Some aspects in dynamic multilayer networks

Yi Yu
Department of Statistics, University of Warwick

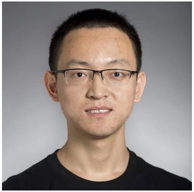
- Static multilayer/multiplex networks
  - Matrices vs. tensors

- Static multilayer/multiplex networks
  - Matrices vs. tensors
- Stationary dynamic networks
  - Temporal dependence modelling

- Static multilayer/multiplex networks
  - Matrices vs. tensors
- Stationary dynamic networks
  - Temporal dependence modelling
- Non-stationary dynamic networks
  - Abrupt changes and smooth changes

- Static multilayer/multiplex networks
  - Matrices vs. tensors
- Stationary dynamic networks
  - Temporal dependence modelling
- Non-stationary dynamic networks
  - Abrupt changes and smooth changes
- Differential privacy
  - Choice of privacy notions in dynamic networks

Fan Wang | Wanshan Li | Oscar Madrid | Alessandro Rinaldo

- Wang, F., Li, W., Madrid Padilla, O. H., **Yu, Y.** & Rinaldo, A. (2025). Multilayer random dot product graphs: Estimation and online change point detection. *JRSS B.*

- Wang, F. & **Yu, Y.** (2026+) Online learning for autoregressive multilayer stochastic block models under stationarity and non-stationarity. *(Ongoing work)*

Models

Adjacency tensor $A \in \mathbb{R}^{n_1 \times n_2 \times L}$ with

$$A_{i,j,l} = \begin{cases} 1, & \text{nodes } i \text{ and } j \text{ are connected in layer } l, \\ 0, & \text{otherwise.} \end{cases}$$

*$[L] = \{1, \ldots, L\}$.

Adjacency tensor $A \in \mathbb{R}^{n_1 \times n_2 \times L}$ with

$$A_{i,j,l} = \begin{cases} 1, & \text{nodes } i \text{ and } j \text{ are connected in layer } l, \\ 0, & \text{otherwise.} \end{cases}$$

Inner product distribution pair Given a sequence of weight matrices $\{W_{(l)}\}_{l=1}^{L} \subset \mathbb{R}^{d \times d}$, let F and $\widetilde{F}$ be two distributions with supports $\mathcal{X}, \widetilde{\mathcal{X}} \subset \mathbb{R}^d$. We say $(F, \widetilde{F})$ is an inner product distribution pair with weights $\{W_{(l)}\}_{l=1}^{L}$ if for all $l \in [L]^*$, $x \in \mathcal{X}$ and $\tilde{x} \in \widetilde{\mathcal{X}}$, it holds that

$$x^\top W_{(l)} \tilde{x} \in [0, 1].$$

$^*[L] = \{1, \ldots, L\}$.

Adjacency tensor $A \in \mathbb{R}^{n_1 \times n_2 \times L}$ with

$$A_{i,j,l} = \begin{cases} 1, & \text{nodes } i \text{ and } j \text{ are connected in layer } l, \\ 0, & \text{otherwise.} \end{cases}$$

Inner product distribution pair Given a sequence of weight matrices $\{W_{(l)}\}_{l=1}^{L} \subset \mathbb{R}^{d \times d}$, let F and $\widetilde{F}$ be two distributions with supports $\mathcal{X}, \widetilde{\mathcal{X}} \subset \mathbb{R}^{d}$. We say $(F, \widetilde{F})$ is an inner product distribution pair with weights $\{W_{(l)}\}_{l=1}^{L}$ if for all $l \in [L]^{*}$, $x \in \mathcal{X}$ and $\tilde{x} \in \widetilde{\mathcal{X}}$, it holds that

$$x^{\top} W_{(l)} \tilde{x} \in [0, 1].$$

Multilayer random dot product graphs with random latent positions. Let $\{X_i\}_{i=1}^{n_1}$ and $\{Y_j\}_{j=1}^{n_2}$ be mutually independent realisations from F and $\widetilde{F}$, respectively. Assume that

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

$^{*}[L] = \{1, \ldots, L\}$.

