Mixture of multilayer stochastic block models for multiview clustering

Kylliann De Santiago, Marie Szafranski, Christophe Ambroise

Fréjus - 2023

General framework

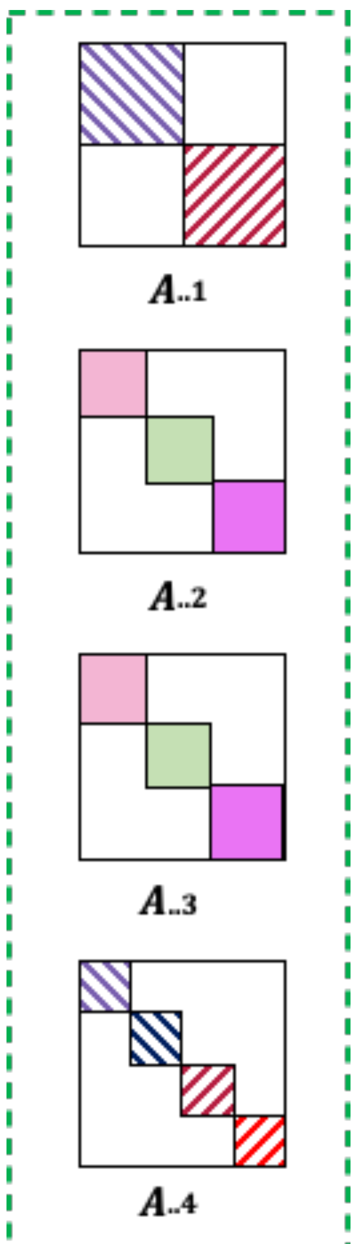
Multi (Layers | Modalities | Views) Learning [2, 3]

