Statistical Learning & Inference Seminars - University of Warwick

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

8th October, 2024

Joint work with:

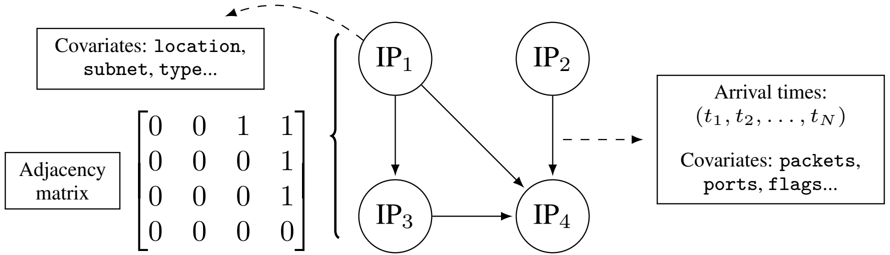
- **Dynamic multiplex graphs**
  - **Maximilian Baum** (PhD student, Imperial College London)
  - Axel Gandy (Chair in Statistics, Imperial College London)
- **Panels of multivariate time series / vector autoregressive processes**
  - **Brendan Martin** (PhD student, Imperial College London)
  - Mihai 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

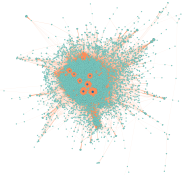
- **Undirected graph** $\mathbb{G} = (V, E)$ where:
  - V is the **node set**, with cardinality $n = |V|$,
  - $E \subseteq V \times V$ is the **edge set**, containing dyads $(i, j) \in V \times V$.
- An edge is drawn if a node $i \in V$ connects to $j \in V$, written $(i, j) \in E$.
- From $\mathbb{G}$, 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** $\kappa : \mathcal{X} \times \mathcal{X} \rightarrow [0, 1]$.
- **Inner product kernels** $\rightarrow$ **random dot product graph** (RDPG, Athreya et al., 2018):

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

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)

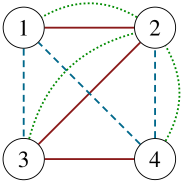
For an integer $d \in \{1, \dots, n\}$ and a binary *symmetric* adjacency matrix $\mathbf{A} \in \{0, 1\}^{n \times n}$, the d-dimensional adjacency spectral embedding (ASE) $\hat{\mathbf{X}} = [\hat{\boldsymbol{x}}_1, \dots, \hat{\boldsymbol{x}}_n]^\intercal$ of $\mathbf{A}$ is

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

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 **Layer 2** **Layer 3**

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \mathbf{A}_2 = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \quad \mathbf{A}_3 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

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)

$(\mathbf{A}, \mathbf{X}, \mathbf{Y}) \sim \text{DMP-RDPG}(\mathcal{F})$ if, conditional on $\mathbf{X}_k$ and $\mathbf{Y}_t$, the matrix $\mathbf{A}_{k,t}$ has entries $A_{i,j,k,t} \sim \text{Bernoulli}(P_{i,j,k,t})$ 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 $(\mathbf{A}, \mathbf{X}, \mathbf{Y}) \sim \mathrm{DMPRDPG}(F_\rho)$ *with* K_n *layers and* T_n *time points. Then, for each* $k \in [K_n]$ *and* $t \in [T_n]$, *there exist sequences of matrices* $\mathbf{W}_X$ *and* $\mathbf{W}_Y \in \mathrm{GL}(d)$ *such that*

$$\|\hat{\mathbf{X}}^k \mathbf{W}_X^{-1} - \mathbf{X}^k\|_{2\to\infty} = O_{\mathbb{P}}\left\{\frac{\log^{1/2}(n)}{\rho_n^{1/2} n^{1/2} T_n^{1/2}}\right\},$$

$$\|\hat{\mathbf{Y}}^t \mathbf{W}_Y^{-1} - \mathbf{Y}^t\|_{2\to\infty} = O_{\mathbb{P}}\left\{\frac{\log^{1/2}(n)}{\rho_n^{1/2} n^{1/2} K_n^{1/2}}\right\}.$$

