Multilayer Spectral Graph Clustering via Convex Layer Aggregation

Pin-Yu Chen\textsuperscript{1,2}  
Alfred Hero\textsuperscript{1}

\textsuperscript{1}Department of Electrical Engineering and Computer Science  
University of Michigan  
\textsuperscript{2}IBM Thomas J. Watson Research Center

pin-yu.chen@ibm.com  
hero@umich.edu

December 7, 2016
Multilayer Graph Clustering / Community Detection

- Common node set + Different types of relation (layers)
- Goal: assign consensus cluster/community label to each node
- Key challenge: How to combine information from different layers?
Multilayer Spectral Graph Clustering via Convex Layer Aggregation

- Multilayer graph: $L$ layers of graphs with common node set
- Layer weight vector $\mathbf{w} = [w_1, \ldots, w_L]$, $w_\ell \geq 0$ and $\sum_{\ell=1}^{L} w_\ell = 1$
Multilayer Spectral Graph Clustering via Convex Layer Aggregation

- $L$ layers of weighted undirected graphs $G_\ell = (\mathcal{V}, \mathcal{E}_\ell)$, $1 \leq \ell \leq L$.
  - $|\mathcal{V}| = n$ and $|\mathcal{E}_\ell| = m_\ell$
- $A^{(\ell)}$: binary adjacency matrix of $G_\ell$
- $W^{(\ell)}$: nonnegative edge weight matrix of $G_\ell$
- Aggregated matrices $A^w = \sum_{\ell=1}^L w_\ell A^{(\ell)}$, $W^w = \sum_{\ell=1}^L w_\ell W^{(\ell)}$

Multilayer SGC Algorithm

Given $\{G_\ell\}_{\ell=1}^L$, layer weight vector $w$, # of clusters $K$

1. Compute graph Laplacian matrix $L^w = S^w - W^w$, $S^w = \text{diag}(W^w 1_n)$
2. Obtain the $K$ smallest eigenvectors $\{y_k\}_{k=1}^K$ of $L^w$. $Y = [y_1 \ y_2 \ \cdots \ y_K]$.
3. Perform K-means on the rows of $Y$ to separate the nodes into $K$ groups

- **Question I:** The effect of $w$ on multilayer SGC? - this talk
- **Question II:** How to select the best $w$ and $K$? - ongoing work
Multilayer Block Model and Multilayer RIM

- Multilayer block model - $K$ clusters & $L$ layers:

$$A^{(\ell)} = \begin{bmatrix}
A_1^{(\ell)} & C_{12}^{(\ell)} & C_{13}^{(\ell)} & \cdots & C_{1K}^{(\ell)} \\
C_{21}^{(\ell)} & A_2^{(\ell)} & C_{23}^{(\ell)} & \cdots & C_{2K}^{(\ell)} \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
C_{K1}^{(\ell)} & C_{K2}^{(\ell)} & \cdots & \cdots & A_K^{(\ell)}
\end{bmatrix}, \quad \ell \in \{1, 2, \ldots, L\}$$

- Similar block model for edge weight matrix $W^{(\ell)}$

Multilayer Random Interconnection Model (RIM) [Chen-Hero’16]

1. $A_k^{(\ell)}$ and $W_k^{(\ell)}$ arbitrary; $1 \leq k \leq K$, $1 \leq \ell \leq L$
2. $[C_{ij}^{(\ell)}]_{uv} \sim \text{Bernoulli}(p_{ij}^{(\ell)}); \quad 1 \leq i, j \leq K, \ i \neq j, \ \forall \ \ell$
3. $[W_{ij}^{(\ell)}]_{uv} \sim \text{common nonnegative bounded distribution with mean } \overline{W}_{ij}^{(\ell)}, \ \forall \ \ell$

Chen-Hero, “Phase Transitions and a Model Order Selection Criterion for Spectral Graph Clustering”, arXiv 2016
**“Signal + Noise” Perspective**

- **Signal**: (aggregated) within-cluster edges (fixed and arbitrary)
- **Noise**: (aggregated) between-cluster edges (varying and random)

Multilayer RIM: correlated signal \( \{W_{k}^{(\ell)}\}_{\ell=1}^{L} \) + independent Bernoulli noise \( \{C_{ij}^{(\ell)}\} \) and edge weight \( \{W_{ij}^{(\ell)}\} \)

How does the noise level in each layer and the layer weight vector \( \mathbf{w} \) affect the performance of multilayer SGC?
Multilayer SGC via Convex Layer Aggregation - Analysis