Milk

Sugar

Pastry

Coffee



Adjacency matrices

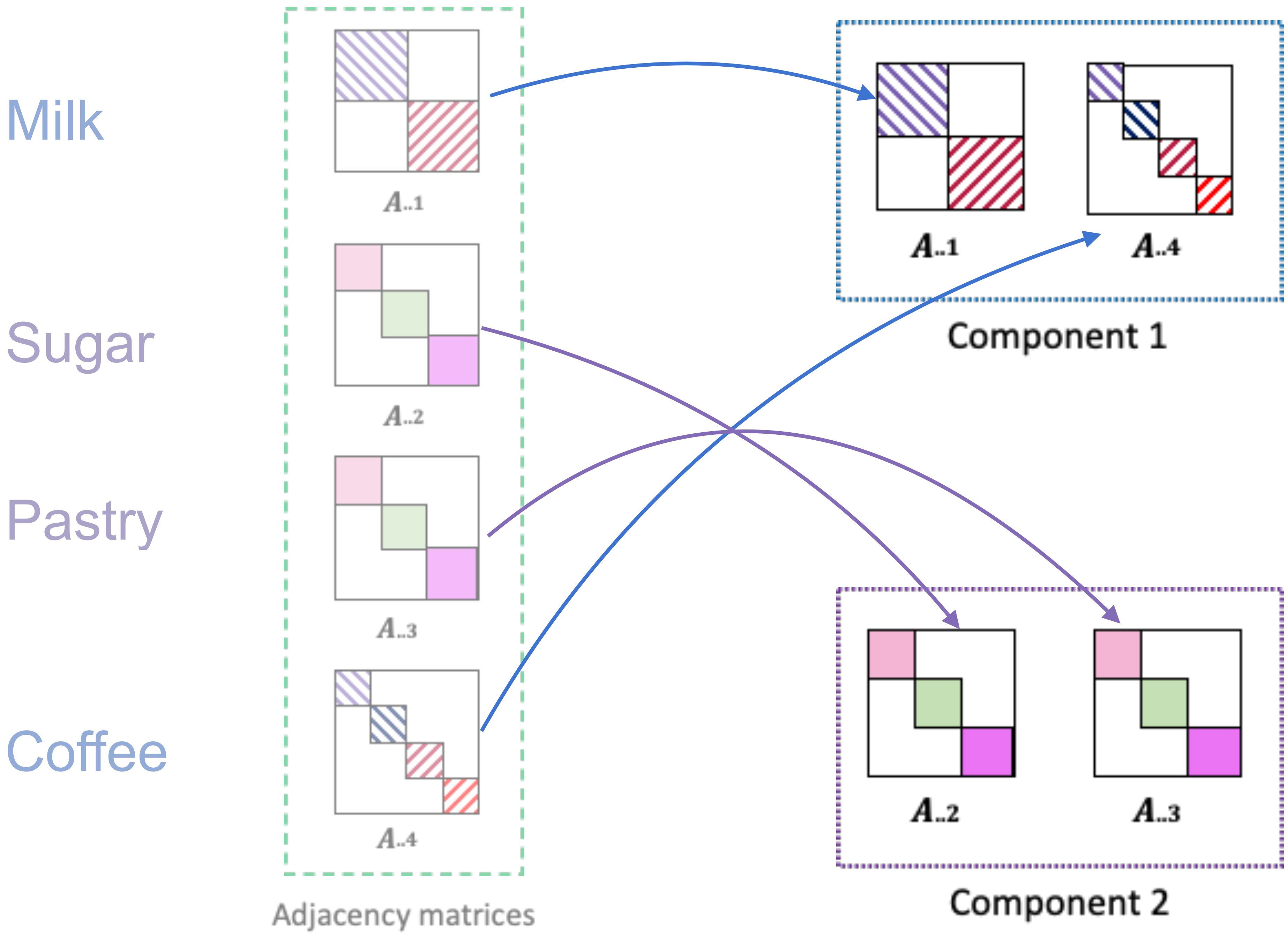
N observations and V views

$$A_{ijv} = \begin{cases} 1, & \text{if observations } (i, j) \\ & \text{ belong to the same cluster,} \\ 0, & \text{otherwise.} \end{cases}$$

Multiple layers, which are adjacency matrices derived from a group of variables used for clustering

General framework

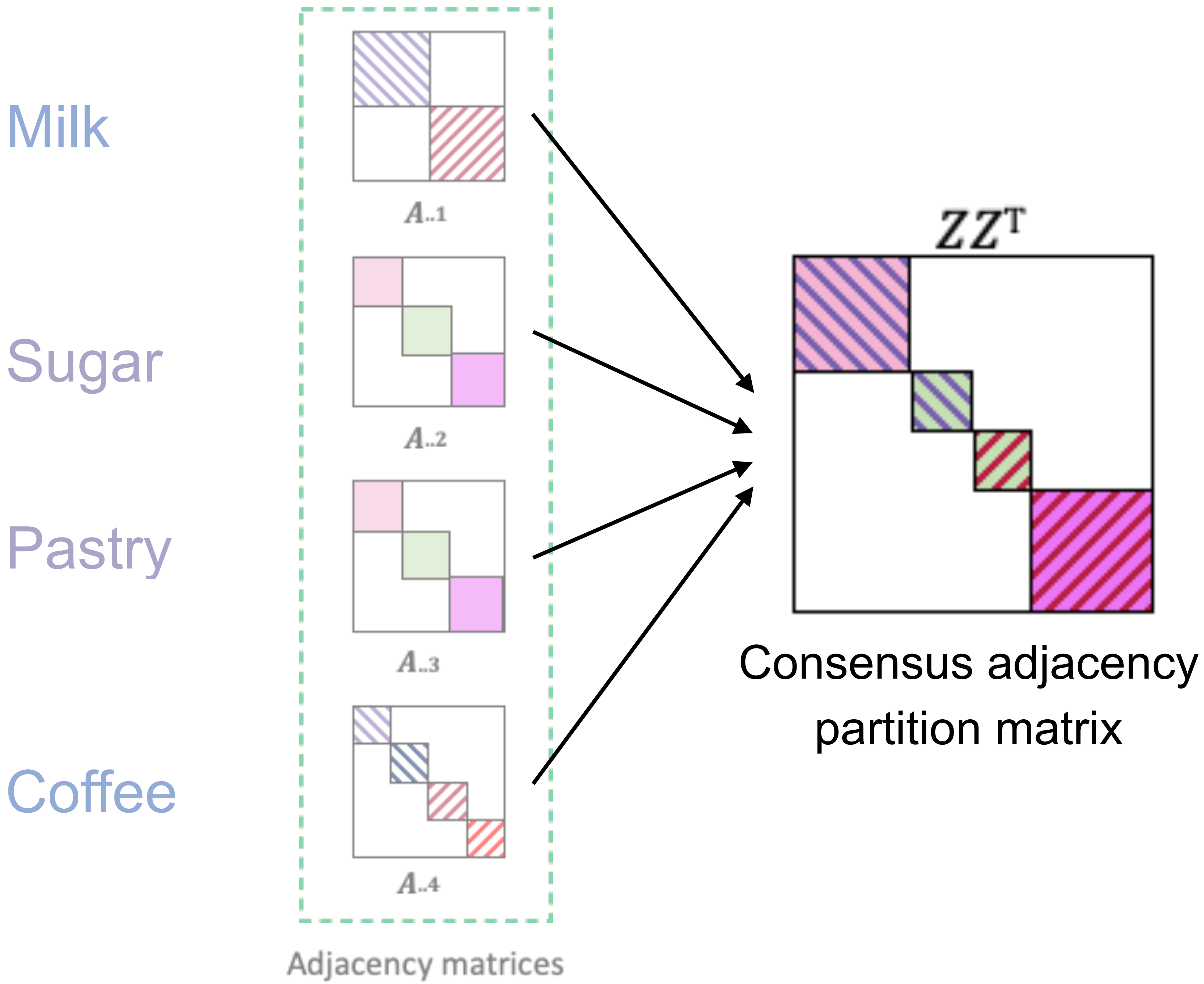
Multi (Layers | Modalities | Views) Learning [2, 3]



