Learning Bipartite Graphs: Heavy Tails and Multiple Components

José Vinícius de M Cardoso, Jiaxi Ying, and Daniel P Palomar

The Hong Kong University of Science and Technology

TL;DR

- We propose estimators for (k-component) bipartite graphs under the assumption of heavy-tailed data
- Code available at `https://mirca.github.io`

Background

- We associate a real-valued random variable x_i to each node i of a graph, such that realizations of $\mathbf{x} = (x_1, \ldots, x_p)^\top$ represent *graph signals*

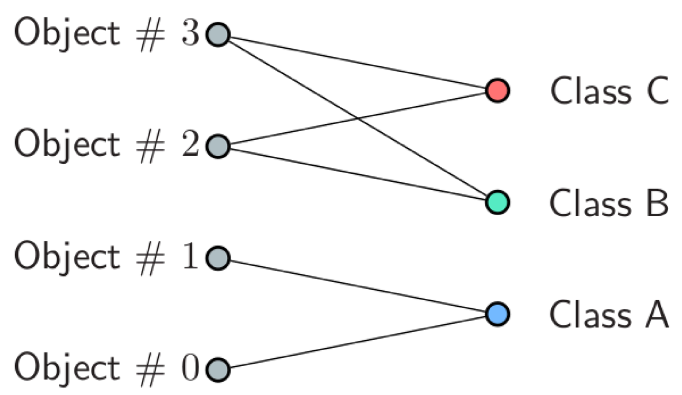

Fig. 1: A bipartite graph with two-components illustrating the modeling of dependencies between a collection of objects and their classes.

- Assuming we are given n data samples of $\mathbf{x}$, $\mathbf{X} \in \mathbb{R}^{n\times p}$, and that $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{L}^\dagger)$, then the MLE of the Laplacian matrix $\mathbf{L}$ is given as:

$$\begin{aligned} \underset{\mathbf{L}\succeq 0}{\text{minimize}} \quad & \operatorname{tr}(\mathbf{LS}) - \log \det{}^*(\mathbf{L}), \\ \text{subject to} \quad & \mathbf{L1} = \mathbf{0},\ L_{ij} = L_{ji} \leqslant 0, \end{aligned} \tag{1}$$

- where $\mathbf{S}$ is a similarity matrix, *e.g.*, sample covariance matrix $\mathbf{S} \propto \mathbf{X}^\top\mathbf{X}$ and $\det^*(\mathbf{L})$ is the product of the positive eigenvalues of $\mathbf{L}$
- For a bipartite graph, we have:

$$\boldsymbol{L} = \begin{bmatrix} \text{Diag}(\boldsymbol{B}\mathbf{1}_q) & -\boldsymbol{B} \\ -\boldsymbol{B}^\top & \text{Diag}(\boldsymbol{B}^\top\mathbf{1}_r) \end{bmatrix}, \tag{2}$$

where $\boldsymbol{B} \in \mathbb{R}_+^{r\times q}$ contains the edge weights between the nodes of objects and the nodes of classes

Heavy Tails

- Returns of financial instruments, such as equities and cryptos, are often heavy-tailed
- SOTA methods may not perform well when the data is not Gaussian distributed