Tensor representation

$$\mathbb{E}\{A|X,Y\} = P = S \times_1 X \times_2 Y \times_3 Q,$$

where

- $X = (X_1, \ldots, X_{n_1})^\top \in \mathbb{R}^{n_1 \times d}$, $Y = (Y_1, \ldots, Y_{n_2})^\top \in \mathbb{R}^{n_2 \times d}$,
- $S \in \mathbb{R}^{d \times d \times d^2}$ with $S_{i,j,l} = \mathbb{1}\{l = (i-1)d + j\}$, and
- $$Q = \begin{pmatrix} (W_{(1)})^1 & \cdots & (W_{(1)})^d \\ \vdots & \vdots & \vdots \\ (W_{(L)})^1 & \cdots & (W_{(L)})^d \end{pmatrix} \in \mathbb{R}^{L \times d^2}$$

Tensor representation

$$\mathbb{E}\{A|X, Y\} = P = S \times_1 X \times_2 Y \times_3 Q,$$

where

- $X = (X_1, \ldots, X_{n_1})^\top \in \mathbb{R}^{n_1 \times d}$, $Y = (Y_1, \ldots, Y_{n_2})^\top \in \mathbb{R}^{n_2 \times d}$,
- $S \in \mathbb{R}^{d \times d \times d^2}$ with $S_{i,j,l} = \mathbb{1}\{l = (i-1)d + j\}$, and
- $$Q = \begin{pmatrix} (W_{(1)})^1 & \cdots & (W_{(1)})^d \\ \vdots & \vdots & \vdots \\ (W_{(L)})^1 & \cdots & (W_{(L)})^d \end{pmatrix} \in \mathbb{R}^{L \times d^2}$$

Flexibility in allowing for both layer-specific and layer-shared latent positions, e.g. with

$$W_{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad W_{(2)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

it holds that $X^\top W_{(1)} Y = X_1 Y_1 + X_3 Y_3$ and $X^\top W_{(2)} Y = X_2 Y_2 + X_3 Y_3$.

Static/single multilayer random dot product graph (MRDPG) estimation

The key ingredient in both algorithms and theory for single MRDPG is the low-rank tensor estimation.

- HOSVD: higher-order singular value decomposition
  De Lathauwer, L., De Moor, B., & Vandewalle, J. (2000). A multilinear singular value decomposition. SIAM journal on Matrix Analysis and Applications, 21(4), 1253-1278.

- H-PCA: heteroskedastic principal component analysis.
  Zhang, A. R., Cai, T. T., & Wu, Y. (2022). Heteroskedastic PCA: Algorithm, optimality, and applications. The Annals of Statistics, 50(1), 53-80.

- TH-PCA: tensor heteroskedastic principal component analysis.
  Han, R., Willett, R., & Zhang, A. R. (2022). An optimal statistical and computational framework for generalized tensor estimation. The Annals of Statistics, 50(1), 1-29.

Under regularity conditions (e.g. eigengap and moment conditions), we can show that

$$\mathbb{P}\left\{\|\widehat{P} - P\|_{\mathrm{F}}^2 \leq C\left(d^2 m + n_1 d + n_2 d + Lm\right)\right\} > 1 - C\left(n_1 \vee n_2 \vee L\right)^{-c},$$

where $C, c > 0$ are absolute constants and

$$m = \mathrm{rank}(Q) = \mathrm{rank}\left\{\begin{pmatrix} (W_{(1)})^1 & \cdots & (W_{(1)})^d \\ \vdots & \vdots & \vdots \\ (W_{(L)})^1 & \cdots & (W_{(L)})^d \end{pmatrix}\right\} \leq L \wedge d^2.$$

Recall that $\mathbb{E}\{A|X, Y\} = P = S \times_1 X \times_2 Y \times_3 Q$.

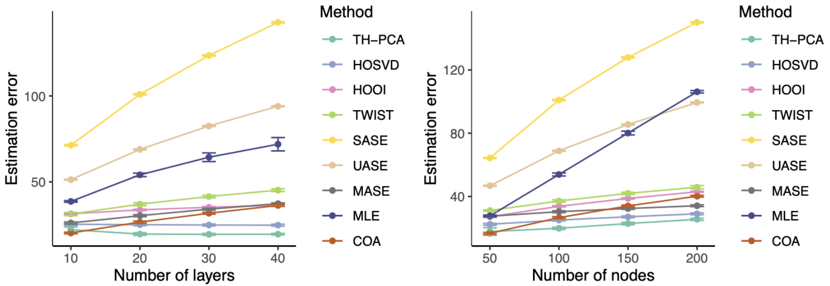
TH-PCA vs. HOSVD

$$\mathbb{P}\left\{\|\widehat{P}-P\|_{\mathrm{F}}^2 \leq C\left(d^2 m+n_1 d+n_2 d+L m\right)\right\}>1-C\left(n_1 \vee n_2 \vee L\right)^{-c}$$