Assuming the layers come from a mixture model

General framework

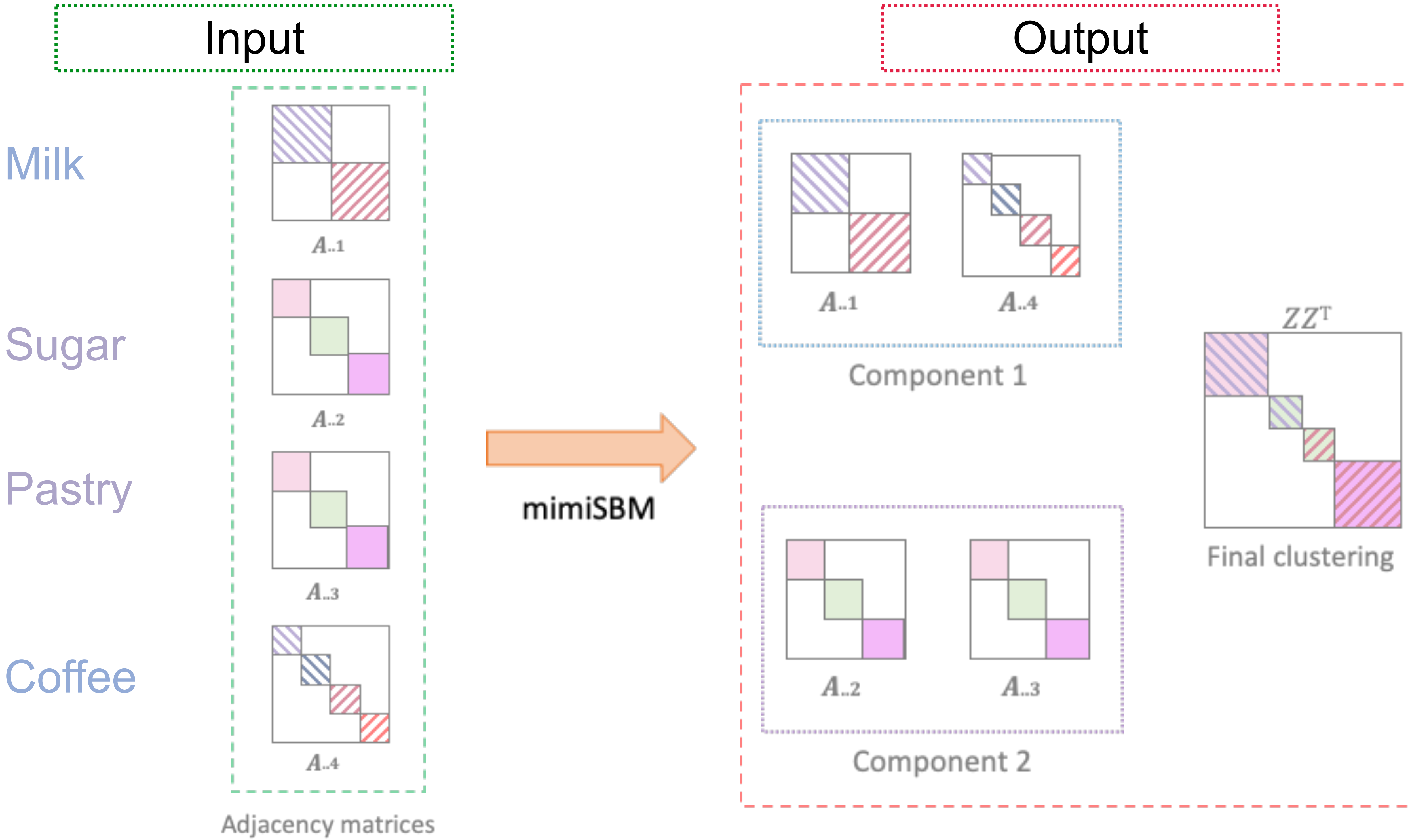
Multi (Layers | Modalities | Views) Learning [2, 3]



