Online Change Point Localisation in Multilayer Random Dot Product Graph Models

Fan Wang

Department of Statistics, University of Warwick

December 13, 2022

Multilayer Networks

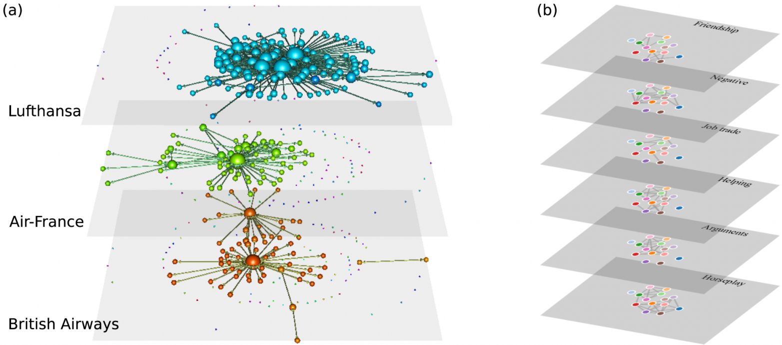

Figure: Visualisation of two multilayer networks. (*a*) Air transportation network from Cardillo et al. (2013). (*b*) Bank-wiring room network from Roethlisberger and Dickson (2003).

A Single Multilayer Random Dot Product Graph (MRDPG)

Adjacency Tensors

Definition (Adjacency tensor)

The adjacency tensor $A \in \mathbb{R}^{n_1 \times n_2 \times L}$ of a multilayer network $\mathcal{G} = (\mathcal{V}_1, \mathcal{V}_2, \mathcal{E}, \mathcal{L})$, is defined as

$$A_{i,j,l} = \begin{cases} 1, & \text{if } (i,j,l) \in \mathcal{E}, \\ 0, & \text{otherwise.} \end{cases}$$

with

- node sets $\mathcal{V}_1 = \{1, \ldots, n_1\}$ and $\mathcal{V}_2 = \{1, \ldots, n_2\}$;
- a layer set $\mathcal{L} = \{1, \ldots, L\}$;
- a edge set $\mathcal{E} \subseteq \{(i,j,l) : i \in \mathcal{V}_1, j \in \mathcal{V}_2, \text{ and } l \in \mathcal{L}\}$.

Latent Position Models

The latent position model is defined as follows.

1. Each node i is mapped to a vector $X_i \in \mathcal{X}$ with some underlying latent space $\mathcal{X} \subset \mathbb{R}^d$.
2. Conditional on latent positions, the i-th and j-th nodes connect independently with probability $K\left(X_i, X_j\right)$ with the function $K : \mathcal{X} \times \mathcal{X} \to [0, 1]$.

The random dot product graph (RDPG) is an especially tractable latent position model with the function K satisfies $K(x, y) = x^\top y$.

MRDPGs

Definition (MRDPGs)

Let $\{X_i\}_{i=1}^{n_1}, \{Y_j\}_{j=1}^{n_2} \subset \mathbb{R}^d$ be mutually independent random vectors generated from F and $\tilde{F}$, respectively.
We say that A is an adjacency tensor of a MRDPG, with random distributed latent positions $\{X_i\}_{i=1}^{n_1}$, $\{Y_j\}_{j=1}^{n_2}$ and fixed weight matrices $\{W_{(l)}\}_{l=1}^{L} \subset \mathbb{R}^{d\times d}$, if

$$\mathbb{P}\{A|\{X_i\}_{i=1}^{n_1}, \{Y_j\}_{j=1}^{n_2}\} = \prod_{i,j,l=1}^{n_1,n_2,L} P_{i,j,l}^{A_{i,j,l}}(1 - P_{i,j,l})^{1-A_{i,j,l}}$$

$$= \prod_{i,j,l=1}^{n_1,n_2,L} \left(X_i^\top W_{(l)} Y_j\right)^{A_{i,j,l}} \left(1 - X_i^\top W_{(l)} Y_j\right)^{1-A_{i,j,l}}.$$

Estimation Methods for MRDPGs

- **The unfolded adjacency spectral embeddings (UASE) method** was proposed by Jones and Rubin-Delanchy (2020);
- **Maximum likelihood estimators (MLEs)** were introduced in Zhang et al. (2020);
- **Convex optimization estimators in combination with a nuclear norm penalty** were proposed by MacDonald et al. (2022).

