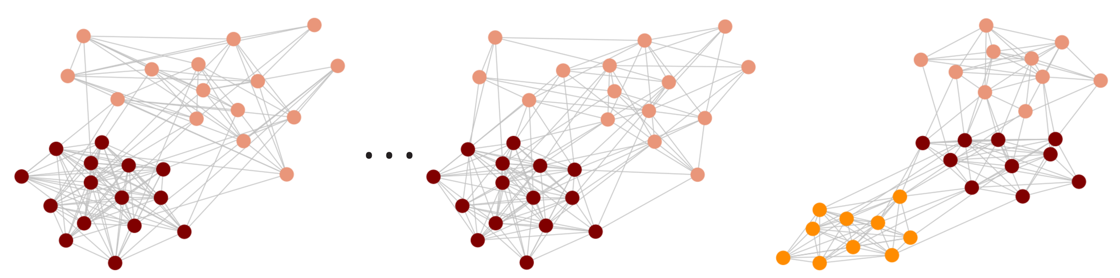
Tracking the Adjacency Spectral Embedding for Streaming Graphs

Paola Bermolen, Marcelo Fiori, Federico Larroca, Bernardo Marenco, Gonzalo Mateos
{paola,mfiori,flarroca,bmarenco}@fing.edu.uy, gmateosb@ur.rochester.edu

Paper Number 1094

Problem statement

- **Given:** Stream of graph observations



- **Model:** Random Dot Product Graph (RDPG) [1]
- **Goal:** Efficiently track nodes latent positions

Contributions and impact:

⇒ Track representation for dynamic graphs

⇒ Scalable and fast computation of nodal representations

⇒ Embedding dynamic networks with varying number of nodes

RDPGs

- Node $i = 1, \dots, n$ has associated latent vector $\mathbf{x}_i \in \mathbb{R}^d$
- Edge (i,j) exists with probability $P_{ij} = \mathbf{x}_i^\top \mathbf{x}_j$
- Notation:
  - $\mathbf{A} \in \{0,1\}^{n \times n}$: adjacency matrix
  - $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^\top \in \mathbb{R}^{n \times d}$
  - $\mathbf{P} = \mathbf{X}\mathbf{X}^\top \in \mathbb{R}^{n \times n}$
- Model is invariant to rotations in $\mathbf{X}$: $\mathbf{P} = \mathbf{X}\mathbf{X}^\top = \mathbf{X}\mathbf{W}(\mathbf{X}\mathbf{W})^\top$ for any orthogonal $\mathbf{W}$
- Expressive model, SBM a special case of RDPG