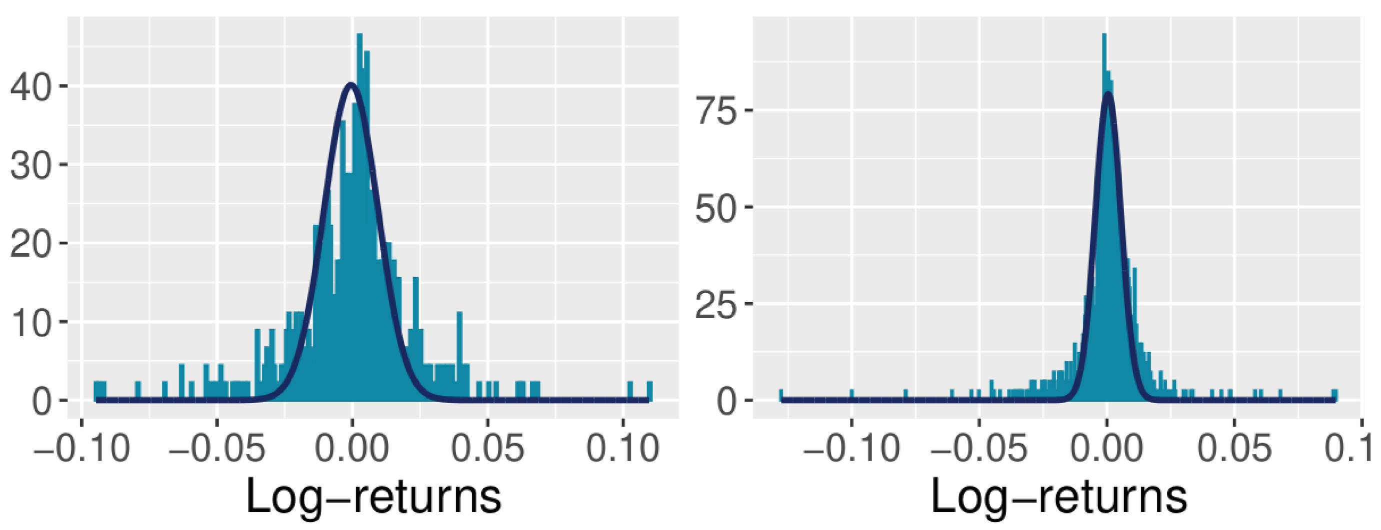

Fig. 2: Histograms of the S&P500 log-returns during different time periods ranging from 2004 to 2020. Solid lines represent Gaussian pdf fits.

State-of-the-art Methods

Bipartite Structure (Nie *et al*, 2017) proposed the following optimization problem to learn a k-component bipartite graph from a given bipartite graph weights $\boldsymbol{A} \in \mathbb{R}^{r\times q}$:

$$\underset{\boldsymbol{B}, \boldsymbol{V}\in\mathbb{R}^{p\times k}}{\text{minimize}} \ \|\boldsymbol{B} - \boldsymbol{A}\|_F^2 + \eta \operatorname{tr}(\boldsymbol{V}^\top\boldsymbol{L}\boldsymbol{V}), \ \text{subject to} \ \boldsymbol{B} \geqslant \mathbf{0},\ \boldsymbol{B}\mathbf{1}_q = \mathbf{1}_r,\ \boldsymbol{V}^\top\boldsymbol{V} = \boldsymbol{I}_k,$$

where $\boldsymbol{L}$ depends on $\boldsymbol{B}$ through (2), $\eta > 0$ is a hyperparameter that promotes the rank of $\boldsymbol{L}$ to be $p - k$, and $\boldsymbol{A}$ can be constructed from the correlation between nodes of objects and classes

Spectral Regularization Properties associated with the spectral decomposition of graph matrices have demonstrated advantages that enable learning graphs with specific structures, such as bipartite and k-component graphs. By leveraging those spectral properties, (Kumar *et al*, 2020) introduced the following formulation:

$$\begin{aligned} \underset{\boldsymbol{w}\geqslant 0, \boldsymbol{V}, \boldsymbol{U}, \boldsymbol{\psi}, \boldsymbol{\lambda}}{\text{minimize}} \ & \operatorname{tr}(\mathcal{L}\boldsymbol{w}\boldsymbol{S}) - \log\det{}^*(\mathcal{L}\boldsymbol{w}) + \tfrac{\gamma}{2}\left\|\mathcal{A}\boldsymbol{w} - \boldsymbol{U}\text{Diag}(\boldsymbol{\psi})\boldsymbol{U}^\top\right\|_F^2 \\ & + \tfrac{\beta}{2}\left\|\mathcal{L}\boldsymbol{w} - \boldsymbol{V}\text{Diag}(\boldsymbol{\lambda})\boldsymbol{V}^\top\right\|_F^2, \\ \text{subject to} \ & \boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I},\ \boldsymbol{U}\in\mathbb{R}^{p\times p},\ \boldsymbol{\psi}\in\mathcal{C}_\psi,\ \boldsymbol{V}^\top\boldsymbol{V} = \boldsymbol{I},\ \boldsymbol{V}\in\mathbb{R}^{p\times p},\ \boldsymbol{\lambda}\in\mathcal{C}_\lambda. \end{aligned}$$

where $\mathcal{L}$ and $\mathcal{A}$ are the Laplacian and adjacency operators and $\boldsymbol{w}$ is the vector of graph weights