Establishing a consensus-based clustering approach

General framework

MIxture of Multilayer Integrator Stochastic Block Model (mimiSBM)



Indicator membership matrices distribution

Probabilistic assumptions on the latent variables

N observations according to K classes $\quad \mathbf{Z} \in \{0,1\}^{N \times K}$

$$\mathbf{Z}_i \sim \mathcal{M}(1, \boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)) \text{ and } \mathbb{P}(\mathbf{Z} \mid \boldsymbol{\pi}) = \prod_{i=1}^{N} \prod_{k=1}^{K} \pi_k^{\mathbf{Z}_{ik}}$$

Indicator membership matrices distribution

Probabilistic assumptions on the latent variables

N observations according to K classes $\mathbf{Z} \in \{0,1\}^{N \times K}$

$$\mathbf{Z}_i \sim \mathscr{M}(1, \boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)) \text{ and } \mathbb{P}(\mathbf{Z} \mid \boldsymbol{\pi}) = \prod_{i=1}^{N} \prod_{k=1}^{K} \pi_k^{\mathbf{Z}_{ik}}$$

V layers according to Q components $\mathbf{W} \in \{0,1\}^{V \times Q}$

$$\mathbf{W}_{\mathbf{v}} \sim \mathscr{M}(1, \boldsymbol{\rho} = (\rho_1, \ldots, \rho_Q)) \text{ and } \mathbb{P}(\mathbf{W} \mid \boldsymbol{\rho}) = \prod_{v=1}^{V} \prod_{s=1}^{Q} \rho_v^{\mathbf{W}_{vs}}$$

Adjacency multilayer distribution

Mixture of Multilayer SBM framework :

