Community Detection on Mixture Multilayer Networks

Via Regularized Tensor Decomposition

Damiano Orlandi

Tensor Decomposition for Big Data Analysis

December 10, 2024

The Annals of Statistics
2021, Vol. 49, No. 6, 3181–3205
https://doi.org/10.1214/21-AOS2079

COMMUNITY DETECTION ON MIXTURE MULTILAYER NETWORKS VIA REGULARIZED TENSOR DECOMPOSITION

BY BING-YI JING[1,*], TING LI[2], ZHONGYUAN LYU[1,†] AND DONG XIA[1,‡]

[1]*Department of Mathematics, The Hong Kong University of Science and Technology,* [*]*majing@ust.hk;* [†]*zlyuab@connect.ust.hk;* [‡]*madxia@ust.hk*

[2]*Department of Applied Mathematics, The Hong Kong Polytechnic University, tingeric.li@polyu.edu.hk*

We study the problem of community detection in multilayer networks, where pairs of nodes can be related in multiple modalities. We introduce a general framework, that is, mixture multilayer stochastic block model (MMSBM), which includes many earlier models as special cases. We propose a tensor-based algorithm (TWIST) to reveal both global/local memberships of nodes, and memberships of layers. We show that the TWIST procedure can accurately detect the communities with small misclassification error as the number of nodes and/or number of layers increases. Numerical studies confirm our theoretical findings. To our best knowledge, this is the first systematic study on the mixture multilayer networks using tensor decomposition. The method is applied to two real datasets: worldwide trading networks and malaria parasite genes networks, yielding new and interesting findings.

Access the Paper Here:

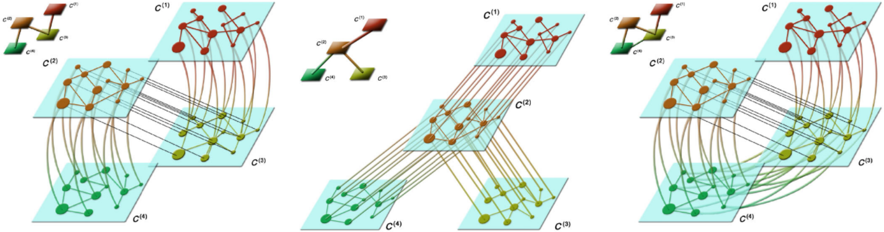
arXiv:2002.04457v1

Contents

1. Multilayer Networks
2. MMSBM
3. TWIST
4. Real-world network application

Multilayer Network: Overview

- Generalized network model that incorporates multiple types of interactions.
- Each layer represents a distinct type of interaction.
- Example:
  - Social networks:
    Facebook = friendships, Twitter = ideology, Linkedin = professionality, Telegram = intimate friendships.

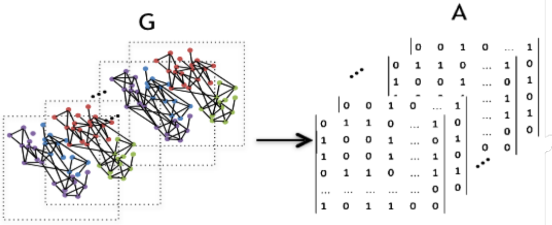


Multilayer Network: Adjacency Tensor

$$A \in \mathbb{R}^{n \times n \times L},$$

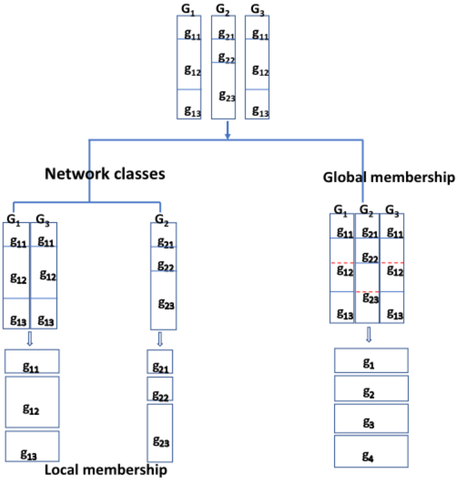
where:

- n: Number of nodes.
- L: Number of layers.
- $A(:,:,\ell) = A_\ell$: Adjacency matrix of the ℓ-th layer.



