}_t & X_{21}^{(t)} & X_{22}^{(t)} & X_{23}^{(t)} & \cdots & \vdots \\ \text{Dividends per share}_t & X_{31}^{(t)} & X_{32}^{(t)} & X_{33}^{(t)} & \cdots & \vdots \\ \vdots & \cdots & \cdots & \cdots & \ddots & \vdots \end{matrix}$$

# Examples of Tensor Data

- Matrix time series

$$\mathbf{M}_t = \begin{matrix} \text{Revenue}_t \\ \text{Assets}_t \\ \text{Dividends per share}_t \\ \vdots \end{matrix} \begin{matrix} \text{Apple} \\ \text{Twitter} \\ \text{Tesla} \\ \dots \end{matrix} = \begin{matrix} X_{11}^{(t)} & X_{12}^{(t)} & X_{13}^{(t)} & \dots \\ X_{21}^{(t)} & X_{22}^{(t)} & X_{23}^{(t)} & \dots \\ X_{31}^{(t)} & X_{32}^{(t)} & X_{33}^{(t)} & \dots \\ \dots & \dots & \dots & \ddots \end{matrix}$$

- Brain imaging data

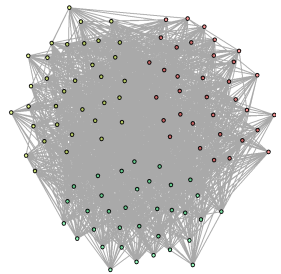
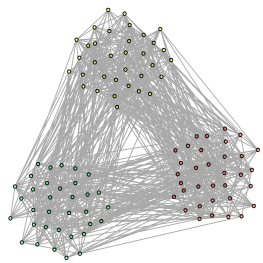
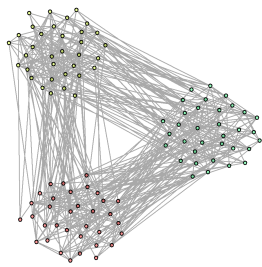




# Special Case: Multilayer Networks

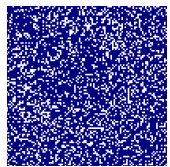
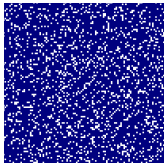
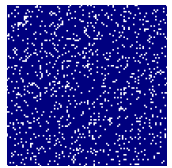
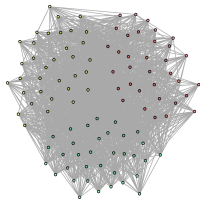
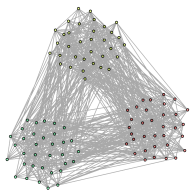
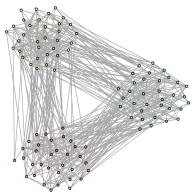
## Setting

Observe  $L$  networks on same  $n$  vertices



# Multilayer Networks

Identify each network with its adjacency matrix:

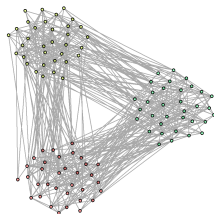
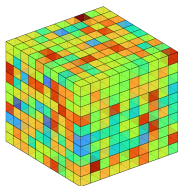


Organize adjacency matrices into  $n \times n \times L$  Tensor!

# This Talk

## Main Question

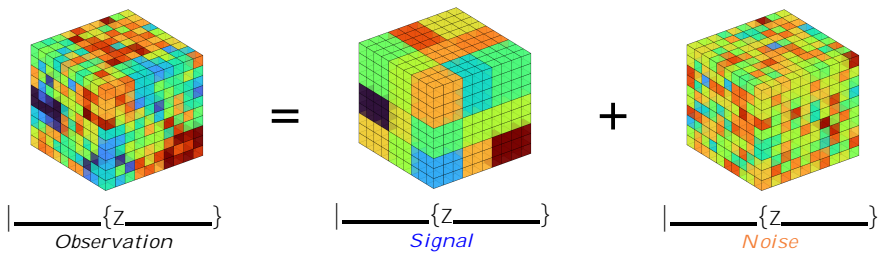
Given a noisy high-dimensional  $p_1 \times p_2 \times p_3$  tensor with underlying community structure, can we consistently estimate the communities in the high dimensional regime  $p_1, p_2, p_3 \asymp p$  as  $p \rightarrow \infty$ ?



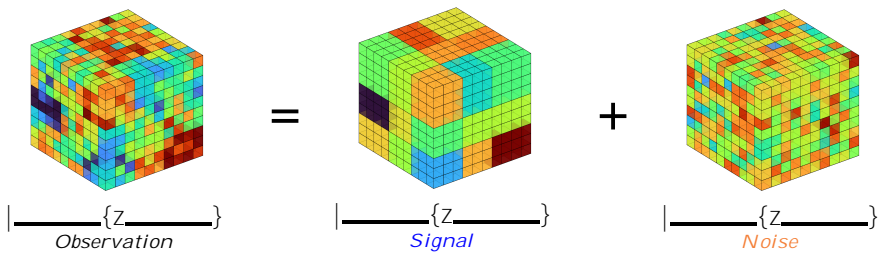
# Outline

- 1 Motivation
- 2 Community Models**
- 3 Estimation Algorithm
- 4  $\ell_{2,1}$  Tensor Perturbation
- 5 Data Analysis
- 6 Conclusion

# Tensor Blockmodels



# Tensor Blockmodels



## Interpretation

- There are  $r_k$  communities for each mode ( $k = 1, 2, 3$ ), and each node belongs to one community.
- Entry  $i_1, i_2, i_3$  of the signal tensor is determined by community memberships of nodes  $i_1, i_2$ , and  $i_3$ .

# Tensor Blockmodels

## Interpretation

- There are  $r_k$  communities for each mode ( $k = 1, 2, 3$ ), and each node belongs to one community.
- Entry  $i_1, i_2, i_3$  of the signal tensor is determined by community memberships of nodes  $i_1, i_2$ , and  $i_3$ .

Example I:

- Multilayer network with airports, airports, and airlines as the modes.

# Tensor Blockmodels

## Interpretation

- There are  $r_k$  communities for each mode ( $k = 1, 2, 3$ ), and each node belongs to one community.
- Entry  $i_1, i_2, i_3$  of the signal tensor is determined by community memberships of nodes  $i_1, i_2$ , and  $i_3$ .

### Example I:

- Multilayer network with airports, airports, and airlines as the modes.
- Each airport belongs to one of  $r_1$  communities ( $r_1 = r_2$ ).



# Tensor Blockmodels

## Interpretation

- There are  $r_k$  communities for each mode ( $k = 1, 2, 3$ ), and each node belongs to one community.
- Entry  $i_1, i_2, i_3$  of the signal tensor is determined by community memberships of nodes  $i_1, i_2$ , and  $i_3$ .

### Example I:

- Multilayer network with airports, airports, and airlines as the modes.
- Each airport belongs to one of  $r_1$  communities ( $r_1 = r_2$ ).
- Each airline belongs to one of  $r_3$  communities.

# Tensor Blockmodels

## Interpretation

- There are  $r_k$  communities for each mode ( $k = 1, 2, 3$ ), and each node belongs to one community.
- Entry  $i_1, i_2, i_3$  of the signal tensor is determined by community memberships of nodes  $i_1, i_2$ , and  $i_3$ .

### Example I:

- Multilayer network with airports, airports, and airlines as the modes.
- Each airport belongs to one of  $r_1$  communities ( $r_1 = r_2$ ).
- Each airline belongs to one of  $r_3$  communities.
- $i_1, i_2, i_3$  entry of the signal tensor is governed by mean associated to memberships of airport  $i_1$ , airport  $i_2$ , and airline  $i_3$ .

# Tensor Blockmodels

## Interpretation

- There are  $r_k$  communities for each mode ( $k = 1, 2, 3$ ), and each node belongs to one community.
- Entry  $i_1, i_2, i_3$  of the signal tensor is determined by community memberships of nodes  $i_1, i_2$ , and  $i_3$ .

Example II:

- Multilayer network with airports, airports, and time as the modes.