Tensor Based Estimation Methods

Low-rank tensor estimation:

- **The higher order SVD (HOSVD)** method was introduced by De Lathauwer et al. (2000b);
- **The higher order orthogonal iteration (HOOI)** method was introduced by De Lathauwer et al. (2000a).
- **The tensor heteroskedastic principal component analysis (TH-PCA)** algorithm proposed by Han et al. (2022) who applied **the heteroskedastic principal component analysis (H-PCA)** algorithm introduced in Zhang et al. (2018) to accommodate heteroskedastic noise.

We use the TH-PCA algorithm as the main algorithm for estimating a single MRDPG.

Theoretical Results

Theorem (Estimation error bound)

Under some regularity conditions, it holds with a high probability that

$$\|\widehat{P} - P\|_{\mathrm{F}}^2 \lesssim d^2 m + n_1 d + n_2 d + Lm$$

where

- $\widehat{P}$ *is the output of the TH-PCA algorithm;*
- d *is the dimension of the latent position;*
- m *is the rank of the matrix related to the weight matrices* $\{W_{(l)}\}_{l=1}^{L}$.

Theoretical Comparison

We emphasise three points here.

- There is **no restriction** of n_1, n_2, L, d and m.
- It achieves the **minimax optimal rate** of estimation error (the lower bound shown in Zhang and Xia (2018)).
- It has a **sharper** estimation error bound than all other methods. For example, the convex optimization estimators with high probability have the following estimation error bound

$$\|\hat{P}^{\mathrm{COE}} - P\|_{\mathrm{F}}^2 \lesssim L(n_1 \vee n_2)d.$$

Numerical Comparison

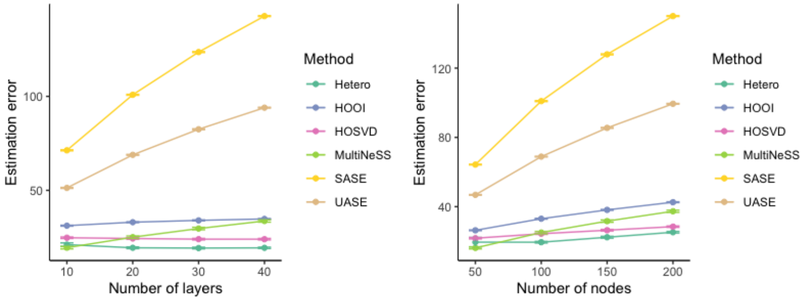

Figure: Multilayer stochastic block models. Left panel: $n_1 = 100$ and $L \in \{10, 20, 30, 40\}$. Right panel: $L = 20$ and $n \in \{50, 100, 150, 200\}$. Each result is shown in the form of mean and standard deviation.

Dynamic Multilayer Random Dot Product Graphs (D-MRDPGs)

D-MRDPGs

Definition (D-MRDPGs)

Let $\{X_i(t)\}_{i=1}^{n_1}, \{Y_j(t)\}_{j=1}^{n_2} \subset \mathbb{R}^d$ be mutually independent random vectors generated from $F(t)$ and $\tilde{F}(t)$, respectively.
We say that $\{A(t)\}_{t\in\mathbb{N}^*}$ is a sequence of independent adjacency tensors of D-MRDPGs, with random distributed latent positions $\{X_i(t)\}_{i=1}^{n_1}$ $\{Y_j(t)\}_{j=1}^{n_2}$ and fixed weight matrices $\{W_l(t)\}_{l=1}^{L}$.