If we use HOSVD instead of TH-PCA, we would have the following instead

$$d^2 m+n_1^2+n_2^2+L^2 .$$

- COA
  MacDonald, P. W., Levina, E. & Zhu, J. (2022). Latent space models for multiplex networks with shared structure. *Biometrika.*
- MLE
  Zhang, X., Xue, S. & Zhu, J. (2020). A flexible latent space model for multilayer networks. *ICML.*
- MASE
  Arroyo, J., Athreya, A., Cape, J., Chen, G., Priebe, C. E., & Vogelstein, J. T. (2021). Inference for multiple heterogeneous networks with a common invariant subspace. *JMLR.*
- UASE
  Jones, A. & Rubin-Delanchy, P. (2020). The multilayer random dot product graph. *arXiv preprint arXiv:2007.10455.*
- SASE
  Sussman, D. L., Tang, M., Fishkind, D. E. & Priebe, C. E. (2012). A consistent adjacency spectral embedding for stochastic blockmodel graphs. *JASA.*
- TWIST
  Jing, B. Y., Li, T., Lyu, Z. & Xia, D. (2021). Community detection on mixture multilayer networks via regularized tensor decomposition. *AoS.*
- HOOI
  Zhang, A. & Xia, D. (2018). Tensor svd: Statistical and computational limits. *IEEE Trans. Inf. Theory.*



Results of estimating a probability tensor $\|\hat{P} - P\|_{\mathrm{F}}$. Multilayer stochastic block models, 4 even-sized communities.

$$\mathbb{P}\left\{\|\widehat{P} - P\|_{\mathrm{F}}^2 \leq C(d^2 m + n_1 d + n_2 d + Lm)\right\} > 1 - C(n_1 \vee n_2 \vee L)^{-c}$$

- TWIST: no directly comparable results.
- MLE

$$Ln + d^2 n + d^2 L$$

- UASE, SASE, MASE, COA

$$Lnd$$

Dynamic version...

- Independence.

- Independence.
- Sticky processes
$$X_t \begin{cases} = X_{t-1}, & \text{with probability } \rho, \\ \sim F, & \text{otherwise.} \end{cases}$$
- Time series models with network covariates (not really dynamic networks).
- Latent position processes satisfying some temporal dependence.

Autoregressive AR(1) network models is proposed in Jiang, Li and Yao (2023, JMLR) and is defined as, for adjacency matrices $\{A^t\}_{t\in\mathbb{Z}_+} \subset \{0,1\}^{n\times n}$ that,

$$A_{i,j}^t = A_{i,j}^{t-1}\mathbb{1}\{E_{i,j}^t = 0\} + \mathbb{1}\{E_{i,j}^t = 1\},$$

where $\{E_{i,j}^t\}_{i,j=1,t\in\mathbb{Z}_+}^{n,n}$ are independent random variables with

$$\mathbb{P}\{E_{i,j}^t = 1\} = \alpha_{i,j}^t, \quad \mathbb{P}\{E_{i,j}^t = -1\} = \beta_{i,j}^t \quad \text{and} \quad \mathbb{P}\{E_{i,j}^t = 0\} = 1 - \alpha_{i,j}^t - \beta_{i,j}^t,$$

with $\alpha_{i,j}^t, \beta_{i,j}^t \geq 0$ and $\alpha_{i,j}^t + \beta_{i,j}^t \leq 1$.

Autoregressive AR(1) network models is proposed in Jiang, Li and Yao (2023, JMLR) and is defined as, for adjacency matrices $\{A^t\}_{t\in\mathbb{Z}_+} \subset \{0,1\}^{n\times n}$ that,

$$A^t_{i,j} = A^{t-1}_{i,j}\mathbb{1}\{E^t_{i,j} = 0\} + \mathbb{1}\{E^t_{i,j} = 1\},$$

where $\{E^t_{i,j}\}^{n,n}_{i,j=1,t\in\mathbb{Z}_+}$ are independent random variables with

$$\mathbb{P}\{E^t_{i,j} = 1\} = \alpha^t_{i,j}, \quad \mathbb{P}\{E^t_{i,j} = -1\} = \beta^t_{i,j} \quad \text{and} \quad \mathbb{P}\{E^t_{i,j} = 0\} = 1 - \alpha^t_{i,j} - \beta^t_{i,j},$$

with $\alpha^t_{i,j}, \beta^t_{i,j} \geq 0$ and $\alpha^t_{i,j} + \beta^t_{i,j} \leq 1$.