# Tensor Blockmodels

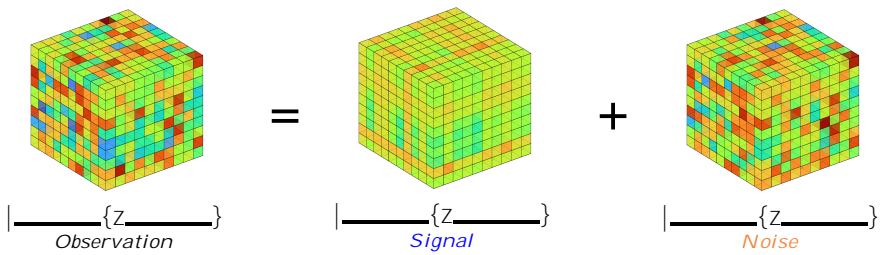
## Interpretation

- There are  $r_k$  communities for each mode ( $k = 1, 2, 3$ ), and each node belongs to one community.
- Entry  $i_1, i_2, i_3$  of the signal tensor is determined by community memberships of nodes  $i_1, i_2$ , and  $i_3$ .

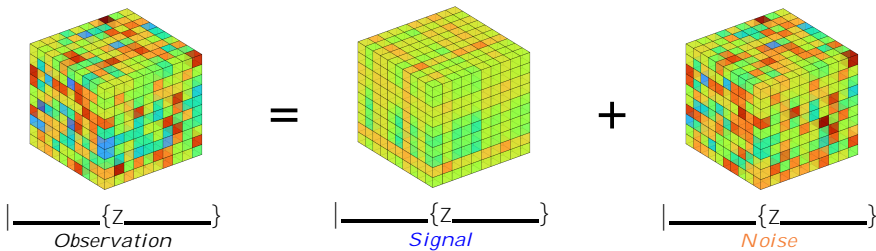
### Example II:

- Multilayer network with airports, airports, and time as the modes.
- Each airport belongs to one of  $r_1$  communities ( $r_1 = r_2$ ).
- Each *time* belongs to one of  $r_3$  communities.
- $i_1, i_2, i_3$  entry of the signal tensor is governed by mean associated to memberships of airport  $i_1$ , airport  $i_2$ , and time  $i_3$ .

# Tensor Mixed-Membership Blockmodels



# Tensor Mixed-Membership Blockmodels



## Interpretation

- There are  $r_k$  communities for each mode and each node belongs to a convex combination of communities.
- Entry  $i_1, i_2, i_3$  of the signal tensor is determined by community membership vectors of nodes  $i_1, i_2, i_3$ .

# Tensor Mixed-Membership Blockmodels

## Interpretation

- There are  $r_k$  communities for each mode and each node belongs to a convex combination of communities.
- Entry  $i_1, i_2, i_3$  of the signal tensor is determined by community membership vectors of nodes  $i_1, i_2, i_3$ .

### Example I:

- Multilayer network with airports, airports, and airlines as the modes.

# Tensor Mixed-Membership Blockmodels

## Interpretation

- There are  $r_k$  communities *for each mode* and each node belongs to *a convex combination of communities*.
- Entry  $i_1, i_2, i_3$  of the signal tensor is determined by community membership vectors of nodes  $i_1, i_2, i_3$ .

### Example I:

- Multilayer network with airports, airports, and airlines as the modes.
- Each airport belongs to convex combination of  $r_1$  communities ( $r_1 = r_2$ ).



# Tensor Mixed-Membership Blockmodels

## Interpretation

- There are  $r_k$  communities for each mode and each node belongs to a convex combination of communities.
- Entry  $i_1, i_2, i_3$  of the signal tensor is determined by community membership vectors of nodes  $i_1, i_2, i_3$ .

### Example I:

- Multilayer network with airports, airports, and airlines as the modes.
- Each airport belongs to convex combination of  $r_1$  communities ( $r_1 = r_2$ ).
- Each *airline* belongs to convex combination of  $r_3$  communities.

# Tensor Mixed-Membership Blockmodels

## Interpretation

- There are  $r_k$  communities for each mode and each node belongs to a convex combination of communities.
- Entry  $i_1, i_2, i_3$  of the signal tensor is determined by community membership vectors of nodes  $i_1, i_2, i_3$ .

### Example I:

- Multilayer network with airports, airports, and airlines as the modes.
- Each airport belongs to convex combination of  $r_1$  communities ( $r_1 = r_2$ ).
- Each *airline* belongs to convex combination of  $r_3$  communities.
- $i_1, i_2, i_3$  entry of the signal tensor is governed by weighted sum of community means associated to airport  $i_1$ , airport  $i_2$ , and airline  $i_3$ .

# Tensor Mixed-Membership Blockmodels

## Interpretation

- There are  $r_k$  communities *for each mode* and each node belongs to *a convex combination of communities*.
- Entry  $i_1, i_2, i_3$  of the signal tensor is determined by community membership vectors of nodes  $i_1, i_2, i_3$ .

### Example II:

- Multilayer network with airports, airports, and time as the modes.
- Each airport belongs to convex combination of  $r_1$  communities ( $r_1 = r_2$ ).
- Each *time* belongs to convex combination of  $r_3$  communities.
- $i_1, i_2, i_3$  entry of the signal tensor is governed by weighted sum of community means.

# Tensor Mixed-Membership Blockmodels

$$\mathcal{T}_{i_1 i_2 i_3} = \sum_{l_1=1}^{X_1} \sum_{l_2=1}^{X_2} \sum_{l_3=1}^{X_3} \mathcal{S}_{l_1 l_2 l_3} \begin{matrix} 1 & 2 & 3 \\ i_1 l_1 & i_2 l_2 & i_3 l_3 \end{matrix} \quad 1 \leq i_k \leq p_k;$$

$$\mathcal{T} = \mathcal{S} \times_1 \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} \times_2 \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} \times_3 \begin{matrix} 1 \\ 2 \\ 3 \end{matrix};$$

$\mathcal{S} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$  is a *Mean Tensor*;

$k \in [0, 1]^{p_k \times r_k}$  are *Mixed – Membership Matrices* ( $\|k_j\|_1 = 1$ )

# Tensor Mixed-Membership Blockmodels

$$\mathcal{T}_{i_1 i_2 i_3} = \sum_{l_1=1}^{r_1} \sum_{l_2=1}^{r_2} \sum_{l_3=1}^{r_3} \mathcal{S}_{l_1 l_2 l_3} \quad 1 \leq i_k \leq p_k;$$

$$\mathcal{T} = \mathcal{S} \times_1 \mathbf{X}_1 \times_2 \mathbf{X}_2 \times_3 \mathbf{X}_3;$$

$\mathcal{S} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$  is a *Mean Tensor*;

$\mathbf{X}_k \in [0, 1]^{p_k \times r_k}$  are *Mixed – Membership Matrices* ( $\|\mathbf{X}_k\|_1 = 1$ )

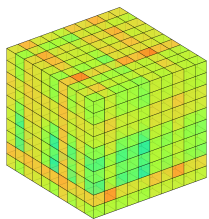
## Interpretation

- There are  $r_k$  communities for each mode.
- $\mathcal{S} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$  encodes the community means.
- $i_k$ 'th row of  $\mathbf{X}_k \in [0, 1]^{p_k \times r_k}$  tells us how much node  $i_k$  belongs to each community.

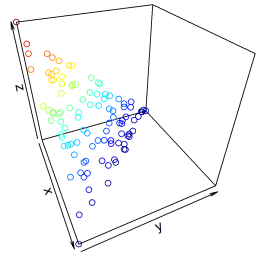
# Tensor Mixed-Membership Blockmodels

## Interpretation

- There are  $r_k$  communities for each mode.
- $\mathcal{S} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$  encodes the community means.
- $i_k$ 'th row of  $\mathbf{S}_k \in [0, 1]^{p_k \times r_k}$  tells us how much node  $i_k$  belongs to each community.



Extract  $\xrightarrow{1}$



Underlying Tensor  $\mathcal{T} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$ .

# Identifiability

- Observation: if  $k \in \{0, 1\}^{p_k}$   $r_k$  then  $\mathcal{T}$  is a *tensor blockmodel*.

# Identifiability

- Observation: if  $\rho_k \in \{0, 1\}^{p_k}$  then  $\mathcal{T}$  is a *tensor blockmodel*.
- If row  $i_k$  of  $\rho_k$  is  $\{0, 1\}$ -valued, we say  $i_k$  is a *pure node*.



# Identifiability

- Observation: if  $i_k \in \{0, 1\}^{p_k}$  then  $\mathcal{T}$  is a *tensor blockmodel*.
- If row  $i_k$  of  $S$  is  $\{0, 1\}$ -valued, we say  $i_k$  is a *pure node*.

## Proposition

*Suppose each mode contains at least one pure node for each community and  $S$  is full rank. Then the model is identifiable up to community relabeling.*

Full rank:  $S$  can be written as matrix in three different ways – each of these is full rank.

