AMS Weekly Seminars - Johns Hopkins University

Low-rank models for dynamic multiplex graphs and vector autoregressive processes

IMPERIAL

Francesco Sanna Passino

Department of Mathematics, Imperial College London

f.sannapassino@imperial.ac.uk

11th April, 2024

Joint work with:

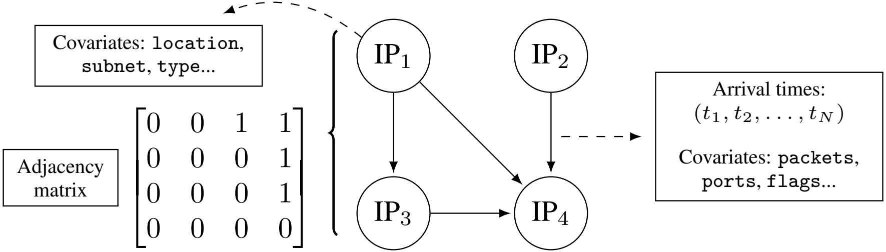
- **Dynamic multiplex graphs**
  - **Maximilian Baum** (PhD student, Imperial College London)
  - Axel Gandy (Chair in Statistics, Imperial College London)
- **Panels of multivariate time series / vector autoregressive processes**
  - **Brendan Martin** (PhD student, Imperial College London)
  - Mihai Cucuringu (Associate Professor of Statistics, University of Oxford)
  - Alessandra Luati (Chair in Statistics, Imperial College London & University of Bologna)

Acknowledgments:

- UK Research and Innovation / EPSRC ECR International Collaboration Grant 2024-2026 on *"Spectral embedding methods and subsequent inference tasks on dynamic multiplex graphs"* (EP/Y002113/1) in collaboration with Professor Carey Priebe (Johns Hopkins University)

Undirected graphs

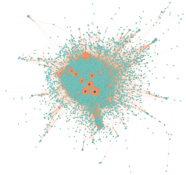
- **Undirected graph** $\mathbb{G} = (V, E)$ where:
  - V is the **node set**, with cardinality $n = |V|$,
  - $E \subseteq V \times V$ is the **edge set**, containing dyads $(i, j) \in V \times V$.
- An edge is drawn if a node $i \in V$ connects to $j \in V$, written $(i, j) \in E$.
- From $\mathbb{G}$, an **adjacency matrix** $\mathbf{A} \in \{0,1\}^{n \times n}$, can be obtained via $A_{i,j} = \mathbb{1}_E\{(i,j)\}$.
- Real-world graphs tend to be more complex. For example:



Latent position models for graphs

- **Latent position models** (Hoff, Raftery, and Handcock, 2002) for **adjacency matrices**:

$$\boldsymbol{x}_i \overset{iid}{\sim} \mathcal{F} \quad \rightarrow \quad \mathbb{P}(A_{i,j} = 1 \mid \boldsymbol{x}_i, \boldsymbol{x}_j) = \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) \quad \rightarrow$$



- LPMs illustrate a powerful idea for network modelling: **expressing edge-specific quantities through unobserved node features** $\boldsymbol{x}_i \in \mathcal{X} \subseteq \mathbb{R}^d$ sampled from a distribution $\mathcal{F}$.
- Node features are "linked" to link probabilities via a **kernel function** $\kappa : \mathcal{X} \times \mathcal{X} \to [0, 1]$.
- **Inner product kernels** $\to$ **random dot product graph** (RDPG, Athreya et al., 2018):

$$A_{i,j} \mid \boldsymbol{x}_1, \dots, \boldsymbol{x}_n \sim \text{Bernoulli}(\boldsymbol{x}_i^\intercal \boldsymbol{x}_j).$$

Random dot product graphs

- RDPGs (and their generalisation, GRDPG, see Rubin-Delanchy et al., 2022) include:
  - Stochastic blockmodels (Holland, Laskey, and Leinhardt, 1983): $\boldsymbol{x}_i = \boldsymbol{\mu}_{z_i}$ for a community $z_i \in \{1, \dots, K\}$, giving a between-community constant connection probability $B_{k\ell} = \boldsymbol{\mu}_k^\intercal \boldsymbol{\mu}_\ell$;
  - Degree-corrected stochastic blockmodels (Karrer and Newman, 2011): $\boldsymbol{x}_i = \rho_i \boldsymbol{\mu}_{z_i}$ for community $z_i \in \{1, \dots, K\}$ and degree-correction parameter $\rho_i \in (0, 1)$.
- The latent positions can be **estimated via the spectral decomposition** of $\mathbf{A}$.

Definition (ASE – Adjacency spectral embedding)

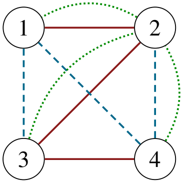
For an integer $d \in \{1, \dots, n\}$ and a binary *symmetric* adjacency matrix $\mathbf{A} \in \{0, 1\}^{n \times n}$, the d-dimensional adjacency spectral embedding (ASE) $\hat{\mathbf{X}} = [\hat{\boldsymbol{x}}_1, \dots, \hat{\boldsymbol{x}}_n]^\intercal$ of $\mathbf{A}$ is

$$\hat{\mathbf{X}} = \boldsymbol{\Gamma} |\boldsymbol{\Lambda}|^{1/2} \in \mathbb{R}^{n \times d},$$

where $\boldsymbol{\Lambda}$ is a $d \times d$ diagonal matrix containing the d largest eigenvalues in magnitude, and $\boldsymbol{\Gamma}$ is a $n \times d$ matrix containing corresponding orthonormal eigenvectors.

Multiplex graphs

- In many real-world applications, edges can have **different types**. For example, links in cyber-security applications occur on different ports. In transportation networks, there are different means of transport between two locations. Edge types are usually called **layers**.
- **Undirected multiplex graph** $\mathbb{G} = (V, \{E_1, \dots, E_K\})$ where:
  - V is the **shared node set** across layers, with cardinality $n = |V|$,
  - $E_k \subseteq V \times V$ is the **edge set** for the k-th layer, containing dyads $(i, j) \in V \times V$.
  - Denote the adjacency matrix for the k-th layer as $\mathbf{A}_k$ [It is **not assumed** that $E_k \cap E_\ell = \varnothing$].