References

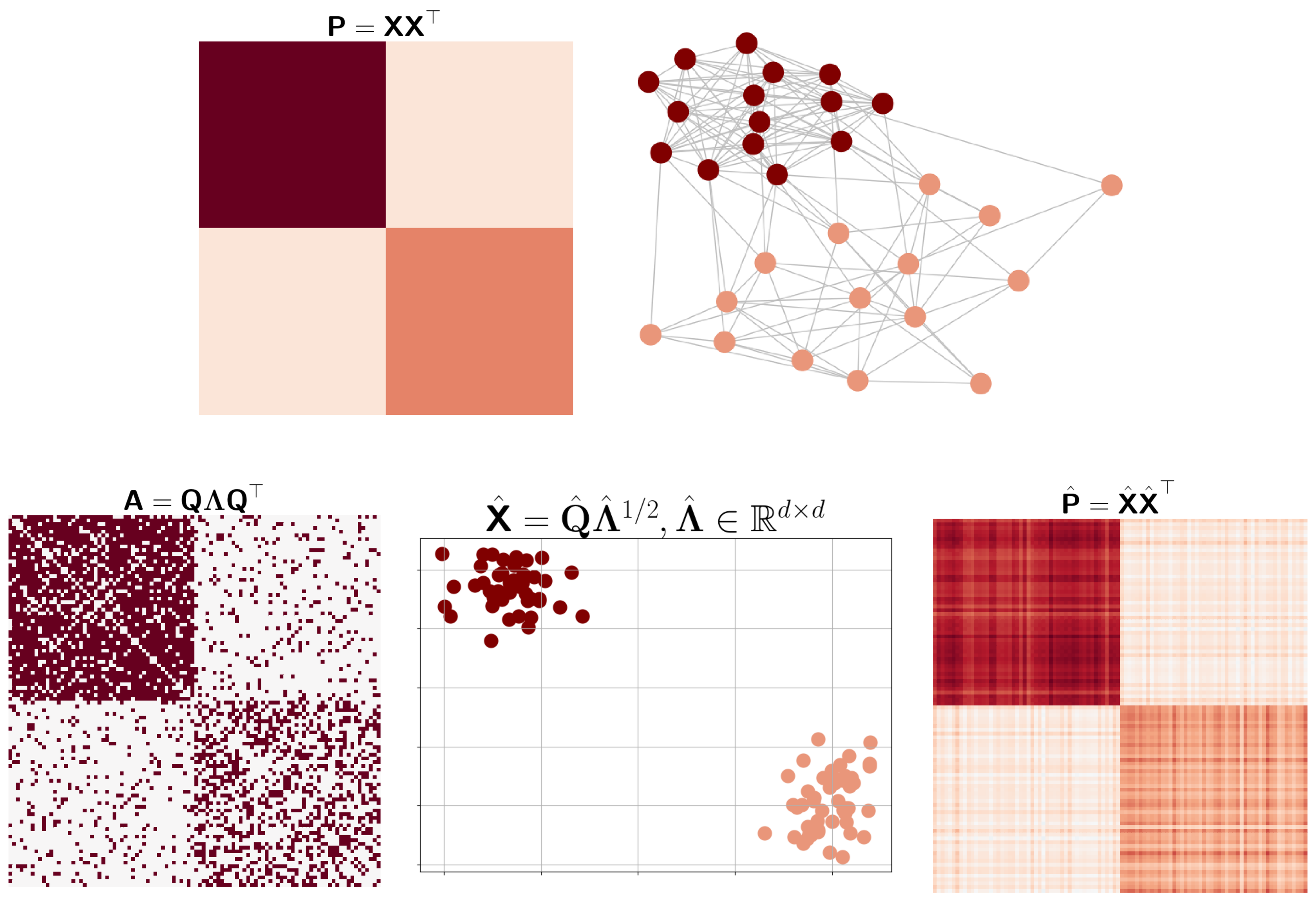
[1] A. Athreya et al. *Statistical Inference on Random Dot Product Graphs: a Survey*. JMLR, 2018

[2] M. Fiori et al. *Algorithmic advances for the adjacency spectral embedding*. EUSIPCO, 2022

[3] M. Brand. *Fast low-rank modifications of the thin singular value decomposition*. Linear algebra and its applications, 2006

[4] K. Levin et al. *Out-of-sample extension of graph adjacency spectral embedding*. ICML, 2018