Proposed Formulations

Gaussian Bipartite Graphs

$$\begin{aligned} \underset{\boldsymbol{B}\geqslant 0}{\text{minimize}} \ & -\log\det\left(\begin{bmatrix} \text{Diag}(\boldsymbol{B}\mathbf{1}_q) + \boldsymbol{J}_{rr} & -\boldsymbol{B} + \boldsymbol{J}_{rq} \\ -\boldsymbol{B}^\top + \boldsymbol{J}_{qr} & \text{Diag}(\boldsymbol{B}^\top\mathbf{1}_r) + \boldsymbol{J}_{qq} \end{bmatrix}\right) \\ & + \operatorname{tr}\left(\boldsymbol{B}\left(\mathbf{1}_q\boldsymbol{s}_{1:r}^\top + \boldsymbol{s}_{r+1:p}\mathbf{1}_r^\top - 2\boldsymbol{S}_{rq}^\top\right)\right). \end{aligned}$$

- **Algorithm**: projected gradient descent with backtracking line search

Student-t Bipartite Graphs

$$p(\boldsymbol{x}) \propto \sqrt{\det{}^*(\boldsymbol{\Theta})}\left(1 + \frac{\boldsymbol{x}^\top\boldsymbol{\Theta}\boldsymbol{x}}{\nu}\right)^{-(\nu+p)/2},$$

$$\begin{aligned} \underset{\boldsymbol{B}\geqslant 0,\ \boldsymbol{B}\mathbf{1}_q = \mathbf{1}_r}{\text{minimize}} \ & -\log\det\left(\text{Diag}(\boldsymbol{B}^\top\mathbf{1}_r) + \boldsymbol{J}_{qq} - (\boldsymbol{B} - \boldsymbol{J}_{rq})^\top(\boldsymbol{I}_r + \boldsymbol{J}_{rr})^{-1}(\boldsymbol{B} - \boldsymbol{J}_{rq})\right) \\ & + \frac{p+\nu}{n}\sum_{i=1}^{n}\log\left(1 + \frac{h_i + \operatorname{tr}(\boldsymbol{B}\boldsymbol{G}_i)}{\nu}\right). \end{aligned}$$

- **Algorithm**: Majorization-Minimization and projected gradient descent with backtracking line search

Multiple Components Student-t Bipartite Graphs

$$\begin{aligned} \underset{\boldsymbol{L}\succeq 0, \boldsymbol{B}}{\text{minimize}} \ & \frac{p+\nu}{n}\sum_{i=1}^{n}\log\left(1 + \frac{h_i + \operatorname{tr}(\boldsymbol{B}\boldsymbol{G}_i)}{\nu}\right) - \log\det{}^*(\boldsymbol{L}), \\ \text{subject to} \ & \boldsymbol{L} = \begin{bmatrix} \boldsymbol{I}_r & -\boldsymbol{B} \\ -\boldsymbol{B}^\top & \text{Diag}(\boldsymbol{B}^\top\mathbf{1}_r) \end{bmatrix},\ \text{rank}(\boldsymbol{L}) = p - k,\ \boldsymbol{B} \geqslant \mathbf{0}, \boldsymbol{B}\mathbf{1}_q = \mathbf{1}_r. \end{aligned}$$

- **Algorithm**: Alternating Direction Method of Multipliers + Majorization-Minimization.

