Tracking the Adjacency Spectral Embedding for Streaming Graphs

Federico Larroca, Paola Bermolen, Marcelo Fiori, Bernardo Marenco and Gonzalo Mateos

Asilomar Conference on Signals, Systems, and Computers
November 2022

Random dot product graphs

- Consider a latent space $\mathcal{X}_d \subset \mathbb{R}^d$ such that for all

$$\mathbf{x}, \mathbf{y} \in \mathcal{X}_d \quad \Rightarrow \quad \mathbf{x}^\top \mathbf{y} \in [0, 1]$$

  $\Rightarrow$ Inner-product distribution $F : \mathcal{X}_d \mapsto [0, 1]$

- Random dot product graphs (RDPGs) are defined as follows:

$$\mathbf{x}_1, \ldots, \mathbf{x}_{N_v} \overset{\text{i.i.d.}}{\sim} F,$$
$$A_{ij} \,\big|\, \mathbf{x}_i, \mathbf{x}_j \sim \text{Bernoulli}(\mathbf{x}_i^\top \mathbf{x}_j)$$

  for $1 \leq i, j \leq N_v$, where $A_{ij} = A_{ji}$ and $A_{ii} \equiv 0$

- A particularly tractable latent position random graph model

  $\Rightarrow$ Vertex positions $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_v}]^\top \in \mathbb{R}^{N_v \times d}$

S. J. Young and E. R. Scheinerman, "Random dot product graph models for social networks," *WAW*, 2007

Estimation of latent positions

- **Q:** Given G from an RDPG, find the 'best' $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_v}]^\top$?
- MLE is well motivated but it is intractable for large N_v

$$\hat{\mathbf{X}}_{ML} = \underset{\mathbf{X}}{\operatorname{argmax}} \prod_{i<j} (\mathbf{x}_i^\top \mathbf{x}_j)^{A_{ij}} (1 - \mathbf{x}_i^\top \mathbf{x}_j)^{1-A_{ij}}$$

- Instead, let $P_{ij} = \mathrm{P}\left((i,j) \in \mathcal{E}\right)$ and define $\mathbf{P} = [P_{ij}] \in [0,1]^{N_v \times N_v}$
  - $\Rightarrow$ The RDPG model specifies that $\mathbf{P} = \mathbf{X}\mathbf{X}^\top$
  - $\Rightarrow$ **Key:** Observed $\mathbf{A}$ is a noisy realization of $\mathbf{P}$ $\left(\mathbb{E}\left[\mathbf{A}\right] = \mathbf{P}\right)$
- Suggests a LS regression approach to find $\mathbf{X}$ s.t. $\mathbf{X}\mathbf{X}^\top \approx \mathbf{A}$

$$\hat{\mathbf{X}}_{LS} = \underset{\mathbf{X}}{\operatorname{argmin}} \|\mathbf{X}\mathbf{X}^\top - \mathbf{A}\|_F^2$$

A. Athreya et al, "Statistical inference on random dot product graphs: A survey," *JMLR*, 2018

Adjacency spectral embedding

- Since $\mathbf{A}$ is real and symmetric, can decompose it as $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$
  - $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_{N_v}]$ is the orthogonal matrix of eigenvectors
  - $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_{N_v})$, with eigvenvalues $\lambda_1 \geq \ldots \geq \lambda_{N_v}$
- Define $\hat{\boldsymbol{\Lambda}} = \text{diag}(\lambda_1^+, \ldots, \lambda_d^+)$ and $\hat{\mathbf{U}} = [\mathbf{u}_1, \ldots, \mathbf{u}_d]$ $(\lambda^+ := \max(0, \lambda))$
- Best rank-d, positive semi-definite (PSD) approximation of $\mathbf{A}$ is $\hat{\mathbf{U}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{U}}^\top$

$\Rightarrow$ Ajacency spectral embedding (ASE) is $\hat{\mathbf{X}}_{LS} = \hat{\mathbf{U}}\hat{\boldsymbol{\Lambda}}^{1/2}$ since

$$\mathbf{A} \approx \hat{\mathbf{U}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{U}}^\top = \hat{\mathbf{U}}\hat{\boldsymbol{\Lambda}}^{1/2}\hat{\boldsymbol{\Lambda}}^{1/2}\hat{\mathbf{U}}^\top = \hat{\mathbf{X}}_{LS}\hat{\mathbf{X}}_{LS}^\top$$