# Identifiability

- Observation: if  $\rho_k \in \{0, 1\}^{p_k}$  then  $\mathcal{T}$  is a *tensor blockmodel*.
- If row  $i_k$  of  $\rho_k$  is  $\{0, 1\}$ -valued, we say  $i_k$  is a *pure node*.

## Proposition

Suppose each mode contains at least one pure node for each community and  $S$  is full rank. Then the model is identifiable up to community relabeling.

Full rank:  $S$  can be written as matrix in three different ways – each of these is full rank.

## Goal

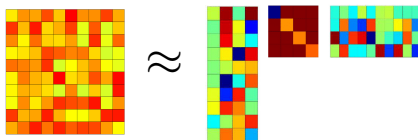
Estimate community membership matrices  $\rho_1, \rho_2,$  and  $\rho_3 \in [0, 1]^{p_k \times r_k}$ .

# Outline

- 1 Motivation
- 2 Community Models
- 3 Estimation Algorithm**
- 4  $\ell_{2,1}$  Tensor Perturbation
- 5 Data Analysis
- 6 Conclusion

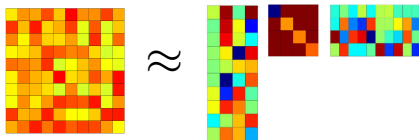
# Brief Detour: Tucker Decompositions and Tensor SVD

- Matrix SVD:



# Brief Detour: Tucker Decompositions and Tensor SVD

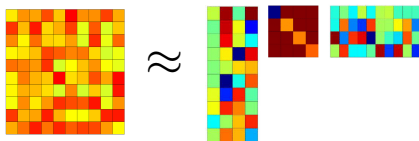
- Matrix SVD:



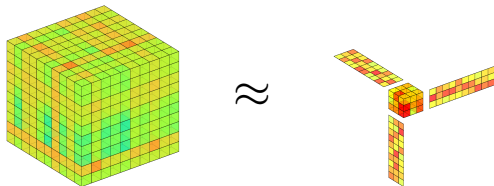
- No unified notion of Tensor SVD!

# Brief Detour: Tucker Decompositions and Tensor SVD

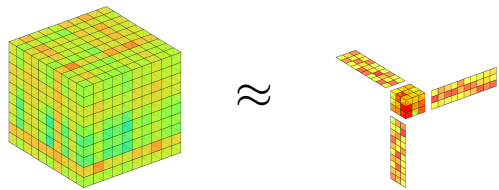
- Matrix SVD:



- No unified notion of Tensor SVD!
- Tensor mixed-membership blockmodel is related to Tucker decomposition



# Brief Detour: Tucker Decomposition and Tensor SVD



## Definition

A tensor  $\mathcal{T}$  is rank  $\mathbf{r} = (r_1, r_2, r_3)$  if  $\mathbf{r}$  is the smallest triplet such that

$$\mathcal{T} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3;$$

$\mathcal{C} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$  is a *core tensor*;

$\mathbf{U}_k \in \mathbb{R}^{p_k \times r_k}$  are orthonormal *loading matrices*.

Note: generalizes matrix SVD to higher-order.

# Spectral Geometry

## Proposition

Suppose  $\mathcal{T} = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$  is a tensor MMBM such that  $\mathcal{S}$  is full rank with a pure node for each community along each mode. Let  $\mathcal{T} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$  be the rank  $(r_1, r_2, r_3)$  Tucker factorization. Then

$$\mathbf{U}_k = \mathbf{U}_k^{(\text{pure})},$$

where  $\mathbf{U}_k^{(\text{pure})} \in \mathbb{R}^{r_k \times r_k}$  consists of the rows of  $\mathbf{U}_k$  corresponding to pure nodes for mode  $k$ .



# Spectral Geometry

## Proposition

Suppose  $\mathcal{T} = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$  is a tensor MMBM such that  $\mathcal{S}$  is full rank with a pure node for each community along each mode. Let  $\mathcal{T} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$  be the rank  $(r_1, r_2, r_3)$  Tucker factorization. Then

$$\mathbf{U}_k = \mathbf{U}_k^{(\text{pure})},$$

where  $\mathbf{U}_k^{(\text{pure})} \in \mathbb{R}^{r_k \times r_k}$  consists of the rows of  $\mathbf{U}_k$  corresponding to pure nodes for mode  $k$ .

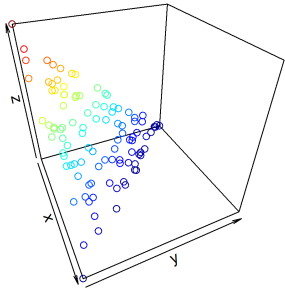
## Geometric Insight

Rows of loading matrices of  $\mathcal{T}$  belong to a simplex with corners determined by pure nodes!

# Spectral Geometry

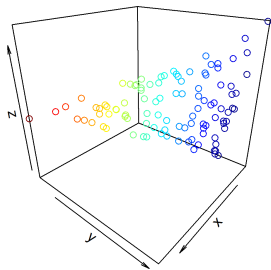
## Geometric Insight

Rows of loading matrices of  $\mathcal{T}$  belong to a simplex with corners determined by pure nodes!



Rows of matrix  $\mathbf{U}_1$ .

$$\xrightarrow{\mathbf{U}_1^{(\text{pure})}}$$



Rows of loading matrix  $\mathbf{U}_1$ .

# Estimation Procedure

## Key Idea

Given an observation  $\mathcal{P} = \mathcal{T} + \text{Noise}$ :

- 1 First estimate the loading matrices  $\mathbf{U}_1$ ,  $\mathbf{U}_2$ , and  $\mathbf{U}_3$ .
- 2 Next estimate pure nodes by finding the corners of the simplex associated to rows of  $\mathbf{U}_k$  (standard corner-finding algorithms exist)
- 3 Estimate  $\mathbf{u}_k$  via plug-in from the equation

$$\mathbf{u}_k = \mathbf{U}_k \mathbf{u}_k^{(\text{pure})} \mathbf{1}.$$

# Estimation Procedure

## Key Idea

Given an observation  $\mathcal{P} = \mathcal{T} + \text{Noise}$ :

- 1 First estimate the loading matrices  $\mathbf{U}_1$ ,  $\mathbf{U}_2$ , and  $\mathbf{U}_3$ .
- 2 Next estimate pure nodes by finding the corners of the simplex associated to rows of  $\mathbf{U}_k$  (standard corner-finding algorithms exist)
- 3 Estimate  $\mathbf{U}_k$  via plug-in from the equation

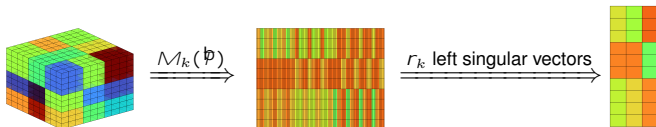
$$\mathbf{U}_k = \mathbf{U}_k \mathbf{U}_k^{(\text{pure})} \mathbf{1}^T.$$

## Problem

Need to estimate the loading matrices!

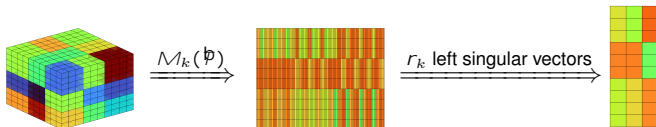
# Higher-Order SVD

- First straightforward guess: use leading  $r_k$  left singular vectors of  $\mathcal{M}_k(\mathcal{P})$



# Higher-Order SVD

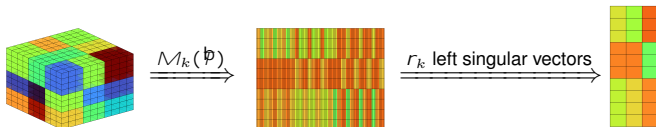
- First straightforward guess: use leading  $r_k$  left singular vectors of  $\mathcal{M}_k(\mathcal{P})$



- Problem: can be suboptimal in high-noise regime!

# Higher-Order SVD

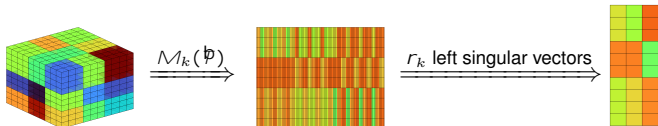
- First straightforward guess: use leading  $r_k$  left singular vectors of  $\mathcal{M}_k(\mathcal{P})$



- Problem: can be suboptimal in high-noise regime!
- Why? Higher-Order SVD *ignores* tensor structure.

