Multilayer Spectral Graph Clustering via Convex Layer Aggregation

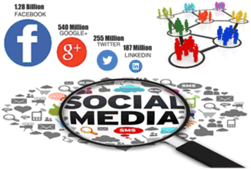
Pin-Yu Chen[1,2]
Alfred Hero[1]

[1]Department of Electrical Engineering and Computer Science
University of Michigan
[2]IBM Thomas J. Watson Research Center

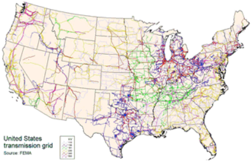
pin-yu.chen@ibm.com
hero@umich.edu

December 7, 2016

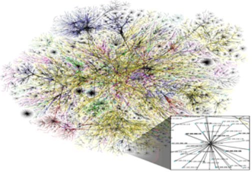
Graphs in Nature and Society



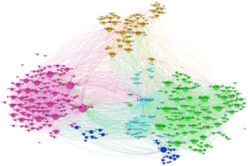
Social Network



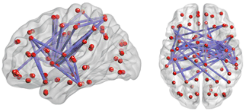
Power Grid



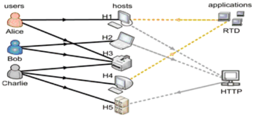
Communication Network



Information System

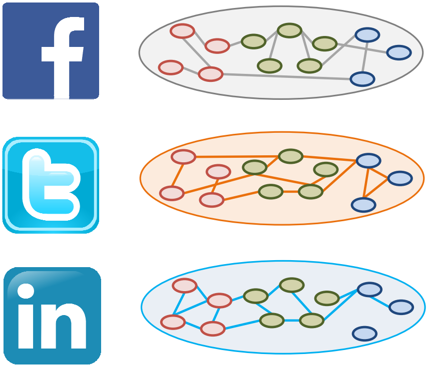


Bio Informatics



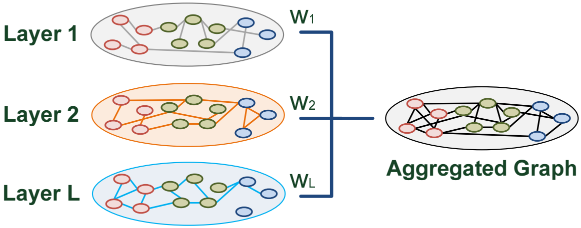
Cyber-Physical Sytem

Multilayer Graph Clustering / Community Detection

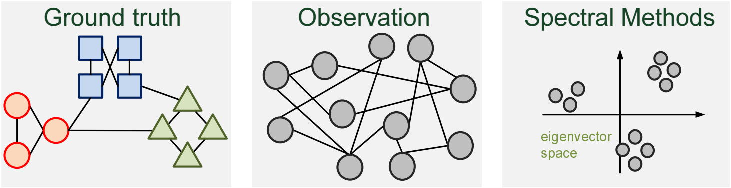


- Common node set + Different types of relation (layers)
- Goal: assign consensus cluster/community label to each node
- Key challenge: How to combine information from different layers?

Multilayer Spectral Graph Clustering via Convex Layer Aggregation



- Multilayer graph: L layers of graphs with common node set
- Layer weight vector $\mathbf{w} = [w_1, \ldots, w_L]$, $\mathbf{w}_\ell \geq 0$ 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 $|\mathcal{E}_\ell| = m_\ell$
- $\mathbf{A}^{(\ell)}$: binary adjacency matrix of G_ℓ
- $\mathbf{W}^{(\ell)}$: nonnegative edge weight matrix of G_ℓ
- Aggregated matrices $\mathbf{A}^{\mathbf{w}} = \sum_{\ell=1}^{L} w_\ell \mathbf{A}^{(\ell)}$, $\mathbf{W}^{\mathbf{w}} = \sum_{\ell=1}^{L} w_\ell \mathbf{W}^{(\ell)}$

Multilayer SGC Algorithm

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

1. Compute graph Laplacian matrix $\mathbf{L}^{\mathbf{w}} = \mathbf{S}^{\mathbf{w}} - \mathbf{W}^{\mathbf{w}}$, $\mathbf{S}^{\mathbf{w}} = \mathrm{diag}(\mathbf{W}^{\mathbf{w}} \mathbf{1}_n)$
2. Obtain the K smallest eigenvectors $\{\mathbf{y}_k\}_{k=1}^{K}$ of $\mathbf{L}^{\mathbf{w}}$. $\mathbf{Y} = [\mathbf{y}_1 \; \mathbf{y}_2 \; \cdots \; \mathbf{y}_K]$.
3. Perform K-means on the rows of $\mathbf{Y}$ to separate the nodes into K groups