Layer 1

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Layer 2

$$\mathbf{A}_2 = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

Layer 3

$$\mathbf{A}_3 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Spectral embedding of multiplex graphs

- Multiplex graphs can mainly be spectrally embedded via two methods: **OMNI** and **UASE**.
- **Omnibus** embedding (OMNI, Levin et al., 2017) – Take ASE of the following block-matrix:

$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_1 & (\mathbf{A}_1+\mathbf{A}_2)/2 & \cdots & (\mathbf{A}_1+\mathbf{A}_K)/2 \\ (\mathbf{A}_2+\mathbf{A}_1)/2 & \mathbf{A}_2 & \cdots & (\mathbf{A}_2+\mathbf{A}_K)/2 \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{A}_K+\mathbf{A}_1)/2 & (\mathbf{A}_K+\mathbf{A}_2)/2 & \cdots & \mathbf{A}_K \end{bmatrix},$$

- **Unfolded** adjacency spectral embedding (UASE, Jones and Rubin-Delanchy, 2020; Gallagher, Jones, and Rubin-Delanchy, 2021) – Obtain the embedding from the singular value the composition of the unfolded matrix $\tilde{\mathbf{A}} \in \{0,1\}^{n\times nK}$, defined as follows:

$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_K \end{bmatrix}.$$

Dynamic multiplex graphs

- In real-world applications, graphs **evolve over time** $\to$ **dynamic multiplex graphs**.
- **Dynamic multiplex graph** $\mathbb{G} = (V, \{E_{k,t}\}_{k=1,\dots,K,\ t=1,\dots,T})$ where:
  - V is the **shared node set** across layers and time points, with cardinality $n = |V|$.
  - $E_{k,t} \subseteq V \times V$ is the **edge set** for the k-th layer at the t-th time.
  - Denote the adjacency matrix for the k-th layer at the t-th time as $\mathbf{A}_{k,t}$.
- We propose a **dynamic multiplex RDPG (DMP-RDPG)** model where each node is represented by latent positions in $\mathcal{X}_k \subseteq \mathbb{R}^d$ and $\mathcal{Y}_t \subseteq \mathbb{R}^d$, $k = 1, \dots, K,\ t = 1, \dots, T$.
  - Positions $\boldsymbol{x}_{i,k} \in \mathcal{X}_k$ are shared across time but are different across layers.
  - Positions $\boldsymbol{y}_{j,t} \in \mathcal{Y}_t$ are shared across layers but vary over time.
  - The connectivity model for nodes i and j at time t in layer k is given by:

$$A_{i,j,k,t} \sim \text{Bernoulli}\left(\boldsymbol{x}_{i,k}^\intercal \boldsymbol{y}_{j,t}\right).$$

DMP-RDPG: THE SETUP

- For K and T fixed, let $\mathcal{X}_1, \dots, \mathcal{X}_K, \mathcal{Y}_1, \dots, \mathcal{Y}_T \subseteq \mathbb{R}^d$ for some shared $d \in \mathbb{N}$, such that $\boldsymbol{x}^\intercal \boldsymbol{y} \in [0,1]$ for any $\boldsymbol{x} \in \mathcal{X}_k$ and $\boldsymbol{y} \in \mathcal{Y}_t$, for all $k = 1, \dots, K$ and $t = 1, \dots, T$.
- Let $\mathcal{F}$ be a distribution over the product space $[\bigotimes_{k=1}^K \mathcal{X}_k^n] \bigotimes [\bigotimes_{t=1}^T \mathcal{Y}_t^n]$.
- Let $\boldsymbol{x}_{1,1}, \dots, \boldsymbol{x}_{n,1}, \dots, \boldsymbol{x}_{n,K}, \boldsymbol{y}_{1,1}, \dots, \boldsymbol{y}_{n,1}, \dots, \boldsymbol{y}_{n,T} \sim \mathcal{F}$. By organising these positions into matrices $\mathbf{X}_k = [\boldsymbol{x}_{1,k}^\intercal, \cdots, \boldsymbol{x}_{n,k}^\intercal]^\intercal$ and $\mathbf{Y}_t = [\boldsymbol{y}_{1,t}^\intercal, \dots, \boldsymbol{y}_{n,t}^\intercal]^\intercal$, the stacked matrices $\mathbf{X} = [\mathbf{X}_1^\intercal \mid \cdots \mid \mathbf{X}_K^\intercal]^\intercal \in \mathbb{R}^{nK \times d}$ and $\mathbf{Y} = [\mathbf{Y}_1^\intercal \mid \cdots \mid \mathbf{Y}_T^\intercal]^\intercal \in \mathbb{R}^{nT \times d}$ can be constructed.
- We define the $n \times n$ probability matrices for each time point and layer as $\mathbf{P}_{k,t} = \mathbf{X}_k \mathbf{Y}_t^\intercal$.
- Consider the following $nK \times nT$ **double unfolding** of the matrices $\mathbf{P}_{k,t}$ and $\mathbf{A}_{k,t}$:

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{1,1} & \dots & \mathbf{P}_{1,T} \\ \vdots & \ddots & \vdots \\ \mathbf{P}_{K,1} & \dots & \mathbf{P}_{K,T} \end{bmatrix} = \mathbf{X}\mathbf{Y}^\intercal, \qquad \mathbf{A} = \begin{bmatrix} \mathbf{A}_{1,1} & \dots & \mathbf{A}_{1,T} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{K,1} & \dots & \mathbf{A}_{K,T} \end{bmatrix}.$$

DMP-RDPG: THE DEFINITION

- This setup leads to the definition of the **dynamic multiplex RDPG (DMP-RDPG)** model.

Definition (DMP-RDPG: Dynamic Multiplex Random Dot Product Graph)

$(\mathbf{A}, \mathbf{X}, \mathbf{Y}) \sim \mathrm{DMP\text{-}RDPG}(\mathcal{F})$ if, conditional on $\mathbf{X}_k$ and $\mathbf{Y}_t$, the matrix $\mathbf{A}_{k,t}$ has entries $A_{i,j,k,t} \sim \mathrm{Bernoulli}(P_{i,j,k,t})$ for all $i, j \in V$, with $P_{i,j,k,t} = \boldsymbol{x}_{i,k}^\intercal \boldsymbol{y}_{j,t}$.

- Key features:
  - Positions $\boldsymbol{x}_{i,k} \in \mathcal{X}_k$ are shared across time but are different across layers.
  - Positions $\boldsymbol{y}_{j,t} \in \mathcal{Y}_t$ are shared across layers but vary over time.
- Given a realisation $\mathbf{A}$ from a DMP-RDPG, the inferential objective is to estimate $\mathbf{X}$ and $\mathbf{Y}$. We propose a **doubly unfolded adjacency spectral embedding** estimator.

Doubly unfolded adjacency spectral embedding (DUASE)

Definition (DUASE – Doubly unfolded adjacency spectral embedding)

For a given integer $d \in \{1, \dots, n\}$ and dynamic multiplex graph adjacency matrices $\mathbf{A}_{k,t} \in \{0,1\}^{n \times n}$, $k = 1, \dots, K$, $t = 1, \dots, T$, consider the doubly unfolded matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{1,1} & \dots & \mathbf{A}_{1,T} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{K,1} & \dots & \mathbf{A}_{K,T} \end{bmatrix}.$$

The d-dimensional doubly unfolded adjacency spectral embedding (DUASE) of $\mathbf{A}$ is:

$$\hat{\mathbf{X}} = \mathbf{U}\mathbf{D}^{1/2} \in \mathbb{R}^{nK \times d}, \qquad \hat{\mathbf{Y}} = \mathbf{V}\mathbf{D}^{1/2} \in \mathbb{R}^{nT \times d},$$

where $\mathbf{D}$ is a $d \times d$ diagonal matrix containing the d largest singular values of $\mathbf{A}$, and $\mathbf{U}$ and $\mathbf{V}$ are $nK \times d$ and $nT \times d$ matrices containing the corresponding singular vectors.

Results: two-to-infinity norm bound for DUASE

- This theorem is an adaptation of Theorem 2 in Jones and Rubin-Delanchy, 2020.

