AMOS: An Automated Model Order Selection Algorithm for Spectral Graph Clustering

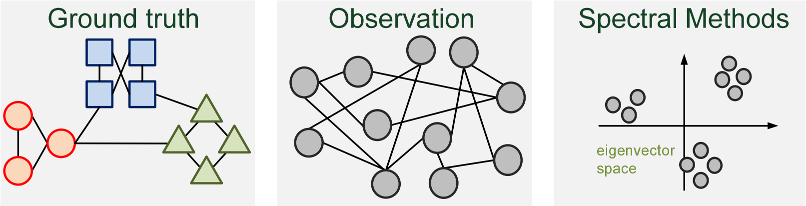
Pin-Yu Chen[1,2], **Thibaut Gensollen**[1], **Alfred Hero**[1]

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

pin-yu.chen@ibm.com
{thibautg,hero}@umich.edu

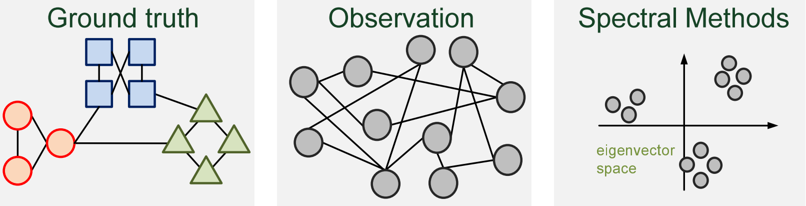
March 8, 2017

Graph Clustering/Community Detection



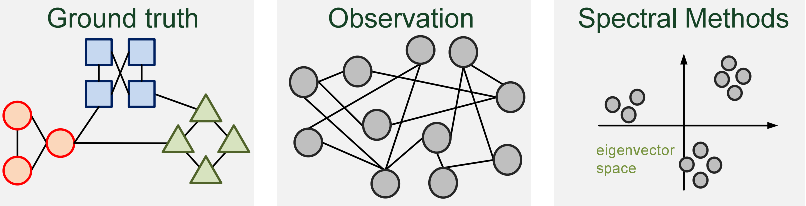
- **Goal:** separate the nodes in the graph into groups of high similarity
- **Applications:** network analysis, unsupervised learning, image segmentation, recommendation systems, ...
- **Challenge I:** unknown number K of clusters (communities)
  1. eigen-spectra based approach [Polito'01,Ng'02,Luxburg'07]
  2. eigenvector based approach [Zelnik-Manor'04]
- **Challenge II:** lack of absolute criterion for clustering reliability
  $\rightarrow$ many clustering evaluation metrics are relative criterion:
  1. cut-based score: min-cut, ratio-cut, ...
  2. modularity

Graph Clustering/Community Detection



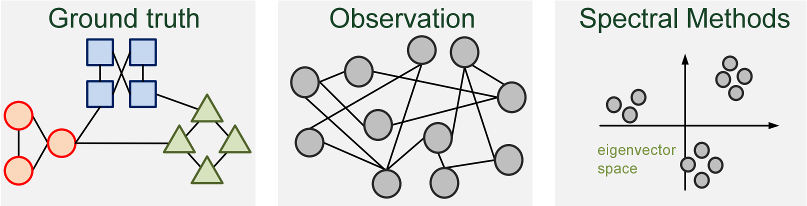
- **Goal:** separate the nodes in the graph into groups of high similarity
- **Applications:** network analysis, unsupervised learning, image segmentation, recommendation systems, ...
- **Challenge I:** unknown number K of clusters (communities)
  1. eigen-spectra based approach [Polito'01,Ng'02,Luxburg'07]
  2. eigenvector based approach [Zelnik-Manor'04]
- **Challenge II:** lack of absolute criterion for clustering reliability
  $\rightarrow$ many clustering evaluation metrics are relative criterion:
  1. cut-based score: min-cut, ratio-cut, ...
  2. modularity