- **Q:** Is the solution unique? Nope, inner-products are rotation invariant

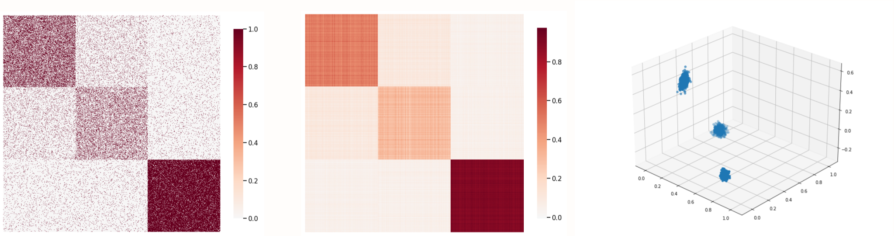
$$\mathbf{P} = \mathbf{XW}(\mathbf{XW})^\top = \mathbf{XX}^\top, \quad \mathbf{WW}^\top = \mathbf{I}_d$$

$\Rightarrow$ RDPG embedding problem is identifiable modulo rotations

Embedding an SBM graph

- Ex: SBM with $N_v = 1500$, $Q = 3$ and mixing parameters

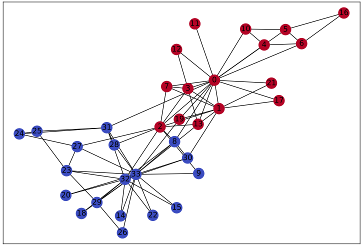
$$\boldsymbol{\alpha} = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \quad \boldsymbol{\Pi} = \begin{bmatrix} 0.5 & 0.1 & 0.05 \\ 0.1 & 0.3 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{bmatrix}$$

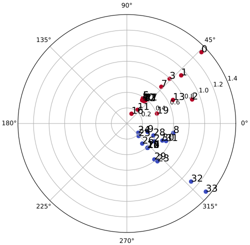


- Sample adjacency $\mathbf{A}$ (left), $\hat{\mathbf{X}}_{LS}\hat{\mathbf{X}}_{LS}^\top$ (center), rows of $\hat{\mathbf{X}}_{LS}$ (right)
- Use embeddings to bring to bear geometric methods of analysis

Interpretability of the embeddings

- **Ex:** Zachary's karate club graph with $N_v = 34$, $N_e = 78$ (left)





- Node embeddings (rows of $\hat{\mathbf{X}}_{LS}$) for $d = 2$ (right)
  - Club's administrator ($i = 0$) and instructor ($j = 33$) are orthogonal
- Interpretability of embeddings a valuable asset for RDPGs

  $\Rightarrow$ Vector magnitudes indicate how well connected nodes are

  $\Rightarrow$ Vector angles indicate nodes' affinity

Streaming Graphs

- **Goal:** track the underlying model of a stream of graphs G_t

  Ex 1: monitoring a wireless network

  Ex 2: evolving social network

- Naive approach: estimate $\hat{\mathbf{X}}_t$ by finding the ASE for each $\mathbf{A}_t$ separately

✗ Computationally expensive

✗ Challenging to align separate embeddings

- ASE does not account for the all-zero diagonal of $\mathbf{A}$. Truly we want to solve

$$\hat{\mathbf{X}} \in \underset{\mathbf{X} \in \mathbb{R}^{N \times d}}{\operatorname{argmin}} \|\mathbf{M} \circ (\mathbf{A} - \mathbf{X}\mathbf{X}^\top)\|_F^2$$

$\Rightarrow$ $\mathbf{M} := \mathbf{1}\mathbf{1}^\top - \mathbf{I}$ is a mask matrix, with zero-diagonal and ones everywhere else

Gradient descent

- Let $f : \mathbb{R}^{N_v \times d} \mapsto \mathbb{R}$ be the objective function $f(\mathbf{X}) = \|\mathbf{M} \circ (\mathbf{A} - \mathbf{X}\mathbf{X}^\top)\|_F^2$
  $\Rightarrow$ Non-convex w.r.t. $\mathbf{X}$, convex w.r.t. $\mathbf{P} = \mathbf{X}\mathbf{X}^\top$
- Gradient descent (GD) method (a.k.a. *factorized GD* or *Procrustes flow*)

$$\mathbf{X}_{t+1} = \mathbf{X}_t - \alpha \nabla f(\mathbf{X}_t), \quad t = 0, 1, 2, \ldots$$