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

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.
- For our main results to hold in the asymptotic regime where K_n and T_n tend to infinity, we further require the existence of the $d\times d$ matrices $\tilde{\boldsymbol{\Delta}}_X = \lim_{n\to\infty} K_n^{-1}\sum_{k=1}^{K_n}\boldsymbol{\Delta}_{X,k}$ and $\tilde{\boldsymbol{\Delta}}_Y = \lim_{n\to\infty} T_n^{-1}\sum_{t=1}^{T_n}\boldsymbol{\Delta}_{Y,t}$. where $\tilde{\boldsymbol{\Delta}}_{X,k} = \mathbb{E}[\boldsymbol{\xi}_{i,k}\boldsymbol{\xi}_{i,k}^\intercal]$ and $\tilde{\boldsymbol{\Delta}}_{Y,t} = \mathbb{E}[\boldsymbol{\nu}_{j,t}\boldsymbol{\nu}_{j,t}^\intercal]$ 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.

Theorem (CLT for DUASE)

Let $(\mathbf{A}, \mathbf{X}, \mathbf{Y}) \sim \mathrm{DMPRDPG}(F_\rho)$ *with* K_n *layers and* T_n *time points. Given latent positions* $\boldsymbol{x} \in \mathcal{X}_k$ *and* $\boldsymbol{y} \in \mathcal{Y}_t$*, then for all* $\boldsymbol{z} \in \mathbb{R}^d$ *and for any fixed* $i \in [n]$, $k \in [K_n]$ *and* $t \in [T_n]$ *there exist sequences of matrices* $\mathbf{W}_X, \mathbf{W}_Y \in \mathrm{GL}(d)$ *(dependent on* n*) such that, for* $n \to \infty$*:*

$$\mathbb{P}\left\{n^{1/2}T_n^{1/2}(\hat{\mathbf{X}}_k\mathbf{W}_X^{-1} - \mathbf{X}_k)_i^\intercal \leq \boldsymbol{z} \mid \boldsymbol{\xi}_{i,k} = \boldsymbol{x}\right\} \to \Phi\left\{\boldsymbol{z}, \tilde{\boldsymbol{\Delta}}_Y^{-1}\mathbf{V}_Y(\boldsymbol{x})\tilde{\boldsymbol{\Delta}}_Y^{-1}\right\},$$

$$\mathbb{P}\left\{n^{1/2}K_n^{1/2}(\hat{\mathbf{Y}}_t\mathbf{W}_Y^{-1} - \mathbf{Y}_t)_i^\intercal \leq \boldsymbol{z} \mid \boldsymbol{\nu}_{i,t} = \boldsymbol{y}\right\} \to \Phi\left\{\boldsymbol{z}, \tilde{\boldsymbol{\Delta}}_X^{-1}\mathbf{V}_X(\boldsymbol{y})\tilde{\boldsymbol{\Delta}}_X^{-1}\right\},$$

where $\Phi(\boldsymbol{z}, \boldsymbol{\Sigma})$ *is the CDF of a* d*-dimensional normal distribution centered at* $\mathbf{0}$ *(the identically zero vector of dimension* d*), with covariance matrix* $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$*, evaluated at* $\boldsymbol{z} \in \mathbb{R}^d$*. The form of* $\mathbf{V}_Y$ *and* $\mathbf{V}_X$ *is analytically available.*

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, $B_{h,\ell,k,t} = \boldsymbol{\mu}_{h,k}^\intercal \boldsymbol{\lambda}_{\ell,t}$ for $\boldsymbol{\mu}_{h,k}, \boldsymbol{\lambda}_{\ell,t} \in \mathbb{R}^d$, which gives:

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

- 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**:

$$\mathbb{P}\left\{n^{1/2}T_n^{1/2}\left(\mathbf{W}_X^{-1}\hat{\boldsymbol{x}}_{i,k} - \boldsymbol{\mu}_{g,k}\right)^\intercal \leq \boldsymbol{q}\right\} \to \Phi\left\{\boldsymbol{q}, \boldsymbol{\Sigma}_{X,g}\right\}, \quad g \in [G],\ \boldsymbol{q} \in \mathbb{R}^d,$$

$$\mathbb{P}\left\{n^{1/2}K_n^{1/2}\left(\mathbf{W}_Y^{-1}\hat{\boldsymbol{y}}_{i,t} - \boldsymbol{\lambda}_{g,t}\right)^\intercal \leq \boldsymbol{q}\right\} \to \Phi\left\{\boldsymbol{q}, \boldsymbol{\Sigma}_{Y,g}\right\}, \quad g \in [G'],\ \boldsymbol{q} \in \mathbb{R}^d.$$