Experimental Results

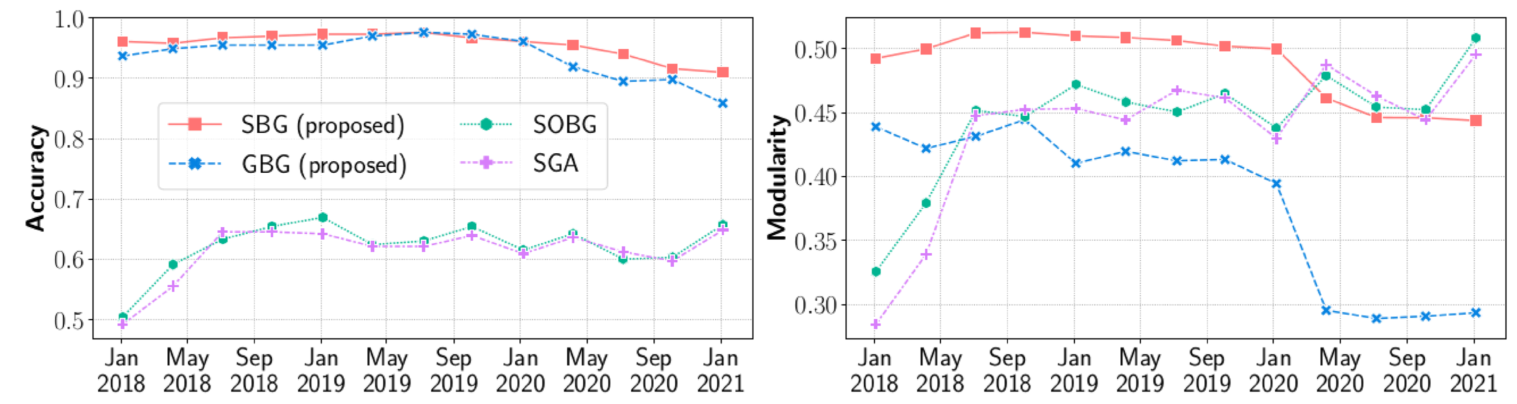

Fig. 3: Performance of the estimators for connected bipartite graphs of S&P500 stocks.

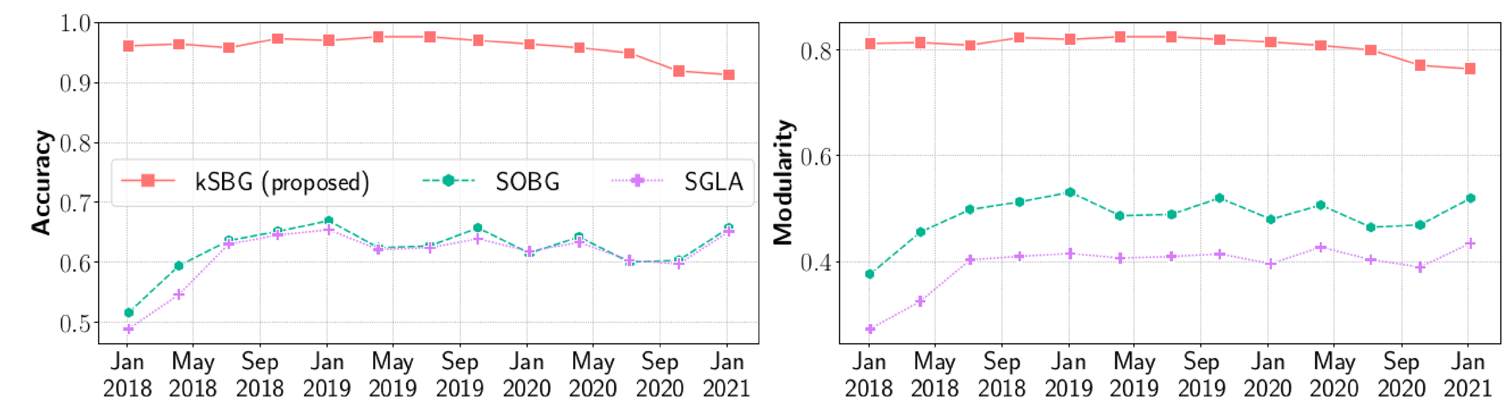

Fig. 4: Performance of the estimators for 8-component bipartite graphs of S&P500 stocks.