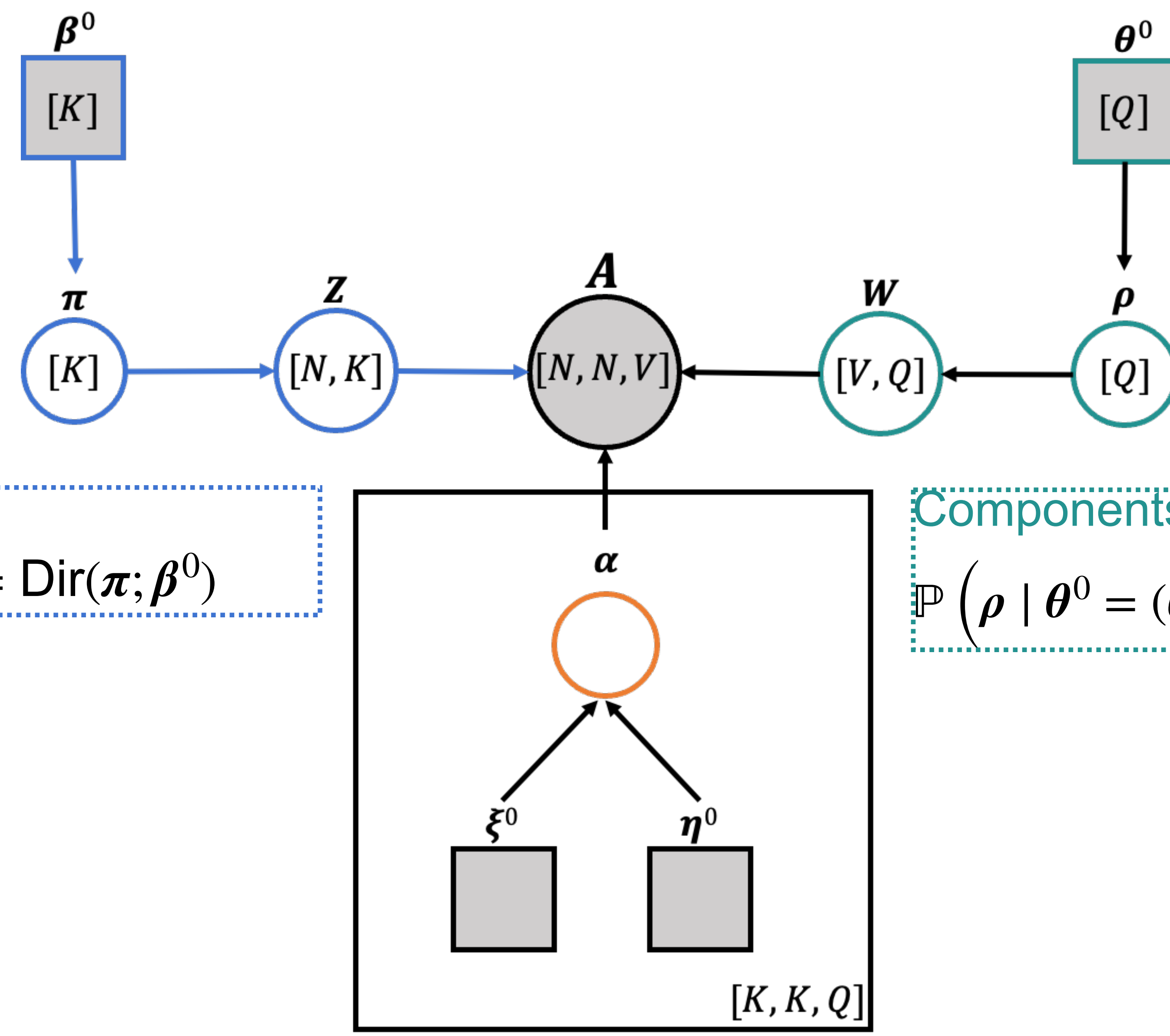
$$A_{ijv} \mid \mathbf{Z}_i = k, \mathbf{Z}_j = l, \mathbf{W}_v = s \sim B\left(\alpha_{kls}\right)$$

Connection between observations according to the multiple layers

$$\mathbb{P}\left(\mathbf{A} \mid \mathbf{Z}, \mathbf{W}, \alpha, \boldsymbol{\pi}, \boldsymbol{\rho}\right) = \prod_{\substack{i=1 \\ i<j}}^{N} \prod_{k,l=1}^{K} \prod_{v=1}^{V} \prod_{s=1}^{Q} \left(\mathbb{P}\left(A_{ijv} \mid \mathbf{Z}_i = k, \mathbf{Z}_j = l, \mathbf{W}_v = s, \alpha, \boldsymbol{\pi}, \boldsymbol{\rho}\right)\right)^{\mathbf{Z}_{ik}\mathbf{Z}_{jl}\mathbf{W}_{vs}}$$

$$= \prod_{\substack{i=1 \\ i<j}}^{N} \prod_{k,l=1}^{K} \prod_{v=1}^{V} \prod_{s=1}^{Q} \left(\alpha_{kls}^{A_{ijv}} \left(1 - \alpha_{kls}\right)^{1-A_{ijv}}\right)^{\mathbf{Z}_{ik}\mathbf{Z}_{jl}\mathbf{W}_{vs}}$$

Bayesian framework of mimi-SBM



Classes :

$\mathbb{P}\left(\boldsymbol{\pi} \mid \boldsymbol{\beta}^0 = (\beta_1^0, \ldots, \beta_K^0)\right) = \text{Dir}(\boldsymbol{\pi}; \boldsymbol{\beta}^0)$

Components :