# Higher-Order SVD

- First straightforward guess: use leading  $r_k$  left singular vectors of  $\mathcal{M}_k(\mathcal{P})$



- Problem: can be suboptimal in high-noise regime!
- Why? Higher-Order SVD *ignores* tensor structure.

## Solution

*Iteratively refine* the initial estimate using tensor structure!



# Higher-Order Orthogonal Iteration

- Given previous iterate  $\mathbf{U}_k^{(t-1)}$  update  $\mathbf{U}_k^{(t)}$  via

$$\mathbf{U}_1^{(t)} = r_1 \text{ left singular vectors of } \mathcal{M}_1 \times_2 \mathbf{U}_2^{(t-1)} \times_3 \mathbf{U}_3^{(t-1)} ;$$

$$\mathbf{U}_2^{(t)} = r_2 \text{ left singular vectors of } \mathcal{M}_2 \times_3 \mathbf{U}_3^{(t-1)} \times_1 \mathbf{U}_1^{(t)} ;$$

$$\mathbf{U}_3^{(t)} = r_3 \text{ left singular vectors of } \mathcal{M}_3 \times_1 \mathbf{U}_1^{(t)} \times_2 \mathbf{U}_2^{(t)} .$$

# Higher-Order Orthogonal Iteration

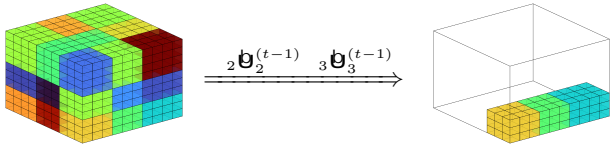
- Given previous iterate  $\mathbf{U}_k^{(t-1)}$  update  $\mathbf{U}_k^{(t)}$  via

$$\mathbf{U}_1^{(t)} = r_1 \text{ left singular vectors of } \mathcal{M}_1 \mathcal{P} \times_2 \mathbf{U}_2^{(t-1)} \times_3 \mathbf{U}_3^{(t-1)} ;$$

$$\mathbf{U}_2^{(t)} = r_2 \text{ left singular vectors of } \mathcal{M}_2 \mathcal{P} \times_3 \mathbf{U}_3^{(t-1)} \times_1 \mathbf{U}_1^{(t)} ;$$

$$\mathbf{U}_3^{(t)} = r_3 \text{ left singular vectors of } \mathcal{M}_3 \mathcal{P} \times_1 \mathbf{U}_1^{(t)} \times_2 \mathbf{U}_2^{(t)} .$$

- Intuition: preserves singular subspace, but greatly reduces noise!

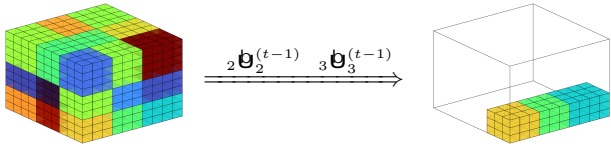


# Higher-Order Orthogonal Iteration

- Given previous iterate  $\mathbf{U}_k^{(t-1)}$  update  $\mathbf{U}_k^{(t)}$  via

$$\begin{aligned} \mathbf{U}_1^{(t)} &= r_1 \text{ left singular vectors of } \mathcal{M}_1 \mathcal{P} \times_2 \mathbf{U}_2^{(t-1)} \times_3 \mathbf{U}_3^{(t-1)} ; \\ \mathbf{U}_2^{(t)} &= r_2 \text{ left singular vectors of } \mathcal{M}_2 \mathcal{P} \times_3 \mathbf{U}_3^{(t-1)} \times_1 \mathbf{U}_1^{(t)} ; \\ \mathbf{U}_3^{(t)} &= r_3 \text{ left singular vectors of } \mathcal{M}_3 \mathcal{P} \times_1 \mathbf{U}_1^{(t)} \times_2 \mathbf{U}_2^{(t)} . \end{aligned}$$

- Intuition: preserves singular subspace, but greatly reduces noise!



- Warm start: use modified spectral initialization to account for heteroskedastic noise.

# Full Estimation Procedure

Given a tensor  $\mathcal{P} = \mathcal{T} + \text{Noise} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$ :

# Full Estimation Procedure

Given a tensor  $\mathcal{P} = \mathcal{T} + \text{Noise} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$ :

- Estimate loading matrices (singular vectors) via HOOI, obtaining estimates  $\{\mathbf{U}_k\}_{k=1}^3$  with  $\mathbf{U}_k \in \mathbb{R}^{p_k \times r_k}$ .
- For  $k = 1, 2, 3$ :

# Full Estimation Procedure

Given a tensor  $\mathcal{T} = \mathcal{T} + \text{Noise} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$ :

- Estimate loading matrices (singular vectors) via HOOI, obtaining estimates  $\{\mathbf{U}_k\}_{k=1}^3$  with  $\mathbf{U}_k \in \mathbb{R}^{p_k \times r_k}$ .
- For  $k = 1, 2, 3$ :
  - Estimate pure nodes via a corner finding algorithm and obtain  $\mathbf{u}_k^{(\text{pure})}$ .

# Full Estimation Procedure

Given a tensor  $\mathcal{T} = \mathcal{T} + \text{Noise} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$ :

- Estimate loading matrices (singular vectors) via HOOI, obtaining estimates  $\{\mathbf{u}_k\}_{k=1}^3$  with  $\mathbf{u}_k \in \mathbb{R}^{p_k \times r_k}$ .
- For  $k = 1, 2, 3$ :
  - Estimate pure nodes via a corner finding algorithm and obtain  $\mathbf{u}_k^{(\text{pure})}$ .
  - Estimate  $\mathbf{b}_k$  via

$$\mathbf{b}_k = \mathbf{u}_k(\mathbf{u}_k^{(\text{pure})})^{-1}.$$

# Theoretical Guarantees

## Theorem (Agterberg and Zhang (2022))

Suppose that:

- (Identifiability) each mode contains at least one pure node for each community and  $S$  is full rank.



# Theoretical Guarantees

## Theorem (Agterberg and Zhang (2022))

Suppose that:

- (Identifiability) each mode contains at least one pure node for each community and  $S$  is full rank.
- (Regime) that  $r_k \asymp r$  and  $p_k \asymp p$  with  $r \leq Cp^{1=2}$

# Theoretical Guarantees

## Theorem (Agterberg and Zhang (2022))

Suppose that:

- (Identifiability) each mode contains at least one pure node for each community and  $S$  is full rank.
- (Regime) that  $r_k \asymp r$  and  $p_k \asymp p$  with  $r \leq Cp^{1=2}$
- (Signal)  $S$  has smallest singular value  $\Delta$  and condition number  $\kappa$ ;

# Theoretical Guarantees

## Theorem (Agterberg and Zhang (2022))

Suppose that:

- (Identifiability) each mode contains at least one pure node for each community and  $S$  is full rank.
- (Regime) that  $r_k \asymp r$  and  $p_k \asymp p$  with  $r \leq Cp^{1=2}$
- (Signal)  $S$  has smallest singular value  $\Delta$  and condition number  $\kappa$ ;
- (Noise) the noise is independent with variance at most  $\sigma^2$ ;

# Theoretical Guarantees

## Theorem (Agterberg and Zhang (2022))

Suppose that:

- (Identifiability) each mode contains at least one pure node for each community and  $S$  is full rank.
- (Regime) that  $r_k \asymp r$  and  $p_k \asymp p$  with  $r \leq Cp^{1=2}$
- (Signal)  $S$  has smallest singular value  $\Delta$  and condition number  $\kappa$ ;
- (Noise) the noise is independent with variance at most  $\sigma^2$ ;
- (SNR)  $\frac{\Delta^2}{\sigma^2} \geq C \frac{r^3 \log(p)}{p^{3/2}}$ .

# Theoretical Guarantees

## Theorem (Agterberg and Zhang (2022))

Suppose that:

- (Identifiability) each mode contains at least one pure node for each community and  $S$  is full rank.
- (Regime) that  $r_k \asymp r$  and  $p_k \asymp p$  with  $r \leq Cp^{1=2}$
- (Signal)  $S$  has smallest singular value  $\Delta$  and condition number  $\kappa$ ;
- (Noise) the noise is independent with variance at most  $\sigma^2$ ;
- (SNR)  $\frac{2}{2} \geq C \frac{2r^3 \log(p)}{p^{3/2}}$ .