Graph Clustering/Community Detection



- **Goal:** separate the nodes in the graph into groups of high similarity
- **Applications:** network analysis, unsupervised learning, image segmentation, recommendation systems, . . .
- **Challenge I:** unknown number K of clusters (communities)
  1. eigen-spectra based approach [Polito'01,Ng'02,Luxburg'07]
  2. eigenvector based approach [Zelnik-Manor'04]
- **Challenge II:** lack of absolute criterion for clustering reliability
  $\rightarrow$ many clustering evaluation metrics are relative criterion:
  1. cut-based score: min-cut, ratio-cut, . . .
  2. modularity

Summary



- Highlights of this talk:
  1. Spectral properties of **Graph Laplacian matrix** under a general network model
  2. An automated model order selection algorithm (AMOS) for spectral graph clustering with statistical clustering reliability guarantees

Block Representation for Clusters in Weighted Graphs

- $G = (\mathcal{V}, \mathcal{E}, \mathbf{W})$: undirected weighted graph of n nodes and m edges
- $\mathbf{A}$: $n \times n$ binary adjacency matrix - $[\mathbf{A}]_{uv} = 1$ if $(u,v) \in \mathcal{E}$
- $\mathbf{W}$: $n \times n$ nonnegative edge weight matrix - $[\mathbf{W}]_{uv} > 0$ if $(u,v) \in \mathcal{E}$
- Block representation of G with K clusters:

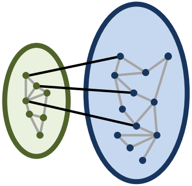
$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{C}_{12} & \mathbf{C}_{13} & \cdots & \mathbf{C}_{1K} \\ \mathbf{C}_{21} & \mathbf{A}_2 & \mathbf{C}_{23} & \cdots & \mathbf{C}_{2K} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_{K1} & \mathbf{C}_{K2} & \cdots & \cdots & \mathbf{A}_K \end{bmatrix}; \ \mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_{12} & \mathbf{W}_{13} & \cdots & \mathbf{W}_{1K} \\ \mathbf{W}_{21} & \mathbf{W}_2 & \mathbf{W}_{23} & \cdots & \mathbf{W}_{2K} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{W}_{K1} & \mathbf{W}_{K2} & \cdots & \cdots & \mathbf{W}_K \end{bmatrix}.$$

- $\mathbf{A}_k$ ($\mathbf{W}_k$): an $n_k \times n_k$ adjacency (weight) matrix of within-cluster edges in cluster k
- $\mathbf{C}_{ij}$ ($\mathbf{W}_{ij}$): an $n_i \times n_j$ adjacency (weight) matrix of between-cluster edges of clusters i and j. $\mathbf{C}_{ij} = \mathbf{C}_{ji}^T$. $\mathbf{W}_{ij} = \mathbf{W}_{ji}^T$.

Random Interconnection Model (RIM)

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

1. $\mathbf{A}_k$ and $\mathbf{W}_k$ arbitrary, $1 \leq k \leq K$ (within-cluster edges)
2. $[\mathbf{C}_{ij}]_{uv} \sim$ Bernoulli(p_{ij}), $1 \leq i,j \leq K$, $i \neq j$ (between-cluster edges)
3. $[\mathbf{W}_{ij}]_{uv} \sim$ common nonnegative bounded distribution with mean $\overline{W}_{ij}$



Block Models with K = 2 clusters

stochastic block model (SBM)

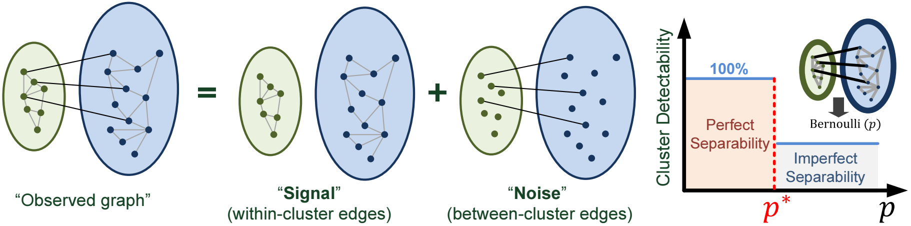
$$\mathbf{A} \sim \begin{array}{c} \left[\begin{array}{cc} \mathrm{Ber}\,(p_1) & \mathrm{Ber}\,(p) \\ & \mathrm{Ber}\,(p_2) \end{array}\right] \begin{array}{c} n_1 \\ n_2 \end{array} \\ \begin{array}{cc} n_1 & \quad\quad n_2 \end{array} \end{array}$$