Community Detection:

- **Global community**
  = dense region of nodes across layers.
- **Local community**
  = dense region of nodes within all individual layers.



Multilayer Networks: Challenges

Why multilayer networks are challenging?

- Dimensionality
- Complex Interdependencies
- Data Sparsity

Contents

1 Multilayer Networks

2 MMSBM

3 TWIST

4 Real-world network application

MMSBM: Overview

Mixture Multilayer Stochastic Block Model

- A probabilistic framework defined for multilayer networks
- Generates:
  - Global communities.
  - Layer labels.
  - Local communities.

MMSBM: Notation

- Let the set of vertices V be:
$$V = \{1, 2, \dots, n\}$$
- Let the network G be:
$$G = \{G_\ell : \ell = 1, \dots, L\}$$
- Let the network labels be:
$$j = 1, 2, \dots, m$$
- Let the number of global communities be:
$$k = 1, 2, \dots, m$$

MMSBM: Notation

- Assume each layer G_ℓ is sampled independently from a mixture of m latent network models with probabilities:

$$\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m), \quad \sum_{j=1}^{m} \pi_j = 1.$$

- Let the probability of a layer G_ℓ having label j be:

$$P(G_\ell = j) = \pi_j, \quad \sum_{j=1}^{m} \pi_j = 1, \quad \text{with } 1 \leq \ell \leq L.$$

MMSBM: SBM

Stochastic Block Model (SBM):
All lables $j \in [m]$, where j represents the j-th label, is described by:

- Membership matrix $Z_j \in \{0,1\}^{n \times K_j}$
- Probability matrix $B_j = \bar{p}\, B_j^0 \in [0,1]^{K_j \times K_j}$, where:
  - $\bar{p} \in (0,1]$ is a parameter that controls the overall density.
  - $\max_j \|B_j^0\|_\infty = 1$

$\mathrm{SBM}(Z_j, B_j)$

MMSBM: Notation

1. **Set of communities**
   - $\mathcal{V}_k$
   - $\mathcal{V}_j := \{\mathcal{V}_k^j\}_{k=1}^{K_j}$

2. **Number of layers**
   - $L_j = |\{\ell : s_\ell = j, 1 \leq \ell \leq L\}|$ = the number of layers generated by SBM(Z_j, B_j). Where $L = \sum_{j=1}^{m} L_j$.

3. **Number of communities across layers**
   - K_j
     Where $\mathring{K} = K_1 + \cdots + K_m$

4. Network labels

- Each $s_\ell \in \{1, 2, \dots, m\}$ indicates the SBM generating the ℓ-th layer. Where $\mathcal{S} = \{s_\ell\}_{\ell=1}^{L}$

5. Adjacency matrix $A_\ell \in \{0,1\}^{n \times n}$ is generated as:

$$A_\ell(i_1, i_2) \mid s_\ell \overset{i.i.d.}{\sim} \mathrm{Bern}(Z_{s_\ell}(i_1, :) B_{s_\ell} Z_{s_\ell}(i_2, :)^T)$$

for all $i_1 \leq i_2 \in [n]$, where $Z(i_1, :)$ denotes the i-th row of Z (exactly one nonzero entry).