- **Question I:** The effect of $\mathbf{w}$ on multilayer SGC? - this talk
- **Question II:** How to select the best $\mathbf{w}$ and K? - ongoing work

Multilayer Block Model and Multilayer RIM

- Multilayer block model - K clusters & L layers:

$$\mathbf{A}^{(\ell)} = \begin{bmatrix} \mathbf{A}_1^{(\ell)} & \mathbf{C}_{12}^{(\ell)} & \mathbf{C}_{13}^{(\ell)} & \cdots & \mathbf{C}_{1K}^{(\ell)} \\ \mathbf{C}_{21}^{(\ell)} & \mathbf{A}_2^{(\ell)} & \mathbf{C}_{23}^{(\ell)} & \cdots & \mathbf{C}_{2K}^{(\ell)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{C}_{K1}^{(\ell)} & \mathbf{C}_{K2}^{(\ell)} & \cdots & \cdots & \mathbf{A}_K^{(\ell)} \end{bmatrix}, \ \ell \in \{1, 2, \ldots, L\}$$

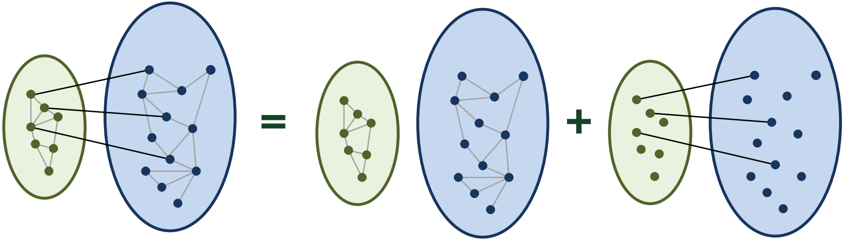
- Similar block model for edge weight matrix $\mathbf{W}^{(\ell)}$

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

1. $\mathbf{A}_k^{(\ell)}$ and $\mathbf{W}_k^{(\ell)}$ arbitrary; $1 \leq k \leq K$, $1 \leq \ell \leq L$
2. $[\mathbf{C}_{ij}^{(\ell)}]_{uv} \sim$ Bernoulli$(p_{ij}^{(\ell)})$; $1 \leq i, j \leq K$, $i \neq j$, $\forall \ \ell$
3. $[\mathbf{W}_{ij}^{(\ell)}]_{uv} \sim$ 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 $\{\mathbf{W}_k^{(\ell)}\}_{\ell=1}^{L}$ + independent Bernoulli noise $\{\mathbf{C}_{ij}^{(\ell)}\}$ and edge weight $\{\mathbf{W}_{ij}^{(\ell)}\}$
- 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. $n_{\min} = \min_k n_k$. $n_{\max} = \max_k n_k$.
- $\mathbf{L}_k^{\mathbf{w}}$: aggregated graph Laplacian matrix of cluster k
- $S_{2:K}(\mathbf{L}_k^{\mathbf{w}}) = \sum_{k=2}^{K} \lambda_k(\mathbf{L}_k^{\mathbf{w}})$. $\lambda_k(\mathbf{L}_k^{\mathbf{w}})$: k-th smallest eigenvalue
- Layer-wise block noise level: $t_{ij}^{(\ell)} = p_{ij}^{(\ell)} \cdot \overline{W}_{ij}^{(\ell)}$. $t_{\max}^{(\ell)} = \max_{i,j} t_{ij}^{(\ell)}$.
- Layer-wise homogeneous RIM: $t_{ij}^{(\ell)} = t^{(\ell)}$; otherwise layer-wise inhomogeneous RIM
- Aggregated noise level under hom-RIM: $t^{\mathbf{w}} = \sum_{\ell=1}^{L} w_\ell t^{(\ell)}$
- Aggregated maximum noise level under inhom-RIM: $t_{\max}^{\mathbf{w}} = \sum_{\ell=1}^{L} w_\ell t_{\max}^{(\ell)}$

Theorem (Summary of Phase Transition Analysis)

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

Multilayer SGC via Convex Layer Aggregation - Analysis

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

Theorem (Summary of Phase Transition Analysis)

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

Analysis under Layer-wise Homogeneous RIM

Theorem (block-wise identical noise $t^{(\ell)}$. $t^{\mathbf{w}} = \sum_{\ell=1}^{L} w_\ell t^{(\ell)}$)