This model implies that

$$\mathbb{P}\{A^t_{i,j} = 1 | A^{t-1}_{i,j} = 0\} = \alpha^t_{i,j} \quad \text{and} \quad \mathbb{P}\{A^t_{i,j} = 0 | A^{t-1}_{i,j} = 1\} = \beta^t_{i,j}.$$

Autoregressive AR(1) network models is proposed in Jiang, Li and Yao (2023, JMLR) and is defined as, for adjacency matrices $\{A^t\}_{t\in\mathbb{Z}_+} \subset \{0,1\}^{n\times n}$ that,

$$A_{i,j}^t = A_{i,j}^{t-1}\mathbb{1}\{E_{i,j}^t = 0\} + \mathbb{1}\{E_{i,j}^t = 1\},$$

where $\{E_{i,j}^t\}_{i,j=1,t\in\mathbb{Z}_+}^{n,n}$ are independent random variables with

$$\mathbb{P}\{E_{i,j}^t = 1\} = \alpha_{i,j}^t, \quad \mathbb{P}\{E_{i,j}^t = -1\} = \beta_{i,j}^t \quad \text{and} \quad \mathbb{P}\{E_{i,j}^t = 0\} = 1 - \alpha_{i,j}^t - \beta_{i,j}^t,$$

with $\alpha_{i,j}^t, \beta_{i,j}^t \geq 0$ and $\alpha_{i,j}^t + \beta_{i,j}^t \leq 1$.

This model implies that

$$\mathbb{P}\{A_{i,j}^t = 1 | A_{i,j}^{t-1} = 0\} = \alpha_{i,j}^t \quad \text{and} \quad \mathbb{P}\{A_{i,j}^t = 0 | A_{i,j}^{t-1} = 1\} = \beta_{i,j}^t.$$

Extensions of this AR(1) network model includes its stationary version, stochastic block model version, change point version, AR(p) version.

$$A_{i,j}^{t} = A_{i,j}^{t-1}\mathbb{1}\{E_{i,j}^{t} = 0\} + \mathbb{1}\{E_{i,j}^{t} = 1\}$$

The motivation of this model is from the edge jittering privacy mechanism studied in Chang, Hu, Kolaczyka, Yao and Yi (2024, AoS) and Karwa, Krivitsky and Slavković (2017, JRSSC) on central DP static networks. To be specific, release

$$Z_{i,j} = X_{i,j}\mathbb{1}\{\varepsilon_{i,j} = 0\} + \mathbb{1}\{\varepsilon_{i,j} = 1\}.$$

Our extensions include a multilayer version and a smoothly varying + abrupt changes version.

Our extensions include a multilayer version and a smoothly varying + abrupt changes version.

DEFINITION A stationary AR(1)-multilayer SBM (MSBM) is a sequence of adjacency tensors $\{A^t\}_{t\in\mathbb{Z}_+} \subset \{0,1\}^{n\times n\times L}$ that

$$A^t_{i,j,l} = A^{t-1}_{i,j,l}\mathbb{1}\{E^t_{i,j,l}=0\} + \mathbb{1}\{E^t_{i,j,l}=1\},$$

where $\{E^t_{i,j,l}\}^{n,n,L}_{i,j,l=1,t\in\mathbb{Z}_+}$ are independent random variables with

$$\mathbb{P}\{E^t_{i,j,l}=1\} = \Theta_{i,j,l} = Z_i W_{:,:,l} Z_j^\top, \quad \mathbb{P}\{E^t_{i,j,l}=-1\} = \Delta_{i,j,l} = Z_i M_{:,:,l} Z_j^\top$$

and

$$\mathbb{P}\{E^t_{i,j,l}=0\} = 1-\Theta_{i,j,l}-\Delta_{i,j,l},$$

with $Z \in \{0,1\}^{n\times K}$ being the community membership matrix, and $W, M \in (0,1)^{K\times K\times L}$ being the connectivity tensors.

The goal is to estimate the transition probability tensors Θ and Δ, as well as to recover the community structures.

The algorithm consists of three key steps, namely

- S1 likelihood formulation and initial estimators,
- S2 low-rank refinement via tensor-based methods, and
- S3 community recovery.

