

# Multi-Way Overlapping Clustering by Bayesian Tensor Decomposition

Zhuofan Wang<sup>1</sup> Fangting Zhou Kejun He Yang Ni

<sup>1</sup>Institute of Statistics and Big Data  
Renmin University of China

July 16, 2024

# Overview

---

1. Background
2. Method
3. Bayesian Inference
4. Simulation
5. Application

# **1. Background**

2. Method

3. Bayesian Inference

4. Simulation

5. Application

# Tensor

---

- Tensors are multi-dimensional arrays.
- $\mathcal{X} \in \mathbb{R}^{p_1 \times p_2 \times \dots \times p_k}$  is an order-k tensor.

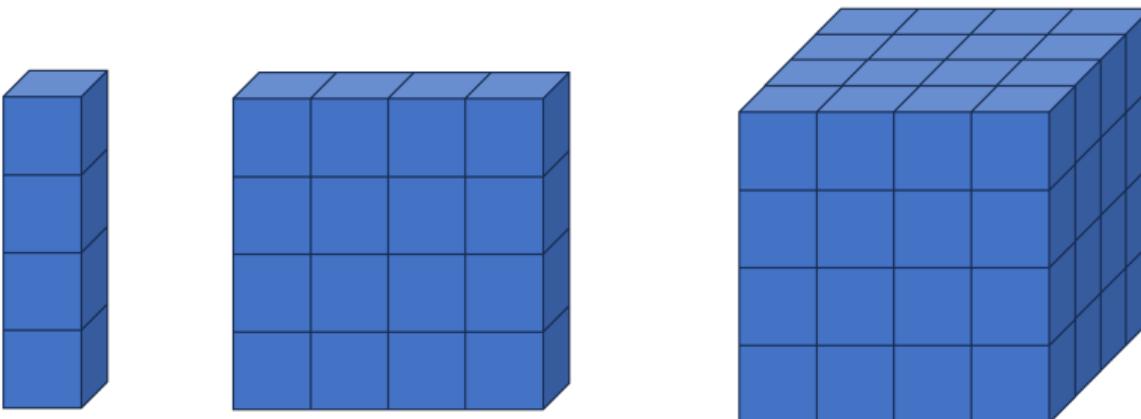


Figure 1: From left to right: order-1 tensor (vector), order-2 tensor(matrix) , order-3 tensor .

# Motivation

The Genotype-Tissue Expression (**GTEx**) project [Melé et al., 2015] was established to characterize genetic effects on the transcriptome across human tissues.

- It collects 17,382 RNA-sequencing samples from 54 **tissues** of 948 post-mortem **donors**.
- The data can be seen as an order-3 tensor: Gene  $\times$  Tissue  $\times$  Donor.
- **Heterogeneity** across tissues and donors.

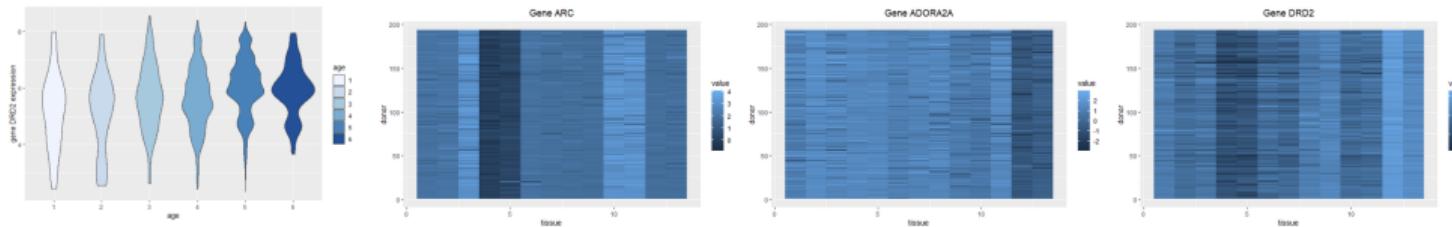
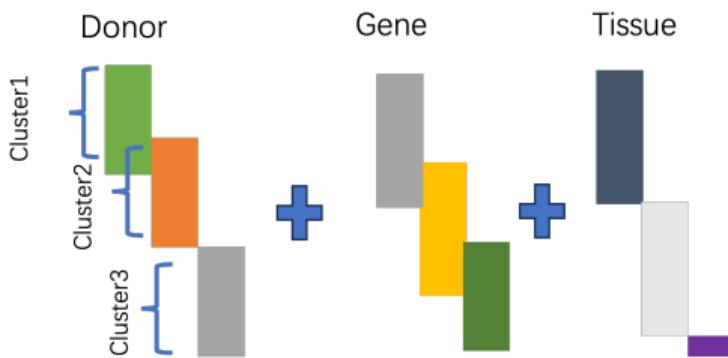
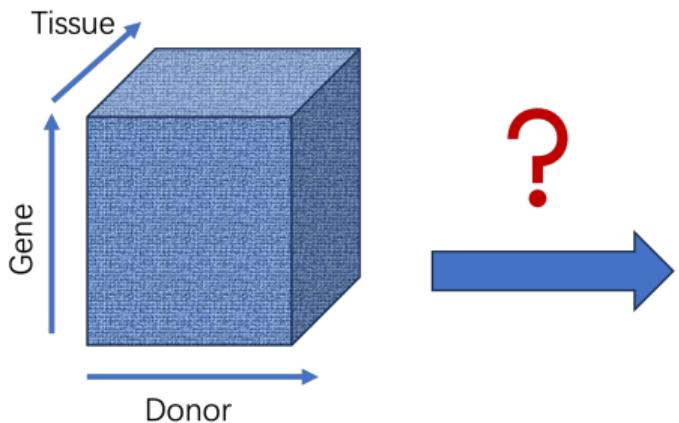