Given a layer weight vector $\mathbf{w}$, *recall* $S_{2:K}(\mathbf{L}) = \sum_{k=2}^{K} \lambda_k(\mathbf{L})$ *and* $\mathbf{Y} = [\mathbf{y}_2 \ \cdots \ \mathbf{y}_K] = [\mathbf{Y}_1^T \ \mathbf{Y}_2^T \ \cdots \ \mathbf{Y}_K^T]^T$. *There exists a critical value* $t^{\mathbf{w}*}$ *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) $\begin{cases} \text{If } t^{\mathbf{w}} < t^{\mathbf{w}*}, & \mathbf{Y}_k = \left[v_1^k \mathbf{1}_{n_k}, v_2^k \mathbf{1}_{n_k}, \ldots, v_{K-1}^k \mathbf{1}_{n_k}\right] \\ \text{If } t^{\mathbf{w}} > t^{\mathbf{w}*}, & \mathbf{Y}_k^T \mathbf{1}_{n_k} = \mathbf{0}_{K-1} \end{cases}$

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

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

- When $t^{\mathbf{w}} < t^{\mathbf{w}*}$, $\mathbf{Y}$ has the following properties:
  1. The columns of $\mathbf{Y}_k$ 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 $\mathbf{Y}_k$ are identical and cluster-wise distinct

Analysis under Layer-wise Inhomogeneous RIM

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

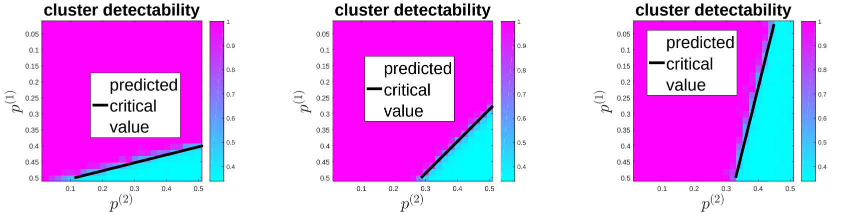
Theorem (block-wise non-identical noise $t_{ij}^{(\ell)}$. $t_{\max}^{(\ell)} = \max_{i,j} t_{ij}^{(\ell)}$)

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

If $t_{\max}^{\mathbf{w}} < t^{\mathbf{w}*}$, $\|\sin \mathbf{\Theta}(\mathbf{Y}, \widetilde{\mathbf{Y}})\|_F \leq \min_{t^{\mathbf{w}} \leq t_{\max}^{\mathbf{w}}} \frac{\|\mathbf{L}^{\mathbf{w}} - \widetilde{\mathbf{L}}^{\mathbf{w}}\|_F}{n \delta_{t^{\mathbf{w}}}}$,
where $\delta_{t^{\mathbf{w}}}$ is some constant.

Simulation - Two-Layer Correlated Erdos-Renyi Graph

- 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

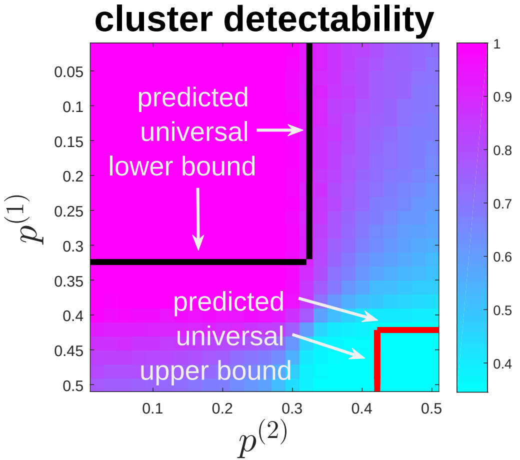
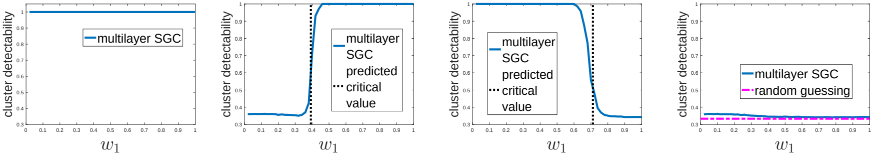


Figure: Two-layer correlated graphs. Averaged over 20 uniformly selected layer weight vectors $\mathbf{w} \in \mathcal{W}_2$. $n_1 = n_2 = n_3 = 200$, $q_{11} = 0.3$, $q_{10} = 0.2$, $q_{01} = 0.1$, and $q_{00} = 0.4$.

Simulation - Two-Layer Correlated Erdos-Renyi Graph

- 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 $\mathbf{w}$, cluster connectivity $(S_{2:K}(\mathbf{L}_k^{\mathbf{w}}))$, and noise level $t_{ij}^{(\ell)}$ 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 $\mathbf{w}$ and number of clusters K