  $\Rightarrow$ Step size $\alpha > 0$ and $\nabla f(\mathbf{X}) = 4\left[\mathbf{M} \circ (\mathbf{X}\mathbf{X}^\top - \mathbf{A})\right]\mathbf{X}$, for symmetric $\mathbf{A}$ and $\mathbf{M}$
- Convergence: if $\mathbf{X}_0$ is close to the solution, iterations converge with linear rate to $\hat{\mathbf{X}}$

Proposition. There exist $\delta > 0$ and $0 < \kappa < 1$ such that, if $\|\mathbf{X}_0 - \hat{\mathbf{X}}\|_F \leq \delta$, then

$$d(\mathbf{X}_t, \hat{\mathbf{X}}) \leq \kappa^t d(\mathbf{X}_0, \hat{\mathbf{X}}), \quad \text{for all } t > 0,$$

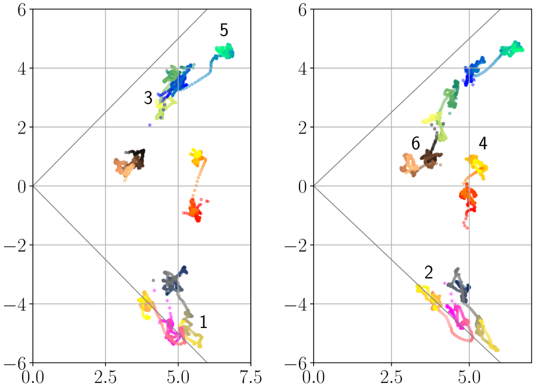
where $d(\mathbf{X}, \hat{\mathbf{X}}) := \min_{\mathbf{W} \in \mathcal{O}^{d \times d}} \|\mathbf{X}\mathbf{W} - \hat{\mathbf{X}}\|_F^2$ accounts for the rotational ambiguity.

Y. Chi *et al.*, "Nonconvex optimization meets low-rank matrix factorization: An overview," *TSP*, 2019

Tracking via warm restarts

Idea: Update $\hat{\mathbf{X}}_t$ through GD initialized with the previous estimate $\hat{\mathbf{X}}_{t-1}$

- Example:
  - $N_v = 6$ Wi-Fi APs in a Uruguayan school
  - Hourly measurements over 4 weeks (655 graphs)
  - AP 4 was moved at $t \approx 310$



M. Fiori *et al.*, "Algorithmic Advances for the Adjacency Spectral Embedding," *EUSIPCO*, 2022

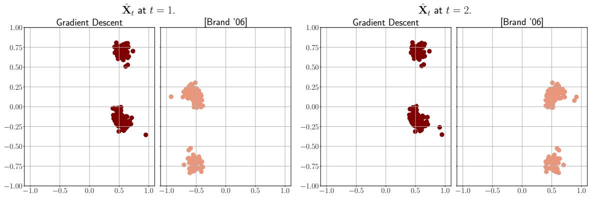
Our approach in context

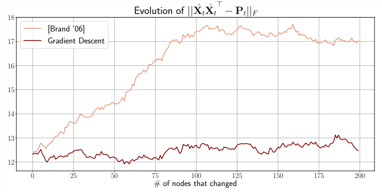
Q Isn't this the classic problem of recursively updating eigenvalues/vectors?

A Yes, but

- ✗ Computationally expensive except for specific types of changes (e.g. rank-1)
- ✗ Available methods accumulate error and/or still produce unaligned estimates

■ Example: an SBM with two communities, at each t a random node changes affiliation





M. Brand, "Fast low-rank modifications of the thin singular value decomposition," *Linear Algebra and its Applications*, 2006

Varying number of nodes

- Dynamic graphs typically include deletions/additions of nodes
  - Deletions are easy to handle, but additions?
- Assume a single node $i = N_v + 1$ is added

