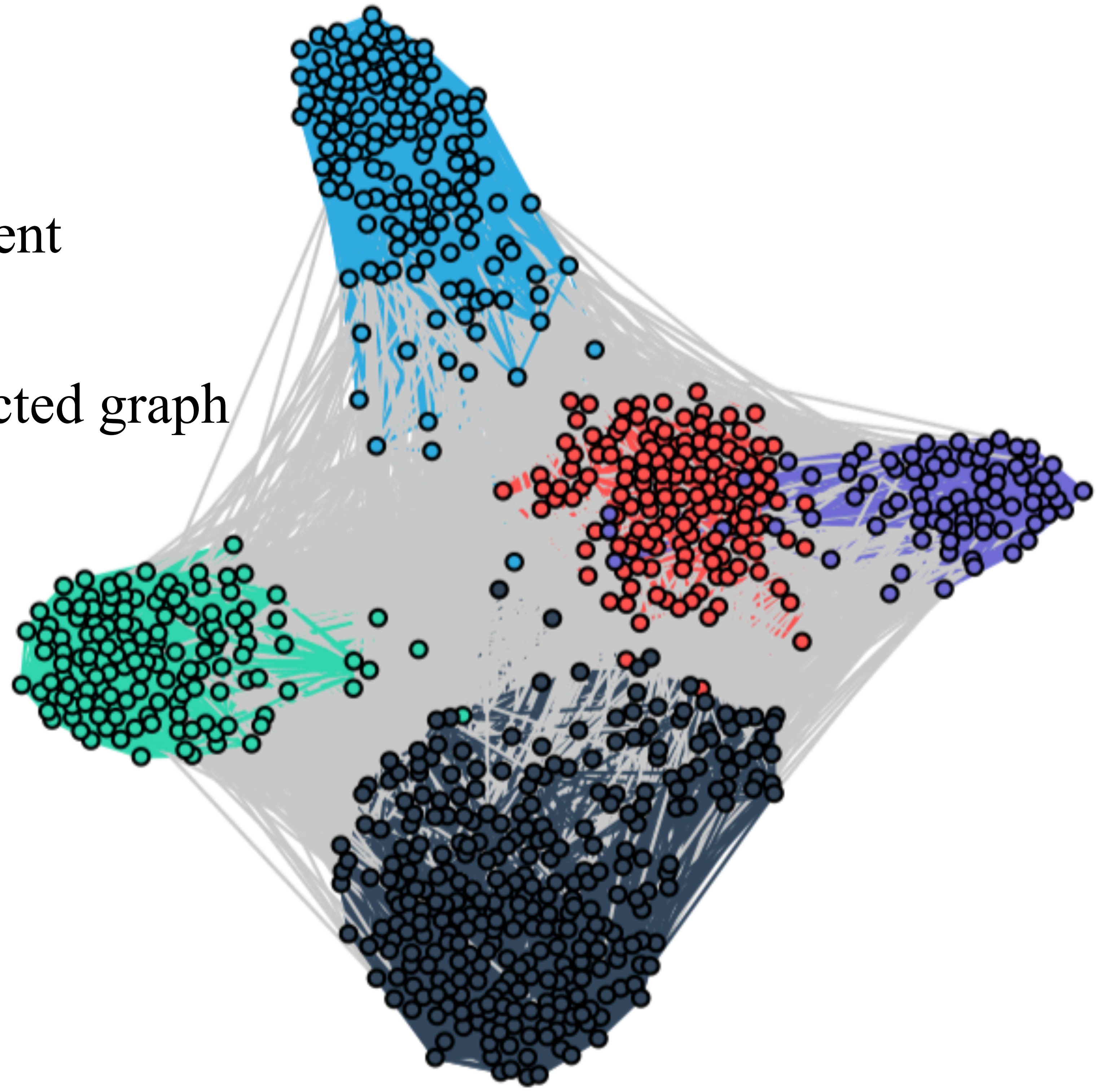
# A Unified Framework for Structured Graph Learning via Spectral Constraints

Sandeep Kumar , Jiaxi Ying, Jos´e Vin´ıcius de M. Cardoso, Daniel P. Palomar
HKUST

- an **simple undirected** weighted graph is matched to the variables
- each vertex corresponds to a variable
- an edge is present between two vertices if the corresponding random variables are conditionally dependent



**Gaussian graphical modeling**

$x = [x_1, x_2, \cdots, x_p]'$ be a p-dimensional zero mean multivariate random variable associated with an undirected graph

**Definition 1.** Let $\Theta$ be a $p \times p$ symmetric positive semidefinite matrix with rank $p-k > 0$. Then $\mathbf{x} = [x_1, x_2, \ldots, x_p]^\top$ is an improper GMRF (IGMRF) of rank $p-k$ with parameters $(\boldsymbol{\mu}, \Theta)$, assuming $\boldsymbol{\mu} = \mathbf{0}$ without loss of generality, whenever its probability density is

$$p(\mathbf{x}) = (2\pi)^{\frac{-(p-k)}{2}} (\mathrm{gdet}(\Theta))^{\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\mathbf{x}^\top \Theta \mathbf{x}\right)\right), \tag{4}$$

$$\Theta_{ij} \neq 0 \iff \{i, j\} \in \mathcal{E} \; \forall \, i \neq j,$$
$$\Theta_{ij} = 0 \iff x_i \perp x_j | \mathbf{x}/\{x_i, x_j\}.$$

$$\text{maximize}_{\Theta \in \mathcal{S}^p_{++}} \log\det(\Theta) - \mathrm{tr}(\Theta S) - \alpha h(\Theta)$$

S be the sample covariance matrix (SCM) calculated from n number of observations

$\Theta$ denotes the graph matrix to be estimated

$\mathcal{S}^p_{++}$ denotes the set of positive definite matrices

p is the number of nodes (vertices) in the graph

h(·) is a regularization term

α > o is the regularization hyperparameter

Laplacian matrix

$$\mathcal{S}_\Theta = \left\{ \Theta \in \mathbb{R}^{p \times p} \mid \Theta_{ij} = \Theta_{ji} \leq 0 \text{ for } i \neq j; \Theta_{ii} = -\sum_{j \neq i} \Theta_{ij} \right\}$$

Adjacency matrix

$$W_{ij} = \begin{cases} -\Theta_{ij}, & \text{if } i \neq j, \\ 0, & \text{if } i = j. \end{cases}$$

## A General Framework for Graph Learning under Spectral Constraints

$$\begin{array}{ll}\underset{\Theta}{\text{maximize}} & \log\ \text{gdet}(\Theta) - \text{tr}\big(\Theta S\big) - \alpha h(\Theta), \\ \text{subject to} & \Theta \in \mathcal{S}_{\Theta},\ \lambda(\mathcal{T}(\Theta)) \in \mathcal{S}_{\mathcal{T}},\end{array}$$

$S_\theta$ is the Laplacian matrix structural constraint set

T (·) is a transformation on the matrix Θ

λ ( T (Θ)) denotes the eigenvalues of T (Θ) with an increasing order

Structured Graph Learning via Laplacian Spectral Constraints

$$\begin{array}{ll}\underset{\Theta,\boldsymbol{\lambda},U}{\text{maximize}} & \log\ \text{gdet}(\Theta) - \text{tr}\big(\Theta S\big) - \alpha h(\Theta), \\ \text{subject to} & \Theta \in \mathcal{S}_{\Theta},\ \Theta = U\text{Diag}(\boldsymbol{\lambda})U^{\top},\ \boldsymbol{\lambda} \in \mathcal{S}_{\lambda},\ U^{\top}U = I,\end{array}$$

Structured Graph Learning via Adjacency Spectral Constraints

$$\begin{array}{ll}\underset{\Theta,\boldsymbol{\psi},V}{\text{maximize}} & \log\ \text{gdet}(\Theta) - \text{tr}\big(\Theta S\big) - \alpha h(\Theta), \\ \text{subject to} & \Theta \in \mathcal{S}_{\Theta},\ \mathcal{W}(\Theta) = V\text{Diag}(\boldsymbol{\psi})V^{\top},\ \boldsymbol{\psi} \in \mathcal{S}_{\psi},\ V^{\top}V = I,\end{array}$$

Structured Graph Learning via Joint **Laplacian and Adjacency** Spectral Constraints

$$\begin{array}{ll}\underset{\Theta,\boldsymbol{\lambda},\boldsymbol{\psi},U,V}{\text{maximize}} & \log\ \text{gdet}(\Theta) - \text{tr}\big(\Theta S\big) - \alpha h(\Theta), \\ \text{subject to} & \Theta \in \mathcal{S}_{\Theta},\ \Theta = U\text{Diag}(\boldsymbol{\lambda})U^{\top},\ \mathcal{W}(\Theta) = V\text{Diag}(\boldsymbol{\psi})V^{\top}, \\ & \boldsymbol{\lambda} \in \mathcal{S}_{\lambda},\ U^{\top}U = I,\ \boldsymbol{\psi} \in \mathcal{S}_{\psi},\ \ V^{\top}V = I,\end{array}$$

## Structured Graph Learning via Laplacian Spectral Constraints

$\mathcal{T}(\Theta) = \Theta$

$$\begin{array}{ll} \underset{\Theta,\boldsymbol{\lambda},U}{\text{maximize}} & \log\ \text{gdet}(\Theta) - \text{tr}\big(\Theta S\big) - \alpha h(\Theta), \\ \text{subject to} & \Theta \in \mathcal{S}_\Theta,\ \Theta = U\text{Diag}(\boldsymbol{\lambda})U^\top,\ \boldsymbol{\lambda} \in \mathcal{S}_\lambda,\ U^\top U = I, \end{array}$$

k-component graph

The multiplicity of the zero eigenvalue of a Laplacian matrix gives the number of connected components of a graph G

**Theorem 1.** *(Chung, 1997) The eigenvalues of any Laplacian matrix can be expressed as:*

$$\mathcal{S}_\lambda = \left\{\boldsymbol{\lambda} \in \mathbb{R}^p | \{\lambda_j = 0\}_{j=1}^k,\ c_1 \leq \lambda_{k+1} \leq \cdots \leq \lambda_p \leq c_2\right\}, \tag{9}$$

*where $k \geq 1$ denotes the number of connected components in the graph, and $c_1 > 0, c_2 > 0$ are constants that depend on the number of edges and their weights (Spielman and Teng, 2011).*

d-regular graph — all the nodes have the same weighted degree

$$\mathcal{S}_\lambda = \{\boldsymbol{\lambda} \in \mathbb{R}^p | \lambda_1 = 0, c_1 \leq \lambda_2 \leq \cdots \leq \lambda_p \leq c_2\},\ \text{diag}(\Theta) = d\mathbf{1}.$$

k-component d-regular graph

$$\mathcal{S}_\lambda = \left\{\boldsymbol{\lambda} \in \mathbb{R}^p | \{\lambda_j = 0\}_{j=1}^k,\ c_1 \leq \lambda_{k+1} \leq \cdots \leq \lambda_p \leq c_2\right\},\ \text{diag}(\Theta) = d\mathbf{1}.$$