DMP-SBM + DUASE: A SIMULATION

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

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

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

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

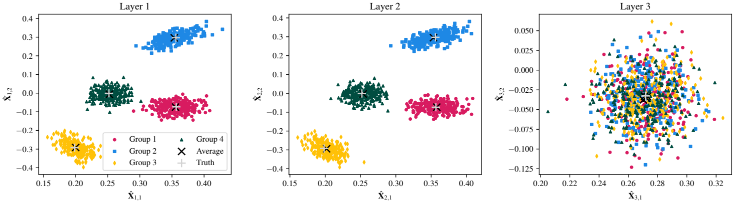


Figure 1. Left DUASE of a simulated DMP-SBM against the true latent positions (after orthogonal Procrustes rotation).

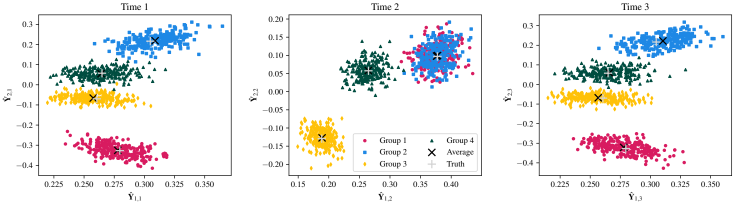


Figure 2. Right DUASE of a simulated DMP-SBM against the true latent positions (after orthogonal Procrustes rotation).

Connections with literature on Euclidean mirrors

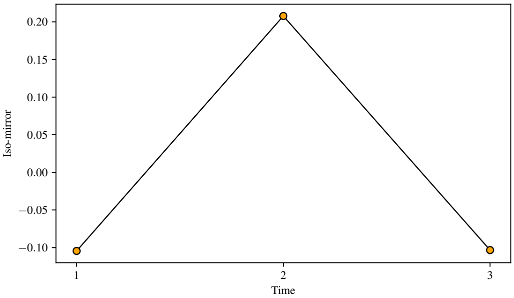
- Consider the right DUASE embedding $\hat{\mathbf{Y}} = [\hat{\mathbf{Y}}_1^\intercal \mid \cdots \mid \hat{\mathbf{Y}}_T^\intercal]^\intercal \in \mathbb{R}^{nT\times d}$ and extract the sequence of **aligned** time-specific embeddings $\hat{\mathbf{Y}}_1, \dots, \hat{\mathbf{Y}}_T$.
- Athreya et al., 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**.

Iso-mirror on SBM with DUASE

(a) Iso-mirror for $T = 3$ time indices



(b) Iso-mirror for $K = 3$ layers

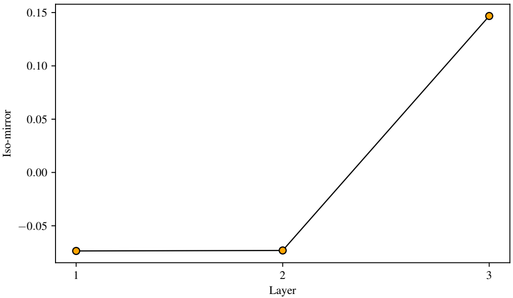


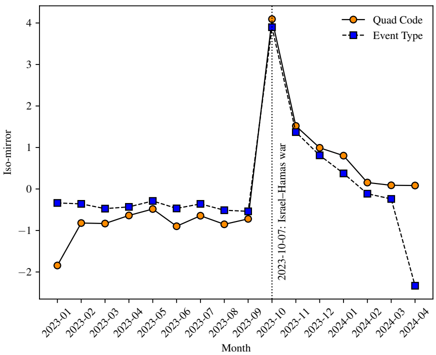
Figure 3. Iso-mirrors calculated from DUASE on the simulated DMP-SBM in previous slides.