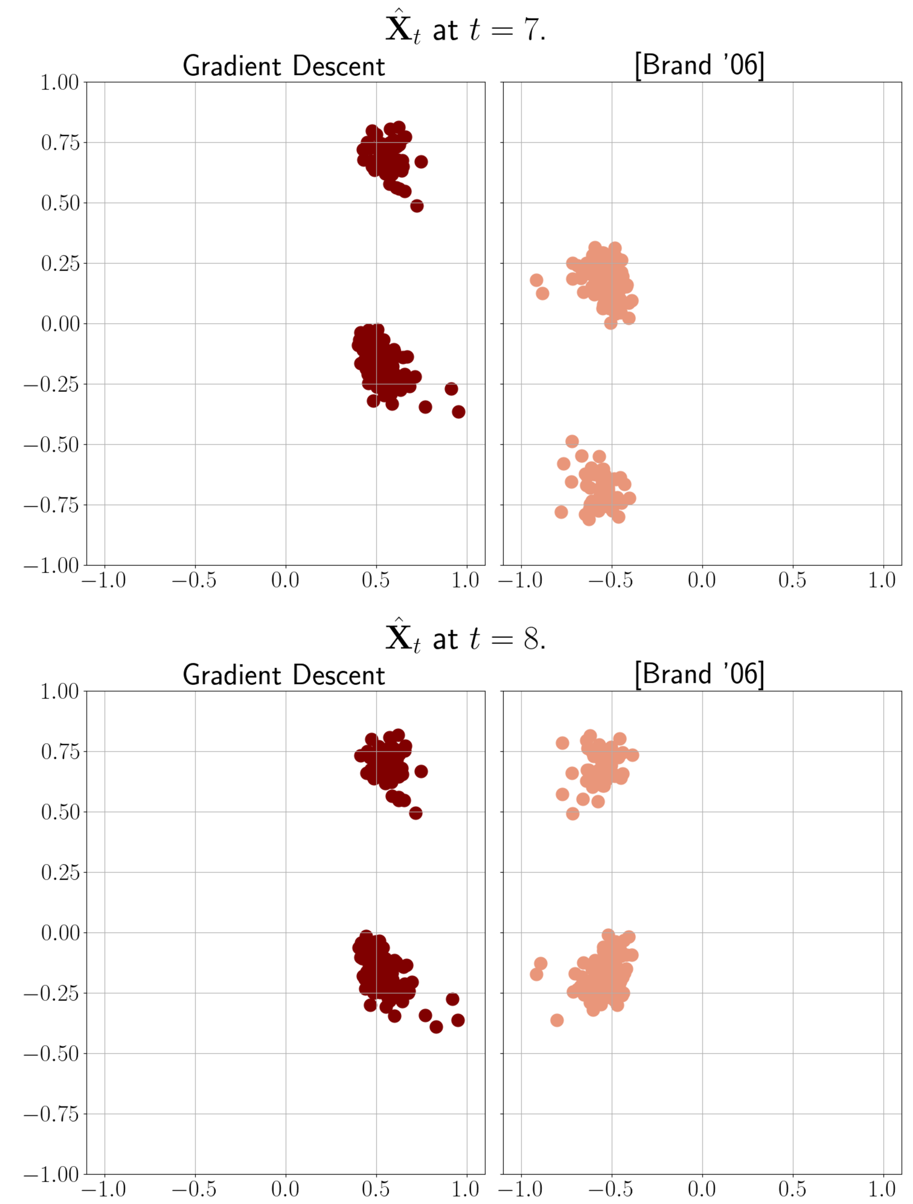
Adjacency Spectral Embedding



- Estimate $\mathbf{X}$ via LS regression: $\hat{\mathbf{X}}_{LS} = \mathrm{argmin}_{\mathbf{X}} \|\mathbf{X}\mathbf{X}^\top - \mathbf{A}\|_F^2 \Rightarrow \hat{\mathbf{X}}_{LS} = \hat{\mathbf{Q}}\hat{\mathbf{\Lambda}}^{1/2}$
- **Q:** Scalability for large graphs? Streaming settings for dynamic graphs? Missing data in $\mathbf{A}$?
- **Gradient descent (GD) approach**: Estimate $\mathbf{X}_{t+1} = \mathbf{X}_t - \alpha \nabla f(\mathbf{X}_t)$ with $f(\mathbf{X}) = \|\mathbf{M} \circ (\mathbf{A} - \mathbf{X}\mathbf{X}^\top)\|_F^2$
- Mask matrix $\mathbf{M} := \mathbf{1}\mathbf{1}^\top - \mathbf{I}$ accounts for all-zero diagonal of $\mathbf{A}$
- Convergence: if $\mathbf{X}_0$ is close to the solution, the iteration converges with linear rate to $\hat{\mathbf{X}}_{LS}$ [2]