Figure 2: Heterogeneity across donors grouped by age (left); Heterogeneity across tissues (right).

# Motivation (Cont'd)

How to find the underlying clusters on each mode?



# Literature review I

---

- **Biclustering [Hartigan, 1972]**

jointly cluster both features and observations .

- Cheng and Church [2000]: early work to apply biclustering on gene expression data;
- Lazzeroni and Owen [2002] a plaid model;
- Lee et al. [2010] regularized singular value decomposition (SVD);
- Zhou et al. [2022b] Bayesian multinomial matrix factorization; Zhou et al. [2022a] multi-omic data.
- Only applicable to matrices.

# Literature review II

---

- **Multi-way clustering**

- Tucker decomposition-based [Tucker, 1966]: Tensor block model [Wang and Zeng, 2019]; Higher-order Lloyd algorithm [HLloyd, Han et al., 2020]; A convex surrogate [Chi et al., 2020].
- CANDECOMP/PARAFAC (CP) decomposition-based [Carroll and Chang, 1970]: A semi-nonnegative CP decomposition [MultiCluster, Wang et al., 2019].

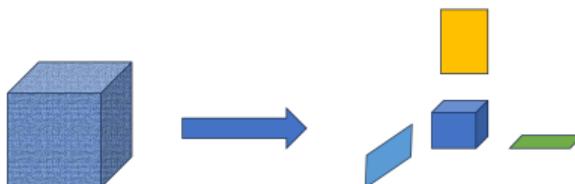


Figure 3: Tucker decomposition

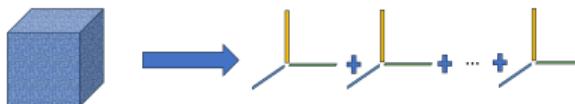


Figure 4: CP decomposition

# Our contribution

---

- Limitations.
  - Do not allow overlapping clusters.
  - Pre-specify the number of clusters.
  - Unable to capture over/under-express information.
- Our contributions.
  - Allow overlapping: an element can be in more than one cluster, or in none at all.
  - Automatically identify the number of clusters.
  - Characterize over/under-expression.

1. Background

## 2. Method

3. Bayesian Inference

4. Simulation

5. Application

# Methodology I: Trichotomize

---

We propose a Bayesian multi-way clustering (BayMC) approach.

- Suppose the RNA-seq data is collected from  $G$  genes across  $T$  tissues and  $D$  donors. Denote the mRNA measurement from the  $g$ -th gene,  $t$ -th tissue and  $d$ -th donor as  $y_{dtg}$  (log-transformed).
- According to Parmigiani et al. [2002], we use probability of expression model (POE):

$$\begin{aligned}y_{dtg} \sim & I(z_{dtg} = -1)U(\mu_t + \mu_g - k_g^-, \mu_t + \mu_g) \\& + I(z_{dtg} = 0)N(\mu_t + \mu_g, \sigma_g^2) \\& + I(z_{dtg} = 1)U(\mu_t + \mu_g, \mu_t + \mu_g + k_g^+)\end{aligned}$$

- The latent indicator  $z_{dtg} = -1, 0$ , and  $1$  represent the case of under, normal, and over-expression of gene  $g$  respectively.

## Methodology II: Decomposition

- To reduce the dimensionality and achieve the purpose of clustering, we introduce lower-dimensional matrices  $\mathbf{C}_1 \in \{0, 1\}^{D \times R}$ ,  $\mathbf{C}_2 \in \{0, 1\}^{T \times R}$ , and  $\mathbf{C}_3 \in \{-1, 0, 1\}^{G \times R}$ .  $R$  is the number of clusters.  $c_{ij}^r$  denotes the  $j$ -th row and  $r$ -th column in  $\mathbf{C}_i$ .
- We link  $z_{dtg}$  with  $\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3$  by a latent multi-class logistic model:

$$z_{dtg} \sim \text{Categorical} \left\{ M^{-1} \exp(\theta_{dtg}^-), M^{-1}, M^{-1} \exp(\theta_{dtg}^+) \right\},$$

where  $M$  is a normalizing constant

- Let  $\omega_{3g}^{r-} (\omega_{3g}^{r+}) > 0$  denote the weight parameter,

$$\theta_{dtg}^- = \sum_{r=1}^R c_{1d}^r c_{2t}^r \omega_{3g}^{r-} I(c_{3g}^r = -1) + b^-, \quad (1)$$