Applications on real-world multiplex networks

- **PLOVER dataset** (Halterman et al., 2023)
  - Consider the Political Language Ontology for Verifiable Event Records (PLOVER) dataset, consisting in $624{,}888$ political interaction events between $n = 104$ countries.
  - We group the events across $T = 16$ months ranging between January 2023 and April 2024. Each event is associated with one of $K = 16$ event types based on the PLOVER categories.
  - Each of the event types is further grouped into $K^* = 4$ macro-groups called *quad categories*: material cooperation, verbal cooperation, verbal conflict, and material conflict.
- **FinDKG dataset** (Li and Sanna Passino, 2024)
  - The graph has a total of $241{,}948$ edges between $n = 13{,}637$ nodes, with $K = 15$ different connection types related to financial concepts, such as “Raise”, “Invests_In” or “Produce”.
  - Nodes represent financial institutions, politicians, businessmen, countries, financial concepts, and commodities.

PLOVER DATASET

(a) Iso-mirror on right DUASE $\hat{\mathbf{Y}}$



(b) Iso-mirror on $\hat{\mathbf{X}}$, grouped by event type

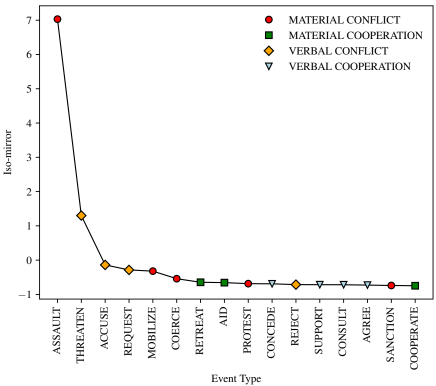


Figure 4. Iso-mirror across time and event types on the POLECAT data.

PLOVER DATASET

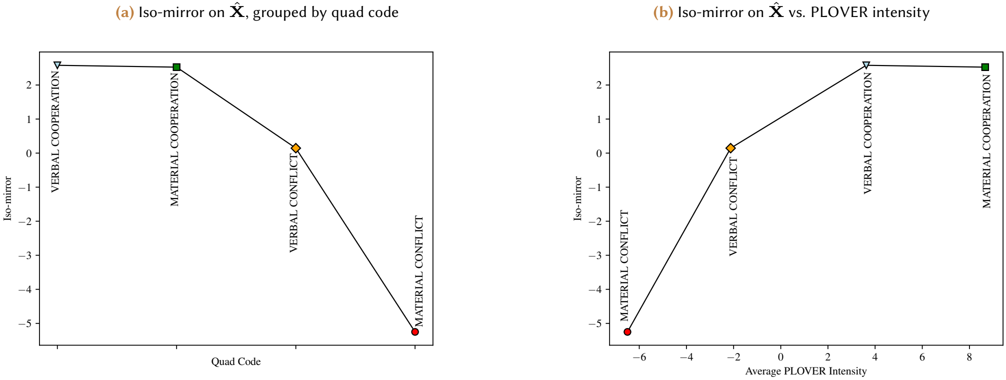
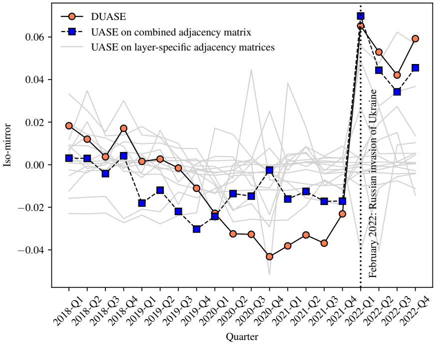


Figure 5. Iso-mirror across time and event types on the POLECAT data, grouped by quad-code.

FinDKG dataset

(a) Iso-mirror for right DUASE $\hat{\mathbf{Y}}$



(b) Iso-mirror for left DUASE $\hat{\mathbf{X}}$

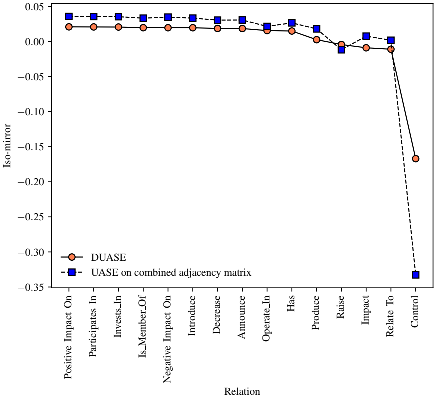


Figure 6. Iso-mirror across time and event types on the FinDKG data.