## Structured Graph Learning via Adjacency Spectral Constraints

$\mathcal{T}(\Theta) = W$

$$\begin{array}{ll} \underset{\Theta,\boldsymbol{\psi},V}{\text{maximize}} & \log\ \text{gdet}(\Theta) - \text{tr}\big(\Theta S\big) - \alpha h(\Theta), \\ \text{subject to} & \Theta \in \mathcal{S}_\Theta,\ \mathcal{W}(\Theta) = V\text{Diag}(\boldsymbol{\psi})V^\top,\ \boldsymbol{\psi} \in \mathcal{S}_\psi,\ V^\top V = I, \end{array}$$

General bipartite graph

**Theorem 2.** *(Van Mieghem, 2010, Ch.5, Thm. 22) A graph is bipartite if and only if the spectrum of the associated adjacency matrix is symmetric about the origin*

$$\mathcal{S}_\psi = \{\boldsymbol{\psi} \in \mathbb{R}^p | \psi_1 \geq \psi_2 \geq \cdots \geq \psi_p,\ \psi_i = -\psi_{p-i+1},\ i = 1, 2, \ldots, p\}. \tag{14}$$

Connected bipartite graph

**Theorem 3.** *A graph is connected bipartite if and only if the spectrum of the associated adjacency matrix is symmetric about the origin with non-repeated extreme eigenvalues*

$$\mathcal{S}_\psi = \{\boldsymbol{\psi} \in \mathbb{R}^p | \psi_1 > \psi_2 \geq \cdots \geq \psi_{p-1} > \psi_p,\ \psi_i = -\psi_{p-i+1},\ i = 1, 2, \ldots, p\}. \tag{15}$$

## Structured Graph Learning via Joint Laplacian and Adjacency Spectral Constraints

$$\begin{array}{ll} \underset{\Theta,\boldsymbol{\lambda},\boldsymbol{\psi},U,V}{\text{maximize}} & \log\ \text{gdet}(\Theta) - \text{tr}\big(\Theta S\big) - \alpha h(\Theta), \\ \text{subject to} & \Theta \in \mathcal{S}_\Theta,\ \Theta = U\text{Diag}(\boldsymbol{\lambda})U^\top,\ \mathcal{W}(\Theta) = V\text{Diag}(\boldsymbol{\psi})V^\top, \\ & \boldsymbol{\lambda} \in \mathcal{S}_\lambda,\ U^\top U = I,\ \boldsymbol{\psi} \in \mathcal{S}_\psi,\ \ V^\top V = I, \end{array}$$

k-component bipartite graph

$$\mathcal{S}_\lambda = \left\{\boldsymbol{\lambda} \in \mathbb{R}^p | \{\lambda_j = 0\}_{j=1}^k,\ c_1 \leq \lambda_{k+1} \leq \cdots \leq \lambda_p \leq c_2\right\},$$
$$\mathcal{S}_\psi = \{\boldsymbol{\psi} \in \mathbb{R}^p | \psi_i = -\psi_{p-i+1},\ \psi_1 \geq \psi_2 \geq \cdots \geq \psi_p,\ i = 1, 2, \ldots, p\}.$$

k-component regular bipartite graph

$$\mathcal{S}_\lambda = \left\{\boldsymbol{\lambda} \in \mathbb{R}^p | \{\lambda_j = 0\}_{j=1}^k,\ c_1 \leq \lambda_{k+1} \leq \cdots \leq \lambda_p \leq c_2\right\},$$
$$\mathcal{S}_\psi = \{\boldsymbol{\psi} \in \mathbb{R}^p | \psi_i = d - \lambda_i,\ \psi_1 \geq \psi_2 \geq \cdots \geq \psi_p,\ i = 1, 2, \ldots, p\}.$$

## Block Majorization-Minimization Framework

non-convex, NP-hard problems.

Original Problem

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in \mathcal{X}, \end{array}$$

the optimization variable x is partitioned into m blocks as $X = (x_1, x_2, \cdots, x_m)$

At the t-th iteration, each block x_i is updated in a cyclic order by solving the following:

$$\begin{array}{ll} \underset{\mathbf{x}_i}{\text{minimize}} & g_i\left(\mathbf{x}_i | \mathbf{x}_1^{(t)}, \ldots, \mathbf{x}_{i-1}^{(t)}, \mathbf{x}_{i+1}^{(t-1)}, \ldots, \mathbf{x}_m^{(t-1)}\right), \\ \text{subject to} & \mathbf{x}_i \in \mathcal{X}_i, \end{array}$$

where $g_i\left(\mathbf{x}_i | \mathbf{y}_i^{(t)}\right)$ with $\mathbf{y}_i^{(t)} \triangleq \left(\mathbf{x}_1^{(t)}, \ldots, \mathbf{x}_{i-1}^{(t)}, \mathbf{x}_i^{(t-1)}, \mathbf{x}_{i+1}^{(t-1)}, \ldots, \mathbf{x}_m^{(t-1)}\right)$ is a majorization function of $f(\mathbf{x})$ at $\mathbf{y}_i^{(t)}$ satisfying

$$g_i\left(\mathbf{x}_i | \mathbf{y}_i^{(t)}\right) \text{ is continuous in } \left(\mathbf{x}_i, \mathbf{y}_i^{(t)}\right), \forall\, i,$$

$$g_i\left(\mathbf{x}_i^{(t)} | \mathbf{y}_i^{(t)}\right) = f\left(\mathbf{x}_1^{(t)}, \ldots, \mathbf{x}_{i-1}^{(t)}, \mathbf{x}_i^{(t)}, \mathbf{x}_{i+1}^{(t-1)}, \ldots, \mathbf{x}_m^{(t-1)}\right),$$

$$g_i\left(\mathbf{x}_i | \mathbf{y}_i^{(t)}\right) \geq f\left(\mathbf{x}_1^{(t)}, \ldots, \mathbf{x}_{i-1}^{(t)}, \mathbf{x}_i, \mathbf{x}_{i+1}^{(t-1)}, \ldots, \mathbf{x}_m^{(t-1)}\right), \forall\, \mathbf{x}_i \in \mathcal{X}_i, \forall\, \mathbf{y}_i \in \mathcal{X}, \forall\, i,$$

$$g_i'\left(\mathbf{x}_i; \mathbf{d}_i | \mathbf{y}_i^{(t)}\right) \mathbf{B}(t) ig|_{\mathbf{x}_i = \mathbf{x}_i^{(t)}} = f'\left(\mathbf{x}_1^{(t)}, \ldots, \mathbf{x}_{i-1}^{(t)}, \mathbf{x}_i, \mathbf{x}_{i+1}^{(t-1)}, \ldots, \mathbf{x}_m^{(t-1)}; \mathbf{d}\right),$$

$$\forall\; \mathbf{d} = (\mathbf{0}, \ldots, \mathbf{d}_i, \ldots, \mathbf{0}) \text{ such that } \mathbf{x}_i^{(t)} + \mathbf{d}_i \in \mathcal{X}_i,\ \forall\, i,$$

## Laplacian Spectral Constraints

$$\begin{aligned} \underset{\Theta,\boldsymbol{\lambda},U}{\text{maximize}} \quad & \log\ \text{gdet}(\Theta) - \text{tr}\big(\Theta S\big) - \alpha h(\Theta), \\ \text{subject to} \quad & \Theta \in \mathcal{S}_\Theta,\ \Theta = U\text{Diag}(\boldsymbol{\lambda})U^\top,\ \boldsymbol{\lambda} \in \mathcal{S}_\lambda,\ U^\top U = I, \end{aligned}$$

$h(\Theta) = \sum_{i>j} \phi(\Theta_{ij})$ with $\phi(x) \triangleq \log(\epsilon + |x|)$.

$$\begin{aligned} \underset{\Theta,\boldsymbol{\lambda},U}{\text{minimize}} \quad & -\log\ \text{gdet}(\Theta) + \text{tr}\left(\Theta S\right) + \alpha \sum_{i>j} \phi(\Theta_{ij}), \\ \text{subject to} \quad & \Theta \in \mathcal{S}_\Theta,\ \Theta = U\text{Diag}(\boldsymbol{\lambda})U^\top,\ \boldsymbol{\lambda} \in \mathcal{S}_\lambda,\ U^\top U = I. \end{aligned}$$

**Definition 2.** The linear operator $\mathcal{L} : \mathbb{R}^{p(p-1)/2} \to \mathbb{R}^{p\times p}$, $\mathbf{w} \mapsto \mathcal{L}\mathbf{w}$, is defined as

$$[\mathcal{L}\mathbf{w}]_{ij} = \begin{cases} -w_{i+d_j} & i > j, \\ [\mathcal{L}\mathbf{w}]_{ji} & i < j, \\ -\sum_{j\neq i} [\mathcal{L}\mathbf{w}]_{ij} & i = j, \end{cases}$$

where $d_j = -j + \frac{j-1}{2}(2p - j)$.

The adjoint operator $\mathcal{L}^*$ of $\mathcal{L}$ is defined so as to satisfy $\langle \mathcal{L}\mathbf{x}, Y\rangle = \langle \mathbf{x}, \mathcal{L}^* Y\rangle$, $\forall \mathbf{x} \in \mathbb{R}^{p(p-1)/2}$ and $Y \in \mathbb{R}^{p\times p}$.