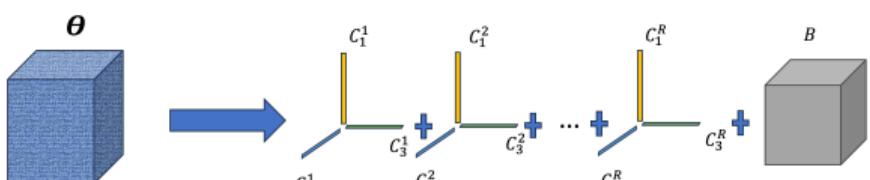
$$\theta_{dtg}^+ = \sum_{r=1}^R c_{1d}^r c_{2t}^r \omega_{3g}^{r+} I(c_{3g}^r = 1) + b^+ \quad (2)$$

## Methodology II: Decomposition

Set  $\tilde{\mathbf{C}}_3^+ = \omega^+ * I(\mathbf{C}_3 = 1)$  and  $\tilde{\mathbf{C}}_3^- = \omega^- * I(\mathbf{C}_3 = -1)$  as the Hadamard product of  $\omega^+$  and  $\omega^-$  respectively with  $I(\mathbf{C}_3 = 1)$  and  $I(\mathbf{C}_3 = -1)$ . Denote the  $r$ -th column of  $\mathbf{C}_i$  as  $\mathbf{c}_i^r$ . Then (1)(2) can be denoted as

$$\Theta^- = \sum_{r=1}^R \mathbf{c}_1^r \circ \mathbf{c}_2^r \circ \tilde{\mathbf{c}}_3^{r-} + \mathbf{B}^- \quad (3)$$

$$\Theta^+ = \sum_{r=1}^R \mathbf{c}_1^r \circ \mathbf{c}_2^r \circ \tilde{\mathbf{c}}_3^{r+} + \mathbf{B}^+ \quad (4)$$



The proposed model (3)(4) coincides with the CP (CANDECOMP/PARAFAC) decomposition for its simplicity of representation and meaningful interpretation of clustering.

# Identifiability

Unlike the matrix decomposition, the condition to achieve uniqueness of CP decomposition is weaker. We present two sufficient conditions.

- Kruskal's condition [Kruskal, 1977] :

$$k_{\mathbf{C}_1} + k_{\mathbf{C}_2} + \min\{k_{\tilde{\mathbf{C}}_3^+}, k_{\tilde{\mathbf{C}}_3^-}\} \geq 2R + 2.$$

$k_{\mathbf{C}_1}$ ,  $k_{\mathbf{C}_2}$ ,  $k_{\tilde{\mathbf{C}}_3^+}$ , and  $k_{\tilde{\mathbf{C}}_3^-}$  denote the  $k$ -ranks of matrices  $\mathbf{C}_1$ ,  $\mathbf{C}_2$ ,  $\tilde{\mathbf{C}}_3^+$ , and  $\tilde{\mathbf{C}}_3^-$ , respectively.

- Our proposition (condition for binary case)

## Proposition

For model (3), if there exist integer matrices  $\mathbf{U}_1 \in \mathbb{Z}^{R \times D}$  and  $\mathbf{U}_2 \in \mathbb{Z}^{R \times T}$  such that  $\mathbf{U}_1 \mathbf{C}_1 = \mathbf{I}_R$  and  $\mathbf{U}_2 \mathbf{C}_2 = \mathbf{I}_R$ , then  $\mathbf{C}_1$ ,  $\mathbf{C}_2$ , and  $\mathbf{C}_3$  are uniquely identifiable up to column permutation.

# Identifiability

---

## Remark

The condition used in Proposition 1 is mild since it can be satisfied, taking  $\mathbf{U}_1 \mathbf{C}_1 = \mathbf{I}_R$  as an example, if for any  $r = 1, \dots, R$ , there exists  $d = 1, \dots, D$  such that  $\mathbf{c}_d = \mathbf{e}_r$  where  $\mathbf{c}_d$  is the  $d$ -th row of  $\mathbf{C}_1$  and  $\mathbf{e}_r$  is a unit vector with 1 at its  $r$ -th entry (in this case,  $\mathbf{U}_1$  is just a binary matrix that acts to pick out those  $R$  rows of  $\mathbf{C}_1$ ). This implies that, the condition can be satisfied if for any cluster  $r$ , there exists at least one member of this cluster that does not belong to any other clusters.