Vector autoregressive models for multivariate time series

- **Panels of multivariate time series** $\{\mathbf{X}_t = (X_{1,t}, \dots, X_{N,t})^\intercal,\ X_{i,t} \in \mathbb{R}\}_{t\in\mathbb{Z}}$ 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 $[N] = \{1, \dots, N\}$.
- **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 structure $\mathbf{\Sigma}_\varepsilon$.
- For large N, modelling via the VAR framework becomes **prohibitive** as the number of model parameters grows as $O(N^2)$ 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)

For some fixed $q \in [Q]$, 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 is

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

in which $(\mathbf{A}^{(r)}, \mathbf{Y}^{(r)}) \sim \mathrm{SBM}(\mathbf{B}^{(r)}, \boldsymbol{\pi}^{(r)})$, $r \in [Q]$, and $\tilde{\mathbf{\Phi}}^{(r)}, r \in [Q]$, is an $N \times N$ matrix of fixed weights. Defining $\mathbf{\Phi}^{(r)} = \mathbf{A}^{(r)} \odot \tilde{\mathbf{\Phi}}^{(r)}$ and $\mathbf{\Phi} = (\mathbf{\Phi}^{(1)} | \cdots | \mathbf{\Phi}^{(Q)})$, we write $\mathbf{X}_t^{(q)} \sim \mathrm{NIRVAR}(\mathbf{\Phi})$.

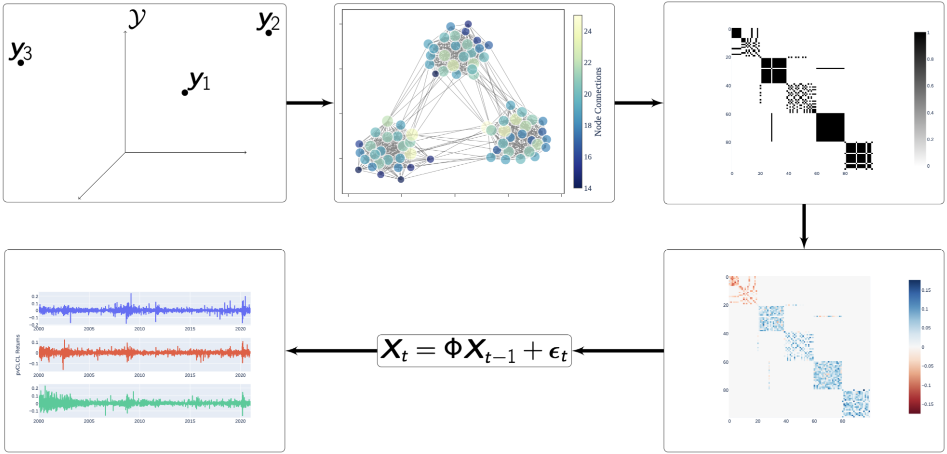
NIRVAR model

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

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

  which allows the interpretation of the NIRVAR model as a **restricted VAR whose restrictions are determined by the graph**. Note that **the restrictions are static**.
- Equivalently, we can define the coefficient matrices $\mathbf{\Phi}^{(r)}$ as the adjacency matrix resulting from a **weighted SBM** (**WSBM**; Gallagher, Jones, and Rubin-Delanchy, 2021), for a suitable choice of the weight distribution. A **zero-inflated** component is added to the distribution to obtain sparsity.

NIRVAR: the model



NIRVAR: ESTIMATION

- Recovering the **complete 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 $\hat{\mathbf{Y}} \in \mathbb{R}^{N \times d}$ 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:
$$\hat{A}_{i,j} = \mathbb{1}\{\hat{z}_i = \hat{z}_j\},$$
     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).
- If multiple features are used, the clustering can be done on **feature-specific embeddings** $\hat{\mathbf{Y}}^{(q)} \in \mathbb{R}^{N \times d}$, and the same procedure is followed to obtain $\hat{\mathbf{\Phi}}^{(q)}$.
- The **one-step-ahead prediction equation** becomes: $\mathbf{X}_{t+1}^{(q)} = \sum_{r=1}^{Q} \hat{\mathbf{\Phi}}^{(r)} \mathbf{X}_t^{(r)}$.