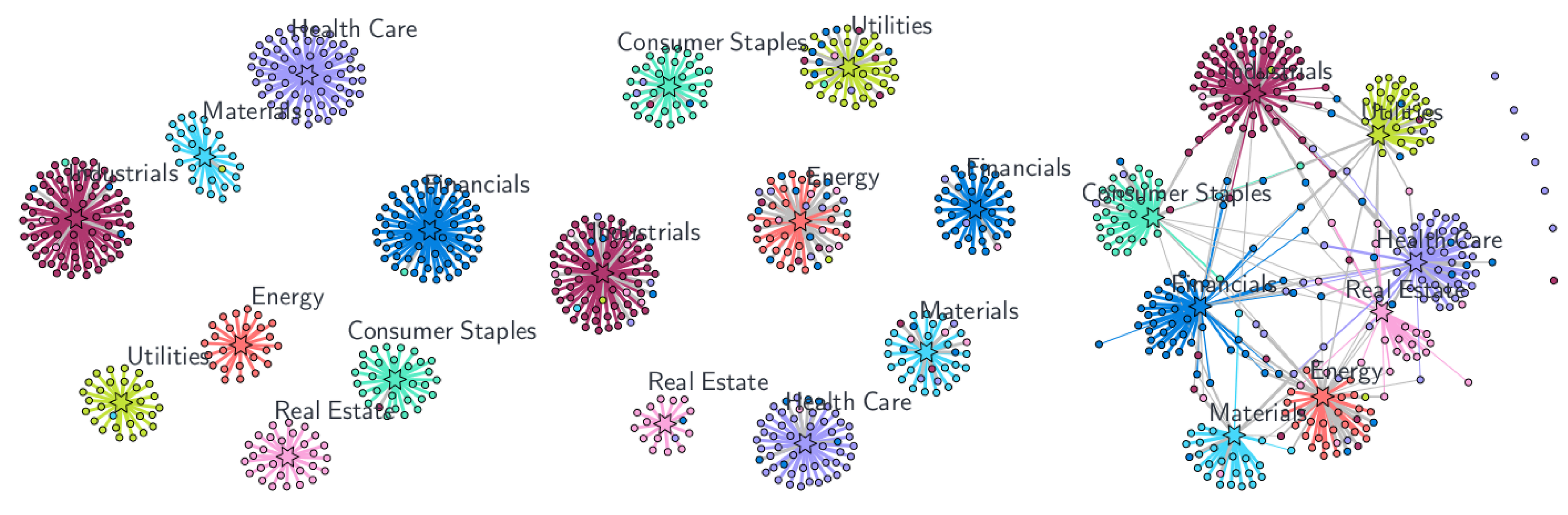

Fig. 5: From left to right: Proposed method (acc = 0.97, mod = 0.82), (Nie *et al* 2017) (acc = 0.75, mod = 0.61) (Kumar *et al* 2020) (acc = 0.77, mod = 0.56)

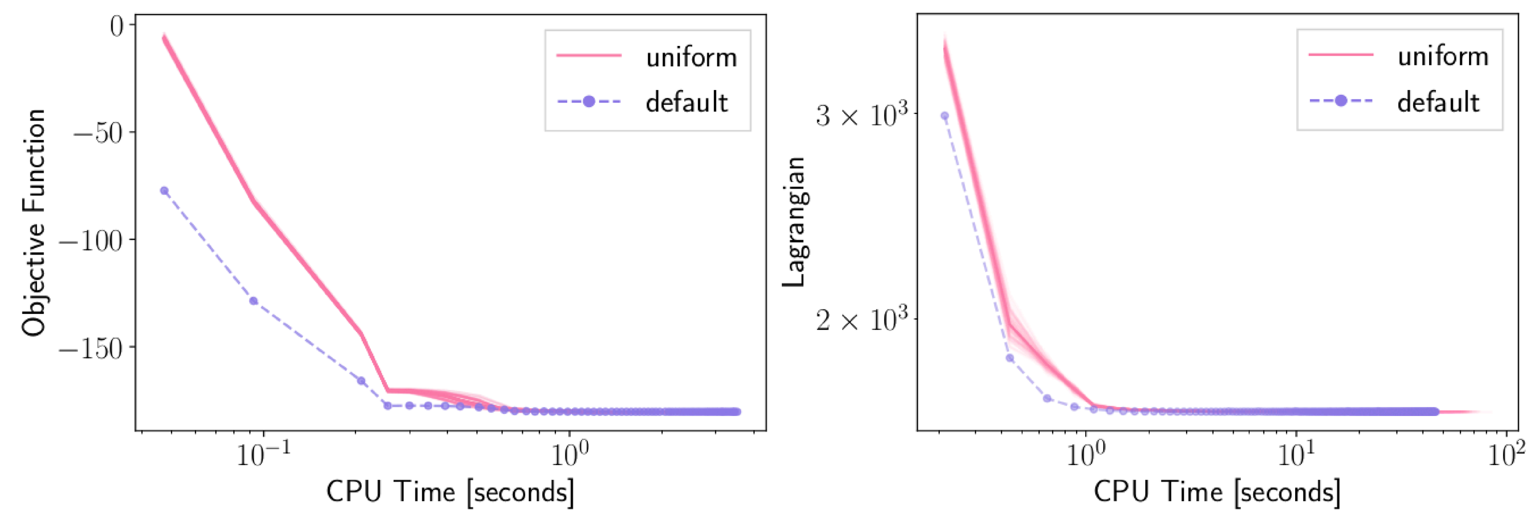

Fig. 6: Convergence trend of the proposed algorithms for different initial points

References

S. Kumar, *et al.* A unified framework for structured graph learning via spectral constraints. Journal of Machine Learning Research, 21:1–60, 2020.

F. Nie, *et al.* Learning a structured optimal bipartite graph for co-clustering. Advances on Neural Information Processing Systems (NeurIPS'17), page 4132–4141, 2017.

Acknowledgments

This work was supported by the Hong Kong GRF 16207820 research grant.