Change Point Analysis

Given D-MRDPGs $\{A(t)\}_{t\in\mathbb{N}^*} \subset \mathbb{R}^{n_1\times n_2\times L}$, for each $t \in \mathbb{N}^*$, let

$$H(t) = \left(\{X_1(t)\}^\top W_1(t) Y_1(t), \ldots, \{X_1(t)\}^\top W_{(L)}(t) Y_1(t)\right)^\top$$

denote a L-dimensional random vector at time point t, with distribution $\mathcal{H}(t)$.

Assumption (No change point)

Given a D-MRDPGs $\{A(t)\}_{t\in\mathbb{N}^*} \subset \mathbb{R}^{n_1\times n_2\times L}$, assume that

$$\mathcal{H}(1) = \mathcal{H}(2) = \ldots$$

One change point analysis: Assume that there exists an integer $\Delta \geq 1$ such that

$$\mathcal{H}(1) = \cdots = \mathcal{H}(\Delta) \neq \mathcal{H}(\Delta+1) = \mathcal{H}(\Delta+2) = \cdots .$$

The Nonparametric Distributional Change

- **The univariate case:**
  - the Kolmogorov–Smirnov (KS) distance between distribution functions.
- **The multivariate case:**
  - the supremum norm of the differences between the densities.
  - transforming the change to the change in the univariate mean.

In our context, the density may not exist.

The Expectation of the Kernel Density Estimator

The expectation of the kernel density estimator (KDE):
Given a kernel function $\mathcal{K} : \mathbb{R}^L \to \mathbb{R}$ and a bandwidth $h > 0$, for $t \in \mathbb{N}^*$, let $G_t : [0,1]^L \to \mathbb{R}$ with

$$G_t(\cdot) = \mathbb{E}\left\{ h^{-L}\mathcal{K}\left(\frac{\cdot - P_{1,2,:}(t)}{h} \right) \right\}.$$

- The expectation of KDE is a **Lebesgue probability density** regardless of whether the nonparametric multivariate distribution admits a Lebesgue density
- The expectation of KDE is often able to capture important **topological properties** of the underlying distribution or of its support shown in Fasy et al. (2014).

One Change Point Assumption

Assumption (One change point)

Given D-MRDPGs $\{A(t)\}_{t \in \mathbb{N}^*} \subset \mathbb{R}^{n_1 \times n_2 \times L}$. *Assume that there exists an integer* $\Delta \geq 1$ *such that*

$$G_1 = \cdots = G_\Delta \neq G_{\Delta+1} = G_{\Delta+2} = \cdots .$$

Let the jump size be

$$\kappa = \sup_{z \in [0,1]^L} |G_\Delta(z) - G_{\Delta+1}(z)| > 0.$$

Online Change Point Detection Algorithm

Algorithm 3 D-MRDPG change point detection

INPUT: $\{A(t)\}_{t \in \mathbb{N}^*} \subset \mathbb{R}^{n_1 \times n_2 \times L}$, $\{\tau_{s,t}\}_{1 \leq s < t} \subset \mathbb{R}$.

$t \leftarrow 1$, FLAG $\leftarrow 0$

while FLAG $= 0$ **do**

  $t \leftarrow t + 1$

  FLAG $\leftarrow 1 - \prod_{s=1}^{t-1} \{\widehat{D}_{s,t} \leq \tau_{s,t}\}$

end while

OUTPUT: t.

Statistics

We emphasise three points for computing the statistic $\widehat{D}_{s,t}$ for any $1 \leq s < t$.

- It is an extension of the **CUSUM statistic** using the **KDEs**.
- We estimate the **average** of the probability tensors instead of estimating the **single** probability tensor at each time point.
- We only use the **first** singular vector for spectral estimations of the average of the probability tensors.