S1 Likelihood formulation and initial estimators

The likelihood can be written as

$$\prod_{s=1}^{t} \prod_{1\leq i\leq j\leq n, l\in[L]} \Theta_{i,j,l}^{A^s_{i,j,l}(1-A^{s-1}_{i,j,l})}(1-\Theta_{i,j,l})^{(1-A^s_{i,j,l})(1-A^{s-1}_{i,j,l})}\Delta_{i,j,l}^{(1-A^s_{i,j,l})A^{s-1}_{i,j,l}}(1-\Delta_{i,j,l})^{A^s_{i,j,l}A^{s-1}_{i,j,l}}.$$

The MLEs serve as the initial estimators that

$$\widehat{\Theta}^t_{i,j,l} = \frac{\sum_{s=1}^t A^s_{i,j,l}(1-A^{s-1}_{i,j,l})}{\sum_{s=1}^t (1-A^{s-1}_{i,j,l})} \quad \text{and} \quad \widehat{\Delta}^t = \frac{(1-A^s_{i,j,l})A^{s-1}_{i,j,l}}{\sum_{s=1}^t A^{s-1}_{i,j,l}}.$$

S1 Likelihood formulation and initial estimators

The likelihood can be written as

$$\prod_{s=1}^{t} \prod_{1 \le i \le j \le n, l \in [L]} \Theta_{i,j,l}^{A^s_{i,j,l}(1-A^{s-1}_{i,j,l})} (1-\Theta_{i,j,l})^{(1-A^s_{i,j,l})(1-A^{s-1}_{i,j,l})} \Delta_{i,j,l}^{(1-A^s_{i,j,l})A^{s-1}_{i,j,l}} (1-\Delta_{i,j,l})^{A^s_{i,j,l}A^{s-1}_{i,j,l}}.$$

The MLEs serve as the initial estimators that

$$\widehat{\Theta}^t_{i,j,l} = \frac{\sum_{s=1}^t A^s_{i,j,l}(1 - A^{s-1}_{i,j,l})}{\sum_{s=1}^t (1 - A^{s-1}_{i,j,l})} \quad \text{and} \quad \widehat{\Delta}^t = \frac{(1 - A^s_{i,j,l})A^{s-1}_{i,j,l}}{\sum_{s=1}^t A^{s-1}_{i,j,l}}.$$

The initial estimators only depend on three counts, therefore can be estimated in an online fashion.

S2 Low-rank refinement via tensor-based methods

This is done by the tensor heteroskedastic principal component analysis (TH-PCA) proposed in Han, Willett and Zhang (2022, AoS).

S3 Community recovery

$$A_{i,j,l}^{t} = A_{i,j,l}^{t-1}\mathbb{1}\{E_{i,j,l}^{t} = 0\} + \mathbb{1}\{E_{i,j,l}^{t} = 1\}$$
$$\mathbb{P}\{E_{i,j,l}^{t} = 1\} = \Theta_{i,j,l} = Z_i W_{:,:,l} Z_j^{\top}, \quad \mathbb{P}\{E_{i,j,l}^{t} = -1\} = \Delta_{i,j,l} = Z_i M_{:,:,l} Z_j^{\top}$$
$$\mathbb{P}\{E_{i,j,l}^{t} = 0\} = 1 - \Theta_{i,j,l} - \Delta_{i,j,l}$$

Transition probability tensors

THEOREM (informal) Under regularity conditions and denoting that

$$K = \mathrm{rank}(\mathcal{M}_1(W)) = \mathrm{rank}(\mathcal{M}_1(M)), \quad r = \min\{\mathrm{rank}(\mathcal{M}_3(W)), \mathrm{rank}(\mathcal{M}_3(M))\},$$

it holds that

$$\|\widetilde{\Theta}^t - \Theta\|_{\mathrm{F}} + \|\widetilde{\Delta}^t - \Delta\|_{\mathrm{F}} \lesssim \sqrt{\frac{nK + Lr + rK^2}{t}} \vee \frac{n\sqrt{L}}{t}.$$

Remarks

- The first term is due to the low-rank estimation variance.
- The second term is due to the bias from the initial estimator.
- The estimation error for the initial estimator
$$\frac{n\sqrt{L}}{\sqrt{t}}.$$
- The estimation error for a matrix-based method
$$\sqrt{\frac{nLK}{t}} \vee \frac{n\sqrt{L}}{t}.$$

$$\|\widetilde{\Theta}^t - \Theta\|_{\mathrm{F}} + \|\widetilde{\Delta}^t - \Delta\|_{\mathrm{F}} \lesssim \sqrt{\frac{nK + Lr + rK^2}{t}} \vee \frac{n\sqrt{L}}{t}.$$

We also obtain a lower bound that is

$$\inf_{(\widehat{\Theta},\widehat{\Delta})} \sup_{(\Theta,\Delta)\in\mathcal{P}} \mathbb{E}_{\Theta,\Delta}\left\{\|\widehat{\Theta} - \Theta\|_{\mathrm{F}}^2 + \|\widehat{\Delta} - \Delta\|_{\mathrm{F}}^2\right\} \gtrsim \frac{rK^2 + n\log(K) + Lr}{T},$$

where $\mathcal{P}$ collects the distributions satisfying the conditions required by the upper bound result.