**Definition 3.** The adjoint operator $\mathcal{L}^* : \mathbb{R}^{p\times p} \to \mathbb{R}^{p(p-1)/2}$, $Y \mapsto \mathcal{L}^* Y$, is defined by

$$[\mathcal{L}^* Y]_k = Y_{i,i} - Y_{i,j} - Y_{j,i} + Y_{j,j}, \quad k = i - j + \frac{j-1}{2}(2p - j),$$

where $i, j \in \mathbb{Z}^+$ satisfy $k = i - j + \frac{j-1}{2}(2p - j)$ and $i > j$.

$$w = [w_1, w_2, \cdots, w_6]$$

$$\mathcal{L}\mathbf{w} = \begin{bmatrix} \sum_{i=1,2,3} w_i & -w_1 & -w_2 & -w_3 \\ -w_1 & \sum_{i=1,4,5} w_i & -w_4 & -w_5 \\ -w_2 & -w_4 & \sum_{i=2,4,6} w_i & -w_6 \\ -w_3 & -w_5 & -w_6 & \sum_{i=3,5,6} w_i \end{bmatrix}.$$

The operator $\mathcal{L}^*$ on a $4 \times 4$ symmetric matrix $Y$ returns a vector

$$\mathcal{L}^* Y = \begin{bmatrix} Y_{11} - Y_{21} - Y_{12} + Y_{22} \\ Y_{11} - Y_{31} - Y_{13} + Y_{33} \\ Y_{11} - Y_{41} - Y_{14} + Y_{44} \\ Y_{22} - Y_{32} - Y_{23} + Y_{33} \\ Y_{22} - Y_{42} - Y_{24} + Y_{44} \\ Y_{33} - Y_{43} - Y_{34} + Y_{44} \end{bmatrix}.$$

**Lemma 1.** *The operator norm* $\|\mathcal{L}\|_2$ *is* $\sqrt{2p}$*, where* $\|\mathcal{L}\|_2 = \sup_{\|\mathbf{x}\|=1} \|\mathcal{L}\mathbf{x}\|_F$ *with* $\mathbf{x} \in \mathbb{R}^{p(p-1)/2}$.

$$\mathcal{L}\mathbf{w} = \begin{bmatrix} \sum_{i=1,2,3} w_i & -w_1 & -w_2 & -w_3 \\ -w_1 & \sum_{i=1,4,5} w_i & -w_4 & -w_5 \\ -w_2 & -w_4 & \sum_{i=2,4,6} w_i & -w_6 \\ -w_3 & -w_5 & -w_6 & \sum_{i=3,5,6} w_i \end{bmatrix}.$$

### 8.1. Proof of Lemma 1

*Proof.* We define an index set $\Omega_t$:

$$\Omega_t := \left\{ l \; |[\mathcal{L}\mathbf{x}]_{tt} = \sum_{l \in \Omega_t} x_l \right\}, \quad t \in [1, p]. \tag{74}$$

For any $\mathbf{x} \in \mathbb{R}^{\frac{p(p-1)}{2}}$, we have

$$\|\mathcal{L}\mathbf{x}\|_F^2 = 2 \sum_{k=1}^{\frac{p(p-1)}{2}} x_k^2 + \sum_{i=1}^{p} ([\mathcal{L}\mathbf{x}]_{ii})^2 \tag{75}$$

$$= 4 \sum_{k=1}^{\frac{p(p-1)}{2}} x_k^2 + \sum_{t=1}^{p} \sum_{i,j \in \Omega_t,\; i \neq j} x_i x_j \tag{76}$$

$$\leq 4 \sum_{k=1}^{\frac{p(p-1)}{2}} x_k^2 + \frac{1}{2} \sum_{t=1}^{p} \sum_{i,j \in \Omega_t,\; i \neq j} x_i^2 + x_j^2 \tag{77}$$

$$= (4 + 2(|\Omega_t| - 1)) \sum_{k=1}^{\frac{p(p-1)}{2}} x_k^2 \tag{78}$$

$$= 2p \|\mathbf{x}\|^2, \tag{79}$$

$$
\begin{array}{ll}
\underset{\mathbf{w},\boldsymbol{\lambda},U}{\text{minimize}} & -\log \operatorname{gdet}(\operatorname{Diag}(\boldsymbol{\lambda})) + \operatorname{tr}(S\mathcal{L}\mathbf{w}) + \alpha \sum_i \phi(w_i), \\
\text{subject to} & \mathbf{w} \geq 0,\ \mathcal{L}\mathbf{w} = U\operatorname{Diag}(\boldsymbol{\lambda})U^\top,\ \boldsymbol{\lambda} \in \mathcal{S}_\lambda,\ U^\top U = I,
\end{array}
$$

**Relax the problem**

$$
\begin{array}{ll}
\underset{\mathbf{w},\boldsymbol{\lambda},U}{\text{minimize}} & -\log \operatorname{gdet}(\operatorname{Diag}(\boldsymbol{\lambda})) + \operatorname{tr}(S\mathcal{L}\mathbf{w}) + \alpha \sum_i \phi(w_i) + \frac{\beta}{2}\|\mathcal{L}\mathbf{w} - U\operatorname{Diag}(\boldsymbol{\lambda})U^\top\|_F^2, \\
\text{subject to} & \mathbf{w} \geq 0,\ \boldsymbol{\lambda} \in \mathcal{S}_\lambda,\ U^\top U = I.
\end{array}
$$

**Example: a k-component graph structure**

$$
\mathcal{S}_\lambda = \{\ c_1 \leq \lambda_{k+1} \leq \cdots \leq \lambda_p \leq c_2\}.
$$

**Update for U**

$$\underset{\mathbf{w}\geq 0}{\text{minimize}} \quad f(\mathbf{w}) = f_1(\mathbf{w}) + f_2(\mathbf{w}),$$

where $f_1(\mathbf{w}) = \frac{1}{2}\|\mathcal{L}\mathbf{w}\|_F^2 - \mathbf{c}^\top\mathbf{w}$ and $f_2(\mathbf{w}) = \frac{\alpha}{\beta}\sum_i \log(\epsilon + w_i)$

$\mathbf{c} = \mathcal{L}^*\left(U\mathrm{Diag}(\boldsymbol{\lambda})U^\top - \beta^{-1}S\right)$

**Lemma 2.** *$f_1(\mathbf{w})$ in (32) is strictly convex.*

*Proof.* From the definition of operator $\mathcal{L}$ and the property of its adjoint $\mathcal{L}^*$, we have

$$\|\mathcal{L}\mathbf{w}\|_F^2 = \langle\mathcal{L}\mathbf{w}, \mathcal{L}\mathbf{w}\rangle = \langle\mathbf{w}, \mathcal{L}^*\mathcal{L}\mathbf{w}\rangle = \mathbf{w}^\top\mathcal{L}^*\mathcal{L}\mathbf{w} > 0, \quad \forall\, \mathbf{w} \neq \mathbf{0}.$$

The above result implies that $f_1(\mathbf{w})$ is strictly convex.

**Lemma 3.** *The function $f(\mathbf{w})$ in (32) is majorized at $\mathbf{w}^{(t)}$ by the function*

$$g\left(\mathbf{w}|\mathbf{w}^{(t)}\right) = f\left(\mathbf{w}^{(t)}\right) + \left(\mathbf{w} - \mathbf{w}^{(t)}\right)^\top \nabla f\left(\mathbf{w}^{(t)}\right) + \frac{L_1}{2}\left\|\mathbf{w} - \mathbf{w}^{(t)}\right\|^2,$$

*where $\mathbf{w}^{(t)}$ is the update from previous iteration and $L_1 = \|\mathcal{L}\|_2^2 = 2p$.*

*Proof.* $f_1(\mathbf{w})$ in (32) is strictly convex and the majorization function can be

$$g_1\left(\mathbf{w}|\mathbf{w}^{(t)}\right) = f_1\left(\mathbf{w}^{(t)}\right) + \left(\mathbf{w} - \mathbf{w}^{(t)}\right)^\top \nabla f_1\left(\mathbf{w}^{(t)}\right) + \frac{L_1}{2}\left\|\mathbf{w} - \mathbf{w}^{(t)}\right\|^2,$$

According to the definition of $\|\mathcal{L}\|_2$ in Lemma 1, we get $L_1 = \|\mathcal{L}\|_2^2 = 2p$. In addition, $f_2(\mathbf{w}) = \alpha\sum_i \log(\epsilon + w_i)$ is concave and thus can be majorized at $\mathbf{w}^{(t)}$ by the first order Taylor expansion

$$g_2\left(\mathbf{w}|\mathbf{w}^{(t)}\right) = f_2\left(\mathbf{w}^{(t)}\right) + \left(\mathbf{w} - \mathbf{w}^{(t)}\right)^\top \nabla f_2\left(\mathbf{w}^{(t)}\right), \tag{36}$$

Totally, we can conclude that $g\left(\mathbf{w}|\mathbf{w}^{(t)}\right)$ in (34) is the majorization function of $f(\mathbf{w})$ in (32). More details about majorization function can be seen in (Sun et al., 2016; Song et al., 2015). □

$$\underset{\mathbf{w}\geq 0}{\text{minimize}} \quad g\left(\mathbf{w}|\mathbf{w}^{(t)}\right) = \tfrac{1}{2}\mathbf{w}^\top\mathbf{w} - \mathbf{w}^\top\mathbf{a}, \tag{37}$$

where $\mathbf{a} = \mathbf{w}^{(t)} - \frac{1}{L_1}\nabla f\left(\mathbf{w}^{(t)}\right)$ and $\nabla f\left(\mathbf{w}^{(t)}\right) = \mathcal{L}^*\left(\mathcal{L}\mathbf{w}^{(t)}\right) - \mathbf{c} + \mathbf{b}$ with $\mathbf{b} = \frac{\alpha}{\beta}[1/(\epsilon + w_1^{(t)}), \ldots, 1/(\epsilon + w_{p(p-1)/2}^{(t)})]^\top$.

**Lemma 4.** *By the KKT optimality conditions we can obtain the optimal solution to (37) as*

$$\mathbf{w}^{(t+1)} = \left(\mathbf{w}^{(t)} - \frac{1}{L_1}\nabla f\left(\mathbf{w}^{(t)}\right)\right)^+, \tag{38}$$

*where $(x)^+ \triangleq \max(x, 0)$ and and $f$ is defined in (32).*

**Update for U**

$$\underset{U}{\text{maximize}} \quad \mathrm{tr}(U^\top \mathcal{L}\mathbf{w} U \mathrm{Diag}(\boldsymbol{\lambda}))$$
$$\text{subject to} \quad U^\top U = I_q.$$

an optimization problem on the orthogonal Stiefel manifold $S_t(p,q) = \{U \in \mathbb{R}^{p\times q} : U^T U = I_q\}$

**Lemma 5.** *From the KKT optimality conditions the solution to (40) is given by*

$$U^{(t+1)} = \textit{eigenvectors}\left(\mathcal{L}\mathbf{w}^{(t+1)}\right)[k+1:p], \tag{41}$$

*that is, the* $p-k$ *principal eigenvectors of the matrix* $\mathcal{L}\mathbf{w}^{(t+1)}$ *with the corresponding eigenvalues in an increasing order.*

**Update for $\lambda$**

$$\underset{c_1 \leq \lambda_1 \leq \cdots \leq \lambda_q \leq c_2}{\text{minimize}} \quad -\sum_{i=1}^{q} \log \lambda_i + \frac{\beta}{2} \|\boldsymbol{\lambda} - \mathbf{d}\|_2^2,$$

where $\lambda = [\lambda_1, \ldots, \lambda_q]^\top$ and $d = [d_1, \ldots, d_q]^\top$ with $d_i$ being the i-th diagonal element of $Diag(U^T(Lw)U)$