Theorem ($2 \to \infty$ norm bound for DUASE)

Let $(\mathbf{A}, \mathbf{X}, \mathbf{Y}) \sim \mathrm{DMP\text{-}RDPG}(\mathcal{F}_\rho)$, *and consider the DUASE estimates* $\hat{\mathbf{X}}$ *and* $\hat{\mathbf{Y}}$ *of* $\mathbf{X}$ *and* $\mathbf{Y}$. *Then there exist a sequence of orthogonal matrices* $\mathbf{W}$ *such that*

$$\|\hat{\mathbf{X}}_k\mathbf{W} - \mathbf{X}_k\|_{2\to\infty} = \mathcal{O}_{\mathbb{P}}\left(\frac{\log^{1/2}(n)}{\rho^{1/2}n^{1/2}}\right), \quad \|\hat{\mathbf{Y}}_t\mathbf{W} - \mathbf{Y}_t\|_{2\to\infty} = \mathcal{O}_{\mathbb{P}}\left(\frac{\log^{1/2}(n)}{\rho^{1/2}n^{1/2}}\right),$$

for each $k = 1, \dots, K$ *and* $t = 1, \dots, T$.

- $\|\mathbf{M}\|_{2\to\infty} = \sup_{\|\boldsymbol{z}\|_2=1} \|\mathbf{M}\boldsymbol{z}\|_\infty$; $\|\mathbf{M}\| = \mathcal{O}_{\mathbb{P}}(f)$ if for any $\alpha > 0$ there exist a constant $C > 0$ and an integer n^* such that for all $n \geq n^*$, $\mathbb{P}\{\|\mathbf{M}\| \leq Cf(n)\} \geq 1 - n^{-\alpha}$.

Underlying assumptions and sparsity considerations

- A global sparsity factor $\rho_n \in (0,1)$ is used to control the asymptotic connection density of the network as the number of nodes in the network n tends to infinity.
- We assume that the sequence ρ_n converges either to 0 or to some constant c.
- The desired sparsity regime is enforced by defining $\boldsymbol{x}_{i,k} = \rho^{1/2}\boldsymbol{\xi}_{i,k}$, $\boldsymbol{\xi}_{i,k} \sim F_{X,k}$ for some distribution $F_{X,k}$ on $\mathbb{R}^d$ and $\boldsymbol{y}_{j,t} = \rho^{1/2}\boldsymbol{\nu}_{j,t},\ \boldsymbol{\nu}_{j,t} \sim F_{Y,t}$. Therefore, the joint distribution $\mathcal{F}$ factorises into the product of n identical marginals for each $k=1,\dots,K,\ t=1,\dots,T$.
- We adopt the notation $\mathcal{F}_\rho$ to refer to this scaled distribution.
- The following assumptions on $\mathcal{F}_\rho$ must hold for each $k=1,\dots,K,\ t=1,\dots,T$:

$$\|\mathbf{X}_k^\intercal \mathbf{X}_k - n\Delta_{X,k}\| = \mathcal{O}_{\mathbb{P}}(n^{1/2}\log^{1/2}(n)), \quad \|\mathbf{Y}_t^\intercal \mathbf{Y}_t - n\Delta_{Y,t}\| = \mathcal{O}_{\mathbb{P}}(n^{1/2}\log^{1/2}(n)),$$

where $\Delta_{X,k} = \mathbb{E}[\boldsymbol{\xi}_{i,k}\boldsymbol{\xi}_{i,k}^\intercal]$ and $\Delta_{Y,t} = \mathbb{E}[\boldsymbol{\nu}_{j,t}\boldsymbol{\nu}_{j,t}^\intercal]$ are the second moment matrices of $F_{X,k}$ and $F_{Y,t}$ respectively.

Results: central limit theorem for DUASE

- This theorem mirrors Theorem 3 in Jones and Rubin-Delanchy, 2020 and Proposition 3 in Gallagher, Jones, and Rubin-Delanchy, 2021.

Theorem (CLT for DUASE)

Let $(\mathbf{A}, \mathbf{X}, \mathbf{Y}) \sim \mathrm{DMP\text{-}RDPG}(\mathcal{F}_\rho)$. Given $\boldsymbol{x} \in \mathcal{X}_k$ and $\boldsymbol{y} \in \mathcal{Y}_t$ then for all $\boldsymbol{z} \in \mathbb{R}^d$ and for any fixed node i, there exist a sequence of orthogonal matrices $\mathbf{W}$ such that

$$\mathbb{P}\left\{n^{1/2}(\hat{\mathbf{X}}_k\mathbf{W} - \mathbf{X}_k)_i^\intercal \leq \boldsymbol{z} \mid \boldsymbol{\xi}_{i,k} = \boldsymbol{x}\right\} \to \Phi\left\{\boldsymbol{z}; \boldsymbol{\Sigma}_{Y,k}(\boldsymbol{x})\right\},$$
$$\mathbb{P}\left\{n^{1/2}(\hat{\mathbf{Y}}_t\mathbf{W} - \mathbf{Y}_t)_i^\intercal \leq \boldsymbol{z} \mid \boldsymbol{\nu}_{i,t} = \boldsymbol{y}\right\} \to \Phi\left\{\boldsymbol{z}; \boldsymbol{\Sigma}_{X,t}(\boldsymbol{y})\right\}.$$

where $\Phi(\boldsymbol{z}; \boldsymbol{\Sigma})$ is the CDF of a d-dimensional multivariate normal distribution centred at $\mathbf{0}_d$ with covariance matrix $\boldsymbol{\Sigma}$ evaluated at $\boldsymbol{z}$. The form of $\boldsymbol{\Sigma}_{Y,k}$ and $\boldsymbol{\Sigma}_{X,t}$ is analytically available.

Some key results

- There are some key results needed to prove the two theorems in the previous slides.

Proposition (Control of singular values of $\mathbf{P}$)

Let $\sigma_\ell(\mathbf{P})$ denote the ℓ-th non-zero singular value of $\mathbf{P}$ for $\ell \in \{1, \dots, d\}$. Then:

$$\frac{\sigma_\ell(\mathbf{P})}{\rho n} \to \sqrt{\lambda_\ell(\Delta_X \Delta_Y)},$$

where $\Delta_X = \sum_{k=1}^K \Delta_{X,k}$ and $\Delta_Y = \sum_{t=1}^T \Delta_{Y,t}$. Consequently, $\sigma_\ell(\mathbf{P}) = \Omega_{\mathbb{P}}(\rho n)$ and $\sigma_\ell(\mathbf{P}) = \mathcal{O}_{\mathbb{P}}\left(K^{1/2} T^{1/2} \rho n\right)$.

- K and T must both grow at a rate slower than n: $\lim_{n\to\infty} \max\{K, T\}/n = 0$.

Some key results

- There are some key results needed to prove the two theorems in the previous slides.

Proposition

For any $\alpha > 0$ there exists a constant $C > 0$ and an integer $n^* > 0$ such that for all $n > n^*$:

$$\mathbb{P}\left\{\|\mathbf{A}-\mathbf{P}\| \leq C\rho^{1/2}\max\{T,K\}^{1/2}n^{1/2}\log^{1/2}(n)\right\} \geq 1-n^{-\alpha}.$$

In other words:

$$\|\mathbf{A}-\mathbf{P}\| = \mathcal{O}_{\mathbb{P}}\left(\rho^{1/2}\max\{T,K\}^{1/2}n^{1/2}\log^{1/2}(n)\right).$$

- K and T must both grow at a rate slower than n: $\lim_{n\to\infty}\max\{K,T\}/n = 0$.