$\mathbb{P}\left(\boldsymbol{\rho} \mid \boldsymbol{\theta}^0 = (\theta_1^0, \ldots, \theta_Q^0)\right) = \text{Dir}(\boldsymbol{\rho}; \boldsymbol{\theta}^0)$

$$\mathbb{P}(\boldsymbol{\alpha} \mid \boldsymbol{\eta}^0 = (\eta_{kls}^0), \boldsymbol{\xi}^0 = (\xi_{kls}^0)) = \prod_{k,k<l} \prod_{s} \text{Beta}(\alpha_{kls}; \eta_{kls}^0, \xi_{kls}^0)$$

Dir(.) : conjugate prior for the multinomial distribution.

Bayesian Framework

Marginal Likelihood of observed data (evidence)

$$\mathbb{P}\left(\mathbf{A}\right) = \sum_{\mathbf{Z}} \sum_{\mathbf{W}} \int\int\int \mathbb{P}\left(\mathbf{A}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\rho}\right) \; d\boldsymbol{\alpha} \; d\boldsymbol{\pi} \; d\boldsymbol{\rho}$$

Bayesian Framework

Marginal Likelihood of observed data (evidence)

$$\mathbb{P}\left(\mathbf{A}\right) = \sum_{\mathbf{Z}} \sum_{\mathbf{W}} \iiint \mathbb{P}\left(\mathbf{A}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\rho}\right) \; d\boldsymbol{\alpha} \; d\boldsymbol{\pi} \; d\boldsymbol{\rho}$$

Challenging problems :

- Integrals are difficult or impossible to compute analytically
- Sums over $\mathbf{Z}$ and $\mathbf{W}$ are often intractable

Variational distribution

ELBO and KL-divergence

Approximating complex posterior with simpler distributions

Given a variational distribution q over $\{\mathbf{Z}, \mathbf{W}, \boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\rho}\}$, we can decompose the marginal log-likelihood into :

$$\log \mathbb{P}(\mathbf{A}) = \mathscr{L}\left(q(\,.\,)\right) + \mathrm{KL}\left(q(\,.\,) \| \; \mathbb{P}(\,.\,|\,A)\right)$$

Evidence Lower BOund (ELBO)

KL-divergence

Variational distribution

ELBO : mean-field approximation [7]

Typically selected from an easier-to-handle family of distributions

By the mean-field approximation, assume that q can be factorized as :

$$
\begin{aligned}
q(\mathbf{Z}, \mathbf{W}, \boldsymbol{\alpha}, \boldsymbol{\pi}, \boldsymbol{\rho}) &= \prod_{i=1}^{N} q(\mathbf{Z}_i) \prod_{v=1}^{V} q(\mathbf{W}_{\mathbf{v}}) \prod_{s=1}^{Q} \prod_{k,k \leq l}^{K} q(\alpha_{kls})\ q(\boldsymbol{\pi})\ q(\boldsymbol{\rho}) \\
&= \mathrm{Dir}(\boldsymbol{\pi}; \boldsymbol{\beta})\ \mathrm{Dir}(\boldsymbol{\rho}; \boldsymbol{\theta}) \prod_{i=1}^{N} \mathscr{M}(\mathbf{Z}_i; 1, \boldsymbol{\tau}_i) \prod_{v=1}^{V} \mathscr{M}(\mathbf{W}_{\mathbf{v}}; 1, \boldsymbol{\nu}_v) \\
&\quad \prod_{s=1}^{Q} \prod_{k,k \leq l}^{K} \mathrm{Beta}(\alpha_{kls}; \eta_{kls}, \xi_{kls})
\end{aligned}
$$

Variational distribution

ELBO : model selection criterion

Integrated Likelihood variational bayes (ILvb) :

$$\mathscr{L}\left(q(.)\right) = \log\left\{\frac{\Gamma\left(\sum_{k=1}^{K}\beta_k^0\right)\prod_{k=1}^{K}\Gamma\left(\beta_k\right)}{\Gamma\left(\sum_{k=1}^{K}\beta_k\right)\prod_{k=1}^{K}\Gamma\left(\beta_k^0\right)}\right\} + \log\left\{\frac{\Gamma\left(\sum_{q=1}^{Q}\theta_q^0\right)\prod_{q=1}^{Q}\Gamma\left(\theta_q\right)}{\Gamma\left(\sum_{q=1}^{Q}\theta_q\right)\prod_{q=1}^{Q}\Gamma\left(\theta_q^0\right)}\right\}$$