MMSBM: Tucker Decomposition

Tensor decomposition:

$$\mathbb{E}(A \mid S) = B \times_1 \bar{Z} \times_2 \bar{Z} \times_3 \bar{W},$$

whose multilinear product equals:

$$\mathbb{E}(A(i_1, i_2, i_3) \mid \mathcal{S}) = \sum_{j_1=1}^{\mathring{K}} \sum_{j_2=1}^{\mathring{K}} \sum_{j_3=1}^{m} B(j_1, j_2, j_3) \bar{Z}(i_1, j_1) \bar{Z}(i_2, j_2) \bar{W}(i_3, j_3),$$

where:

- $B \in \mathbb{R}^{\mathring{K} \times \mathring{K} \times m}$ is a 3-way probability tensor whose j-th frontal slice is:

  $$B(:,:,j) = \mathrm{diag}(0_{K_1}, \ldots, 0_{K_{j-1}}, B_j, 0_{K_{j+1}}, \ldots, 0_{K_m}), \quad 1 \leq j \leq m,$$

  with 0_K being a $K \times K$ zero matrix.
- $\bar{Z} = (Z_1, Z_2, \ldots, Z_m) \in \{0,1\}^{n \times \mathring{K}}$: the global membership matrix
- $W = (e_{s_1}, e_{s_2}, \ldots, e_{s_L})^\top \in \{0,1\}^{L \times m}$: the network label matrix
  With each row of W having exactly one nonzero entry, and $e_j \in \mathbb{R}^m$ being the j-th canonical basis vector.

MMSBM: Global and Local Memberships

Global and Local Communities:

- Each Z_j is the local membership matrix:

$$\bar{Z} = [Z_1, Z_2, \ldots, Z_m].$$

- Two nodes i_1 and i_2 belong to the same global community if:

$$\bar{Z}(i_1, :) = \bar{Z}(i_2, :).$$

- Recall that K is the number of global communities that is the number of distinct rows of $\mathbf{Z}$. Then, $\max_j K_j \leq K \leq \prod_j K_j$.
- Denote $\mathcal{V} = \{\mathcal{V}_k\}_{k=1}^{K}$ the global community clusters such that $\bigcup_{k=1}^{K} \mathcal{V}_k = \mathcal{V}$. Therefore,

$$\{i_1, i_2\} \in \mathcal{V}_k \iff \{i_1, i_2\} \in \mathcal{V}_j^k \quad \text{for some} \quad k_j \in [K], \forall j \in [m].$$

MMSBM: SVD of $\overline{Z}$

Let $r = \text{rank}(\overline{Z})$, then the SVD is given by:

$$\overline{Z} = \bar{U}\bar{D}\bar{R}^T$$

where:

- $\bar{U} \in \mathbb{R}^{n \times r}$: **Orthogonal matrix**
  - Orthonormal columns.
  - Represents the basis vectors of the row space of $\overline{Z}$.
  - Captures the **global community structure** of the nodes.

MMSBM: SVD of $\bar{Z}$

- $\bar{D} \in \mathbb{R}^{r \times r}$: **Diagonal matrix**
  - $\bar{D} = \mathrm{diag}(\sigma_1(\bar{Z}), \ldots, \sigma_r(\bar{Z})) \in \mathbb{R}^{r \times r}$, with $\sigma_1(\bar{Z}) \geq \cdots \geq \sigma_r(\bar{Z}) > 0$.
  - Singular value representing the contribution of each component.
- $\bar{R}^\top \in \mathbb{R}^{r \times \mathring{K}}$: **Orthogonal matrix**
  - Orthonormal columns.
  - Represents the basis vectors of the column space of $\bar{Z}$.
  - Reflects the **local membership structures** across layers.

MMSBM: Global and Local Memberships

- Let note that $\bar{Z}$ cannot be full rank in general.
  - Overlapping and rudundant information.
  - The redundancy arises because multiple local community structures across layers can map to fewer global community structures.
- 1. So $\max_j K_j \leq r \leq \min\{\overline{K} - (m-1), \overline{K}\}$
  - The lower bound ensures that r captures at least the most complex single-layer structure, while the upper bound accounts for redundancies when combining the m layers.