Dynamic multiplex stochastic blockmodels (DMP-SBM)

- DMP-RDPG to define a **dynamic multiplex stochastic blockmodel (DMP-SBM)**.
- Assume $A_{i,j,k,t} \sim \mathrm{Bernoulli}(B_{z_{i,k},z'_{j,t},k,t})$, where $z_{i,k} \in \{1,\dots,G\}$ and $z'_{j,t} \in \{1,\dots,G'\}$ are group labels for nodes i and j in the k-th layer and t-th time point respectively, and $\mathbf{B} \in [0,1]^{G\times G'\times K\times T}$ is a tensor of probabilities of connections between groups.
- Under a DMP-RDPG representation, $B_{h,\ell,k,t} = \boldsymbol{\mu}_{h,k}^\intercal \boldsymbol{\mu}'_{\ell,t}$ for $\boldsymbol{\mu}_{h,k}, \boldsymbol{\mu}'_{\ell,t} \in \mathbb{R}^d$, which gives:

$$A_{i,j,k,t} \sim \mathrm{Bernoulli}(\boldsymbol{\mu}_{z_{i,k},k}^\intercal \boldsymbol{\mu}'_{z'_{j,t},t}).$$

- The indicators $z_{i,k}$ and $z'_{j,t}$ can be estimated via **Gaussian mixture modelling** on the output of **DUASE**, using the theoretical guarantees provided by our **DUASE-CLT**.

DMP-SBM + DUASE: A SIMULATION

- Simulate a DMP-SBM with $G = 3$, $G' = 4$ and the following connection probabilities:

$$\mathbf{B}_{1,1} = \begin{bmatrix} 0.08 & 0.02 & 0.18 & 0.10 \\ 0.02 & 0.20 & 0.04 & 0.10 \\ 0.18 & 0.04 & 0.02 & 0.02 \\ 0.10 & 0.10 & 0.02 & 0.06 \end{bmatrix}, \quad \mathbf{B}_{1,2} = \begin{bmatrix} 0.16 & 0.16 & 0.04 & 0.10 \\ 0.16 & 0.16 & 0.04 & 0.10 \\ 0.04 & 0.04 & 0.09 & 0.02 \\ 0.10 & 0.10 & 0.02 & 0.06 \end{bmatrix}, \quad \mathbf{B}_{1,3} = \begin{bmatrix} 0.08 & 0.02 & 0.18 & 0.10 \\ 0.02 & 0.20 & 0.04 & 0.10 \\ 0.18 & 0.04 & 0.02 & 0.02 \\ 0.10 & 0.10 & 0.02 & 0.06 \end{bmatrix}$$

$$\mathbf{B}_{2,1} = \begin{bmatrix} 0.08 & 0.02 & 0.18 & 0.10 \\ 0.02 & 0.20 & 0.04 & 0.10 \\ 0.18 & 0.04 & 0.02 & 0.02 \\ 0.10 & 0.10 & 0.02 & 0.06 \end{bmatrix}, \quad \mathbf{B}_{2,2} = \begin{bmatrix} 0.16 & 0.16 & 0.04 & 0.10 \\ 0.16 & 0.16 & 0.04 & 0.10 \\ 0.04 & 0.04 & 0.09 & 0.02 \\ 0.10 & 0.10 & 0.02 & 0.06 \end{bmatrix}, \quad \mathbf{B}_{2,3} = \begin{bmatrix} 0.08 & 0.02 & 0.18 & 0.10 \\ 0.02 & 0.20 & 0.04 & 0.10 \\ 0.18 & 0.04 & 0.02 & 0.02 \\ 0.10 & 0.10 & 0.02 & 0.06 \end{bmatrix}$$

$$\mathbf{B}_{3,1} = \begin{bmatrix} 0.08 & 0.08 & 0.08 & 0.08 \\ 0.08 & 0.08 & 0.08 & 0.08 \\ 0.08 & 0.08 & 0.08 & 0.08 \\ 0.08 & 0.08 & 0.08 & 0.08 \end{bmatrix}, \quad \mathbf{B}_{3,2} = \begin{bmatrix} 0.08 & 0.08 & 0.08 & 0.08 \\ 0.08 & 0.08 & 0.08 & 0.08 \\ 0.08 & 0.08 & 0.08 & 0.08 \\ 0.08 & 0.08 & 0.08 & 0.08 \end{bmatrix}, \quad \mathbf{B}_{3,3} = \begin{bmatrix} 0.08 & 0.08 & 0.08 & 0.08 \\ 0.08 & 0.08 & 0.08 & 0.08 \\ 0.08 & 0.08 & 0.08 & 0.08 \\ 0.08 & 0.08 & 0.08 & 0.08 \end{bmatrix}$$

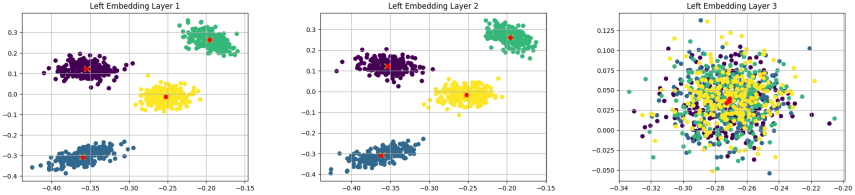


Figure 1. Left DUASE of a simulated DMP-SBM against the true latent positions (after orthogonal Procrustes rotation).

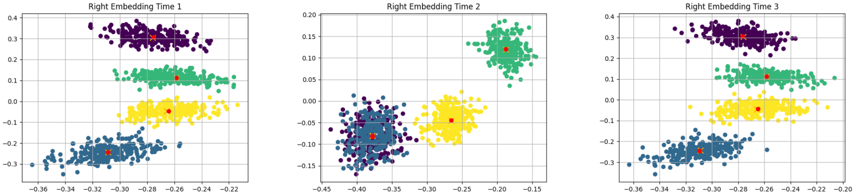


Figure 2. Right DUASE of a simulated DMP-SBM against the true latent positions (after orthogonal Procrustes rotation).

Possible connections with literature on Euclidean mirrors

- Consider the right DUASE embedding $\hat{\mathbf{Y}} = [\hat{\mathbf{Y}}_1^\intercal \mid \cdots \mid \hat{\mathbf{Y}}_T^\intercal]^\intercal \in \mathbb{R}^{nT\times d}$ and extract the sequence of **aligned** time-specific embeddings $\hat{\mathbf{Y}}_1, \dots, \hat{\mathbf{Y}}_T$.
- Athreya et al., 2024 calculates a dissimilarity (distance) matrix $\hat{\mathcal{D}}_\phi$ with entries:

$$\hat{\mathcal{D}}_\phi(t,s) = \hat{d}_{\mathrm{MV}}(\hat{\mathbf{Y}}_t, \hat{\mathbf{Y}}_s) = \min_{\mathbf{Q}\in\mathbb{O}(d)} \frac{1}{\sqrt{n}} \|\hat{\mathbf{Y}}_t - \hat{\mathbf{Y}}_s\mathbf{Q}\|_2,$$

  where $\mathbb{O}(d)$ is the orthogonal group with signature d.
- Since DUASE has **temporal stability**, one can simply set $\hat{\mathcal{D}}_\phi(t,s) = n^{-1/2}\|\hat{\mathbf{Y}}_t - \hat{\mathbf{Y}}_s\|_2$.
- Apply CMDS to $\hat{\mathcal{D}}_\phi$ to estimate $\hat{\psi}(t) \in \mathbb{R}^c$ for $t=1,\dots,T$, for some $c \in \mathbb{N}$.
- Apply ISOMAP to the points in $\mathrm{CMDS}(\hat{\mathcal{D}}_\phi) = \{\hat{\psi}(t),\ t=1,\dots,T\} \subseteq \mathbb{R}^c$ to obtain a 1-dimensional curve, which can be plotted against the time indices $t=1,\dots,T$.
- This gives a **joint (across layers) Euclidean mirror for dynamic multiplex networks**.