$$+\sum_{k\le l}^{K}\sum_{q=1}^{Q}\log\left\{\frac{\Gamma\left(\eta_{klq}^0+\xi_{klq}^0\right)\Gamma\left(\eta_{klq}\right)\Gamma\left(\xi_{klq}\right)}{\Gamma\left(\eta_{klq}+\xi_{klq}\right)\Gamma\left(\eta_{klq}^0\right)\Gamma\left(\xi_{klq}^0\right)}\right\} - \sum_{i}^{N}\sum_{k}^{K}\tau_{ik}\log\tau_{ik} - \sum_{v}^{V}\sum_{q}^{Q}\nu_{vq}\log\nu_{vq}$$

where $\Gamma(\,\cdot\,)$ is the Gamma function

Variational Bayes

EM optimization

- Variational Bayes Expectation step (VBE-step) :
  - $q(\mathbf{Z}_i), \forall i \in \{1,\ldots,N\}$
  - $q(\mathbf{W}_\mathbf{v}), \forall v \in \{1,\ldots,V\}$
- Maximization step (M-step) :
  - $q(\boldsymbol{\pi})$
  - $q(\boldsymbol{\rho})$
  - $q(\boldsymbol{\alpha})$

Until convergence

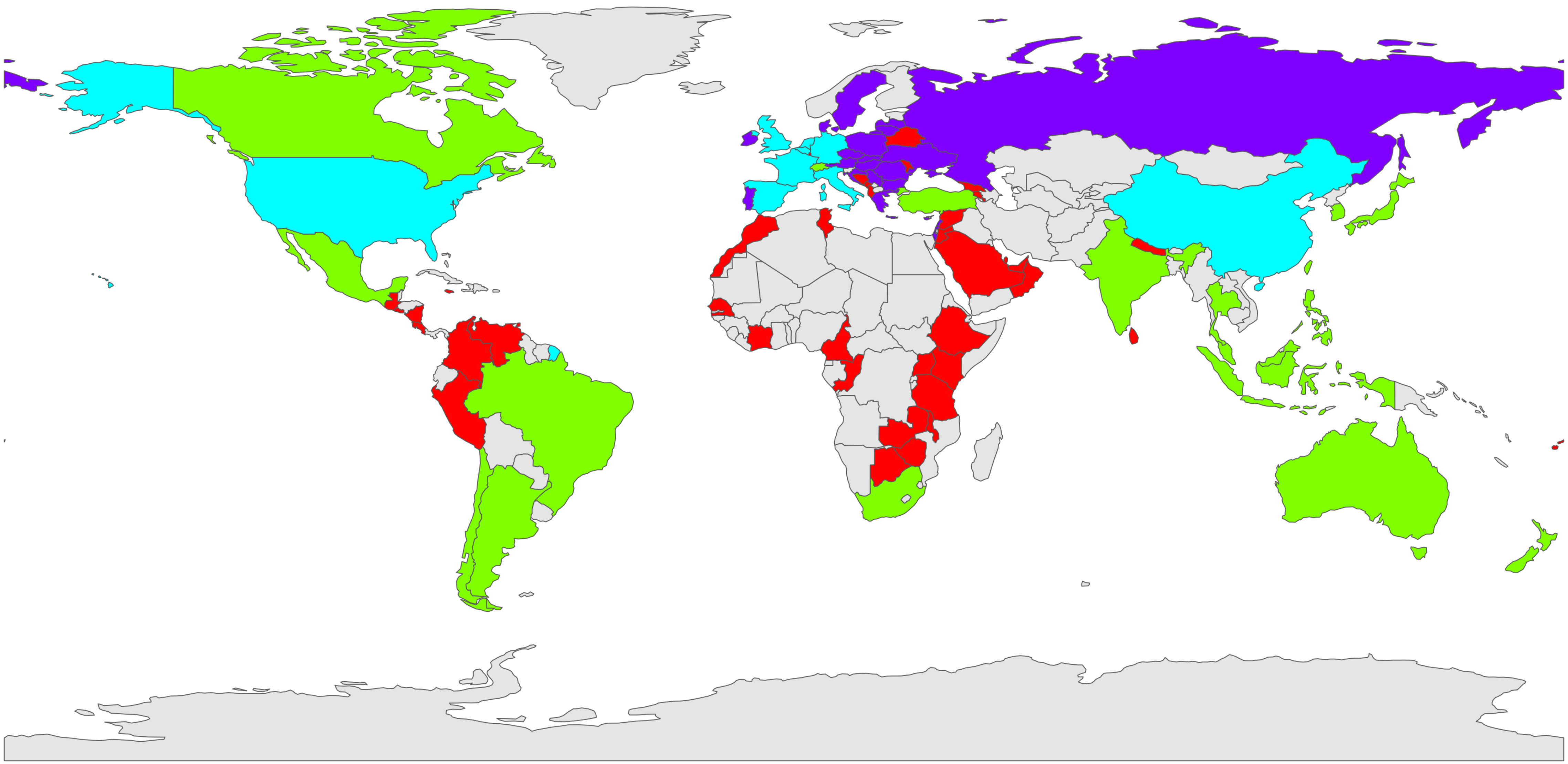
Worldwide Food Trading Networks

Dataset

- A global food trading dataset compiled by *De Dominico et al.* [11]
- This dataset comprises economic networks that feature various products, with 99 countries as nodes and edges denoting trade connections for specific food items
- Each layer reflects the international trade interactions involving **30** distinct food products

Worldwide Food Trading Networks

Countries clustering



Clustering world map: countries are colored according to the clusters defined by the model.

Worldwide Food Trading Networks

Food clustering

View component 1	View component 2
Beverages_non_alcoholic, Food_prep_nes, Chocolate_products_nes , Crude_materials, Fruit_prepared_nes, Beverages_distilled_alcoholic, Pastry, Sugar_confectionery, Wine	Cheese_whole_cow_milk, Cigarettes, Flour_wheat, Beer_of_barley, Cereals_breakfast, Coffee_green, Milk_skimmed_dried, Juice_fruit_nes, Maize, Macaroni, Oil_palm, Milk_whole_dried, Oil_essential_nes, Rice_milled, Sugar_refined, Tea Spices_nes, Vegetables_preserved_nes, Water_ice_etc, Vegetables_fresh_nes, Tobacco_unmanufactured

Processed or prepared food products

basic food items, and non-food products