1. Background

2. Method

**3. Bayesian Inference**

4. Simulation

5. Application

# Indian Buffet Process [IBP, Griffiths and Ghahramani, 2005]

---

- Bayesian nonparametric prior on binary matrices.
- Finite number of rows and potentially an unbounded number of columns.
- Generative process ( $m$  is a parameter)
  - For the first row, select the first  $\text{Poisson}(m)$  entries to be 1.
  - For the  $i$ -th row,  $i \geq 2$ ,
    - Let  $m_r$  be the column sum of the  $r$ -th column from the current matrix with  $i - 1$  rows.  
For  $m_r > 0$ , set the  $r$ -th entry of the  $i$ -th row to 1 with probability  $m_r/i$ .
    - Additionally set the next  $\text{Poisson}(m/i)$  number of entries to be 1.

## IBP (Cont'd)

---

- An alternative (but equivalent) generative process: First assume the binary matrix  $\mathbf{A} = [\alpha_{ir}]$  to be generated has  $n$  rows and  $\tilde{R}$  columns. Conditional on  $\tilde{R}$ ,  $\alpha_{ir} | \pi_r \stackrel{\text{ind}}{\sim} \text{Ber}(\pi_r)$  and  $\pi_r \sim \text{Beta}(m/\tilde{R}, 1)$ ,  $r = 1, \dots, \tilde{R}$ .
- The probability mass function for  $\mathbf{A}$ :

$$p(\mathbf{A}) = \prod_{r=1}^{\tilde{R}} \frac{m\Gamma(s_r + m/\tilde{R})\Gamma(n - s_r + 1)}{\tilde{R}\Gamma(n + 1 + m/\tilde{R})}, \quad (5)$$

where  $s_r = \sum_{i=1}^n \alpha_{ir}$  is the sum of the  $r$ -th column of  $\mathbf{A}$ .

- Let  $\tilde{R} \rightarrow \infty$  and remove the columns where all the entries are zeros.

$$p(\mathbf{A}) = \frac{m^R \exp(-mH_n)}{R!} \prod_{r=1}^R \frac{\Gamma(s_r)\Gamma(n - s_r + 1)}{\Gamma(n + 1)},$$

where  $H_n = \sum_{i=1}^n 1/i$  is the  $n$ -th Harmonic number.

# Prior

---

- Each element of  $\mathbf{C}_1$  is assumed to follow a beta-Bernoulli distribution, i.e.,  
 $c_{1d}^r \stackrel{\text{ind}}{\sim} \text{Ber}(\rho)$  with  $\rho \sim \text{Beta}(a_\rho, b_\rho)$ ;
- Matrix  $\mathbf{C}_2$  follows an IBP,  $\mathbf{C}_2 \sim \text{IBP}(m)$ , which automatically determines the number  $R$  of clusters;
- Each element of  $\mathbf{C}_3$  follows the Dirichlet-categorical distribution, i.e.,  $c_{3g}^r \stackrel{\text{ind}}{\sim}$  Categorical  $(\gamma)$  with  $\gamma = (\gamma_{-1}, \gamma_0, \gamma_1) \sim \text{Dirichlet}(\psi_{-1}, \psi_0, \psi_1)$ ;
- Assume independently  $\omega_{gr}^+, \omega_{gr}^- \sim \text{Gamma}(a_\omega, b_\omega)$ ;
- $b^+, b^- \sim N(\mu_b, \sigma_b^2)$ ;
- $m \sim \text{Gamma}(a_m, b_m)$ .

# MCMC

---

- **Property of IBP** : The conditional probability for  $\alpha_{ir} = 1$  is  
 $p(\alpha_{ir} = 1 | \alpha_{(-i)r}) = s_{(-i)r}/n$  provided  $s_{(-i)r} > 0$ .
- Update existing (non-empty) columns  $r = 1, \dots, R$  of  $\mathbf{C}_2$ . In particular, we sample the binary  $c_{2t}^r$ ,  $r = 1, \dots, R$ , from the full conditional distribution,

$$p(c_{2t}^r | \cdot) \propto p(c_{2t}^r | \mathbf{c}_{2,-t}^r) p(\mathbf{z}_{(2)t} | c_{2t}^r, \mathbf{C}_{-2}^+, \mathbf{C}_{-2}^-, b^+, b^-).$$

- Propose new clusters. After all existing columns are updated, we propose to add new columns. We draw  $R^* \sim \text{Poi}(m/T)$ . If  $R^* = 0$ , we proceed to the next step. Otherwise, we propose a set of new parameters for the new clusters from their prior distributions. We accept new columns and the associated new parameters with probability  $\min(1, \gamma)$  with

$$\gamma = \frac{p(\mathbf{z}_{(2)t} | \mathbf{c}_{2t}, \mathbf{1}_{R^*}, \mathbf{C}_{-2}^+, \mathbf{C}_{-2}^-, \mathbf{C}_{-2}^{*+}, \mathbf{C}_{-2}^{*-}, b^+, b^-)}{p(\mathbf{z}_{(2)t} | \mathbf{c}_{2t}, \mathbf{C}_{-2}^+, \mathbf{C}_{-2}^-, b^+, b^-)}.$$