Remarks

- The first term corresponds to estimate the low-rank core tensor.
- The second term corresponds to the clustering error.
- The last term corresponds to estimating the loading matrices.
- The gap between the upper and lower bounds K vs. $\log(K)$ correspond to the statistical-computational tradeoff.

Community recovery

Theorem (informal) Under the condition

$$\frac{1}{tL} + \frac{n}{t^2} \lesssim 1,$$

we achieve consistency recovery.

Remark

For the single-layer case $L = 1$, Jiang et al. (2022) requires

$$\frac{1}{t} + \frac{n}{t^2} + \frac{1}{n} \lesssim 1.$$

Non-stationarity

Recall in the stationary case, we assume that

$$A_{i,j,l}^{t} = A_{i,j,l}^{t-1} \mathbb{1}\{E_{i,j,l}^{t} = 0\} + \mathbb{1}\{E_{i,j,l}^{t} = 1\},$$

where

$$\mathbb{P}\{E_{i,j,l}^{t} = 1\} = \Theta_{i,j,l} = Z_i W_{:,:,l} Z_j^{\top}, \quad \mathbb{P}\{E_{i,j,l}^{t} = -1\} = \Delta_{i,j,l} = Z_i M_{:,:,l} Z_j^{\top}$$

and

$$\mathbb{P}\{E_{i,j,l}^{t} = 0\} = 1 - \Theta_{i,j,l} - \Delta_{i,j,l}.$$

Recall in the stationary case, we assume that

$$A^t_{i,j,l} = A^{t-1}_{i,j,l}\mathbb{1}\{E^t_{i,j,l} = 0\} + \mathbb{1}\{E^t_{i,j,l} = 1\},$$

where

$$\mathbb{P}\{E^t_{i,j,l} = 1\} = \Theta_{i,j,l} = Z_i W_{:,:,l} Z_j^\top, \quad \mathbb{P}\{E^t_{i,j,l} = -1\} = \Delta_{i,j,l} = Z_i M_{:,:,l} Z_j^\top$$

and

$$\mathbb{P}\{E^t_{i,j,l} = 0\} = 1 - \Theta_{i,j,l} - \Delta_{i,j,l}.$$

Moving on to the non-stationary case, we assume that

$$\mathbb{P}\{E^t_{i,j,l} = 1\} = \Theta^t_{i,j,l} = Z_i W^t_{:,:,l} Z_j^\top, \quad \mathbb{P}\{E^t_{i,j,l} = -1\} = \Delta^t_{i,j,l} = Z_i M^t_{:,:,l} Z_j^\top$$

and

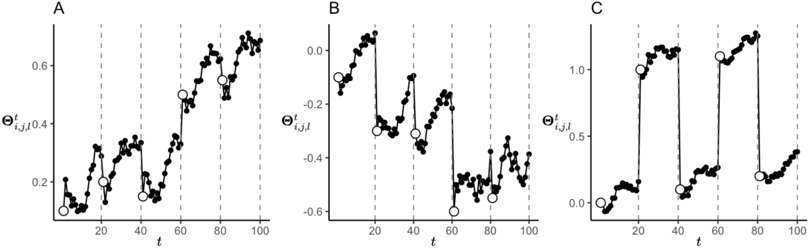
$$\mathbb{P}\{E^t_{i,j,l} = 0\} = 1 - \Theta^t_{i,j,l} - \Delta^t_{i,j,l}.$$

The non-stationarity we allow includes

- abrupt changes - a roughly monotone interval endpoint values, and
- smooth changes - a roughly monotone trend within each interval.

The non-stationarity we allow includes

- abrupt changes - a roughly monotone interval endpoint values, and
- smooth changes - a roughly monotone trend within each interval.



Panels A and B are allowed, but not Panel C.

The main ingredient in the algorithm is an adaptive window selection procedure inspired by Huang and Wang (2024, OR).

Under regularity results, we recover the stationary results within the stationary segments.

Differential privacy

- Local DP
- Central DP
- Federate DP, user-level DP, ...

- Local DP
- Central DP
- Federate DP, user-level DP, ...

- Local DP
- Central DP
- Federate DP, user-level DP, ...

In the context of networks

- Edge DP
- Node DP
- ...

- Local DP
- Central DP
- Federate DP, user-level DP, ...

In the context of networks

- Edge DP
- Node DP
- ...