MMSBM: Global and Local Memberships

2. If $K_j \equiv K$, the maximum rank of $\overline{Z}$ is $\min\{mK - m + 1, \overline{K}\}$.
   - when K_j values are uniform, simplifying the theoretical bound on r. It emphasizes the role of m (number of SBMs) in reducing the rank due to redundancies across layers.
3. If a matrix $Z^* = \overline{K} \times K$ is defined and contains the $\overline{K}$ distinct rows of $\overline{Z}$, then

r essentially equals the rank of Z^*.

   - By working with Z^*, we avoid redundant computations, focusing only on unique global community structures. It connects r directly to the intrinsic structure of global communities.

For further details check lemma 2.2

MMSBM: Tucker Decomposition

Tucker Decomposition of $\mathbb{E}(A|S)$:

$$\mathbb{E}(A|S) = \overline{\mathcal{C}} \times_1 \bar{U} \times_2 \bar{U} \times_3 \bar{W}$$

- C: Core tensor.
- $\bar{U}$: Global memberships matrix.
- $\bar{W}$: Network label factor matrix.

MMSBM: Tucker Decomposition

Core Tensor $\mathcal{C} \in \mathbb{R}^{r \times r \times m}$:

$$\mathcal{C} = B \times_1 (DR^\top) \times_2 (DR^\top) \times_3 D_L^{1/2}$$

- $B \in \mathbb{R}^{\mathring{K} \times \mathring{K} \times m}$: A probability tensor that captures interactions between global communities.
- $D \in \mathbb{R}^{r \times r}$: Diagonal matrix from the SVD of Z, representing singular values.
- $R \in \mathbb{R}^{\mathring{K} \times r}$: Matrix capturing global-to-local community relationships.

MMSBM: Tucker Decomposition

Node Membership Matrix $U \in \mathbb{R}^{n \times r}$:

- Represents global node memberships.
- Rows correspond to nodes, columns represent global communities.
- Orthogonal matrix derived from the SVD of : $U \in \mathbb{R}^{n \times r}$.

Layer Membership Matrix W:

$$W = W_D D_L^{-1/2} \in \mathbb{R}^{L \times m}$$

- $W_D \in \mathbb{R}^{L \times m}$: Diagonal matrix containing layer information.
- $D_L \in \mathbb{R}^{m \times m}$: Diagonal matrix of layer weights, e.g., $\mathrm{diag}(L_1, L_2, \ldots, L_m)$.

Contents

1. Multilayer Networks
2. MMSBM
3. TWIST
4. Real-world network application

TWIST: Overview

1. Low-rank approximation of:
$$\mathbb{E}(A|S)$$
2. Identification of global memberships:
$$\{\bar{V}_k\}_{k=1}^{\bar{K}}$$
3. Identification of network classes:
$$\{s_\ell\}_{\ell=1}^{L}$$
4. Recovery of local memberships:
$$V_j = \{V_k^j : k \in [K_j]\}$$

Step 1: Low-rank approximation

Step 1: Low-rank approximation Apply regularized tensor power iterations to the adjacency tensor A to obtain a low-rank approximation.

Algorithm 1 Regularized power iterations for sparse tensor decomposition

Input: $\mathbf{A} \in \{0, 1\}^{n \times n \times L}$, warm initialization $\widehat{U}^{(0)}$ and $\widehat{W}^{(0)}$
maximum iterations $\text{iter}_{\max}$ and regularization parameters $\delta_1, \delta_2 > 0$.