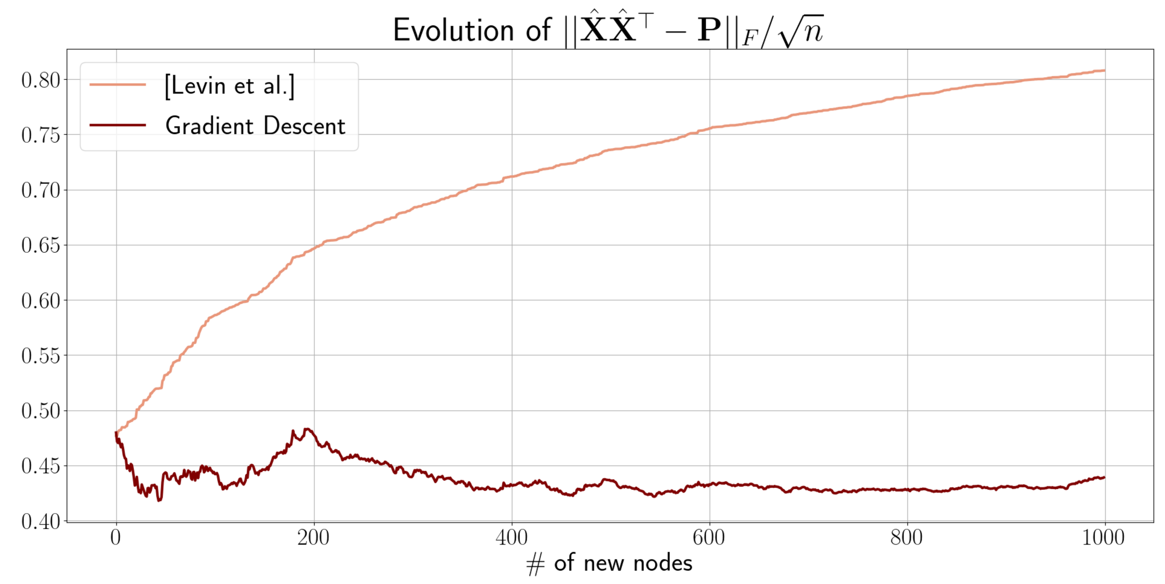
Changes in memberships

- **Alternative approach**: Update SVD decomposition using [3]
- **Limitation**: Does not preserve alignment. Larger estimation error
- Setup: Initially, two-community SBM with 200 nodes. At each timestep, one random node's community assignment is changed



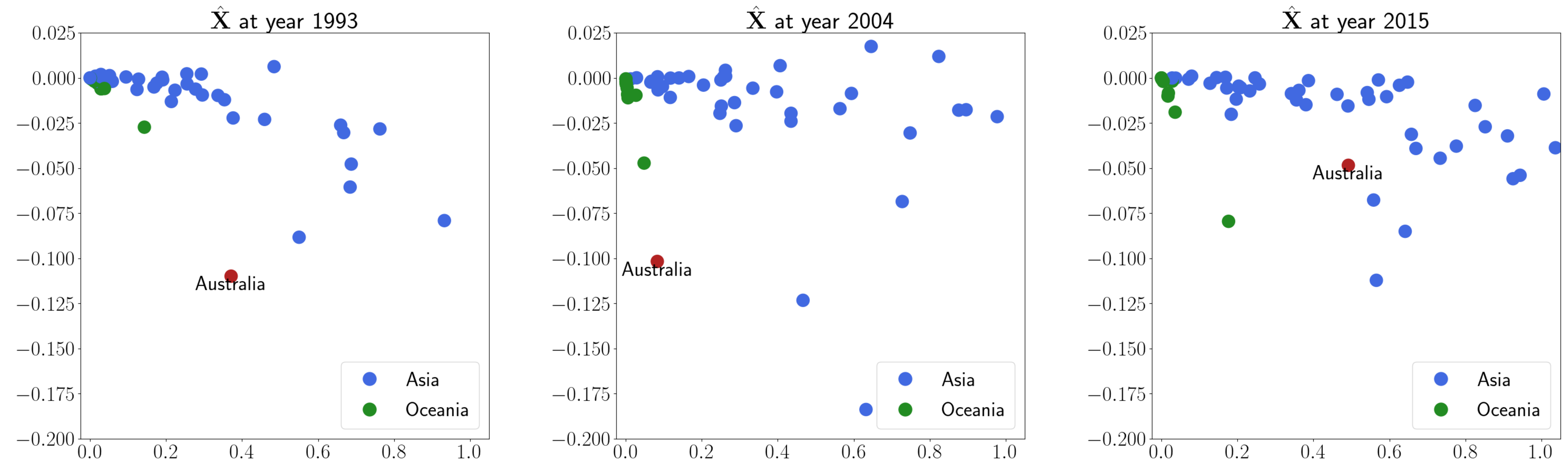
Node addition and deletion

- Project to subspace spanned by columns of $\hat{\mathbf{X}}$, then run GD
- **Alternative approach**: Compute out-of-sample extension [4]
- **Limitation**: Larger estimation error
- Setup: ER with p = 0.1. Initially, 100 nodes. 1000 new nodes are sequentially added



Real data

- Dataset of yearly football matches between national teams [5]
- Australia played in the Oceania Football Confederation until 2005, when it joined the Asian Football Confederation



- **Reproducibility** ⇒ Try it @ `github.com/git-artes`

References

[5] Y. Li and G. Mateos. *Networks of international football: Community structure, evolution and globalization of the game*. Applied Network Science, 2022