- $n_k$: # of nodes in cluster $k$. $n_{\text{min}} = \min_k n_k$. $n_{\text{max}} = \max_k n_k$.
- $L^w_k$: aggregated graph Laplacian matrix of cluster $k$.
- $S_{2:K}(L^w_k) = \sum_{k=2}^{K} \lambda_k(L^w_k)$. $\lambda_k(L^w_k)$: $k$-th smallest eigenvalue
- Layer-wise block noise level: $t^{(\ell)}_{ij} = p^{(\ell)}_{ij} \cdot \overline{W}_{ij}$. $t_{\text{max}} = \max_{i,j} t^{(\ell)}_{ij}$.
- Layer-wise homogeneous RIM: $t^{(\ell)}_{ij} = t^{(\ell)}$; otherwise layer-wise inhomogeneous RIM
- Aggregated noise level under hom-RIM: $t^w = \sum_{\ell=1}^{L} w_\ell t^{(\ell)}$
- Aggregated maximum noise level under inhom-RIM: $t^w_{\text{max}} = \sum_{\ell=1}^{L} w_\ell t^{(\ell)}_{\text{max}}$

**Theorem (Summary of Phase Transition Analysis)**

1. Given $w$, under the layer-wise hom-RIM, there exists a threshold $t^{w*}$ s.t. the clusters can be detected when $t^w < t^{w*}$, and undetectable when $t^w > t^{w*}$
2. Given $w$, under the layer-wise inhom-RIM, high cluster detectability can be guaranteed if $t^w_{\text{max}} < t^{w*}$
3. $t^w_{\text{LB}} \leq t^{w*} \leq t^w_{\text{UB}}$. $t^w_{\text{LB}} = \frac{\min_k \in \{1,2,\ldots,K\} S_{2:K}(L^w_k)}{(K-1)n_{\text{max}}}$; $t^w_{\text{UB}} = \frac{\min_k \in \{1,2,\ldots,K\} S_{2:K}(L^w_k)}{(K-1)n_{\text{min}}}$
4. (Universal lower bound) For any $w$, $t^{w*} \geq \frac{\min_{\ell \in \{1,2,\ldots,L\}} \min_k \in \{1,2,\ldots,K\} S_{2:K}(L^w_{k\ell})}{(K-1)n_{\text{max}}}$
Multilayer SGC via Convex Layer Aggregation - Analysis

- $n_k$: # of nodes in cluster $k$. $n_{\text{min}} = \min_k n_k$. $n_{\text{max}} = \max_k n_k$.
- $L^w_k$: aggregated graph Laplacian matrix of cluster $k$.
- $S_{2:K}(L^w_k) = \sum_{k=2}^K \lambda_k(L^w_k)$. $\lambda_k(L^w_k)$: $k$-th smallest eigenvalue.
- Layer-wise block noise level: $t_{ij}^{(\ell)} = p_{ij}^{(\ell)} \cdot W_{ij}^{(\ell)}$. $t_{\text{max}}^{(\ell)} = \max_{i,j} t_{ij}^{(\ell)}$.
- Layer-wise homogeneous RIM: $t_{ij}^{(\ell)} = t^{(\ell)}$; otherwise layer-wise inhomogeneous RIM.
- Aggregated noise level under hom-RIM: $t^w = \sum_{\ell=1}^L w_{\ell} t^{(\ell)}$.
- Aggregated maximum noise level under inhom-RIM: $t_{\text{max}}^w = \sum_{\ell=1}^L w_{\ell} t_{\text{max}}^{(\ell)}$.

**Theorem (Summary of Phase Transition Analysis)**

1. **Given $w$, under the layer-wise hom-RIM, there exists a threshold $t^w_*$ s.t. the clusters can be detected when $t^w < t^w_*$, and undetectable when $t^w > t^w_*$.**
2. **Given $w$, under the layer-wise inhom-RIM, high cluster detectability can be guaranteed if $t_{\text{max}}^w < t^w_*$.**
3. $t_{\text{LB}}^w \leq t^w_* \leq t_{\text{UB}}^w$. $t_{\text{LB}}^w = \frac{\min_k \in \{1,2,\ldots,K\} S_{2:K}(L^w_k)}{(K-1)n_{\text{max}}}$; $t_{\text{UB}}^w = \frac{\min_k \in \{1,2,\ldots,K\} S_{2:K}(L^w_k)}{(K-1)n_{\text{min}}}$.
4. (Universal lower bound) For any $w$, $t^w_* \geq \frac{\min_{\ell \in \{1,2,\ldots,L\}} \min_k \in \{1,2,\ldots,K\} S_{2:K}(L^{(\ell)}_k)}{(K-1)n_{\text{max}}}$. 

P.-Y. Chen
GlobalSIP 2016
December 7, 2016 8 / 14
Theorem (block-wise identical noise $t^{(\ell)} \cdot t^w = \sum_{\ell=1}^{L} w_{\ell} t^{(\ell)}$)

Given a layer weight vector $w$, recall $S_{2:K}(L) = \sum_{k=2}^{K} \lambda_k(L)$ and $Y = [y_2 \cdots y_K] = [Y_1^T \ Y_2^T \cdots \ Y_K^T]^T$. There exists a critical value $t^w*$ such that the following holds almost surely as $n_k \to \infty \ \forall \ k$ and \( \frac{n_{\min}}{n_{\max}} \to c > 0$:

(a) (separability) \[
\begin{cases}
\text{If } t^w < t^w*, & Y_k = [v_1^k 1_{n_k}, v_2^k 1_{n_k}, \ldots, v_{K-1}^k 1_{n_k}] \\
\text{If } t^w > t^w*, & Y_k^T 1_{n_k} = 0_{K-1}
\end{cases}
\]