Output: $\widehat{U}$ and $\widehat{W}$

Set counter $\text{iter} = 0$.
while $\text{iter} < \text{iter}_{\max}$ **do**
Regularization: $\widetilde{U}^{(\text{iter})} \leftarrow \mathcal{P}_{\delta_1}(\widehat{U}^{(\text{iter})})$ and $\widetilde{W}^{(\text{iter})} \leftarrow \mathcal{P}_{\delta_2}(\widehat{W}^{(\text{iter})})$ by (3.1).
$\text{iter} \leftarrow \text{iter} + 1$
Set $\widehat{U}^{(\text{iter})}$ to be the top r left singular vectors of $\mathcal{M}_1(\mathbf{A} \times_2 \widetilde{U}^{(\text{iter}-1)\top} \times_3 \widetilde{W}^{(\text{iter}-1)\top})$.
set $\widehat{W}^{(\text{iter})}$ to be the top m left singular vectors of $\mathcal{M}_3(\mathbf{A} \times_1 \widetilde{U}^{(\text{iter}-1)\top} \times_2 \widetilde{U}^{(\text{iter}-1)\top})$.
end while
Return $\widehat{U} \leftarrow \widehat{U}^{(\text{iter})}$ and $\widehat{W} \leftarrow \widehat{W}^{(\text{iter})}$.

Step 1: Low-rank approximation

- **Warm Initialization for $\widehat{U}^{(0)}$ and $\widehat{W}^{(0)}$:**
  - Spectral method:
$$\widehat{U}^{(0)} = \mathrm{SVD}(\sum_{\ell} A_\ell).$$
  - Ensure faster convergence and avoids local minima. Check section 5.5 and Lemma 5.6 for more details
- **Regularized Power Iterations:**
  - Manages sparse tensors by reducing the influence of rows with large norm:
$$\mathcal{P}_\delta(U) = \mathrm{SVD}_r(U^*),$$
where $U^*(i,:) = U(i,:) \cdot \min\{\delta, \|U(i,:)\|_F\} / \|U(i,:)\|_F$
  - Regularization parameters:
$$\delta_1 = 2\sqrt{r} \cdot \max_i \deg_i \cdot \left(\sum_i \deg_i^2\right)^{-1/2} \text{ where the node degree } \deg_i = \sum_{j,\ell} A(i,j,\ell).$$
$$\delta_2 = 2\sqrt{m} \cdot \max_\ell \mathrm{neg}_\ell \cdot \left(\sum_\ell \mathrm{neg}_\ell^2\right)^{-1/2} \text{, where the layer degree } \mathrm{neg}_\ell = \sum_{i,j} A(i,j,\ell).$$

Steps 2 and 3: K-means Clustering

Step 2: Identification of global memberships

- K-means clustering on the rows of $\hat{U}$.
- Output:

$$\hat{V} = \{\hat{V}_k\}_{k=1}^{\bar{K}}.$$

Step 3: Identification of network classes

- K-means or Sup-norm clustering on the rows of $\hat{W}$
- output:

$$\hat{S} = \{\hat{s}_\ell \in [m]\}_{\ell=1}^{L}.$$

Step 4: Clustering for Local Memberships

Algorithm 2 Network clustering by sup-norm K-means

Input: $\widehat{W}$, number of clusters m and threshold $\varepsilon \in (0, 1)$
Output: Network labels $\widehat{\mathbb{S}} = \{\hat{s}_l\}_{l=1}^{L}$

Initiate $\mathcal{C} \leftarrow \{1\}$, $\hat{s}_1 \leftarrow 1$, $k \leftarrow 1$ and $l \leftarrow 2$.
while $l \leq L$ **do**
  Compute $j \leftarrow \arg\min_{j \in \mathcal{C}} \|\widehat{W}(l,:) - \widehat{W}(j,:)\|$
  if $\|\widehat{W}(l,:) - \widehat{W}(j,:)\| > \varepsilon$ **then**
    $k \leftarrow k + 1$; $\hat{s}_l \leftarrow k$; $\mathcal{C} \leftarrow \mathcal{C} \cup \{l\}$
  else
    $\hat{s}_l \leftarrow \hat{s}_j$
  end if
  $l \leftarrow l + 1$
end while
if $k > m$ (or $k < m$) **then**
  Set $\varepsilon \leftarrow 2\varepsilon$ (or set $\varepsilon \leftarrow \varepsilon/2$); Rerun the algorithm.
else
  Output $\widehat{\mathbb{S}} = \{\hat{s}_l\}_{l=1}^{L}$
end if

Challenges: Estimate of $r, m, \overline{K}$

- **Challenge:** estimate rank r of A and m number of classes if unknown.
  - Start with large ranks
  - Identify statistically significant components (scree plot)
  - Refine r and m based on core tensor structure.