Let  $\hat{b}_k \in [0, 1]^{p_k}$  denote the estimated memberships with  $t \asymp \log\left(\frac{r^{3/2}}{p^{1/2}}\right)$  iterations for HOOI. Then there exist permutation matrices  $\{\mathcal{P}_k\}$  such that with probability at least  $1 - p^{-10}$  it holds that

$$\max_{i, p_k} \|\hat{b}_k - \mathcal{P}_k \mathbf{1}_i\| \leq C \frac{\kappa r^{3=2} p^{\overline{\log(p)}}}{(\Delta/\sigma)p}$$

# Comparison to Matrix Setting

- For matrix mixed-membership blockmodel, it was previously shown that

$$\max_i \| \mathbf{b}(\text{Matrix}) - (\text{Matrix}) \mathcal{P}_i \| = \Theta \frac{1}{\text{SNR} \times \sqrt{p}} .$$

# Comparison to Matrix Setting

- For matrix mixed-membership blockmodel, it was previously shown that

$$\max_i \left\| \mathbf{b}^{(\text{Matrix})} - \sum_k (\text{Matrix}) \mathcal{P}_k \right\| = \Theta \frac{1}{\text{SNR} \times \sqrt{p}} .$$

- Using HOOI, our results show that

$$\max_i \left\| \mathbf{b}_k - \sum_k \mathcal{P}_k \right\| = \Theta \frac{1}{\text{SNR} \times p} .$$

# Comparison to Matrix Setting

- For matrix mixed-membership blockmodel, it was previously shown that

$$\max_i \left\| \mathbf{b}^{(\text{Matrix})} - \mathbf{P}_i^{(\text{Matrix})} \right\| = \Theta \frac{1}{\text{SNR} \times \sqrt{p}} .$$

- Using HOOI, our results show that

$$\max_i \left\| \mathbf{b}_k - \mathbf{P}_k \right\| = \Theta \frac{1}{\text{SNR} \times p} .$$

## Key Takeaway

Higher-order structures *improve* estimation guarantees relative to the matrix setting.

Note: higher-order SVD results in row-wise error  $O(1)$ .



# Outline

- 1 Motivation
- 2 Community Models
- 3 Estimation Algorithm
- 4  $\ell_{2,1}$  Tensor Perturbation**
- 5 Data Analysis
- 6 Conclusion

# Proof Idea

- Recall the bound:

$$\max_i \| \mathbf{b}_k - \mathcal{P}_k \mathbf{P}_k \mathbf{b}_i \| = \Theta \frac{1}{\text{SNR} \times p} .$$

# Proof Idea

- Recall the bound:

$$\max_i \| \mathbf{b}_k - \mathcal{P}_k \mathbf{b}_i \| = \Theta \frac{1}{\text{SNR} \times p} .$$

- Since  $\mathbf{b}_k = \mathbf{u}_k + \mathbf{u}_k^{(\text{pure})}$ , we need to ensure that:

# Proof Idea

- Recall the bound:

$$\max_i \| \mathbf{b}_k - \mathcal{P}_k \mathbf{b}_i \| = \Theta \frac{1}{\text{SNR} \times p} .$$

- Since  $\mathbf{b}_k = \mathbf{u}_k + \mathbf{u}_k^{(\text{pure})}$ , we need to ensure that:
  - The pure nodes are found correctly with high probability;

# Proof Idea

- Recall the bound:

$$\max_i \|\mathbf{b}_k - \mathcal{P}_k \mathbf{b}_i\| = \Theta \frac{1}{\text{SNR} \times p}.$$

- Since  $\mathbf{b}_k = \mathbf{U}_k \mathbf{U}_k^{(\text{pure})}{}^T$ , we need to ensure that:
  - The pure nodes are found correctly with high probability;
  - $\mathbf{U}_k$  is sufficiently close to  $\mathbf{U}_k$ .

# Proof Idea

- Recall the bound:

$$\max_i \|\mathbf{b}_k - \mathcal{P}_k \mathbf{b}_i\| = \Theta \frac{1}{\text{SNR} \times p}.$$

- Since  $\mathbf{b}_k = \mathbf{u}_k \mathbf{u}_k^{(\text{pure})T}$ , we need to ensure that:
  - The pure nodes are found correctly with high probability;
  - $\mathbf{u}_k$  is sufficiently close to  $\mathbf{U}_k$ .

Need to show that *each row of*  $\mathbf{u}_k$  *is sufficiently close to*  $\mathbf{U}_k$  *with high probability.*

# Technical Tool: $\ell_{2,\infty}$ Perturbation Bound

## Theorem (Agterberg and Zhang (2022))

Let  $\mathcal{T}$  be a rank  $(r_1, r_2, r_3)$  tensor with Tucker decomposition  $\mathcal{T} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$ . Suppose that

- (Regime) that  $r_k \asymp r$  and  $p_k \asymp p$  with  $r \leq Cp^{1/2}$

# Technical Tool: $\ell_{2,\infty}$ Perturbation Bound

## Theorem (Agterberg and Zhang (2022))

Let  $\mathcal{T}$  be a rank  $(r_1, r_2, r_3)$  tensor with Tucker decomposition  $\mathcal{T} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$ . Suppose that

- (Regime) that  $r_k \asymp r$  and  $p_k \asymp p$  with  $r \leq Cp^{1/2}$
- (Signal)  $\mathcal{C}$  has smallest singular value  $\lambda$  and condition number  $\kappa$ ;



# Technical Tool: $\ell_{2,\infty}$ Perturbation Bound

## Theorem (Agterberg and Zhang (2022))

Let  $\mathcal{T}$  be a rank  $(r_1, r_2, r_3)$  tensor with Tucker decomposition  $\mathcal{T} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$ . Suppose that

- (Regime) that  $r_k \asymp r$  and  $p_k \asymp p$  with  $r \leq Cp^{1/2}$
- (Signal)  $\mathcal{C}$  has smallest singular value  $\lambda$  and condition number  $\kappa$ ;
- (Noise) the noise is independent with variance at most  $\sigma^2$

# Technical Tool: $\ell_{2,\infty}$ Perturbation Bound

## Theorem (Agterberg and Zhang (2022))

Let  $\mathcal{T}$  be a rank  $(r_1, r_2, r_3)$  tensor with Tucker decomposition  $\mathcal{T} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$ . Suppose that

- (Regime) that  $r_k \asymp r$  and  $p_k \asymp p$  with  $r \leq Cp^{1=2}$
- (Signal)  $\mathcal{C}$  has smallest singular value  $\lambda$  and condition number  $\kappa$ ;
- (Noise) the noise is independent with variance at most  $\sigma^2$
- (SNR)  $\lambda/\sigma \geq C\kappa p^{3=4} \sqrt{\log(p)}$ .

# Technical Tool: $\ell_{2,\infty}$ Perturbation Bound

## Theorem (Agterberg and Zhang (2022))

Let  $\mathcal{T}$  be a rank  $(r_1, r_2, r_3)$  tensor with Tucker decomposition  $\mathcal{T} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$ . Suppose that

- (Regime) that  $r_k \asymp r$  and  $p_k \asymp p$  with  $r \leq Cp^{1/2}$
- (Signal)  $\mathcal{C}$  has smallest singular value  $\lambda$  and condition number  $\kappa$ ;
- (Noise) the noise is independent with variance at most  $\sigma^2$
- (SNR)  $\lambda/\sigma \geq C\kappa p^{3/4} \sqrt{\log(p)}$ .

Let  $\mathbf{u}_k^{(t)} \in \mathbb{R}^{p_k}$  denote the estimated loadings (singular vectors) from HOOI with  $t \asymp \log \frac{p}{\epsilon}$ . Then there exists an orthogonal matrix  $\mathbf{W}_k^{(t)}$  such that

$$\max_{1 \leq i \leq p_k} \|\mathbf{u}_k^{(t)} - \mathbf{U}_k \mathbf{W}_k^{(t)}\|_i \leq C \frac{\sqrt{r_k \log(p)}}{\lambda/\sigma}.$$

# Previous Bounds

$$\|\mathbf{g}_k^{(t)} - \mathbf{U}_k \mathbf{W}_k^{(t)}\|_F \leq C \frac{\sqrt{p_k}}{\text{SNR}} \quad (\text{Previous Work})$$

$$\max_i \|\mathbf{g}_k^{(t)} - \mathbf{U}_k \mathbf{W}_k^{(t)}\|_i \leq C \frac{r_k \log(p)}{\text{SNR}} \quad (\text{Our Work})$$

# Previous Bounds

$$\|\mathbf{g}_k^{(t)} - \mathbf{U}_k \mathbf{W}_k^{(t)}\|_F \leq C \frac{\sqrt{p_k}}{\text{SNR}} \quad (\text{Previous Work})$$

$$\max_i \|\mathbf{g}_k^{(t)} - \mathbf{U}_k \mathbf{W}_k^{(t)}\|_i \leq C \frac{r_k \log(p)}{\text{SNR}} \quad (\text{Our Work})$$

⇒ Errors are *spread out* along each row!