# Detailed MCMC procedure I

---

1. Update  $\mathbf{C}_2$  as illustrated in the main manuscript.
2. Update elements in  $\mathbf{C}_1$ . For each  $d$  and  $r$ , update  $c_{1d}^r$  by

$$p(c_{1d}^r | \cdot) \propto p(c_{1d}^r | \rho) \prod_{t=1}^T \prod_{g=1}^G p\left(z_{dtg} | \{c_{1d}^r, c_{2t}^r, c_{3g}^r, w_{gr}^-, w_{gr}^+\}_{r=1}^R, b^-, b^+\right).$$

3. Update elements in  $\mathbf{C}_3$ . For each  $g$  and  $r$ , update  $c_{3g}^r$  by

$$p(c_{3g}^r | \cdot) \propto p(c_{3g}^r | \gamma) \prod_{d=1}^D \prod_{t=1}^T p\left(z_{dtg} | \{c_{1d}^r, c_{2t}^r, c_{3g}^r, w_{gr}^-, w_{gr}^+\}_{r=1}^R, b^-, b^+\right).$$

## Detailed MCMC procedure II

---

4. Update  $\omega^+$  and  $\omega^-$ . We will only illustrate the details of  $\omega^+$  since you can derive a similar procedure for  $\omega^-$ . For each element in  $\omega^+$ , we perform a truncated random walk Metropolis-Hastings procedure by proposing  $w_{gr}^{*+} \sim N(w_{gr}^+, \sigma_w^2) I(w_{gr}^{*+} > 0)$ . We accept it with probability  $\min(\alpha, 1)$  where

$$\alpha = \frac{p(w_{gr}^{*+}) \prod_{d=1}^D \prod_{t=1}^T p(z_{dtg} | \left\{ c_{1d}^r, c_{2t}^r, c_{3g}^r, w_{gr}^-, w_{gr}^{*+} \right\}_{r=1}^R, b^-, b^+) p(w_{gr}^+ | w_{gr}^{*+})}{p(w_{gr}^+) \prod_{d=1}^D \prod_{t=1}^T p(z_{dtg} | \left\{ c_{1d}^r, c_{2t}^r, c_{3g}^r, w_{gr}^-, w_{gr}^+ \right\}_{r=1}^R, b^-, b+) p(w_{gr}^{*+} | w_{gr}^+)}.$$

## Detailed MCMC procedure III

---

5. Update  $b^+$  and  $b^-$ . We also only give the exact update procedure of  $b^+$  by considering a random walk M-H step. Draw  $b^{*+} \sim N(b^+, \sigma_b^2)$  and accept it with probability  $\min(\alpha, 1)$  where

$$\alpha = \frac{p(b^{*+}) \prod_{d=1}^D \prod_{t=1}^T \prod_{g=1}^G p\left(z_{dtg} \mid \left\{c_{1d}^r, c_{2t}^r, c_{3g}^r, w_{gr}^-, w_{gr}^{*+}\right\}_{r=1}^R, b^-, b^{*+}\right)}{p(b^+) \prod_{d=1}^D \prod_{t=1}^T \prod_{g=1}^G p\left(z_{dtg} \mid \left\{c_{1d}^r, c_{2t}^r, c_{3g}^r, w_{gr}^-, w_{gr}^{*+}\right\}_{r=1}^R, b^-, b^+\right)}.$$

6. Update  $\mathbf{Z}$ . We update  $z_{dtg}$  with full conditional probability

$$p(z_{dtg} \mid \cdot) \propto p\left(z_{dtg} \mid \left\{c_{1d}^r, c_{2t}^r, c_{3g}^r, w_{gr}^-, w_{gr}^{*+}\right\}_{r=1}^R, b^+, b^-\right) p\left(y_{dtg} \mid z_{dtg}, \mu_t, \mu_g, \sigma_g^2, \kappa_g^+, \kappa_g^-\right)$$

## Detailed MCMC procedure IV

---

7. Update  $\kappa^+$  and  $\kappa^-$ . We update  $\kappa_g^+$  with M-H step by proposing a new  $\kappa_g^{*+}$  with  $\kappa_g^{*+} \sim \text{Gamma}(a_\kappa, b_\kappa)$ . We accept the newly proposed  $\kappa_g^{*+}$  with probability  $\min(\alpha, 1)$  where

$$\alpha = \frac{p(y_{dtg} | z_{dtg}, \mu_t, \mu_g, \sigma_g^2, \kappa_g^{*+}, \kappa_g^-)}{p(y_{dtg} | z_{dtg}, \mu_t, \mu_g, \sigma_g^2, \kappa_g^+, \kappa_g^-)}.$$