Vector autoregressive models for multivariate time series

- **Panels of multivariate time series** $\{\mathbf{X}_t = (X_{1,t}, \ldots, X_{N,t})^\intercal,\ X_{i,t} \in \mathbb{R}\}_{t\in\mathbb{Z}}$ exhibiting **co-movement between components** are central to many scientific disciplines such as environmental science, econometrics, and neuroscience.
- Often, $X_{i,t}$ depends not only on its own past values, but also on the **past values of a subset of other panel components**, $\{X_{j,s} : j \subseteq [N],\ s < t\}$, where $[N] = \{1, \ldots, N\}$.
- **Vector autoregression (VAR)** is a widely used model for multivariate time series:

$$\mathbf{X}_t = \mathbf{\Phi}\mathbf{X}_{t-1} + \boldsymbol{\varepsilon}_t,\ \boldsymbol{\varepsilon}_t \sim \mathbb{N}_N(\mathbf{0}_N, \mathbf{\Sigma}_\varepsilon),$$

  where $\mathbf{\Phi} \in \mathbb{R}^{N\times N}$ is a matrix of **coefficients**, for some covariance $\mathbf{\Sigma}_\varepsilon$.
- For large N, modelling via the VAR framework becomes prohibitive as the number of model parameters grows as $\mathcal{O}(N^2)$ and can quickly exceed the number of observations.

Related literature

- **Factor models** – The large panel of time series are modelled as stemming from a relatively small number of common **latent factors** (Stock and Watson, 2002).
- **Factors with sparse regression** – Fan, Masini, and Medeiros, 2023 combine the dimensionality reduction of factor modelling with the parsimony of sparse linear regression and give a novel test for covariance structure. Their proposed model is called the **Factor Augmented Regression Model (FARM)**.
- **Network VAR** – Knight et al., 2020 introduce generalise network autoregression (**GNAR**) which, given an observed network, fits a flexible network autoregressive model. Barigozzi, Cho, and Owens, 2023 propose an L1-regularised Yule-Walker method for estimating a factor adjusted, idiosyncratic VAR model (**FNETS**).
- **Community detection**: Guðmundsson and Brownlees, 2021 use estimated VAR coefficients to **embed and cluster** the panel components.

Network-informed restricted vector autoregressive (NIRVAR) model

Definition (NIRVAR – Network-informed restricted vector autoregressive model)

Consider a multiplex network $\mathbb{G} = (V, \{E_1, \dots, E_Q\})$ over $V = [N]$ with adjacency matrices $\mathbf{A}^{(q)} = \{A_{ij}^{(q)}\}$, $q = 1, \dots, Q$. Let $\{\mathbf{X}_t^{(q)}\}_{t\in\mathbb{Z}}$ denote a zero mean, second order stationary stochastic process where $\mathbf{X}_t^{(q)} = (X_{1,t}^{(q)}, \cdots, X_{N,t}^{(q)})^\intercal \in \mathbb{R}^N$ and $q \in [Q]$. The NIRVAR model for the q^{th} feature of time series $i \in [N]$ is:

$$X_{i,t}^{(q)} = \sum_{r=1}^{Q} \sum_{j=1}^{N} A_{ij}^{(r)} \tilde{\Phi}_{ij}^{(r)} X_{j,t-1}^{(r)} + \varepsilon_{i,t}^{(q)}, \qquad \varepsilon_{i,t}^{(q)} \sim \mathbb{N}(0, \sigma^2).$$

The $N \times N$ coefficient matrices $\tilde{\mathbf{\Phi}}^{(q)} = \{\tilde{\Phi}_{ij}^{(q)}\}$, $q = 1, \dots, Q$ are distributed according to a distribution $\mathcal{W}^{(q)}$, such that $\rho(\tilde{\mathbf{\Phi}}^{(q)}) < 1$ with probability one to ensure stationarity.

NIRVAR MODEL

- Defining $\mathbf{\Phi}^{(q)} = \mathbf{A}^{(q)} \odot \tilde{\mathbf{\Phi}}^{(q)}$, we can write the NIRVAR model as

$$\mathbf{X}_t^{(q)} = \sum_{r=1}^{Q} \mathbf{\Phi}^{(r)} \mathbf{X}_{t-1}^{(r)} + \boldsymbol{\varepsilon}_t^{(q)}, \qquad \boldsymbol{\varepsilon}_t^{(q)} \sim \mathbb{N}_N(\mathbf{0}_N, \sigma^2 \mathbf{I}_{N\times N}),$$

  which allows the interpretation of the NIRVAR model as a restricted VAR whose restrictions are determined by the graph. Note that $\mathcal{G}^{(q)}$, and hence the restrictions are static.
- In NIRVAR, we define the coefficient matrices $\mathbf{\Phi}^{(q)}$ as the adjacency matrix resulting from a **weighted SBM (WSBM)** (Gallagher et al., 2023), for a suitable choice of the distribution of the weights. A **zero-inflated** component is added to the distribution to obtain sparsity.
- For simplicity, assume $Q = 1$ (only one *feature*).

Weighted stochastic blockmodels

- Let $\mathcal{Z} = \{1, \dots, K\}$ be a sample space with distribution $\mathcal{F}$, and $\{H(z_1, z_2) : z_1, z_2 \in \mathcal{Z}\}$ be a family of real valued distributions satisfying:
  1. $H(z_1, z_2) = H(z_2, z_1)$ for all $z_1, z_2 \in \mathcal{Z}$.
  2. There exists $\phi : \mathcal{Z} \to \mathbb{R}^d$ s.t., for all $z_1, z_2 \in \mathcal{Z}$, if $A \sim H(z_1, z_2)$ then $\mathbb{E}(A) = \phi(z_1)^\intercal \phi(z_2)$.

Definition (WSBM – Weighted SBM)

Let $\mathbf{B} \in \mathbb{R}^{K\times K}$ and $\mathbf{C} \in \mathbb{R}_+^{K\times K}$ denote block mean and block variance matrices. A symmetric matrix, $\mathbf{A} \in \mathbb{R}^{N\times N}$ is distributed as a WSBM if $z_1, \dots, z_N \sim F$ and for $i < j$,

$$A_{ij} \mid z_i, z_j \sim H(z_i, z_j),$$

where the mean and variance of $H(k, \ell)$ are given by $B_{k,\ell}$ and $C_{k,\ell}$ respectively.