Note: result also depends on *spread-outedness* of the tensor.

# Optimality

$$\lambda/\sigma = \text{SNR} \geq C\kappa p^{3=4} \sqrt{\log(p)} \quad (\text{Our Requirement})$$

$$\lambda/\sigma = \text{SNR} \geq Cp^{3=4} \quad (\text{Computational Feasibility})$$

# Optimality

$$\lambda/\sigma = \text{SNR} \geq C\kappa p^{3=4} \sqrt{\frac{p}{\log(p)}} \quad (\text{Our Requirement})$$

$$\lambda/\sigma = \text{SNR} \geq Cp^{3=4} \quad (\text{Computational Feasibility})$$

⇒ Optimal *up to factors of the condition number  $\kappa$  and logarithmic terms.*

# Optimality

$$\lambda/\sigma = \text{SNR} \geq C\kappa p^{3=4} \sqrt{\log(p)} \quad (\text{Our Requirement})$$

$$\lambda/\sigma = \text{SNR} \geq Cp^{3=4} \quad (\text{Computational Feasibility})$$

⇒ Optimal *up to factors of the condition number  $\kappa$  and logarithmic terms.*

⇒ Tensor MMBM SNR condition is essentially optimal.

Note:  $\frac{p^{3/2}}{r^{3/2}}$



# Proof Strategy

- Want to analyze the  $i$ 'th row of the random variable  
 $\mathbf{u}_1^{(t)} - \mathbf{U}_1 \mathbf{W}_1^{(t)}$

# Proof Strategy

- Want to analyze the  $i$ 'th row of the random variable  $\mathbf{U}_1^{(t)} - \mathbf{U}_1 \mathbf{W}_1^{(t)}$
- Show error is nearly linear combination of  $i$ 'th row of the noise tensor and previous iterations.

# Proof Strategy

- Want to analyze the  $i$ 'th row of the random variable  $\mathbf{U}_1 \mathbf{W}_1^{(t)}$
- Show error is nearly linear combination of  $i$ 'th row of the noise tensor and previous iterations.

## Problem

Previous iterations are not *independent* of  $i$ 'th row of the noise tensor.

# Proof Strategy

## Problem

Previous iterations are not *independent* of  $i$ 'th row of the noise tensor.

## Solution

Introduce special *leave-one-out sequences* that:

- Are *independent* from  $i$ 'th row of the noise.
- Are *sufficiently close* to the true sequence  $\mathbf{u}_k^{(t)}$ .

# Proof Strategy

## Problem

Previous iterations are not *independent* of  $i$ 'th row of the noise tensor.

## Solution

Introduce special *leave-one-out sequences* that:

- Are *independent* from  $i$ 'th row of the noise.
- Are *sufficiently close* to the true sequence  $\mathbf{u}_k^{(t)}$ .

Proof tracks *three separate* leave-one-out sequences (one for each mode) and the true sequence simultaneously by leveraging independence between leave-one-out sequences and noise.

# Outline

- 1 Motivation
- 2 Community Models
- 3 Estimation Algorithm
- 4  $\ell_{2,1}$  Tensor Perturbation
- 5 Data Analysis**
- 6 Conclusion

# Analyzing Flight Network Data

- Setting: observe time-series of counts of flights between US airports from January 2016-September 2021
- Results in a tensor  $\mathcal{P} \in \mathbb{R}^{343 \times 343 \times 69}$  (airports  $\times$  airports  $\times$  months)

# Analyzing Flight Network Data

- Setting: observe time-series of counts of flights between US airports from January 2016-September 2021
- Results in a tensor  $\mathcal{P} \in \mathbb{R}^{343 \times 343 \times 69}$  (airports  $\times$  airports  $\times$  months)
- To estimate the number of communities, we use the “elbow method” (look for an elbow in the plot of the singular values). Results in  $\mathbf{b} = (3, 3, 4)$ .



# Analyzing Flight Network Data

- Setting: observe time-series of counts of flights between US airports from January 2016-September 2021
- Results in a tensor  $\mathcal{P} \in \mathbb{R}^{343 \times 343 \times 69}$  (airports  $\times$  airports  $\times$  months)
- To estimate the number of communities, we use the “elbow method” (look for an elbow in the plot of the singular values). Results in  $\mathbf{b} = (3, 3, 4)$ .

# Analyzing Flight Network Data

- Setting: observe time-series of counts of flights between US airports from January 2016-September 2021
- Results in a tensor  $\mathcal{P} \in \mathbb{R}^{343 \times 343 \times 69}$  (airports  $\times$  airports  $\times$  months)
- To estimate the number of communities, we use the “elbow method” (look for an elbow in the plot of the singular values). Results in  $\mathbf{b} = (3, 3, 4)$ .
- Report community membership intensities for each community associated to pure nodes.

# Analyzing Flight Network Data: Airport Mode

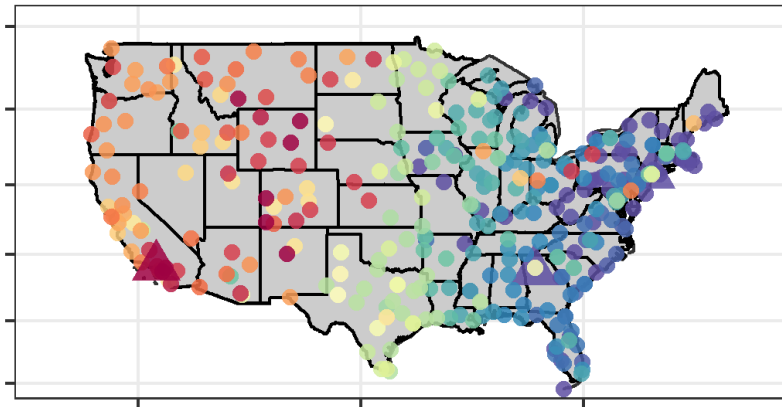


Figure: Community associated to the pure node LAX = Los Angeles. Red means higher membership intensity (closer to 1).

# Analyzing Flight Network Data: Airport Mode

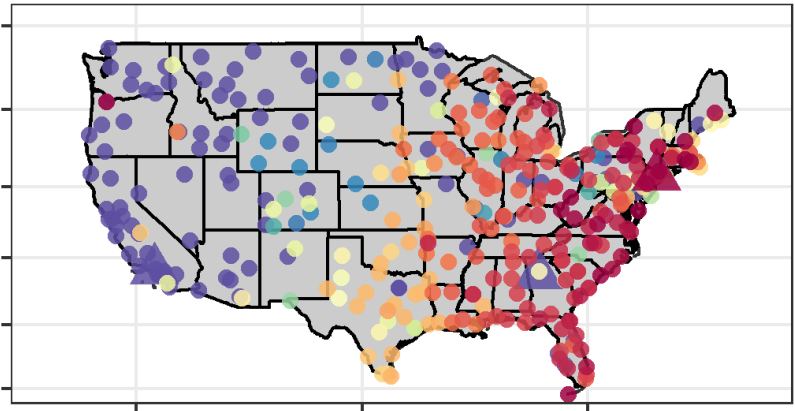


Figure: Community associated to the pure node LGA= New York. Red means higher membership intensity (closer to 1).