8. Update  $\mu$ ,  $\eta$ , and  $\sigma^2$ . For  $g = 1, 2, \dots, G$ , update  $\sigma_g^2$  with conjugate posterior:

$$\sigma_g^2 | \cdot \sim \text{IG}(\alpha'_\sigma, \beta'_\sigma) I(\sigma_g < \min(\kappa_g^-, \kappa_g^+) / \kappa_0),$$

where  $a'_\sigma = a_\sigma + 0.5 \sum_{dt} I(z_{dtg} = 0)$ ,  $b'_\sigma = b_s + 0.5 \sum_{dt} I(z_{dtg} = 0)(y_{dtg} - \mu_t - \mu_g)^2$ .  
And update each  $\mu_g$  with

$$\mu_g | \cdot \sim N(m'_\mu, \sigma'^2_\mu),$$

## Detailed MCMC procedure V

---

where  $\sigma_\mu'^2 = 1 / \left\{ 1/\sigma_\mu^2 + \sum_{dt} I(z_{dtg} = 0) / \sigma_g^2 \right\}$ , and  
 $m'_\mu = \sigma_\mu'^2 \sum_{dt} I(z_{dtg} = 0) y_{dtg} / \sigma_g^2$ . Update each  $\eta_t$  with

$$\eta_t | \cdot \sim N(m'_\eta, \sigma_\eta'^2),$$

where  $\sigma_\eta'^2 = 1 / \left\{ 1/\sigma_\eta^2 + \sum_{dg} I(z_{dtg} = 0) / \sigma_g^2 \right\}$ , and  
 $m'_\eta = \sigma_\eta'^2 \sum_{dg} I(z_{dtg} = 0) y_{dtg} / \sigma_g^2$ .

9. Update  $m$ . With a gamma prior, we have a conjugate posterior of  
 $m \sim \text{Gamma}(a_m + R, b_m + H_T)$  where  $H_T = \sum_{i=1}^T 1/i$  is the  $T$ -th harmonic number.
10. Update  $\rho$  with  $\rho \sim \text{Beta}(a_\rho + \sum I(\mathbf{C}_1 = 0), b_\rho + \sum I(\mathbf{C}_1 = 1))$ .
11. Update  $\gamma$  with  
 $\gamma \sim \text{Dirichlet}(\psi_{-1} + \sum I(\mathbf{C}_3 = -1), \psi_0 + \sum I(\mathbf{C}_3 = 0), \psi_1 + \sum I(\mathbf{C}_3 = 1))$ .

# Posterior inference

---

- Calculate the maximum a posteriori estimate  $\hat{R}$  of  $R$  from the marginal posterior distribution.
- Conditional on estimated  $\hat{R}$ , find the point estimate of  $\mathbf{C}_2$ : For any matrices  $\mathbf{C}_2$  and  $\tilde{\mathbf{C}}_2 \in \{0, 1\}^{T \times \hat{R}}$ , we define a distance  $d(\mathbf{C}_2, \tilde{\mathbf{C}}_2) = \min_{\pi} D(\mathbf{C}_2, \pi(\tilde{\mathbf{C}}_2))$ , where  $\pi(\tilde{\mathbf{C}}_2)$  denotes a permutation of the columns of  $\tilde{\mathbf{C}}_2$  and  $D(\cdot, \cdot)$  is the Hamming distance between the two matrices. A point estimator  $\hat{\mathbf{C}}_2$  of  $\mathbf{C}_2$  is then obtained as

$$\hat{\mathbf{C}}_2 = \arg \min_{\tilde{\mathbf{C}}_2} \int d(\mathbf{C}_2, \tilde{\mathbf{C}}_2) dp(\mathbf{C}_2 | \text{data}).$$

The integral as well as the optimization can be approximated using the available posterior samples.

- Conditional on  $\hat{\mathbf{C}}_2$ , we continue to run the Markov chain for a while and the point estimates of other parameters are obtained as the posterior means computed from the new Monte Carlo samples.

1. Background

2. Method

3. Bayesian Inference

**4. Simulation**

5. Application

# Simulation I

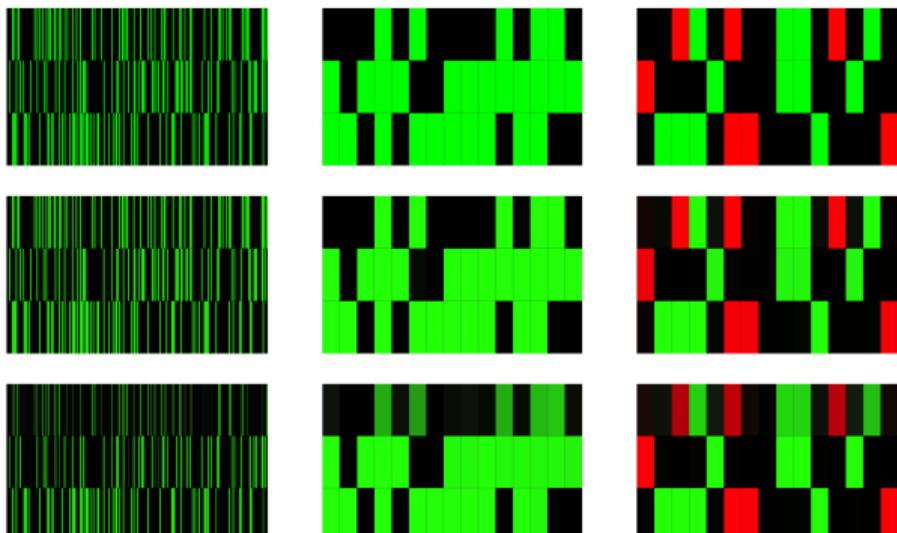
- Compare BayMC with three alternative approaches: (i) a sparse unified matrix factorization [UMF, Zhou et al., 2022a], (ii) the HLlloyd method [Han et al., 2020], and (iii) the MultiCluster method [Wang et al., 2019].
- Consider two settings: overlapped and non-overlapped.