NIRVAR: EMBEDDING

- Let $\mathfrak{X}^{(q)} = (\boldsymbol{x}_1^{(q)}, \dots, \boldsymbol{x}_T^{(q)})$ be the $N \times T$ design matrix of feature q, where $\boldsymbol{x}_t^{(q)} = (x_{1,t}^{(q)}, \dots, x_{N,t}^{(q)})'$ is a realisation of the random variable $\mathbf{X}_t^{(q)}$.
- To construct an embedding $\hat{\boldsymbol{y}}_i^{(q)} \in \mathbb{R}^d$ we use **unfolded adjacency spectral embedding** (UASE, Jones and Rubin-Delanchy, 2020), which obtains embeddings $\hat{\mathbf{Y}}^{(q)} \in \mathbb{R}^{N\times d}$ by considering the SVD of $\tilde{\mathbf{A}} = (\mathbf{A}^{(1)}|\cdots|\mathbf{A}^{(Q)})$.
- UASE has key **stability properties** (Gallagher, Jones, and Rubin-Delanchy, 2021): it assigns the same position, up to noise, to vertices behaving similarly for a given feature (**cross-sectional stability**) and a constant position, up to noise, to a single vertex behaving similarly across different features (**longitudinal stability**).
- We consider the SVD of $\tilde{\mathbf{S}} = (\mathbf{S}^{(1)}|\cdots|\mathbf{S}^{(Q)}) \in \mathbb{R}^{N\times NQ}$ where $\mathbf{S}^{(q)} = \mathfrak{X}^{(q)}\mathfrak{X}^{(q)\intercal}/T \in \mathbb{R}^{N\times N}$ is the **sample covariance matrix** for feature q.

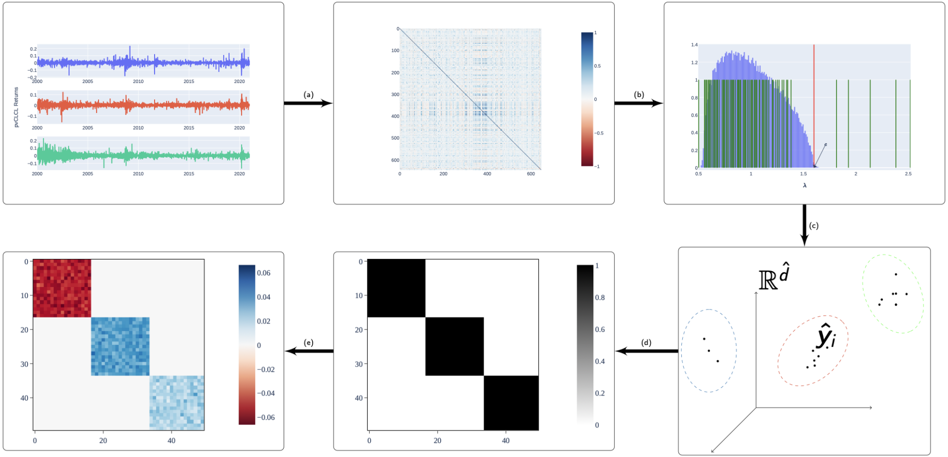
NIRVAR: EMBEDDING

- Estimate the dimension of the latent space as the number of singular values of $\tilde{\mathbf{S}}$ that are greater than the **Marčencko-Pastur distribution cutoff** (Marčenko and Pastur, 1967).
- Let $\mathbf{D} \in \mathbb{R}^{\hat{d}\times\hat{d}}$ and $\mathbf{U} \in \mathbb{R}^{N\times\hat{d}}$ be the matrices containing the $\hat{d}$ largest singular values and corresponding left singular vectors of $\tilde{\mathbf{S}}$. The corresponding **joint** embedding is:

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

- Cluster each $\hat{\boldsymbol{y}}_i \in \mathbb{R}^{\hat{d}}$ into $K = \hat{d}$ groups using a **Gaussian mixture model**.
- **Feature-specific embeddings** $\hat{\mathbf{Y}}^{(q)} \in \mathbb{R}^{N\times\hat{d}}$ can be obtained by **unstacking** $\mathbf{V}\mathbf{D}^{1/2} \in \mathbb{R}^{NQ\times\hat{d}}$ into Q equal blocks $(\hat{\mathbf{Y}}^{(1)};\dots;\hat{\mathbf{Y}}^{(Q)})$, where $\mathbf{V}$ contains the right singular vectors of $\tilde{\mathbf{S}}$ corresponding to the $\hat{d}$ largest singular values.