# Analyzing Flight Network Data: Airport Mode

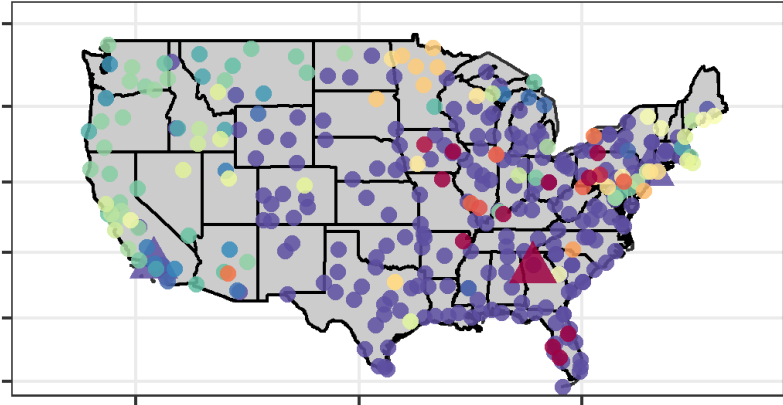
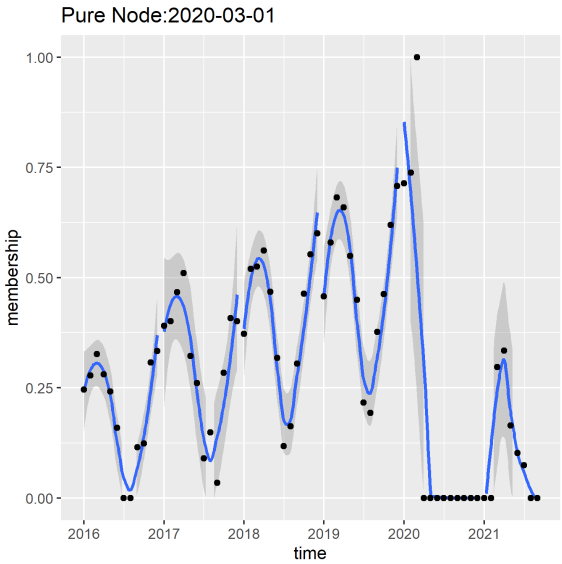
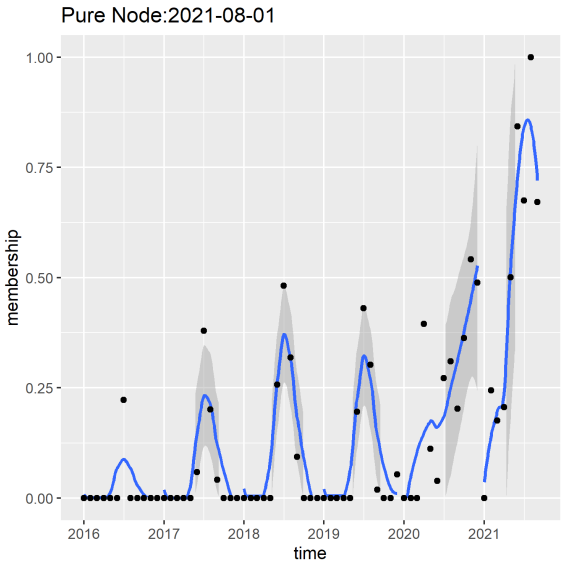


Figure: Community associated to the pure node ATL = Atlanta. Red means higher membership intensity (closer to 1).

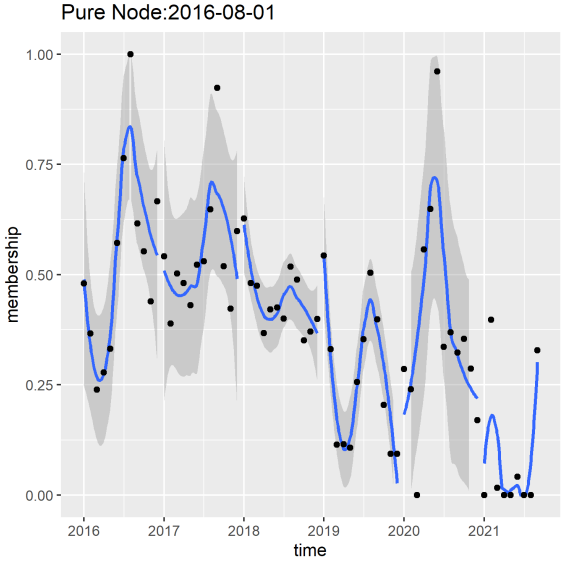
# Analyzing Flight Network Data: Time Mode



# Analyzing Flight Network Data: Time Mode

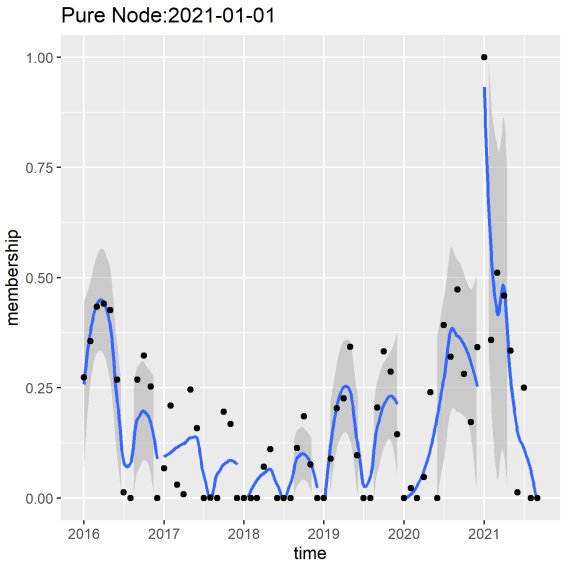


# Analyzing Flight Network Data: Time Mode





# Analyzing Flight Network Data: Time Mode



# Outline

- 1 Motivation
- 2 Community Models
- 3 Estimation Algorithm
- 4  $\ell_{2,1}$  Tensor Perturbation
- 5 Data Analysis
- 6 Conclusion

# Conclusion

## Main Question

Given a noisy high-dimensional  $p_1 \times p_2 \times p_3$  tensor with underlying community structure, can we consistently estimate the communities in the high dimensional regime  $p_1, p_2, p_3 \asymp p$  as  $p \rightarrow \infty$ ?

# Conclusion

## Main Question

Given a noisy high-dimensional  $p_1 \times p_2 \times p_3$  tensor with underlying community structure, can we consistently estimate the communities in the high dimensional regime  $p_1, p_2, p_3 \asymp p$  as  $p \rightarrow \infty$ ?

## Answer

Yes! The maximum row-wise error rate yields an improvement of order  $\sqrt{p}$  for a  $p \times p \times p$  tensor relative to a  $p \times p$  matrix!

# Future and Ongoing Work

- Multilayer networks:
  - Ameliorating degree heterogeneity (Agterberg et al., 2022)
  - More general community models with estimation and testing guarantees with multilayer networks
  - Estimation accuracy in sparse network regimes
  - Network time series

# Future and Ongoing Work

- Multilayer networks:
  - Ameliorating degree heterogeneity (Agterberg et al., 2022)
  - More general community models with estimation and testing guarantees with multilayer networks
  - Estimation accuracy in sparse network regimes
  - Network time series
- Tensor data analysis:
  - Statistical inference for low-rank tensors by building upon  $\ell_{2,1}$  tensor perturbation bound
  - Perturbation bounds for tensor models with other notions of low-rank and applications to other types of tensor data
  - Robustness and sparsity

# Future and Ongoing Work

- Multilayer networks:
  - Ameliorating degree heterogeneity (Agterberg et al., 2022)
  - More general community models with estimation and testing guarantees with multilayer networks
  - Estimation accuracy in sparse network regimes
  - Network time series
- Tensor data analysis:
  - Statistical inference for low-rank tensors by building upon  $\ell_{2,1}$  tensor perturbation bound
  - Perturbation bounds for tensor models with other notions of low-rank and applications to other types of tensor data
  - Robustness and sparsity
- Spectral methods and nonconvex algorithms:
  - Entrywise guarantees for other nonconvex matrix and tensor algorithms under different noise models
  - Inference with the outputs of nonconvex procedures
  - Heterogeneous missingness mechanisms

# References I

Joshua Agterberg and Anru Zhang. Estimating Higher-Order Mixed Memberships via the  $\ell_{2,\infty}$  Tensor Perturbation Bound, December 2022. arXiv:2212.08642 [math, stat].

Joshua Agterberg, Zachary Lubbets, and Jesús Arroyo. Joint Spectral Clustering in Multilayer Degree-Corrected Stochastic Blockmodels, December 2022. arXiv:2212.05053 [math, stat].



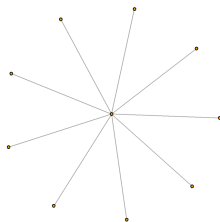
Thank you!

🐦: @JAgterberger

# Correcting For Hubs

## Observation

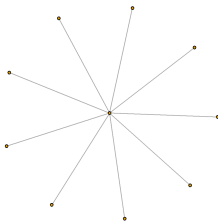
One community in the airport data was a *hub community*.



# Correcting For Hubs

## Observation

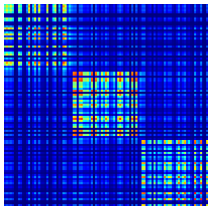
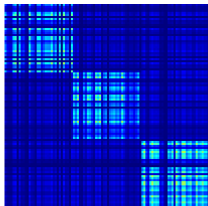
One community in the airport data was a *hub community*.