		BayMC	HLlloyd	UMF	MultiCluster
Overlapped	Error of $\mathbf{C}_1$	<b>0.013</b> (0.009)	0.568 (0.010)	-	0.566 (0.014)
	Error of $\mathbf{C}_2$	<b>0.029</b> (0.021)	0.939 (0.019)	1.602 (0.011)	1.040 (0.011)
	Error of $\mathbf{C}_3$	<b>0.088</b> (0.022)	1.154 (0.009)	0.730 (0.033)	0.981 (0.017)
	Error of $\mathbf{C}_1$	0.019 (0.010)	<b>≤0.001</b> (≤0.001)	-	0.052 (0.029)
	Error of $\mathbf{C}_2$	0.011 (0.007)	<b>≤0.001</b> (≤0.001)	0.965 (0.034)	0.013 (0.013)
	Error of $\mathbf{C}_3$	<b>0.067</b> (0.006)	0.893 (0.010)	0.553 (0.031)	0.872 (0.019)

Table 1: Reported are the relative Hamming distances between the estimated and true  $\mathbf{C}_1$ ,  $\mathbf{C}_2$ , and  $\mathbf{C}_3$ .

## Simulation II

- Additional results: Consider 50% missing data.



(a) True  $C_1$  vs. Es-  
timated  $C_1$ .      (b) True  $C_2$  vs. Es-  
timated  $C_2$ .      (c) True  $C_3$  vs. Es-  
timated  $C_3$ .

Figure: Overlapped case for BayMC. The green, black, and red cells represent 1, 0, and  $-1$ , respectively. The first row: true values. The second row: estimations with the complete observations. The third row: average estimations with 50% proportion of missing observations.

**1. Background**

**2. Method**

**3. Bayesian Inference**

**4. Simulation**

**5. Application**

# GTEX RNA-seq dataset

---

- **Quality control.** We use `estimateSizeFactor` (in R package `DESeq2`), which takes into account both sequencing depth and RNA composition, to achieve normalization.
- **Data selection.** We focus on the 13 brain tissues and 15 depression-related genes. These genes are selected according to Atlas of the Developing Human Brain (<http://www.brainspan.org/ish>)

# Results I

- The trace plot of the number of clusters over the course of MCMC indicates 3 clusters.
- The point estimates of the membership matrices of tissues, genes are depicted from left to right in Figure 6, respectively.

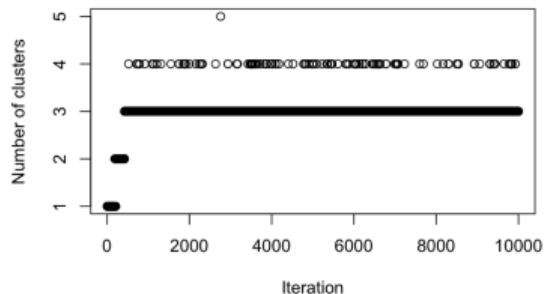


Figure 5: The trace plot of the number of clusters for the proposed BayMC method.

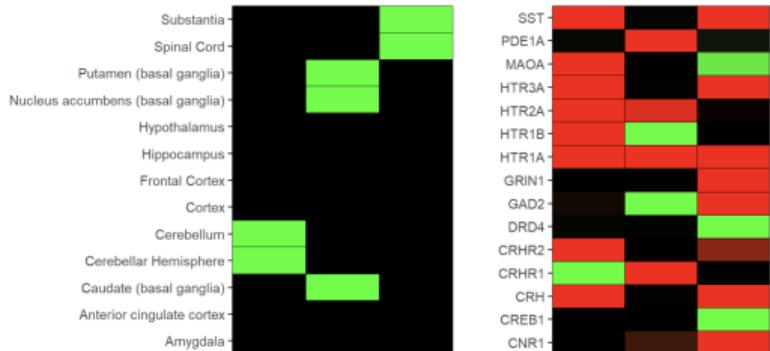


Figure 6: From left to right are the estimated membership matrices of tissues and genes using BayMC.