random interconnection model (RIM)

$$\mathbf{A} \sim \begin{array}{c} \left[\begin{array}{cc} \mathrm{Arbitrary} & \mathrm{Ber}\,(p) \\ & \mathrm{Arbitrary} \end{array}\right] \begin{array}{c} n_1 \\ n_2 \end{array} \\ \begin{array}{cc} n_1 & \quad\quad n_2 \end{array} \end{array}$$

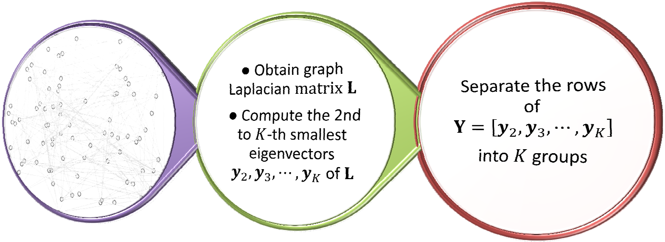
Chen-Hero, "Phase Transitions and a Model Order Selection Criterion for Spectral Graph Clustering", arXiv 2016

“Signal + Noise” Perspective



- **Signal:** within-cluster edges (fixed and arbitrary)
- **Noise:** between-cluster edges (varying and random)
- How does noise affect graph clustering? $\Rightarrow$ phase transition analysis

Spectral Graph Clustering (SGC) for K Clusters



- The graph G is undirected, weighted, and connected
- **spectral graph clustering (SGC) for K clusters**:
  1. Obtain the graph Laplacian matrix $\mathbf{L} = \mathbf{S} - \mathbf{W}$. $\mathbf{S}$ is a diagonal strength (degree) matrix. $(\lambda_k(\mathbf{L}), \mathbf{y}_k)$: k-th smallest eigenpair of $\mathbf{L}$.
  2. Compute the 2nd to the K-th smallest eigenvector of $\mathbf{L}$, $\mathbf{Y} = [\mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_K] \in \mathbb{R}^{n \times (K-1)}$.
  3. K-means clustering on the rows of $\mathbf{Y}$ to obtain K groups.

Phase Transition Analysis of SGC under RIM

- n_k : # of nodes in cluster k. $n_{\min} = \min_k n_k$. $n_{\max} = \max_k n_k$.
- $\mathbf{L}_k$: graph Laplacian matrix of cluster k
- Block noise level: $t_{ij} = p_{ij} \cdot \overline{W}_{ij}$. $t_{\max} = \max_{i,j} t_{ij}$.

Theorem (Homogeneous RIM: $t_{ij} = t$)

Let $S_{2:K}(\mathbf{L}) = \sum_{k=2}^{K} \lambda_k(\mathbf{L})$ *and* $\mathbf{Y} = [\mathbf{y}_2 \ \cdots \ \mathbf{y}_K] = [\mathbf{Y}_1^T \ \mathbf{Y}_2^T \ \cdots \ \mathbf{Y}_K^T]^T$, $\mathbf{Y}_k \in \mathbb{R}^{n_k \times (K-1)}$. *When one sweeps* t, *there exists a critical value* t^* *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) $\begin{cases} \text{If } t < t^*, \ \mathbf{Y}_k = \left[v_1^k \mathbf{1}, v_2^k \mathbf{1}, \ldots, v_{K-1}^k \mathbf{1}\right] = \mathbf{1}\mathbf{v}_k^T \\ \text{If } t > t^*, \ \mathbf{Y}_k^T \mathbf{1}_{n_k} = \mathbf{0}_{K-1} \end{cases}$