---

**Algorithm 1:** Updating rule for $\boldsymbol{\lambda}$

---

**Input:** $d_1, d_2, \ldots, d_q$, $\beta$, $c_1$, and $c_2$;

$\lambda_i = \frac{1}{2}\left(d_i + \sqrt{d_i^2 + 4/\beta}\right)$, $i = 1, 2, \ldots, q$;

**if** $\boldsymbol{\lambda}$ *satisfies* $c_1 \leq \lambda_1 \leq \cdots \leq \lambda_q \leq c_2$ **then**

  **return** $\lambda_1, \ldots, \lambda_q$;

**while not** $c_1 \leq \lambda_1 \leq \cdots \leq \lambda_q \leq c_2$ **do**

  **if** $c_1 \geq \lambda_1 \geq \cdots \geq \lambda_r$ *with at least one inequality strict and* $r \geq 1$ **then**

    $\lambda_1 = \cdots = \lambda_r = c_1$;

  **else if** $\lambda_s \geq \cdots \geq \lambda_q \geq c_2$ *with at least one inequality strict and* $s \leq q$ **then**

    $\lambda_s = \cdots = \lambda_q = c_2$;

  **else if** $\lambda_i \geq \cdots \geq \lambda_m$ *with at least one inequality strict and* $1 \leq i \leq m \leq q$ **then**

    $\bar{d}_{i \to m} = \frac{1}{m-i+1} \sum_{j=i}^{m} d_j$;

    $\lambda_i = \cdots = \lambda_m = \frac{1}{2}\left(\bar{d}_{i \to m} + \sqrt{\bar{d}_{i \to m}^2 + 4/\beta}\right)$;

**end**

**Output:** $\lambda_1, \ldots, \lambda_q$.

---

**Lemma 6.** *The iterative-update procedure summarized in Algorithm 1 converges to the KKT point of Problem* (44).

$$\begin{array}{ll} \underset{\mathbf{w},\boldsymbol{\lambda},U}{\mathsf{minimize}} & -\log \operatorname{gdet}(\operatorname{Diag}(\boldsymbol{\lambda})) + \operatorname{tr}(S\mathcal{L}\mathbf{w}) + \alpha \sum_i \phi(w_i) + \frac{\beta}{2}\|\mathcal{L}\mathbf{w} - U\operatorname{Diag}(\boldsymbol{\lambda})U^\top\|_F^2, \\ \mathsf{subject\ to} & \mathbf{w} \geq 0,\ \boldsymbol{\lambda} \in \mathcal{S}_\lambda,\ U^\top U = I. \end{array} \tag{29}$$

**Theorem 5.** *The sequence* $\left(\mathbf{w}^{(t)}, U^{(t)}, \boldsymbol{\lambda}^{(t)}\right)$ *generated by Algorithm 2 converges to the set of KKT points of* (29).

*Proof.* The detailed proof is deferred to the Appendix 8.3. □

---

**Algorithm 2:** SGL

---

**Input:** SCM $S$, $k$, $c_1$, $c_2$, $\beta$, $\alpha$, $\mathbf{w}^{(0)}$, $\epsilon > 0$;

$t \leftarrow 0$;

**while** *stopping criteria not met* **do**

  update $\mathbf{w}^{(t+1)}$ as in (38);

  update $U^{(t+1)}$ as in (41);

  update $\boldsymbol{\lambda}^{(t+1)}$ by solving (44) with Algorithm 1;

  $t \leftarrow t + 1$;

**end**

**Output:** $\mathcal{L}\mathbf{w}^{(t+1)}$.

---

# Structured Graph Learning via Adjacency Spectral Constraints

**Definition 4.** The linear operator $\mathcal{A} : \mathbb{R}^{p(p-1)/2} \to \mathbb{R}^{p\times p}$, $\mathbf{w} \mapsto \mathcal{A}\mathbf{w}$, is defined as

$$[\mathcal{A}\mathbf{w}]_{ij} = \begin{cases} w_{i+d_j} & i > j, \\ [\mathcal{A}\mathbf{w}]_{ji} & i < j, \\ 0 & i = j, \end{cases}$$

where $d_j = -j + \frac{j-1}{2}(2p-j)$.

$$\mathcal{A}\mathbf{w} = \begin{bmatrix} 0 & w_1 & w_2 & w_3 \\ w_1 & 0 & w_4 & w_5 \\ w_2 & w_4 & 0 & w_6 \\ w_3 & w_5 & w_6 & 0 \end{bmatrix}.$$

The following is omitted...

**Definition 5.** The adjoint operator $\mathcal{A}^* : \mathbb{R}^{p\times p} \to \mathbb{R}^{p(p-1)/2}$, $Y \mapsto \mathcal{A}^*Y$, is defined as

$$[\mathcal{A}^*Y]_k = Y_{ij} + Y_{ji}, \tag{47}$$

where $i, j \in \mathbb{Z}^+$ satisfy $i - j + \frac{j-1}{2}(2p-j) = k$ and $i > j$.

**Lemma 8.** *The operator norm* $\|\mathcal{A}\|_2$ *is* $\sqrt{2}$, $\|\mathcal{A}\|_2 = \sup_{\|\mathbf{x}\|=1} \|\mathcal{A}\mathbf{x}\|_F$ *with* $\mathbf{x} \in \mathbb{R}^{p(p-1)/2}$.

*Proof.* Directly from the definition of operator norm, we have

$$\|\mathcal{A}\|_2 = \sup_{\|\mathbf{x}\|=1} \|\mathcal{A}\mathbf{x}\|_F = \sup_{\|\mathbf{x}\|=1} \sqrt{2}\,\|\mathbf{x}\| = \sqrt{2}. \tag{48}$$

$$\begin{array}{ll} \underset{\Theta,\boldsymbol{\psi},V}{\text{maximize}} & \log\ \text{gdet}(\Theta) - \text{tr}\big(\Theta S\big) - \alpha h(\Theta), \\ \text{subject to} & \Theta \in \mathcal{S}_\Theta,\ \mathcal{W}(\Theta) = V\text{Diag}(\boldsymbol{\psi})V^\top,\ \boldsymbol{\psi} \in \mathcal{S}_\psi,\ V^\top V = I, \end{array}$$

$$\begin{array}{ll} \underset{\mathbf{w},\boldsymbol{\psi},V}{\text{minimize}} & -\log\det(\mathcal{L}\mathbf{w} + J) + \text{tr}\left(S\mathcal{L}\mathbf{w}\right) + \alpha \sum_i \phi(w_i) + \frac{\gamma}{2}\|\mathcal{A}\mathbf{w} - V\text{Diag}(\boldsymbol{\psi})V^\top\|_F^2, \\ \text{subject to} & \mathbf{w} \geq 0,\ \boldsymbol{\psi} \in \mathcal{S}_\psi,\ V^\top V = I, \end{array}$$

**Theorem 3.** *A graph is connected bipartite if and only if the spectrum of the associated adjacency matrix is symmetric about the origin with non-repeated extreme eigenvalues*

$$\mathcal{S}_\psi = \{\boldsymbol{\psi} \in \mathbb{R}^p | \psi_1 > \psi_2 \geq \cdots \geq \psi_{p-1} > \psi_p,\ \psi_i = -\psi_{p-i+1},\ i = 1, 2, \ldots, p\}. \tag{15}$$

## Experiments

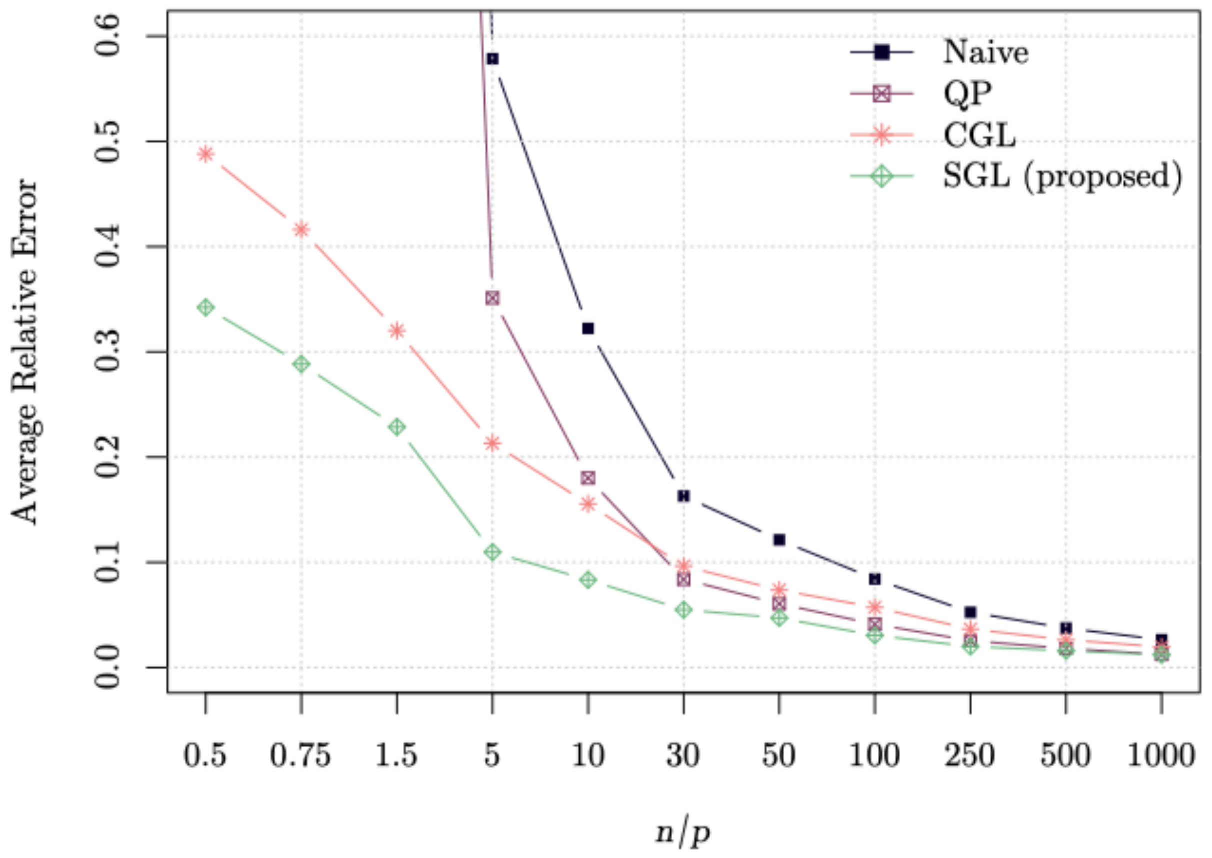
$$\text{Relative Error} = \frac{\left\|\hat{\Theta} - \Theta_{\text{true}}\right\|_F}{\|\Theta_{\text{true}}\|_F}, \quad \text{F-Score} = \frac{2\text{tp}}{2\text{tp} + \text{fp} + \text{fn}},$$