# Results II

---

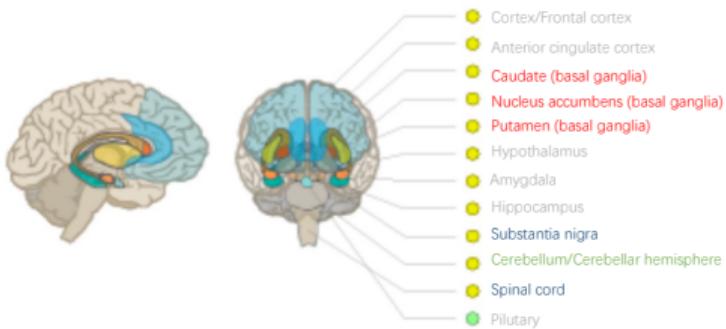


Figure 7: Illustration of brain tissues [Consortium, 2020]. The obtained clustering results can be interpreted by the spatial information in brain for tissues.

Example: Genes underexpressed in cerebellum related cluster are highly correlated to regulation of serotonin secretion (gene ontology (GO) enrichment analyses,  $p = 3.67 \cdot 10^9$  under Bonferroni correction). By conducting hypergeometric test, the results show that females are enriched in this cluster ( $p = 0.076$ ), which may be a factor relevant to the lower incidence of major unipolar depression in males.

# Conclusion

---

- We proposed a novel identifiable multi-way clustering approach for higher-order tensor data.
- With the form of CP decomposition, our model can fully explore the tensor structure, cluster all the modes simultaneously, and characterize the interaction among the modes.
- Using Bayesian hierarchical model and a nonparametric Bayesian prior, our approach can also automatically determine the number of clusters from the posterior samples and allow overlapping clusters.
- Applying the proposed method on GTEx RNA-seq data, we discovered three gene expression modules within brain region.

# References |

---

- J. D. Carroll and J.-J. Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of “eckart-young” decomposition. *Psychometrika*, 35(3):283–319, 1970.
- Y. Cheng and G. M. Church. Biclustering of expression data. In *Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology*, volume 8, pages 93–103, 2000.
- E. C. Chi, B. R. Gaines, W. W. Sun, H. Zhou, and J. Yang. Provable convex co-clustering of tensors. *Journal of Machine Learning Research*, 21(1):8792–8849, 2020.
- G. Consortium. The gtex consortium atlas of genetic regulatory effects across human tissues. *Science*, 369(6509):1318–1330, 2020.
- T. L. Griffiths and Z. Ghahramani. Infinite latent feature models and the indian buffet process. In *Proceedings of the 18th International Conference on Neural Information Processing Systems*, pages 475–482, 2005.
- R. Han, Y. Luo, M. Wang, and A. R. Zhang. Exact clustering in tensor block model: Statistical optimality and computational limit. *arXiv preprint arXiv:2012.09996*, 2020.
- J. A. Hartigan. Direct clustering of a data matrix. *Journal of the American Statistical Association*, 67(337):123–129, 1972.

## References II

---

- J. B. Kruskal. Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and its Applications*, 18(2):95–138, 1977.
- L. Lazzeroni and A. Owen. Plaid models for gene expression data. *Statistica Sinica*, 12(1):61–86, 2002.
- M. Lee, H. Shen, J. Z. Huang, and J. S. Marron. Biclustering via sparse singular value decomposition. *Biometrics*, 66(4):1087–1095, 2010.
- M. Melé, P. G. Ferreira, F. Reverter, D. S. DeLuca, J. Monlong, M. Sammeth, T. R. Young, J. M. Goldmann, D. D. Pervouchine, T. J. Sullivan, R. Johnson, A. V. Segrè, S. Djebali, A. Niarchou, T. G. Consortium, F. A. Wright, T. Lappalainen, M. Calvo, G. Getz, E. T. Dermitzakis, K. G. Ardlie, and R. Guigó. The human transcriptome across tissues and individuals. *Science*, 348(6235):660–665, May 2015. ISSN 0036-8075, 1095-9203.
- G. Parmigiani, E. S. Garrett, R. Anbazhagan, and E. Gabrielson. A statistical framework for expression-based molecular classification in cancer. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):717–736, 2002.
- L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.

## References III

---

- M. Wang and Y. Zeng. Multiway clustering via tensor block models. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pages 715–725, 2019.
- M. Wang, J. Fischer, and Y. S. Song. Three-way clustering of multi-tissue multi-individual gene expression data using semi-nonnegative tensor decomposition. *The Annals of Applied Statistics*, 13(2):1103–1127, 2019.
- F. Zhou, K. He, J. J. Cai, L. A. Davidson, R. S. Chapkin, and Y. Ni. A unified bayesian framework for bi-overlapping-clustering multi-omics data via sparse matrix factorization. *Statistics in Biosciences*, To appear, 2022a.
- F. Zhou, K. He, Q. Li, R. S. Chapkin, and Y. Ni. Bayesian biclustering for microbial metagenomic sequencing data via multinomial matrix factorization. *Biostatistics*, 23(3):891–909, 2022b.

# Thanks!