## Question

Can we still obtain good estimation by *accounting for hubs in the model*?

# Multilayer Degree-Corrected Stochastic Blockmodel



# Multilayer Degree-Corrected Stochastic Blockmodel

## Interpretation

Observe the same communities across the networks, but the means are different and vertices are permitted to differ between and within networks.

# Multilayer Degree-Corrected Stochastic Blockmodel

Each vertex  $i$  belongs to community  $z(i)$ .

# Multilayer Degree-Corrected Stochastic Blockmodel

Each vertex  $i$  belongs to community  $z(i)$ .

Each vertex  $i$  has a layer-dependent degree-correction parameter  $\gamma_i^{(l)}$ .

# Multilayer Degree-Corrected Stochastic Blockmodel

Each vertex  $i$  belongs to community  $z(i)$ .

Each vertex  $i$  has a layer-dependent degree-correction parameter  $\alpha_i^{(l)}$ .

The mean matrix  $P^{(l)}$  satisfies

$$P_{ij}^{(l)} = \underbrace{\alpha_i^{(l)} \alpha_j^{(l)}}_{\text{Degree-Corrections}} \underbrace{B_{z(i)z(j)}^{(l)}}_{\text{Community Mean}} :$$



# Multilayer Degree-Corrected Stochastic Blockmodel

Each vertex  $i$  belongs to community  $z(i)$ .

Each vertex  $i$  has a layer-dependent degree-correction parameter  $\alpha_i^{(l)}$ .

The mean matrix  $P^{(l)}$  satisfies

$$P_{ij}^{(l)} = \alpha_i^{(l)} \frac{B_{z(i)z(j)}^{(l)}}{\sum_{z'} B_{z'}^{(l)}} :$$

Degree-Corrections Community Mean

## Accounting for Hubs

Parameters  $\alpha_i^{(l)}$  allow for high-degree vertices, which allows for hubs.

# Multilayer Degree-Corrected Stochastic Blockmodel

Each vertex  $i$  belongs to community  $z(i)$ .

Each vertex  $i$  has a layer-dependent degree-correction parameter  $\alpha_i^{(l)}$ .

The mean matrix  $P^{(l)}$  satisfies

$$P_{ij}^{(l)} = \alpha_i^{(l)} \frac{1}{|\mathcal{Z}^{(l)}|} \sum_{z \in \mathcal{Z}^{(l)}} B_{z(j)z(j)}^{(l)} :$$

Degree-Corrections Community Mean

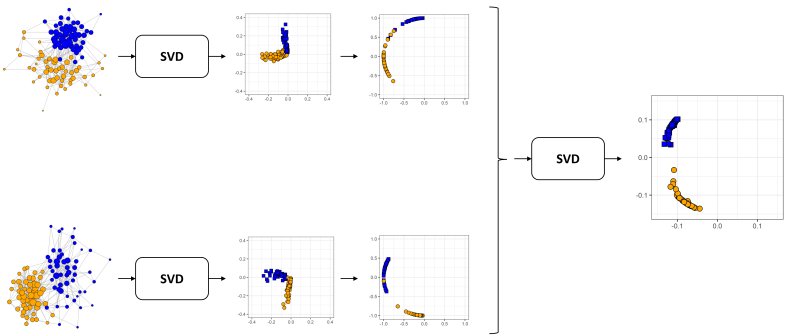
## Accounting for Hubs

Parameters  $\alpha_i^{(l)}$  allow for high-degree vertices, which allows for hubs.

## Goal

Use all  $L$  networks to estimate community memberships.

# DC-MASE: Degree-Corrected Multiple Adjacency Spectral Embedding



# Theoretical Guarantees

Consider a multilayer DCSBM with each network having the same signal strength  $\lambda$ , and suppose each edge probability matrix has rank  $K$ . Let  $\hat{\mathbf{b}}$  denote the output of clustering with  $K$ -means on the rows of the output of DC-MASE, and define

$$\mathcal{E}(\hat{\mathbf{b}}; \mathbf{z}) := \frac{\# \text{ misclustered nodes}}{n}.$$

Then

$$\mathbb{E} \mathcal{E}(\hat{\mathbf{b}}; \mathbf{z}) \leq \frac{2K}{n} \sum_{i=1}^n \exp(-c \lambda_i) \quad \text{SNR-like term}.$$

Note: the same signal strength condition is not required in the main result.

Note: number of misclustered nodes is up to permutation of community labels.

# Improving Estimation?

Has been demonstrated that vanilla spectral clustering (using a slightly different procedure) achieves the error rate:

$$E(\hat{\mathbf{b}}; \mathbf{z}) = \frac{2K}{n} \sum_{i=1}^n \exp(-c_i) \quad \text{SNR-like term} \quad :$$

In contrast, our results show that

$$E(\hat{\mathbf{b}}; \mathbf{z}) = \frac{2K}{n} \sum_{i=1}^n \exp(-c_i L) \quad \text{SNR-like term} \quad :$$

# Improving Estimation?

Has been demonstrated that vanilla spectral clustering (using a slightly different procedure) achieves the error rate:

$$E(\hat{\mathbf{b}}; \mathbf{z}) \leq \frac{2K}{n} \sum_{i=1}^K \exp(-c_i \lambda_i^2) \quad \text{SNR-like term} :$$

In contrast, our results show that

$$E(\hat{\mathbf{b}}; \mathbf{z}) \leq \frac{2K}{n} \sum_{i=1}^K \exp(-cL_i \lambda_i^2) \quad \text{SNR-like term} :$$

## Key Takeaway

Multiple networks improve estimation guarantees (relative to the single network setting).

# Analyzing Flight Network Data

# Multilayer Degree-Corrected Stochastic Blockmodel

Each vertex  $i$  belongs to community  $z(i)$ .



# Multilayer Degree-Corrected Stochastic Blockmodel

Each vertex  $i$  belongs to community  $z(i)$ .

The adjacency matrix  $A^{(l)}$  for layer  $l$  satisfies

$$A_{ij}^{(l)} \sim \text{Bernoulli}(P_{ij}^{(l)}) \quad i, j \text{ independently:}$$

# Multilayer Degree-Corrected Stochastic Blockmodel

Each vertex  $i$  belongs to community  $z(i)$ .

The adjacency matrix  $A^{(l)}$  for layer  $l$  satisfies

$$A_{ij}^{(l)} \sim \text{Bernoulli}(P_{ij}^{(l)}) \quad i, j \text{ independently:}$$

The mean matrix  $P^{(l)}$  satisfies

$$P_{ij}^{(l)} = \underbrace{\frac{1}{d_i} \frac{1}{d_j}}_{\text{Degree-Corrections}} \underbrace{B_{z(i)z(j)}^{(l)}}_{\text{Community Mean}} :$$

# Multilayer Degree-Corrected Stochastic Blockmodel

Each vertex  $i$  belongs to community  $z(i)$ .

The adjacency matrix  $A^{(l)}$  for layer  $l$  satisfies

$$A_{ij}^{(l)} \sim \text{Bernoulli}(P_{ij}^{(l)}) \quad i, j \text{ independently:}$$

The mean matrix  $P^{(l)}$  satisfies

$$P_{ij}^{(l)} = \underbrace{\frac{1}{d_i} \frac{1}{d_j}}_{\text{Degree-Corrections}} \underbrace{B_{z(i)z(j)}}_{\text{Community Mean}} :$$

## Goal

Use all  $L$  networks to estimate community memberships.

# Spectral Geometry

## Observation 1

Each population network is rank  $K$ , with rows of scaled eigenvectors supported on one of  $K$  rays, where  $K$  is the number of communities.

Population adjacency matrix

Rows of scaled eigenvectors of population adjacency matrix, viewed as points in dimension  $K = 3$

# Spectral Geometry

## Observation 2

Projecting each ray to the sphere results in community memberships for a single network.

Rows of scaled eigenvectors of population adjacency matrix, viewed as points in dimension  $K = 3$

Row-normalized scaled eigenvectors of population adjacency matrix, viewed as points in dimension  $K = 3$

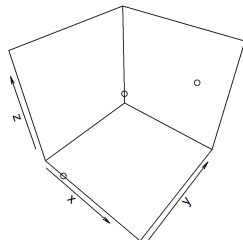
# Spectral Geometry

## Observation 3

$n \times LK$  matrix of concatenated row-normalized embedding has left singular subspace that reveals community memberships for all networks.



$n \times LK$  matrix of concatenated row-normalized embedding.



Rows of left singular vectors viewed as points in dimension  $K = 3$