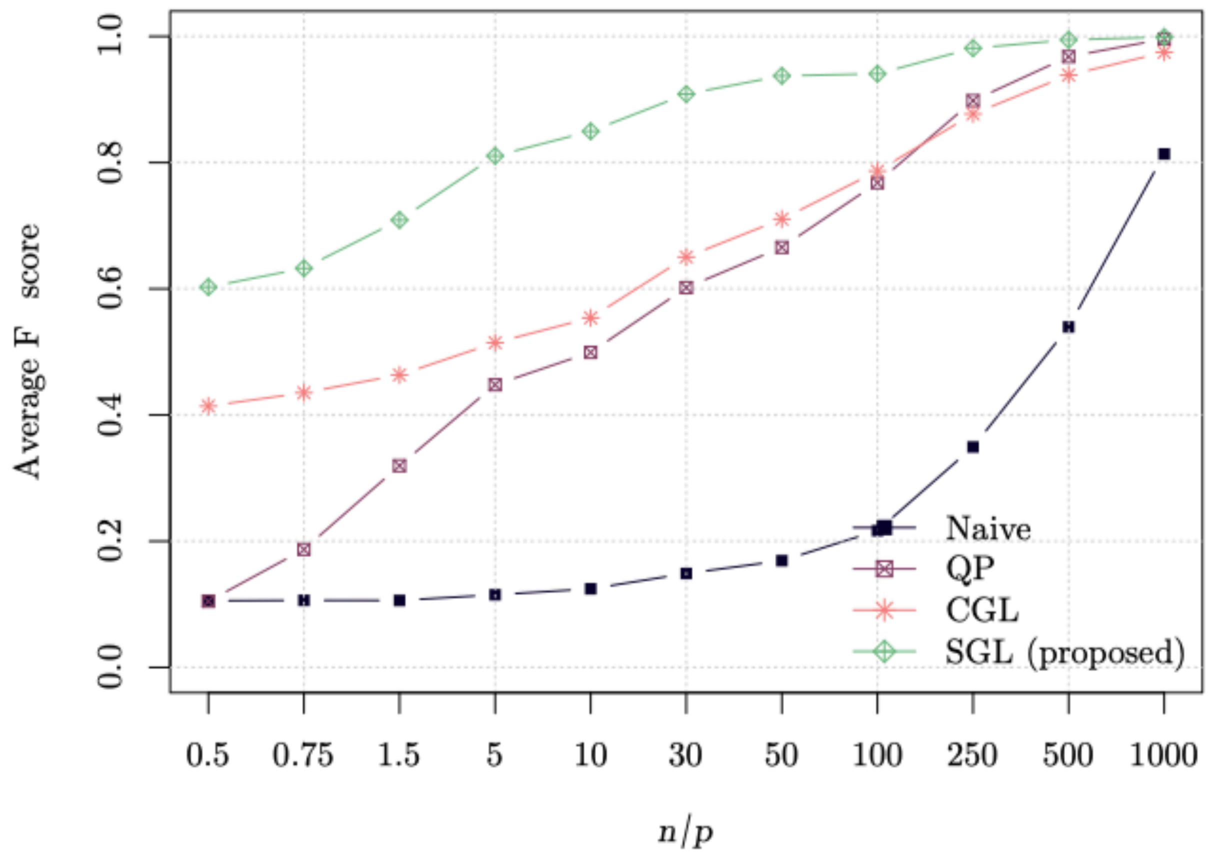
## Performance Evaluation for the SGL Algorithm

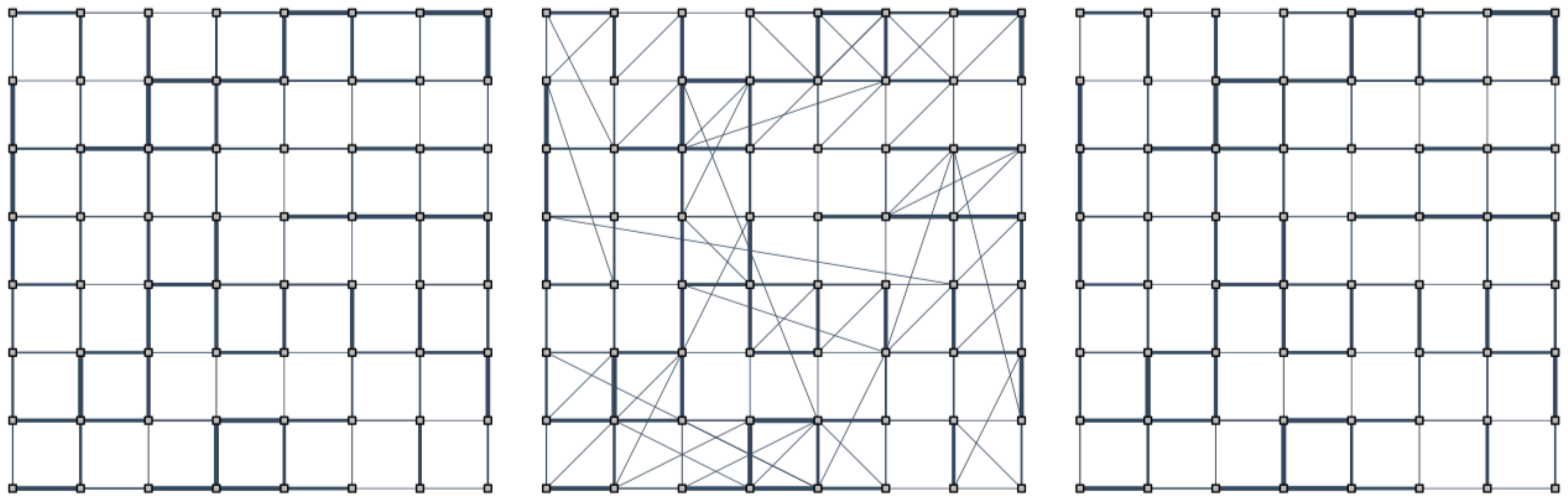
# Grid graph

p = 64

n/p = 100



(a) True grid graph (b) CGL (c) SGL (proposed)

Figure 3: Sample results of learning $\mathcal{G}_{\mathsf{grid}}(64)$ (a) True grid graph, (b) CGL (RE $= 0.0442$, FS $=$ $0.7915$) and (c) SGL with reweighted $\ell_1$-norm (RE $= 0.0378$, FS $= 1$).

# Modular graph

p = 64 nodes and k = 4 modules

$\wp_1 = 0.01$ and $\wp_2 = 0.3$

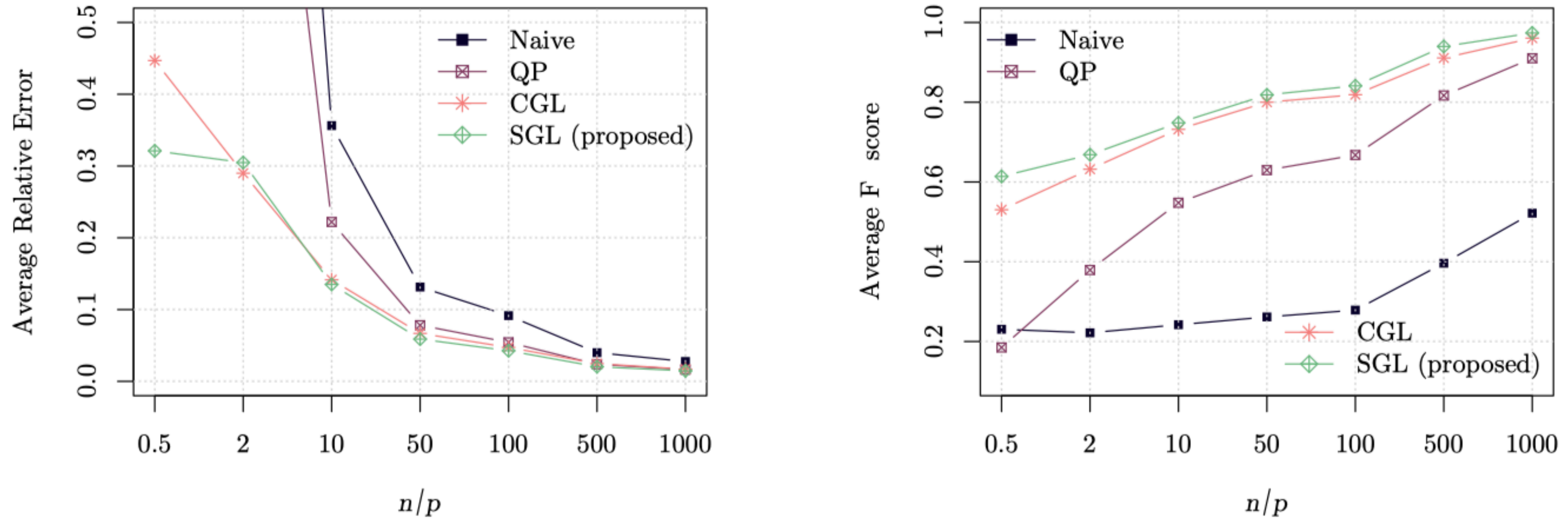



Figure 5: Average performance results for learning Laplacian matrix of a modular graph $\mathcal{G}_{\mathsf{mo}}$ with four modules. We set $\beta = 100, \alpha = 0$ for SGL. It can be observed that SGL outperforms the baseline approaches across almost the whole range of sample size ratios.