Theoretical results

Assumption (Signal-to-noise ratio condition)

Assume that there exists a large enough absolute constant $C_{\mathrm{SNR}} > 0$ such that, for some $\alpha \in (0,1)$, it holds

$$\kappa\sqrt{\Delta} > C_{\mathrm{SNR}} h^{-L-1} \sqrt{\frac{(L^2 \vee d)\log\{(n_1 \vee n_2 \vee \Delta)/\alpha\}}{n_1 \wedge n_2}}.$$

Theorem

Let D-MRDPGs $\{A(t)\}_{t\in\mathbb{N}^*} \subset \mathbb{R}^{n_1\times n_2\times L}$ and $\alpha \in (0,1)$ and $\widehat{\Delta}$ be the output of the algorithm. Under some regularity conditions, Let

- *Under no change point assumption, it holds that $\mathbb{P}_\infty\{\widehat{\Delta} < \infty\} < \alpha$.*
- *Under one change point assumption and signal-to-noise ratio condition, it holds that with absolute constant $C_\epsilon > 0$*

$$\mathbb{P}_\Delta\left\{\Delta < \widehat{\Delta} \le \Delta + C_\epsilon \frac{(L^2 \vee d)\log\left((n_1 \vee n_2 \vee \Delta)/\alpha\right)}{\kappa^2 h^{2L+2}(n_1 \wedge n_2)}\right\} \ge 1 - \alpha.$$

Theoretical Comparison

We emphasise three points here.

- Compared with Padilla et al. (2019) and most dynamic network papers, our **signal-to-noise ratio condition** is **weaker** up to a $\sqrt{\Delta}$ factora and **localisation error** is **sharper** up to a Δ factor.
- We only use **the first singular vectors** for spectral estimation of the average of the probability tensors and do **not** need to estimate the **ranks** of the probability tensors.
- We allow including the **model parameters** including the number of nodes, the dimension of latent position and the magnitude of the change, to **vary** as functions of the location of the change point.

Thank you!

References I

Alessio Cardillo, Jesús Gómez-Gardenes, Massimiliano Zanin, Miguel Romance, David Papo, Francisco del Pozo, and Stefano Boccaletti. Emergence of network features from multiplexity. *Scientific reports*, 3 (1):1–6, 2013.

Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. On the best rank-1 and rank-$(r_1, r_2, ..., r_n)$ approximation of higher-order tensors. *SIAM journal on Matrix Analysis and Applications*, 21(4):1324–1342, 2000a.

Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle. A multilinear singular value decomposition. *SIAM journal on Matrix Analysis and Applications*, 21(4):1253–1278, 2000b.

Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, and Aarti Singh. Confidence sets for persistence diagrams. *The Annals of Statistics*, pages 2301–2339, 2014.

Rungang Han, Rebecca Willett, and Anru R Zhang. An optimal statistical and computational framework for generalized tensor estimation. *The Annals of Statistics*, 50(1):1–29, 2022.

References II

Andrew Jones and Patrick Rubin-Delanchy. The multilayer random dot product graph. *arXiv preprint arXiv:2007.10455*, 2020.

Peter W MacDonald, Elizaveta Levina, and Ji Zhu. Latent space models for multiplex networks with shared structure. *Biometrika*, 109(3): 683–706, 2022.

Oscar Hernan Madrid Padilla, Yi Yu, and Carey E Priebe. Change point localization in dependent dynamic nonparametric random dot product graphs. *arXiv preprint arXiv:1911.07494*, 2019.

Fritz Jules Roethlisberger and William J Dickson. *Management and the Worker*, volume 5. Psychology press, 2003.

Anru Zhang and Dong Xia. Tensor svd: Statistical and computational limits. *IEEE Transactions on Information Theory*, 64(11):7311–7338, 2018.

Anru R Zhang, T Tony Cai, and Yihong Wu. Heteroskedastic pca: Algorithm, optimality, and applications. *arXiv preprint arXiv:1810.08316*, 2018.

Xuefei Zhang, Songkai Xue, and Ji Zhu. A flexible latent space model for multilayer networks. In *International Conference on Machine Learning*, pages 11288–11297. PMLR, 2020.