(b) (noise level bounds) $t_{LB}^w \leq t^w* \leq t_{UB}^w$, where

$$t_{LB}^w = \min_{k \in \{1,2,\ldots,K\}} \frac{S_{2:K}(L^w_k)}{(K-1)n_{\max}}; \quad t_{UB}^w = \min_{k \in \{1,2,\ldots,K\}} \frac{S_{2:K}(L^w_k)}{(K-1)n_{\min}}.$$

- When $t^w < t^w*$, $Y$ has the following properties:
  1. The columns of $Y_k$ are constant vectors
  2. $\sum_{k} n_k v_j^k = 0, \ \forall \ j \in \{1, 2, \ldots, K - 1\}$
  3. The row vectors of $Y_k$ are identical and cluster-wise distinct
Analysis under Layer-wise Inhomogeneous RIM

- $Y$ : eigenvector matrix of the graph Laplacian $L^w$ under the block-wise non-identical noise model
- $\tilde{Y}$ : eigenvector matrix of the graph Laplacian $\tilde{L}^w$ under the block-wise identical noise model with aggregated noise level $t^w$
- $v = [\cos^{-1} \sigma_1(Y^T \tilde{Y}), \ldots, \cos^{-1} \sigma_{K-1}(Y^T \tilde{Y})]^T$: principal angle
- $\Theta(Y, \tilde{Y}) = \text{diag}(v)$. $\sin \Theta(Y, \tilde{Y})$ defined entrywise.

Theorem (block-wise non-identical noise $t^{(\ell)}_{ij}$. $t^{(\ell)}_{\text{max}} = \max_{i,j} t^{(\ell)}_{ij}$)

Given a layer weight vector $w$, let $t^{w*}$ be the critical threshold value for the block-wise identical noise model. Under the same assumption as in the previous theorem, let $t^{w}_{\text{max}} = \sum_{\ell=1}^{L} w^{(\ell)} t^{(\ell)}_{\text{max}}$.

If $t^{w}_{\text{max}} < t^{w*}$, $\| \sin \Theta(Y, \tilde{Y}) \|_F \leq \min_{t^{w} \leq t^{w}_{\text{max}}} \frac{\|L^w - \tilde{L}^w\|_F}{n \delta^{t^w}}$, where $\delta^{t^w}$ is some constant.
Signal: joint within-cluster edge connection probability \( \{q_{xy}\}_{x,y \in \{0,1\}} \) across \( L = 2 \) layers

Noise: layer-wise & cluster-wise independent between-cluster edge connection probability \( \{p^{(\ell)}\}_{\ell = 1}^2 \) under hom-RIM

Phase transitions incurred by noise levels for a given \( \mathbf{w} = [w_1 \ w_2]^T \):

(a) \( (w_1, w_2) = (0.8, 0.2) \)  
(b) \( (w_1, w_2) = (0.5, 0.5) \)  
(c) \( (w_1, w_2) = (0.2, 0.8) \)

Figure: \( n_1 = n_2 = n_3 = 1000 \), \( q_{11} = 0.3 \), \( q_{10} = 0.2 \), \( q_{01} = 0.1 \), and \( q_{00} = 0.4 \).
Simulation - Universal Phase Transition Lower Bound

Figure: Two-layer correlated graphs. Averaged over 20 uniformly selected layer weight vectors $w \in \mathcal{W}_2$. $n_1 = n_2 = n_3 = 200$, $q_{11} = 0.3$, $q_{10} = 0.2$, $q_{01} = 0.1$, and $q_{00} = 0.4$. 
Layer weight vector $\mathbf{w} = [w_1 \ w_2]^T = [w_1 \ 1 - w_1]^T$

Phase transitions incurred by $\mathbf{w}$ for a given noise level $\{p^{(\ell)}\}_{\ell=1}^2$:

(a) low noise
(b) medium noise
(c) medium noise
(d) high noise

Figure: $n_1 = n_2 = n_3 = 1000$, $q_{11} = 0.3$, $q_{10} = 0.2$, $q_{01} = 0.1$, and $q_{00} = 0.4$. From left to right, $(p^{(1)}, p^{(2)}) = (0.2, 0.2), (0.2, 0.5), (0.5, 0.2)$, and $(0.5, 0.5)$, respectively.
Conclusion and Ongoing Work

- Phase transition analysis of multilayer spectral graph clustering (SGC) via convex layer aggregation under layer-wise homogeneous and inhomogeneous RIM
- The effect of layer weight vector $w$, cluster connectivity ($S_{2:K}(L_{k}^w)$), and noise level $t_{ij}^{(l)}$ on multilayer SGC
- Separability (Inseparability) of multilayer SGC w.r.t. the noise level
- Justification of phase transition in two-layer correlated Erdos-Renyi graphs incurred by noise level and layer weight vector
- (Ongoing work) Utilize the established phase transition analysis for selecting layer weight vector $w$ and number of clusters $K$