# Multi-component graph

p = 64, k = 4 and $\wp = 0.5$

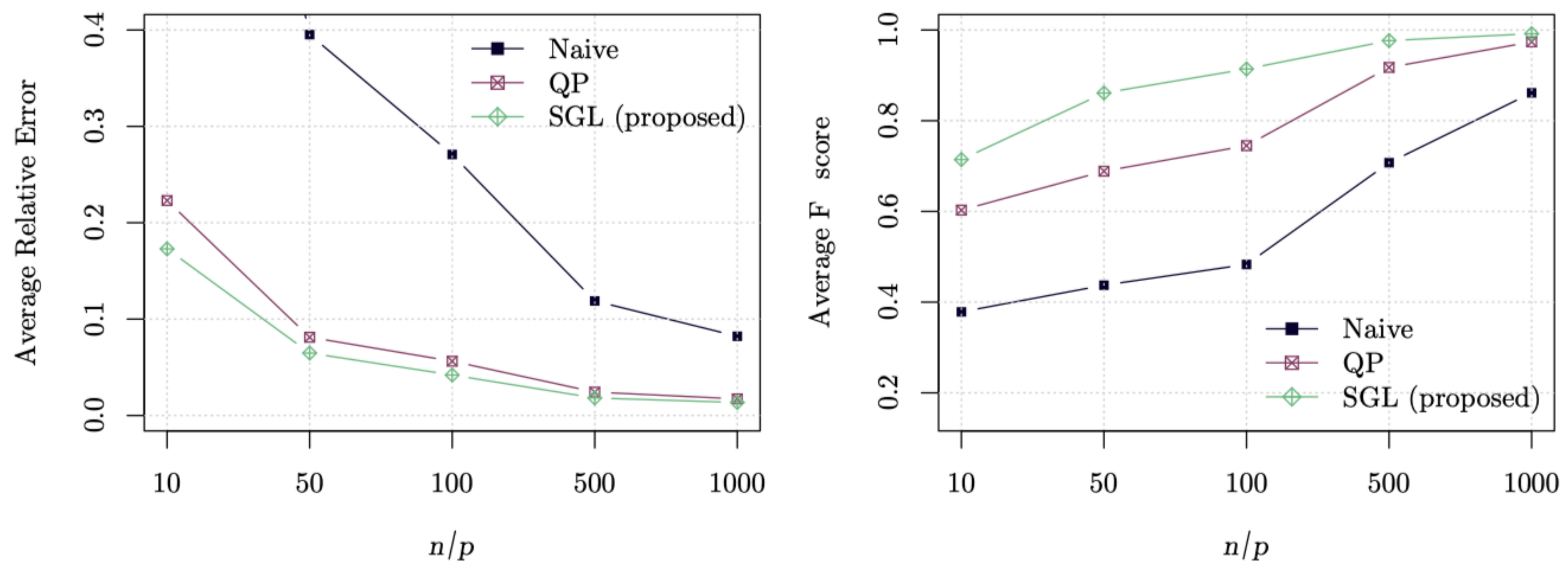



Figure 6: Average performance results as a function of the number of samples for learning the Laplacian matrix of a 4-component graph. SGL significantly outperforms the baseline approaches denoted as Naive and QP.

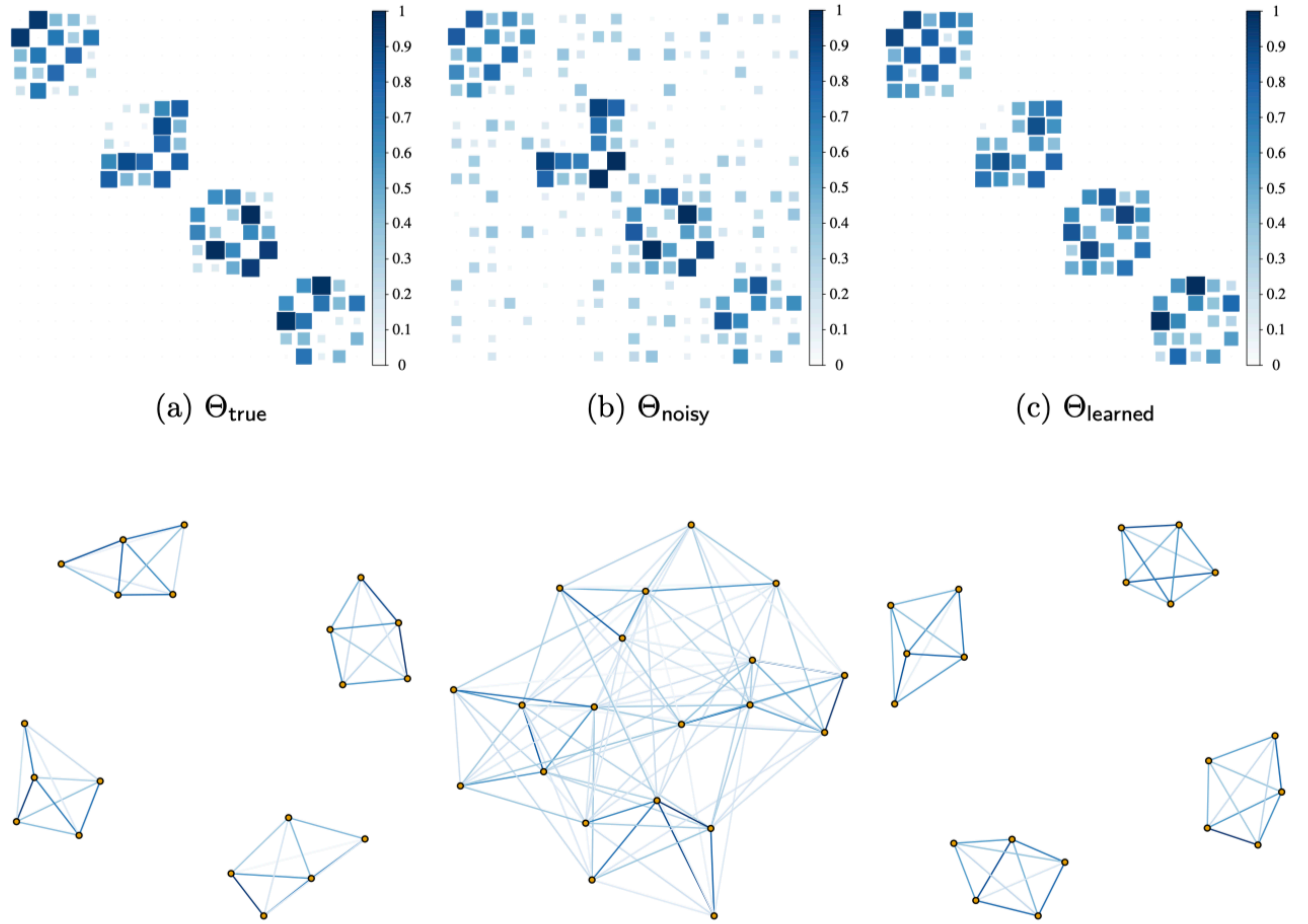
## Multi-component graph: noisy setting

G mc(20, 4, 1) with equal number of nodes across different components

The nodes in each component are fully connected and the edges are drawn randomly uniformly from [0, k].

Erdos-Renyi graph G ER (p, $\wp$ ), where p = 20 is the number of nodes,
$\wp$ = 0.35 is the probability of having an edge between any two pair of nodes,
edge weights are randomly uniformly drawn from [0, $\varkappa$]

n/p = 30, β = 400, α = 0.1, and $\varkappa$ = 0.45.

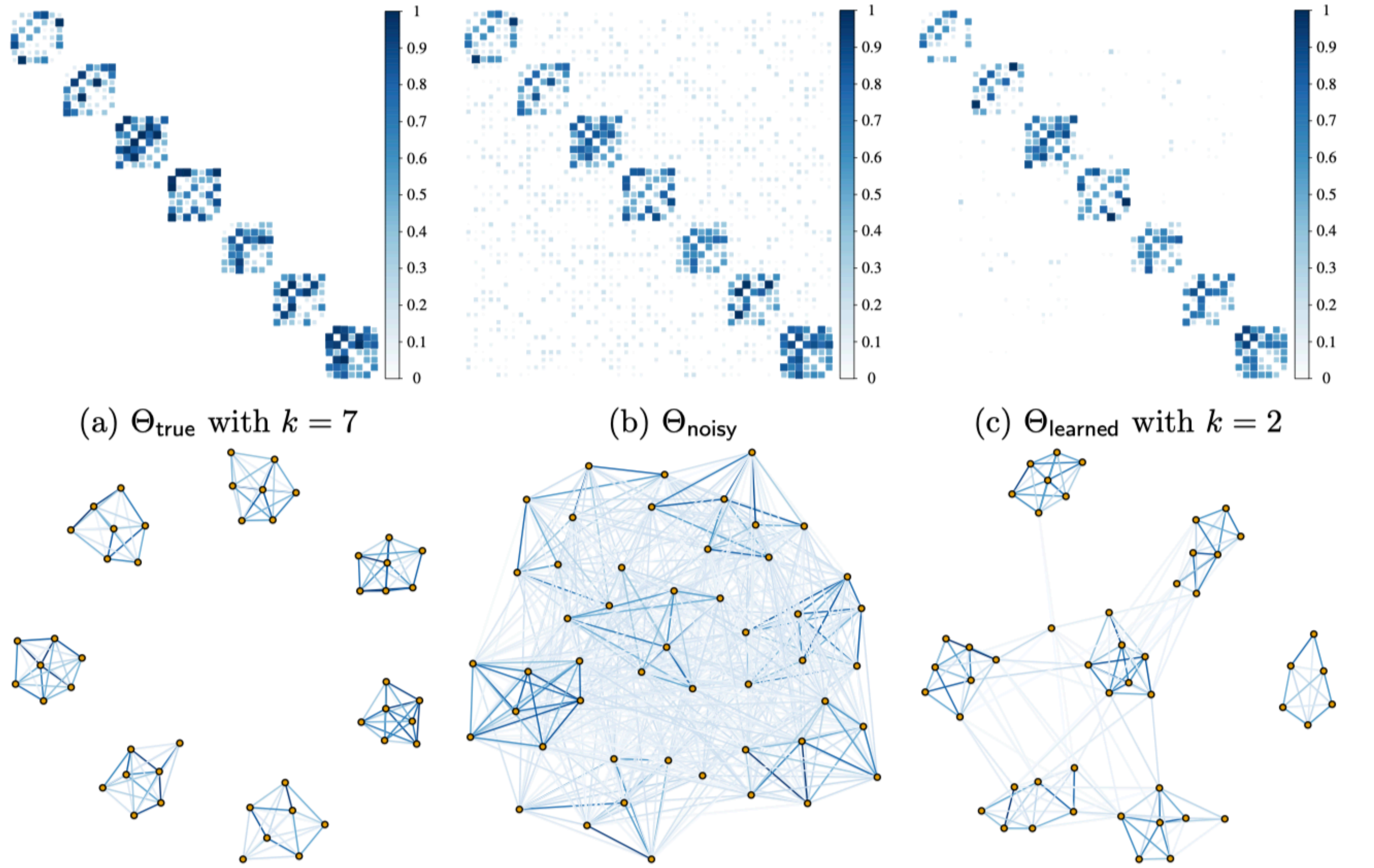


(a) $\Theta_{\text{true}}$ (b) $\Theta_{\text{noisy}}$ (c) $\Theta_{\text{learned}}$

(d) $\mathcal{G}_{\text{true}}$ (e) $\mathcal{G}_{\text{noisy}}$ (f) $\mathcal{G}_{\text{learned}}$

Figure 7: An example of estimating a 4-component graph. Heat maps of the graph matrices: (a) the ground truth graph Laplacian matrix $\Theta_{\text{true}}$, (b) $\Theta_{\text{noisy}}$ after being corrupted by noise, (c) $\Theta_{\text{learned}}$ the learned graph Laplacian with a performance of $(\text{RE}, \text{FS}) = (0.210, 1)$, which means a perfect structure recovery even in a noisy setting that heavily suppresses the ground truth weights. The panels (d), (e), and (f) correspond to the graphs represented by the Laplacian matrices in (a), (b), and (c), respectively.