WSBMs: the setup

- Consider the eigendecomposition $\mathbf{B} = \mathbf{U}_B \mathbf{\Lambda}_B \mathbf{U}_B^\intercal$.
- Constructing the spectral embedding $\mathbf{Y}_B = \mathbf{U}_B \mathbf{\Lambda}_B^{1/2} = [\boldsymbol{\mu}_1^\intercal, \dots, \boldsymbol{\mu}_K^\intercal]^\intercal$, we define the map $\phi : \mathcal{Z} \to \mathbb{R}^K$ as $\phi(k) = \boldsymbol{\mu}_k$. If node i belongs to community $z_i \in \{1, \dots, K\}$, then:

$$\boldsymbol{y}_i = \phi(z_i) = \boldsymbol{\mu}_{z_i}.$$

- Since $\mathbf{B} = \mathbf{Y}_B \mathbf{Y}_B^\intercal$, the expected value of $\mathbf{A}$ has rank $K = d$.
- If $\mathbf{\Phi}$ is a realisation of a WSBM, then $\mathbb{E}(\Phi_{i,j}) = B_{z_i,z_j}$.
- For a realisation of $\mathbf{\Phi}$, consider the eigendecomposition:

$$\mathbf{\Phi} = \mathbf{U}_\Phi \mathbf{\Lambda}_\Phi \mathbf{U}_\Phi^\intercal + \mathbf{U}_{\Phi\perp} \mathbf{\Lambda}_{\Phi\perp} \mathbf{U}_{\Phi\perp}^\intercal.$$

- The corresponding left embedding is:

$$\mathbf{Y}_\Phi = \mathbf{U}_\Phi |\mathbf{\Lambda}_\Phi|^{1/2}.$$

From WSBMs to the NIRVAR covariance matrix

- The VAR covariance matrix $\mathbf{\Gamma} = \mathbb{E}(\mathbf{X}_t\mathbf{X}_t^\intercal)$ satisfies $\mathbf{\Gamma} - \mathbf{\Phi}\mathbf{\Gamma}\mathbf{\Phi}^\intercal = \mathbf{\Sigma}_\varepsilon$ (Lütkepohl, 2005).
- This is a Lyapunov matrix equation. For $\mathbf{\Sigma}_\varepsilon = \sigma^2\mathbf{I}_{N\times N}$, it has the following solution:

$$\mathbf{\Gamma} = \sigma^2 \sum_{k=0}^{\infty} \mathbf{\Phi}^k \mathbf{\Phi}^{\intercal k},$$

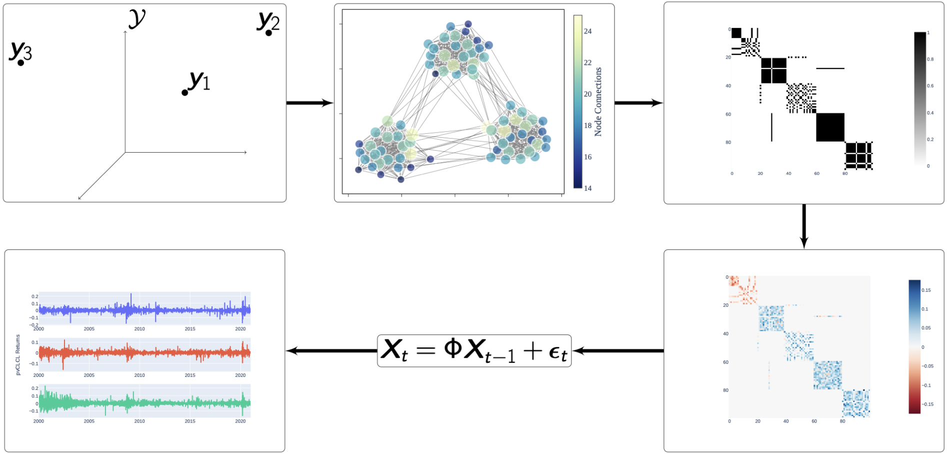
  converging only when $\rho(\mathbf{\Phi}) < 1$ (Smith, 1968; Young, 1981).
- Substituting $\mathbf{\Phi}$ with its spectral decomposition $\mathbf{\Phi} = \mathbf{U}_\Phi\mathbf{\Lambda}_\Phi\mathbf{U}_\Phi^\intercal + \mathbf{U}_{\Phi\perp}\mathbf{\Lambda}_{\Phi\perp}\mathbf{U}_{\Phi\perp}^\intercal$ gives:

$$\mathbf{\Gamma} = \sigma^2\left(\mathbf{I}_{N\times N} - \mathbf{U}_\Phi\mathbf{U}_\Phi^\intercal - \mathbf{U}_{\Phi\perp}\mathbf{U}_{\Phi\perp}^\intercal + \mathbf{U}_\Phi\mathbf{\Lambda}_\Gamma\mathbf{U}_\Phi^\intercal + \mathbf{U}_{\Phi\perp}\mathbf{\Lambda}_{\Gamma\perp}\mathbf{U}_{\Phi\perp}^\intercal\right),$$

  where $\mathbf{\Lambda}_\Gamma$ is a diagonal matrix whose diagonal entries are $\gamma_i = g(\varphi_i) = 1/(1-\varphi_i^2)$ with φ_i being the i^{th} diagonal entry of $\mathbf{\Lambda}_\Phi$ (sorted in order of **magnitude**).
- The rank d spectral embedding of $\mathbf{\Gamma}$ is therefore:

$$\mathbf{Y}_\Gamma = \mathbf{U}_\Phi\mathbf{\Lambda}_\Gamma^{1/2} = \mathbf{U}_\Phi g(\mathbf{\Lambda}_\Phi)^{1/2}.$$

NIRVAR: THE MODEL



$$X_t = \Phi X_{t-1} + \epsilon_t$$

NIRVAR: ESTIMATION

- Recovering the edge set is difficult. Instead, we aim to **recover the community memberships** $z_i,\ i=1,\dots,N$ for each node.
- The two step estimation approach is:
  1. Estimate $\hat{z}_i$ by performing **clustering** on an embedding $\hat{\mathbf{Y}} \in \mathbb{R}^{N\times K}$ of the panel components, and define the binary matrix $\hat{\mathbf{A}} = \{\hat{A}_{i,j}\} \in \{0,1\}^{N\times N}$ with the following entries:

     $$\hat{A}_{i,j} = \mathbb{1}\left\{\hat{z}_i = \hat{z}_j\right\},$$

     where $\mathbb{1}\{\cdot\}$ is the indicator function.
  2. Set $\hat{\Phi}_{i,j} = 0$ if $\hat{A}_{i,j} = 0$ and estimate the remaining **unrestricted parameters** (corresponding to $\hat{A}_{i,j} = 1$) via ordinary least squares (OLS).