In the context of online learning

- Local DP
- Joint DP

Cost of privacy in change point detection

In Li, Berrett and **Y.** (2022, NurIPS), we considered change point detection in independent dynamic Bernoulli networks.

- No privacy

$$\text{jump size}^2 \times \text{sparsity} \times \text{network size} \times \text{spacing} \asymp 1.$$

- Edge local DP

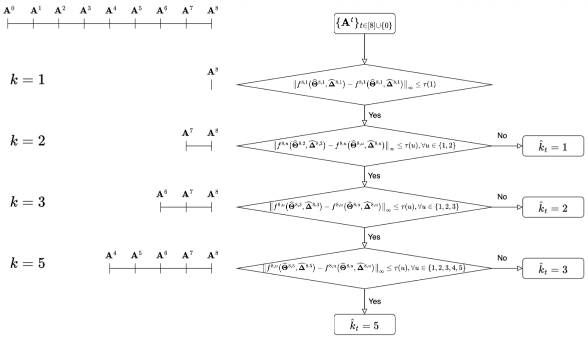
$$\text{jump size}^2 \times \text{sparsity}^2 \times \text{network size} \times \text{spacing} \times \text{privacy}^2 \asymp 1.$$

- Node local DP

$$\text{jump size}^2 \times \text{sparsity}^2 \times \text{network size}^{1/2} \times \text{spacing} \times \text{privacy}^2 \asymp 1.$$

- Edge vs. node privacy
- Central, local vs. federated privacy
- Dependence within networks
- Temporal dependence in dynamic networks

The main ingredient in the algorithm is an adaptive window selection procedure inspired by Huang and Wang (2024, OR).



A privacy mechanism is a randomised algorithm taking an input dataset $X = (X_1, \dots, X_n) \in \mathcal{X}^n$ and producing publishable data Z. Formally, it is a collection of conditional distributions $\mathcal{Q} = \{Q(\cdot|x) : x \in \mathcal{X}^n\}$ such that

$$Z|\{X = x\} \sim Q(\cdot|x).$$

Privacy mechanism Q is called (ϵ, δ)-(central) differentially private (Dwork et al., 2006), with $\epsilon > 0$ and $\delta \geq 0$, if

$$Q(A|x) \leq e^{\epsilon} Q(A|x') + \delta, \tag{1}$$

for all measurable set A, any pair $x = (x_i)_{i=1}^n, x' = (x_i')_{i=1}^n \in \mathcal{X}^n$ such that $\sum_{i=1}^n \mathbf{1}\{x_i \neq x_i'\} \leq 1$. We focus on the regime $\epsilon \in (0, 1]$.

A privacy mechanism is a randomised algorithm taking an input dataset $X = (X_1, \ldots, X_n) \in \mathcal{X}^n$ and producing publishable data Z. Formally, it is a collection of conditional distributions $\mathcal{Q} = \{Q(\cdot|x) : x \in \mathcal{X}^n\}$ such that

$$Z|\{X = x\} \sim Q(\cdot|x).$$

Privacy mechanism Q is called (ϵ, δ)-(central) differentially private (Dwork et al., 2006), with $\epsilon > 0$ and $\delta \geq 0$, if

$$Q(A|x) \leq e^{\epsilon} Q(A|x') + \delta, \tag{1}$$

for all measurable set A, any pair $x = (x_i)_{i=1}^n, x' = (x_i')_{i=1}^n \in \mathcal{X}^n$ such that $\sum_{i=1}^n \mathbf{1}\{x_i \neq x_i'\} \leq 1$. We focus on the regime $\epsilon \in (0, 1]$.

Suppose $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} P$. If an adversary, knowing P, Q and $X_2, \ldots, X_n$, sees $Z \sim Q(\cdot|\{X_i\}_{i=1}^n)$ and wants to test at significance level γ that

$$\mathrm{H}_0 : X_1 = x_1 \quad \text{vs.} \quad \mathrm{H}_1 : X_1 = x_1',$$

then the power of the test is upper bounded by $\min\{1, \gamma e^{\epsilon} + \delta\}$.

For the central differential privacy (CDP), where there is a trusted central data curator having access to all the raw data. For example, when estimating a univariate mean, we can have

$$\widehat{\theta} = Z = \frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{n\varepsilon} W, \quad \text{with } W \sim \mathrm{Lap}(1).$$

The variance of total added noise is of order $(n^2\varepsilon^2)^{-1}$.