## Multi-component graph: model mismatch



(a) $\Theta_{\text{true}}$ with $k = 7$ (b) $\Theta_{\text{noisy}}$ (c) $\Theta_{\text{learned}}$ with $k = 2$

(d) $\mathcal{G}_{\text{true}}$ with $k = 7$ (e) $\mathcal{G}_{\text{noisy}}$ (f) $\mathcal{G}_{\text{learned}}$ with $k = 2$

Figure 8: Heat maps of the graph matrices: (a) the ground truth graph Laplacian of a seven-component graph $\Theta_{\text{true}}$, (b) $\Theta_{\text{noisy}}$ after being corrupted by noise, (c) $\Theta_{\text{learned}}$ the learned graph Laplacian with a performance of $(\text{RE}, \text{FS}) = (0.18, 0.81)$. The panels (d), (e), and (f) correspond to the graphs represented by the Laplacian matrices in (a), (b), and (c), respectively. In Figure 8 (c) and (f) we are essentially getting results corresponding to a two-component graph, which is imperative from the usage of spectral constraints of $k = 2$. It is observed that the learned graph (f) consists of the true graph structure in (d) and some extra edges with very small weights which are due to the inaccurate spectral information. One can use some simple post-processing techniques (e.g., thresholding of elements in the learned matrix $\Theta$), to recover the true component structure.

## Popular multi-component structures

### clustering

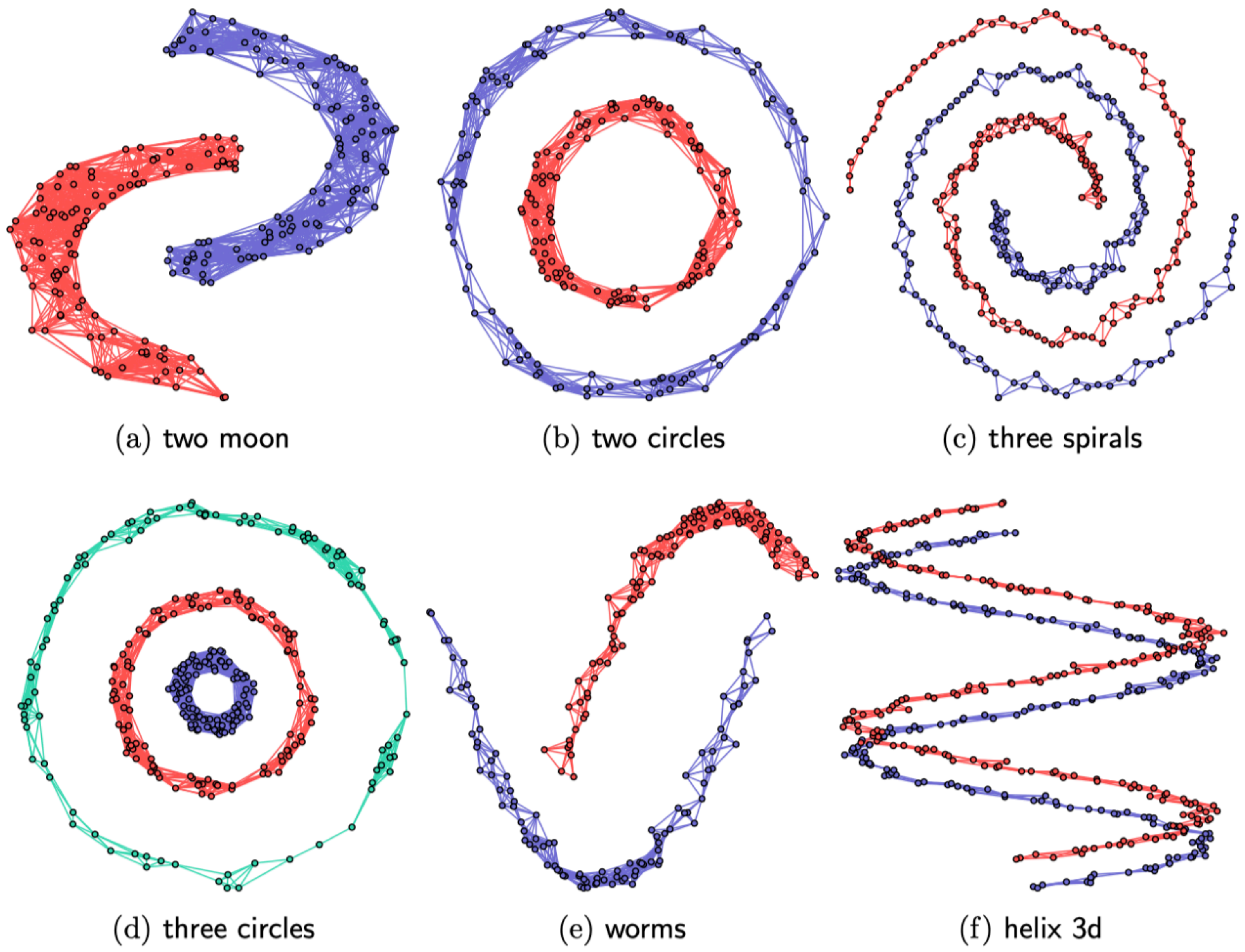



Figure 10: SGL is able to perfectly cluster the data points according to the cluster membership for all the structures.

## Real data: animals data set

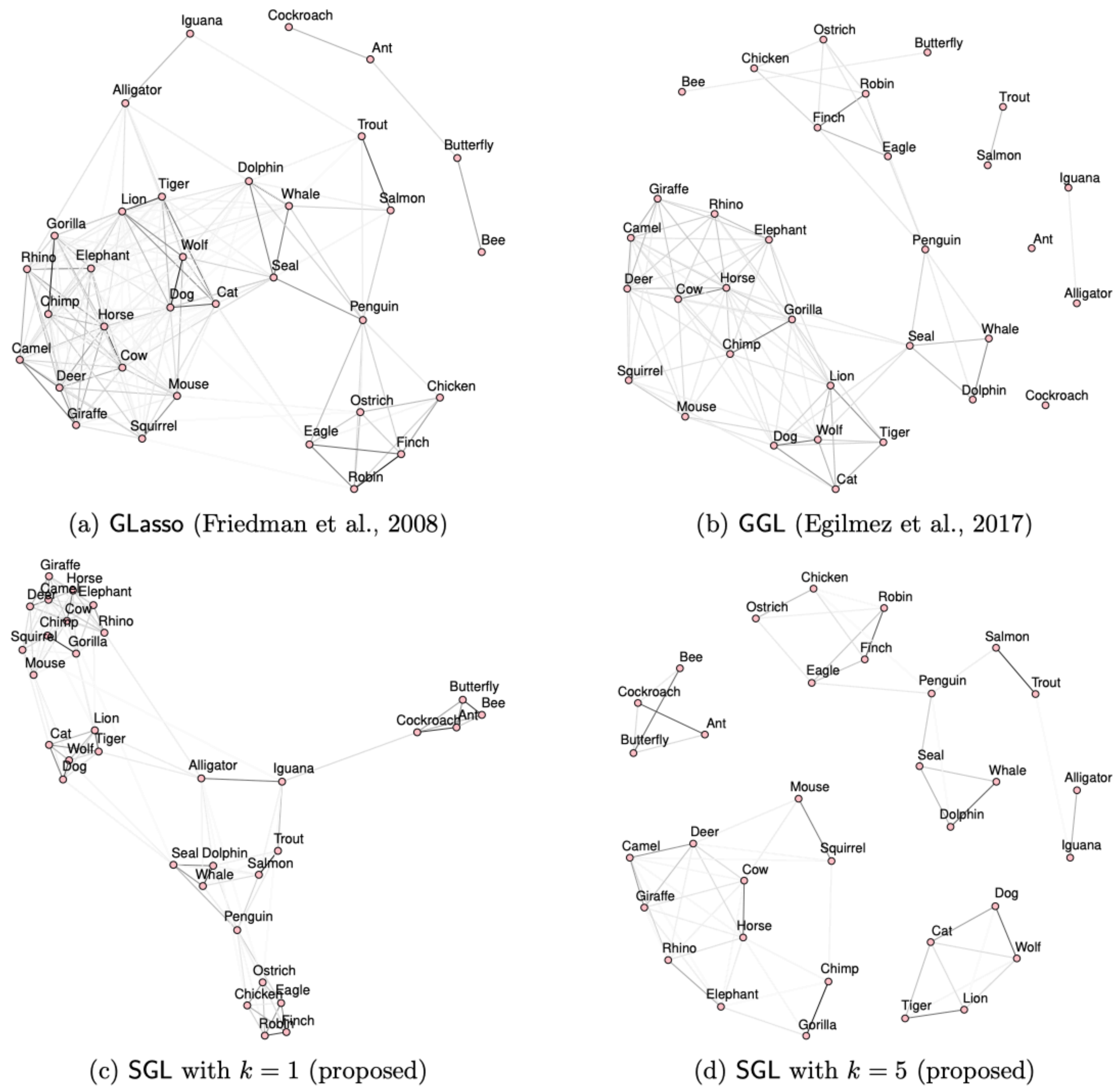
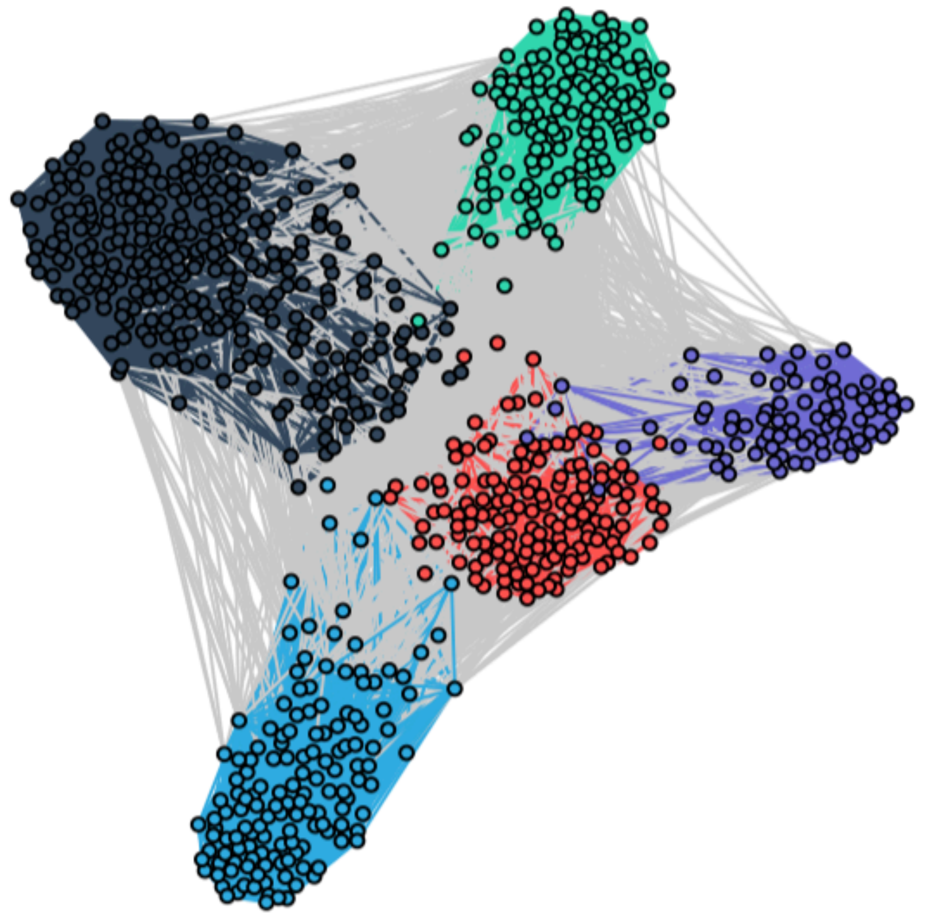
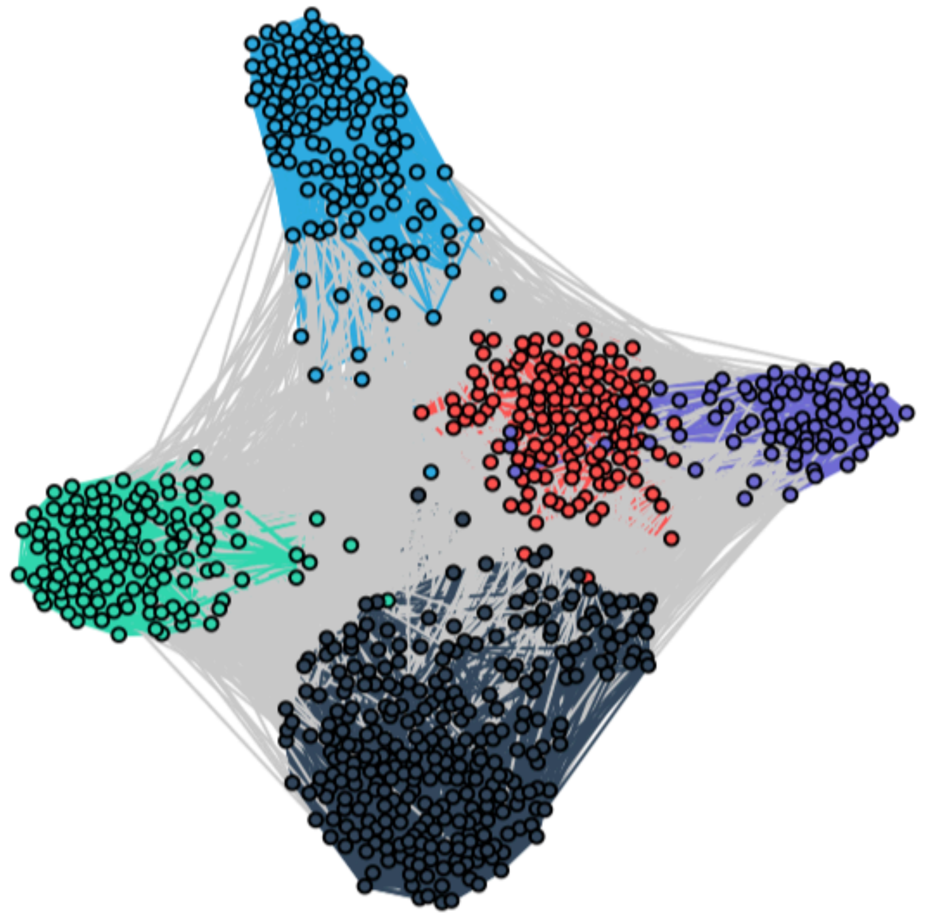