NIRVAR: embedding & estimation



Weighted stochastic blockmodels

- How does $\hat{\boldsymbol{y}}_i^{(q)}$ relate to the ground truth positions $\boldsymbol{\theta}_{z_i}^{(q)}$? We derive a connection between the covariance $\boldsymbol{\Gamma}^{(q)} = \mathbb{E}\{(\mathbf{X}_t^{(q)})(\mathbf{X}_t^{(q)})^\intercal\}$ of a $\mathrm{NIRVAR}(\Phi)$ process and $\boldsymbol{y}_i^{(q)}$ for $Q = 1$.
- If $\boldsymbol{\Phi}$ is symmetric, the rank-d spectral embedding of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are equivalent.

Proposition

Let $\mathbf{X}_t \sim \mathrm{NIRVAR}(\boldsymbol{\Phi})$ where $\boldsymbol{\Phi}$ is assumed to be symmetric. Consider the eigendecomposition $\boldsymbol{\Phi} = \mathbf{U}_{\boldsymbol{\Phi}}\boldsymbol{\Lambda}_{\boldsymbol{\Phi}}\mathbf{U}_{\boldsymbol{\Phi}}^\intercal + \mathbf{U}_{\boldsymbol{\Phi},\perp}\boldsymbol{\Lambda}_{\boldsymbol{\Phi},\perp}\mathbf{U}_{\boldsymbol{\Phi},\perp}^\intercal$, where $\mathbf{U}_{\boldsymbol{\Phi}} \in \mathbb{O}(N \times d)$ and $\boldsymbol{\Lambda}_{\boldsymbol{\Phi}}$ is a $d \times d$ diagonal matrix comprising the d largest eigenvalues in absolute value of $\boldsymbol{\Phi}$. Then the rank d truncated eigendecomposition of the covariance matrix $\boldsymbol{\Gamma} = \mathbb{E}(\mathbf{X}_t\mathbf{X}_t^\intercal)$ is $\boldsymbol{\Gamma} = \mathbf{U}_{\boldsymbol{\Phi}}\boldsymbol{\Lambda}_{\boldsymbol{\Gamma}}\mathbf{U}_{\boldsymbol{\Phi}}^\intercal$ in which $\boldsymbol{\Lambda}_{\boldsymbol{\Gamma}}$ is a $d \times d$ diagonal matrix with diagonal elements $(\lambda_{\boldsymbol{\Gamma}})_i = 1/\{1 - (\lambda_{\boldsymbol{\Phi}})_i^2\}$ where $(\lambda_{\boldsymbol{\Phi}})_i$ is the corresponding diagonal entry of $\boldsymbol{\Lambda}_{\boldsymbol{\Phi}}$.

Macroeconomic applications: FRED-MD dataset

- FRED-MD is a publicly accessible database of monthly observations of macroeconomic variables.
- 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

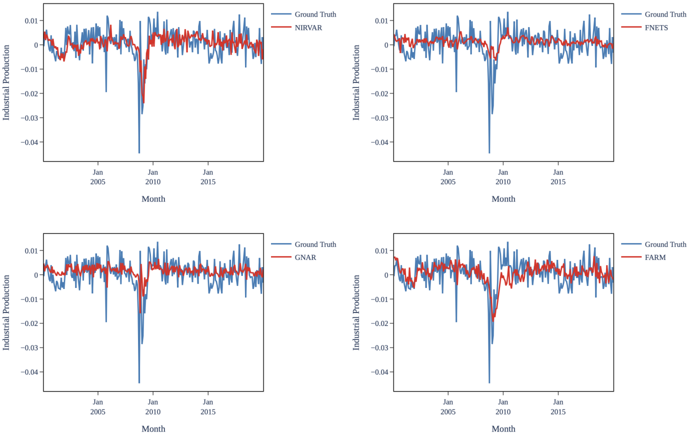


Figure 7. Predicted (log-differenced) IP against the realised (log-differenced) IP for each model.

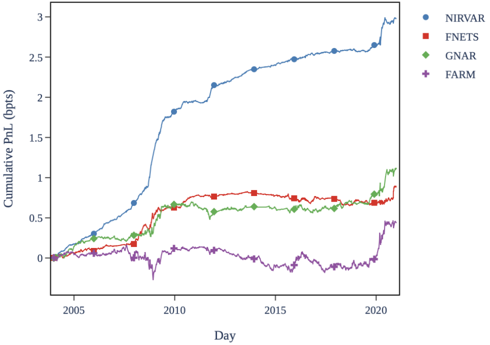
Financial Returns Prediction

- 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.