For the central differential privacy (CDP), where there is a trusted central data curator having access to all the raw data. For example, when estimating a univariate mean, we can have

$$\widehat{\theta} = Z = \frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{n\varepsilon} W, \quad \text{with } W \sim \mathrm{Lap}(1).$$

The variance of total added noise is of order $(n^2\varepsilon^2)^{-1}$.

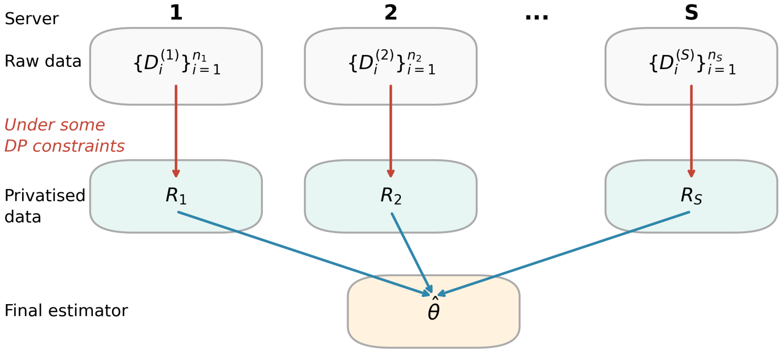
A stronger notion of differential privacy is the local differential privacy (LDP), where data are randomised before collection, that is

$$\mathbb{P}(Z_i \in A | X_i = x) \leq e^{\epsilon}\mathbb{P}(Z_i \in A | X_i = x') + \delta, \quad i \in \{1, \ldots, n\}, \tag{2}$$

for all measurable set A and any pair $x, x' \in \mathcal{X}$. For example, when estimating a univariate mean, we can have

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} Z_i = \frac{1}{n}\sum_{i=1}^{n}\left(X_i + \frac{1}{\varepsilon}W_i\right), \quad \text{with } \{W_i\}_{i=1}^{n} \overset{\text{i.i.d.}}{\sim} \mathrm{Lap}(1).$$

The variance of total added noise is of order $(n\varepsilon^2)^{-1}$.


Server
1
2
...
S
Raw data
$\{D_i^{(1)}\}_{i=1}^{n_1}$
$\{D_i^{(2)}\}_{i=1}^{n_2}$
$\{D_i^{(S)}\}_{i=1}^{n_S}$
Under some
DP constraints
Privatised
data
R_1
R_2
R_S
Final estimator
$\hat{\theta}$

$$\inf_{Q \in \mathcal{Q}_{\varepsilon,\delta}} \inf_{\widehat{\theta}(Z)} \sup_{P_\theta \in \mathcal{P}} \mathbb{E}_{\substack{X \sim P_\theta \\ Z|X \sim Q}} \left\{ \Phi \circ \rho(\widehat{\theta}, \theta) \right\},$$

A simple example: univariate mean estimation measured in squared loss, with S users/servers and n units of data per user.

Setting	Minimax rates	References
No privacy	$1/(Sn)$	Very easy to show
Local item-level	$1/(Sn) \vee$	Duchi et al. (2018)
Local user-level (small n)	$1/(Sn) \vee$	Our result
Local user-level (large n)	$1/(Sn) \vee$	Our result
Central item-level	$1/(Sn) \vee$	Levy et al. (2021)
Central user-level (small n)	$1/(Sn) \vee$	Levy et al. (2021)
Federated	$1/(Sn) \vee$	Our result

$$\inf_{Q \in \mathcal{Q}_{\varepsilon,\delta}} \inf_{\widehat{\theta}(Z)} \sup_{P_\theta \in \mathcal{P}} \mathbb{E}_{\substack{X \sim P_\theta \\ Z|X \sim Q}} \left\{ \Phi \circ \rho(\widehat{\theta}, \theta) \right\},$$

A simple example: univariate mean estimation measured in squared loss, with S users/servers and n units of data per user.

Setting	Minimax rates	References
No privacy	$1/(Sn)$	Very easy to show
Local item-level	$1/(Sn) \vee 1/(Sn\varepsilon^2)$	Duchi et al. (2018)
Local user-level (small n)	$1/(Sn) \vee 1/(Sn\varepsilon^2)$	Our result
Local user-level (large n)	$1/(Sn) \vee e^{-S\varepsilon^2}$	Our result
Central item-level	$1/(Sn) \vee 1/(S^2n^2\varepsilon^2)$	Levy et al. (2021)
Central user-level (small n)	$1/(Sn) \vee 1/(S^2n\varepsilon^2)$	Levy et al. (2021)
Federated	$1/(Sn) \vee 1/(Sn^2\varepsilon^2)$	Our result