Figure 11: Learning the connectivity of the animals data set with (a) GLasso, (b) GGL, (c) SGL with $k = 1$ and (d) SGL with $k = 5$. For all graphs, darker edges denote stronger connections among animals. The methods (a) GGL, (b) GLasso, and (c) SGL $k = 1$ were expected to obtain sparse-connected graphs. But, GGL, GLasso split the graph into multiple components due to the sparsity regularization. While SGL using sparsity regularization along with spectral constraint $k = 1$ (connectedness) yields a sparse-connected graph. (d) SGL with $k = 5$ obtains a graph that depicts a more detailed representation of the network of animals by clustering similar animals within the same component. This highlights the fact that the control of the number of components may yield an improved visualization. Furthermore, the animal data is categorical (non-Gaussian) which does not follow the IGMRF assumption, hence the above result also establishes the capability of SGL under mismatch of the data model.

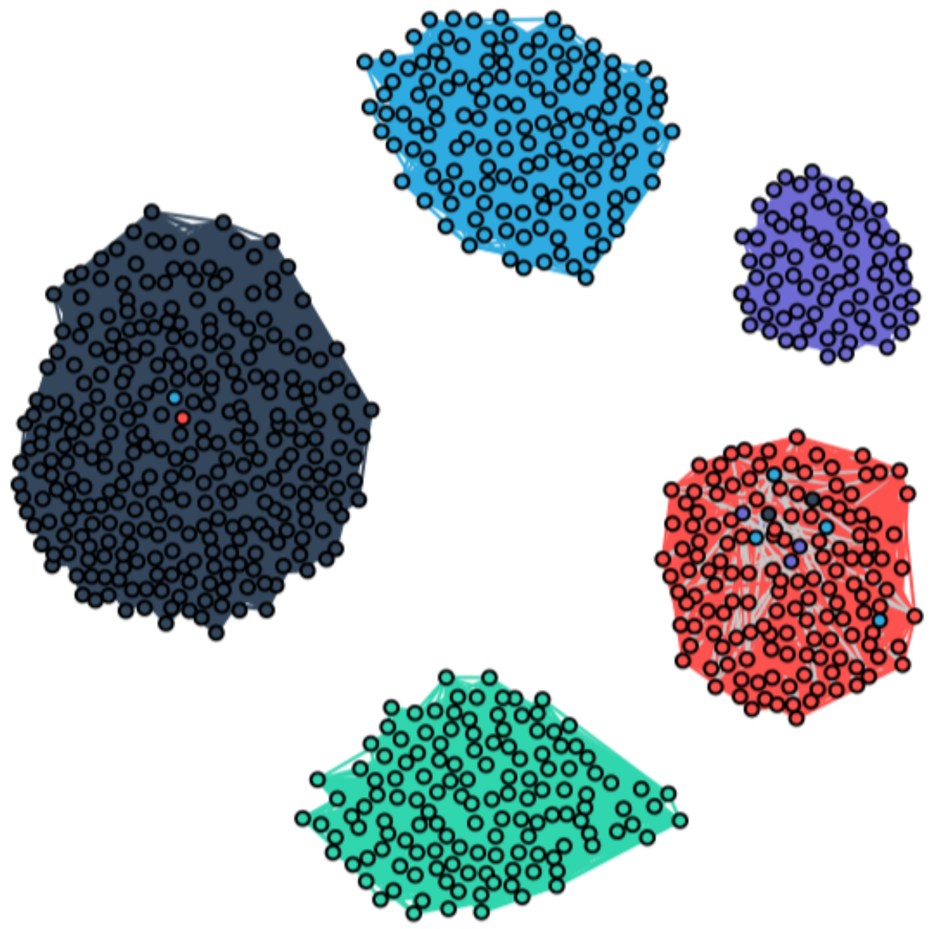
## Cancer Genome data set



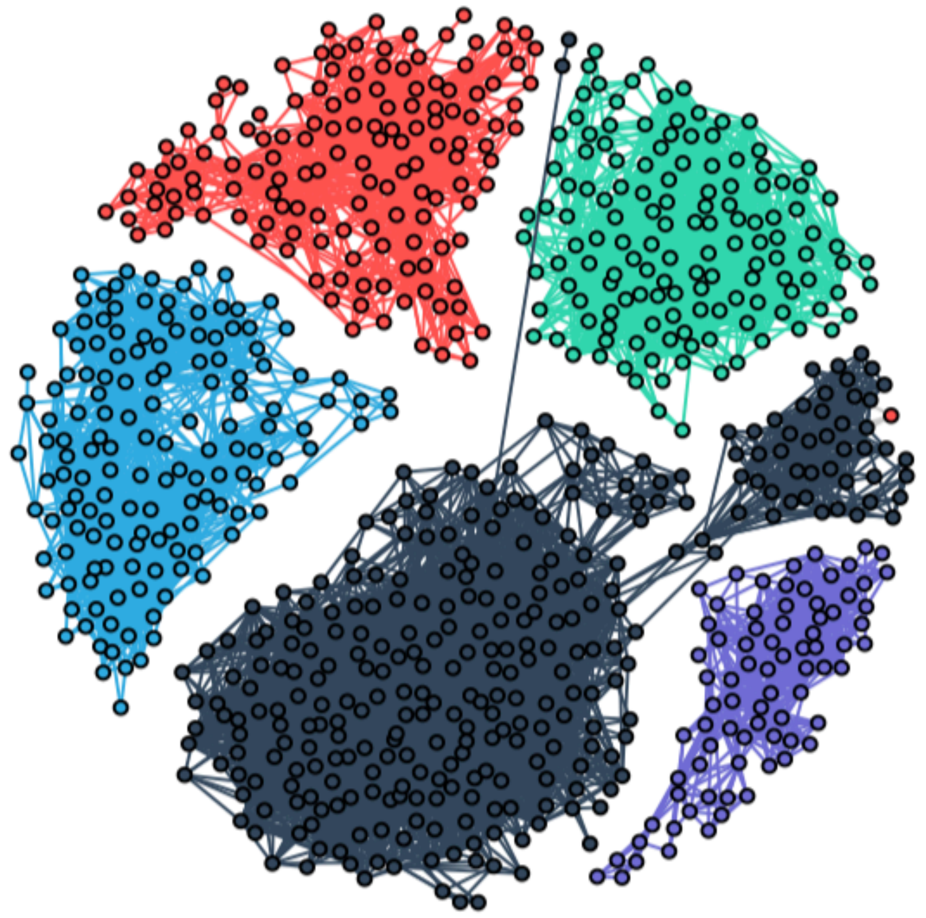
(a) GLasso (Friedman et al., 2008)



(b) GGL (Egilmez et al., 2017)



(c) CLR (Nie et al., 2016)



(d) SGL (proposed)