	NC1	NC2	NP1	NP2	FARM	FNETS	GNAR
Sharpe Ratio	2.50	2.34	**2.82**	2.69	0.22	0.78	0.70
Mean Absolute Error	0.012	0.013	0.012	0.013	0.012	0.012	0.012

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

Financial Returns Prediction



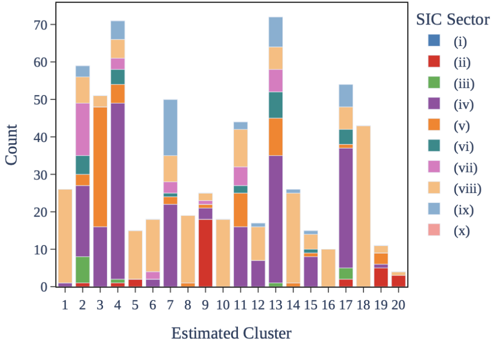


Figure 8. Left: the cumulative PnL in bpts over the backtesting period. Right: comparison of the NIRVAR estimated clusters with the SIC groups on 31/12/2020.

Santander Cycles data

- The first differences of the log daily number of bicycle rides from $N = 774$ Santander Cycles stations in central London from 07/03/2018 until 10/03/2020 ($T = 735$) were obtained using records from TfL Open Data (see `https://cycling.data.tfl.gov.uk/`).
- There were $K = 7$ clusters estimated by NIRVAR, differentiated by their **mean number of bicycle rides** as well as by the change in the number of bicycle rides on **weekdays compared with weekends**.

Model	NIRVAR	FARM	FNETS	GNAR
MSPE	**0.364**	0.370	0.388	0.374

Table 3. Mean squared prediction errors (MSPE) for the Santander cycles data.

Santander Cycles data

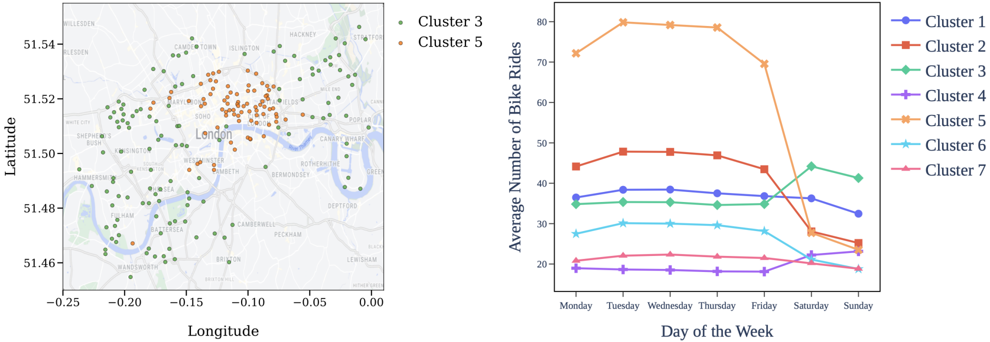


Figure 9. NIRVAR clusters on the Santander Cycles data.

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**.
     - More details can be found in Martin et al. (2024) (preprint available on arXiv).

Martin, B. et al. (2024). "NIRVAR: Network Informed Restricted Vector Autoregression". In: *arXiv e-prints*. arXiv: 2407.13314.

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.
- Guðmundsson, G. S. and C. Brownlees (2021). "Detecting groups in large vector autoregressions". In: *Journal of Econometrics* 225.1, pp. 2–26.
- Halterman, A. et al. (2023). *PLOVER and POLECAT: A New Political Event Ontology and Dataset.* Tech. rep. URL: `osf.io/preprints/socarxiv/rm5dw`.

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].

References III

- Li, X. V. and F. Sanna Passino (July 2024). "FinDKG: Dynamic Knowledge Graphs with Large Language Models for Detecting Global Trends in Financial Markets". In: *arXiv e-prints*, arXiv:2407.10909, arXiv:2407.10909. DOI: 10.48550/arXiv.2407.10909. arXiv: 2407.10909 [q-fin.CP].
- Martin, B. et al. (2024). "NIRVAR: Network Informed Restricted Vector Autoregression". In: *arXiv e-prints*. arXiv: 2407.13314.
- 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.
- 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.
- 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.