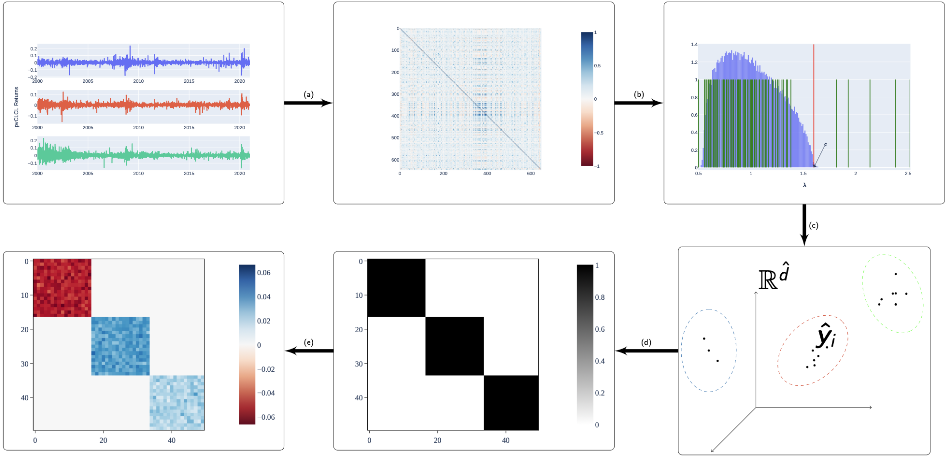
NIRVAR: recovering the communities

- Compute the sample covariance matrix $\mathbf{S}_T$ between panel components.
- Estimate the dimension of the latent space as the number of eigenvalues of $\mathbf{S}_T$ that are greater than the **Marčencko-Pastur distribution cutoff** (Marčenko and Pastur, 1967).
- Let $\mathbf{\Lambda} \in \mathbb{R}^{\hat{d}\times\hat{d}}$ and $\mathbf{U} \in \mathbb{R}^{N\times\hat{d}}$ be the matrices containing the $\hat{d}$ largest eigenvalues and corresponding eigenvectors of $\mathbf{S}_T$. The corresponding embedding is:

$$\hat{\mathbf{Y}} = \mathbf{U}\mathbf{\Lambda}^{1/2} \in \mathbb{R}^{N\times\hat{d}}.$$

- Cluster each $\hat{\boldsymbol{y}}_i \in \mathbb{R}^{\hat{d}}$ into $K = \hat{d}$ groups using a Gaussian Mixture Model.

NIRVAR: ESTIMATION



The NIRVAR estimator: properties

- The NIRVAR estimator is biased whenever $\hat{A}_{i,j} = 0$ but $A_{i,j} = 1$ (model misspecification) and **unbiased** otherwise.

Proposition (Consistency of NIRVAR estimator)

The NIRVAR estimator, $\hat{\gamma}(\hat{\mathbf{A}})$ *is a consistent estimator of* $C\boldsymbol{\gamma}(\mathbf{A})$*, where* C *is the bias of the estimator, and*

$$\sqrt{T}\left\{\hat{\gamma}(\hat{\mathbf{A}}) - C\boldsymbol{\gamma}(\mathbf{A})\right\} \xrightarrow{d} \mathbb{N}_{N^2Q}\left(\mathbf{0}_{N^2Q}, \left\{R(\hat{\mathbf{A}})' \left(\boldsymbol{\Gamma} \otimes \boldsymbol{\Sigma}_{\varepsilon}^{-1}\right) R(\hat{\mathbf{A}})\right\}^{-1}\right),$$

where $\boldsymbol{\Gamma} = \mathbb{E}(\mathbf{X}_t \mathbf{X}_t^\intercal) = \mathrm{plim}(\mathbf{Z}\mathbf{Z}^\intercal)/T$.

The NIRVAR estimator: a reparametrisation

- The model can be written in terms of an unrestricted M-dimensional vector $\boldsymbol{\gamma}(\mathbf{A})$, whose elements are those in the set $\{\mathrm{vec}(\mathbf{A} \odot \tilde{\boldsymbol{\Phi}})_i : \mathrm{vec}(\mathbf{A} \odot \tilde{\boldsymbol{\Phi}})_i \neq 0, i = 1, \ldots, N^2Q\}$, and a $N^2Q \times M$ matrix $R(\mathbf{A})$, where the map, $R : \{0,1\}^{N \times N} \to \{0,1\}^{N^2Q \times M}$, is defined by

$$[R(\mathbf{A})]_{i,j} = \mathrm{vec}(\mathbf{A})_i \mathbb{1}\left\{\sum_{k=1}^{i-1} \mathrm{vec}(\mathbf{A})_k = j - 1\right\}.$$

- The constraints on the model parameters then become:

$$\boldsymbol{\beta} = \mathrm{vec}(\boldsymbol{\Phi}) = R(\mathbf{A})\boldsymbol{\gamma}(\mathbf{A}).$$

- This form is particularly useful to prove results about the NIRVAR model and estimator.

Macroeconomic applications: FRED-MD dataset

- FRED-MD is a publicly accessible database of monthly observations of macroeconomic variables (McCracken and Ng, 2016).
- The prediction task is one-step ahead forecasts of the first order difference of the logarithm of the monthly industrial production (IP) index.
- We backtest NIRVAR, FARM, FNETS, and GNAR from January 2000 - December 2019 using a rolling window framework with a lookback window of 480 observations.

Metric	NIRVAR	FARM	FNETS	GNAR
Overall MSE	**0.0087**	0.0089	0.0096	0.0101

Table 1. Overall MSE of each model for the task of forecasting US industrial production.

Macroeconomic applications: FRED-MD dataset

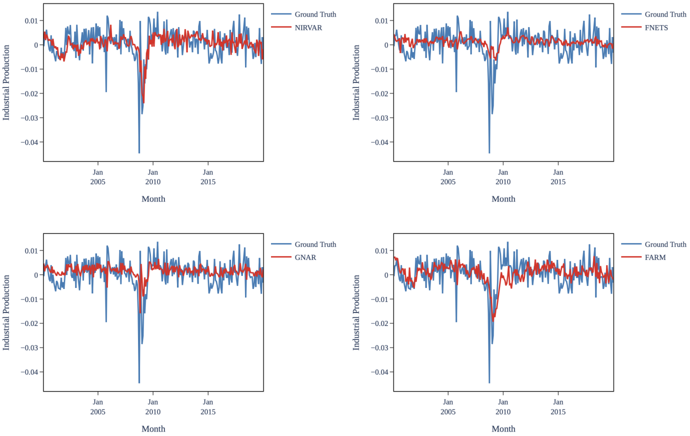


Figure 3. Predicted (log-differenced) IP against the realised (log-differenced) IP for each model.

Financial Returns Prediction

- The previous close-to-close (pvCLCL) market excess returns of 648 financial assets from 03/01/2000 - 31/12/2020 were derived from databases provided by the CRSP.
- The task is to predict the sign of the next day pvCLCL market excess returns (a positive (negative) sign corresponds to a long (short) position in the asset).
- We backtest NIRVAR, FARM, FNETS, and GNAR using a rolling window from 01/01/2004 - 31/12/2020 with a look-back window of four years.

Metric	NIRVAR	FARM	FNETS	GNAR
Sharpe Ratio	**2.82**	0.22	0.78	0.70
Sortino Ratio	**4.80**	0.36	1.39	1.13
Mean Turnover (%)	50.3	**51.1**	50.0	43.0
Maximum Drawdown (%)	**61**	531	107	257
Hit Ratio (%)	**50.7**	48.7	50.2	41.5
Long Ratio (%)	**50.1**	49.0	**50.1**	40.7
Mean Daily PnL (bpts)	**3.00**	0.44	0.89	1.10

Table 2. Statistics on the financial returns predictive performance over the backtesting period.