  $\Rightarrow$ The new $\mathbf{A}_{t+1} \in \{0,1\}^{N_v+1 \times N_v+1}$ has an extra row (column) $\mathbf{a}_{N_v+1} \in \{0,1\}^{N_v}$

  $\Rightarrow$ What about $\hat{\mathbf{x}}_{N_v+1}$?
- Reasonable approximation: project $\mathbf{a}_{N_v+1}$ to the column space of $\hat{\mathbf{X}}_t$

$$\hat{\mathbf{x}}_{N_v+1}^{\text{proj}} = \mathbf{a}_{N_v+1} \hat{\mathbf{X}}_t^{\text{norm}}$$

with $\hat{\mathbf{X}}_t^{\text{norm}}$ the column-wise normalized version of $\hat{\mathbf{X}}_t$

✓ Simple and consistent as $N_v \to \infty$

✓ Preserves alignment

✗ Assumes embeddings do not change over time

✗ Error accumulates as new nodes are added in the finite N_v regime

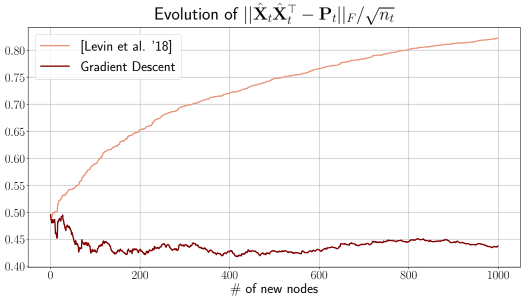
K. Levin *et al.*, "Out-of-sample extension of graph adjacency spectral embedding," *PMLR*, 2018

Varying number of nodes

Idea: Update $\hat{\mathbf{X}}_{t+1}$ using GD where old nodes are initialized at $\hat{\mathbf{x}}_t$ and new ones at $\hat{\mathbf{x}}^{\text{proj}}$

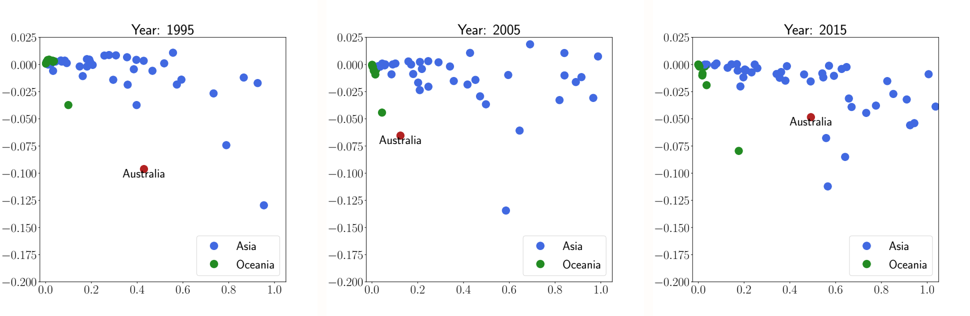
- ✓ Still simple and consistent as $N_v \to \infty$
- ✓ Preserves alignment
- ✓ Embeddings may change over time
- ✓ Constant error as new nodes are added in the finite N_v regime

- Simple example:
  - $G_0 = G_{100,0.1}$. We add new nodes that are also from an ER with $p = 0.1$



Real-world data

- G_t consisting of:
  - Nodes: national football teams
  - (Weighted) edges: number of matches between years $t-3$ and t
- Start at $t = 1930$ ($N_v = 41$) and finish at $t = 2015$ ($N_v = 222$). We use a fixed $d = 7$.
  - **Example**: Australia left the OFC and joined the AFC in 2005
  - Plot: Asia and Oceania's embeddings best 2-d approximation



Y. Li *et al.*, "Networks of international football: Community structure, evolution and globalization of the game," *Applied Network Science*, 2022

Concluding remarks

- ASE to estimate latent nodal positions in RDPGs $\Rightarrow$ Non-convex matrix factorization
- Convergent, first-order gradient descent algorithm for refined formulation
  - $\Rightarrow$ Scalable and fast computation of nodal representations
  - $\Rightarrow$ Track dynamic network representations even when N_v changes
- Future work
  - $\Rightarrow$ Directed case implies constraints on the optimization problem
  - $\Rightarrow$ Consistency, asymptotic normality, stability $(N_v \to \infty)$

https://github.com/git-artes/