- **Challenge:** estimate $\overline{K}$ number of communities.
- **Hierarchical approach:**
  - Initial large groups split into smaller groups iteratively.
  - Stop when a reasonable community structure is reached.

- More details can be found in the supplementary material and in section 5 of the paper.

Contents

1 Multilayer Networks

2 MMSBM

3 TWIST

4 Real-world network application

Real Data Analysis: Worldwide Food Trading Network

Worldwide Food Trading Network

- Data of 30 traded food products in 2010
- $L = 30$ layers and $n = 99$ nodes.
- TWIST applied to analyze layers and extract clusters of countries and products.

Real Data Analysis: Worldwide Food Trading Network

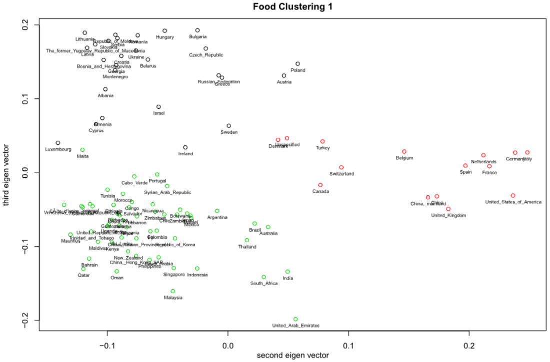
- Layer clustering revealed two major food clusters:
  - Unprocessed food products.
  - Processed food products.
- Regional trading patterns influence proximity effects on processed food trading.

TABLE 1

The resulting two clusters of layers

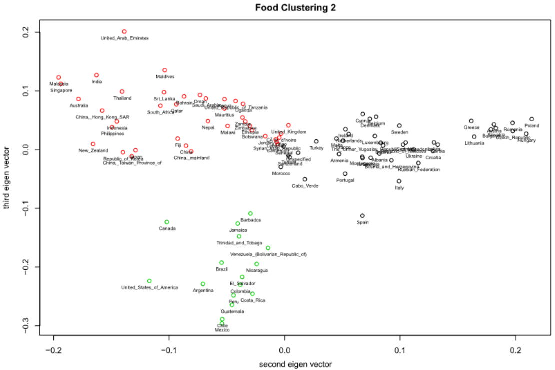
Food cluster 1:	Beverages nonalcoholic, Food prep nes, Chocolate products nes, Crude materials, Fruit prepared nes, Beverages distilled alcoholic, Coffee green, Pastry, Sugar confectionery, Wine, Tobacco unmanufactured
Food cluster 2:	Cheese whole cow milk, Cigarettes, Flour wheat, Beer of barley, Cereals breakfast, 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, Waters ice, etc, Vegetables fresh nes

Visualization: Food Trading Network Results



(a) Embedding of countries for networks in cluster 1.

Visualization: Food Trading Network Results



(b) Embedding of countries for networks in cluster 2.

References

[1] Manlio De Domenico et al. "Mathematical Formulation of Multilayer Networks". In: *Physical Review Journal* (2013).

[2] Bing-Yi Jing et al. "Community Detection on Mixture Multilayer Networks via Regularized Tensor Decomposition". In: *The annals of Statistics* (2021).

[3] Bing-Yi Jing et al. "Supplement to "Community Detection on Mixture Multilayer Networks via Regularized Tensor Decomposition"". In: *The annals of Statistics* (2021).

[4] Toni Vallès-Català et al. "Multilayer Stochastic Block Models Reveal the Multilayer Structure of Complex Networks". In: *Physical Review Journal* (2016).

Thank you for your attention.