Table of members in view components

Conclusion and next steps

Other parts on this work

[X] Algorithm initialization strategy

[X] Equivalence and comparison of selection criteria

[X] Performance on simulated data :

[X] View and individuals clustering

[X] Model selection

[X] Robustness according to perturbed adjacency matrices

[X] Model identifiability and parameter convergence
(Less than a fortnight ago)

Thank you for your attention !

References

[1] Baltrušaitis, T., Ahuja, C., & Morency, L. P. (2018). Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence*, *41*(2), 423-443.

[2] Zhao, J., Xie, X., Xu, X., & Sun, S. (2017). Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, *38*, 43-54.

[3] Monti, S., Tamayo, P., Mesirov, J., & Golub, T. (2003). Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Machine learning*, *52*, 91-118.

[4] Jing, B. Y., Li, T., Lyu, Z., & Xia, D. (2021). Community detection on mixture multilayer networks via regularized tensor decomposition. *The Annals of Statistics*, *49*(6), 3181-3205.

[5] Rebafka, T. (2023). Model-based clustering of multiple networks with a hierarchical algorithm.

[6] Latouche, P., Birmele, E., & Ambroise, C. (2012). Variational Bayesian inference and complexity control for stochastic block models. *Statistical Modelling*, *12*(1), 93-115.

[7] Wainwright, M. J., & Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, *1*(1–2), 1-305.

[8] Allman, E. S., Matias, C., & Rhodes, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables.

[9] Celisse, A., Daudin, J. J., & Pierre, L. (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model.

[10] Tabouy, T., Barbillon, P., & Chiquet, J. (2020). Variational inference for stochastic block models from sampled data. *Journal of the American Statistical Association*, *115*(529), 455-466.

[11] Manlio De Domenico, Vincenzo Nicosia, Alexandre Arenas, and Vito Latora. Structural reducibility of multilayer networks. Nature communications, 6(1): 6864, 2015.