Financial Returns Prediction

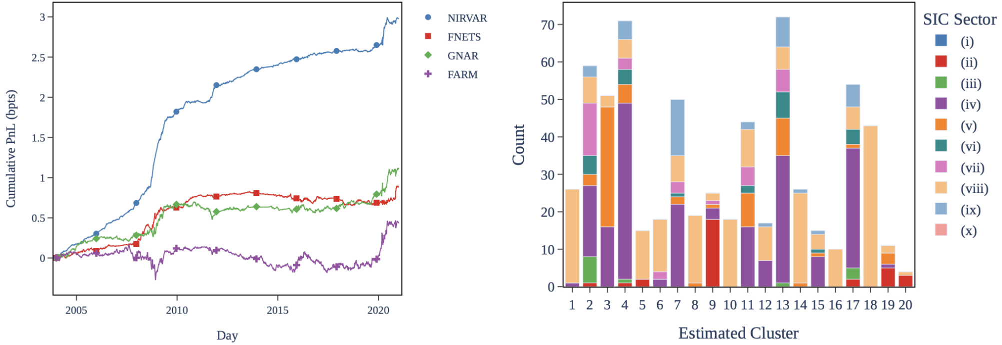


Figure 4. Left: the cumulative PnL in bpts over the backtesting period. Right: comparison of the NIRVAR estimated clusters with the SIC groups on 31/12/2020.

COVID-19 ventilation beds admissions

- Define $Y_{i,t}$ to be the number of COVID-19 patients transferred to mechanical ventilation beds on t, for NHS Trust i, across $N = 140$ trusts in England, for $T = 452$ (Nason, Salnikov, and Cortina-Borja, 2023), between April 2020 and July 2021. Data are available in the UK Government Coronavirus Dashboard (`coronavirus.data.gov.uk`).
- Nason, Salnikov, and Cortina-Borja, 2023 consider the transformation $\tilde{Y}_{i,t} = \log(Y_{i,t} + 1)$.

Model	NIRVAR Covariance	NIRVAR Precision	GNAR(1,[1])	GNAR(1,[1])*	FARM	FNETS
MSPE	3.930 (0.845)	**3.372 (0.911)**	7.414 (2.977)	7.313 (2.752)	4.189 (0.340)	7.646 (3.016)

Table 3. Mean squared prediction errors (MSPE) and standard deviations over 10 predictions for the ventilation beds data.

Conclusion

- This talk introduces two different models and related estimators:
  1. **DMP-RDPG + DUASE for dynamic multiplex graphs**
     - Propose a DMP-RDPG model for dynamic multiplex graphs where $A_{i,j,t,k} \sim \text{Bernoulli}(\boldsymbol{x}_{i,k}^\intercal \boldsymbol{y}_{j,t})$.
     - Positions $\boldsymbol{x}_{i,k} \in \mathcal{X}_k$ are shared across time but are **different across layers**.
     - Positions $\boldsymbol{y}_{j,t} \in \mathcal{Y}_t$ are shared across layers but **vary over time**.
     - Estimates of the latent position are obtained via a **double unfolding** of the adjacency matrices.
  2. **NIRVAR model and estimator for panels of multivariate time series**
     - We model the time series as a VAR process whose parameter matrix is a **realisation of a WSBM**.
     - We introduce an estimation framework that firstly determines the **restrictions** to be placed on the VAR parameters and secondly estimates the remaining unrestricted parameters via OLS.
     - The method can be used to estimate VAR parameters when the underlying network is **unobserved**.

Acknowledgements

- UK Research and Innovation / EPSRC ECR International Collaboration Grant 2024-2026 on *"Spectral embedding methods and subsequent inference tasks on dynamic multiplex graphs"* (EP/Y002113/1) in collaboration with Professor Carey Priebe (Johns Hopkins University)

$$\Downarrow$$

As part of this funding, I will occasionally spend time at JHU in the next two years.
For now, I will be here until 30th April, 2024. If you are interested in these topics, let's chat!

Thank you!

References I

- Athreya, A. et al. (2018). "Statistical Inference on Random Dot Product Graphs: a Survey". In: *Journal of Machine Learning Research* 18.226, pp. 1–92.
- Athreya, A. et al. (2024). "Euclidean mirrors and dynamics in network time series". In: *Journal of the American Statistical Association (to appear).*
- Barigozzi, M., H. Cho, and D. Owens (2023). "FNETS: Factor-adjusted network estimation and forecasting for high-dimensional time series". In: *Journal of Business & Economic Statistics*, pp. 1–13.
- Fan, J., R. P. Masini, and M. C. Medeiros (2023). "Bridging factor and sparse models". In: *The Annals of Statistics* 51.4, pp. 1692–1717.
- Gallagher, I., A. Jones, and P. Rubin-Delanchy (2021). "Spectral embedding for dynamic networks with stability guarantees". In: *Advances in Neural Information Processing Systems.* Vol. 34, pp. 10158–10170.
- Gallagher, I. et al. (2023). "Spectral embedding of weighted graphs". In: *Journal of the American Statistical Association*, pp. 1–10.
- Guðmundsson, G. S. and C. Brownlees (2021). "Detecting groups in large vector autoregressions". In: *Journal of Econometrics* 225.1, pp. 2–26.

References II

- Hoff, P. D, A. E. Raftery, and M. S. Handcock (2002). "Latent space approaches to social network analysis". In: *Journal of the American Statistical Association* 97, pp. 1090–1098.
- Holland, P. W., K. B. Laskey, and S. Leinhardt (1983). "Stochastic blockmodels: First steps". In: *Social Networks* 5.2, pp. 109 –137.
- Jones, A. and P. Rubin-Delanchy (2020). "The multilayer random dot product graph". In: *arXiv e-prints.* arXiv: 2007.10455 [stat.ML].
- Karrer, B. and M. E. J. Newman (2011). "Stochastic blockmodels and community structure in networks". In: *Physical Review E* 83 (1).
- Knight, M. et al. (2020). "Generalized Network Autoregressive Processes and the GNAR Package". In: *Journal of Statistical Software* 96, pp. 1–36.
- Levin, K. et al. (2017). "A central limit theorem for an omnibus embedding of multiple random graphs and implications for multiscale network inference". In: *arXiv e-prints.* arXiv: 1705.09355 [stat.ME].
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis.* Springer Science & Business Media.

References III

- Marčenko, V. A. and L. A. Pastur (1967). "Distribution of eigenvalues for some sets of random matrices". In: *Matematicheskii Sbornik* 114.4, pp. 507–536.
- McCracken, M. W. and S. Ng (2016). "FRED-MD: A monthly database for macroeconomic research". In: *Journal of Business & Economic Statistics* 34.4, pp. 574–589.
- Nason, G., D. Salnikov, and M. Cortina-Borja (2023). "New tools for network time series with an application to COVID-19 hospitalisations". In: *arXiv e-prints*. arXiv: `2312.00530 [stat.ME]`.
- Rubin-Delanchy, P. et al. (2022). "A Statistical Interpretation of Spectral Embedding: The Generalised Random Dot Product Graph". In: *Journal of the Royal Statistical Society Series B* 84.4, pp. 1446–1473.
- Smith, R. (1968). "Matrix Equation XA+BX=C". In: *SIAM Journal on Applied Mathematics* 16.1, pp. 198–201.
- Stock, J. H. and M. W. Watson (2002). "Forecasting using principal components from a large number of predictors". In: *Journal of the American Statistical Association* 97.460, pp. 1167–1179.
- Young, N. (1981). "The rate of convergence of a matrix power series". In: *Linear Algebra and its Applications* 35, pp. 261–278.