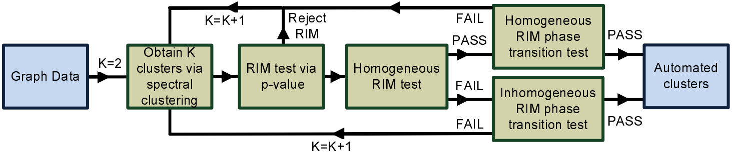
(b) (noise level bounds) $t_{LB} \le t^* \le t_{UB}$, *where*

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

- For inhomogeneous RIM (t_{ij} arbitrary), if $t_{\max} < t^*$, then cluster separability in $\mathbf{Y}$ can be guaranteed

Automated Model Order Selection (AMOS) for SGC

- Utilize phase transition analysis for determining the number of clusters (model order) and evaluating clustering quality (noise estimation)



- Iterating K, obtain clusters $\{\mathcal{C}_k\}_{k=1}^K$ from SGC
  1. Check between-cluster connectivity $\{\widehat{\mathbf{C}}_{ij}\}$ fits the RIM or not (V-test)
  2. If every $\widehat{\mathbf{C}}_{ij}$ fits the RIM, estimate the RIM parameters using $\{\mathcal{C}_k\}_{k=1}^K$
  3. Homogeneous RIM test: homogeneous or inhomogeneous RIM (GLRT)
  4. Homogeneous RIM phase transition test: test $\widehat{t} < \widehat{t}_{\text{LB}}$
  5. Inhomogeneous RIM phase transition test: test $\widehat{t}_{\max} < \widehat{t}_{\text{LB}}$
  6. Stop if item 4 or item 5 is true
- Provide statistical interpretation of clustering reliability
- Efficient incremental SGC [Chen-Zhang-Hasan-Hero KDD-MLG'16]
- AMOS codes: https://github.com/tgensol/AMOS

Performance Evaluation

- Comparative automated graph clustering methods:
  1. Louvain: greedy modularity maximization [Blonde'08]
  2. NB: spectral method using non-backtracking matrix [Krzakala'13]
  3. ST: self-tuning algorithm based on graph Laplacian [Zelnik-Manor'04]

Dataset	Method	NMI	Rand Index	F-measure	Conductance	Normalized Cut
IEEE RTS (3)	AMOS (3)	**.89**	**.96**	**.94**	.046	.068
	Louvain (6)	.74	.84	.67	.144	.169
	NB (3)	.75	.88	.81	.070	.100
	ST (2)	.74	.78	.75	**.021**	**.041**
Hibernia (2)	AMOS (2)	**1.0**	**1.0**	**1.0**	.030	.057
	Louvain (6)	.27	.51	.33	.222	.263
	NB (2)	.73	.89	.90	**.027**	.053
	ST (2)	.88	.96	.97	.028	**.050**
Cogent (2)	AMOS (4)	**.42**	**.63**	.53	**.036**	**.049**
	Louvain (11)	.25	.54	.26	.186	.204
	NB (3)	.26	.54	**.58**	.073	.109
	ST (14)	.34	.55	.29	.148	.164
Minnesota (-)	AMOS (46)	-	-	-	**.074**	**.076**
	Louvain (33)				.290	.299
	NB (35)				.140	.144
	ST (100)				.119	.120
Facebook (-)	AMOS (5)	-	-	-	**.004**	**.004**
	Louvain (17)				.076	.079
	NB (55)				.478	.486
	ST (7)				.006	.007

- AMOS is superior in most of clustering metrics

Conclusion and Ongoing Work

- Phase transition analysis of spectral graph clustering (SGC) under random interconnection model (RIM)
- Cluster separability (inseparability) in the eigenvector matrix $\mathbf{Y}$ of graph Laplacian matrix $\mathbf{L}$ w.r.t. noise level t (between-cluster edges)
- Closed-form expression for upper and lower bounds on t^*
- AMOS: theory-driven automated SGC with statistical clustering reliability guarantees
- Comparing multiple clustering metrics, AMOS outperforms 3 other automated methods in the datasets
- Ongoing work: automated graph clustering for multi-layer graphs

Reference

- U. Luxburg, "A tutorial on spectral clustering," Statistics and Computing, vol. 17, no. 4, pp. 395416, Dec. 2007.
- M. Polito, and P. Perona, "Grouping and dimensionality reduction by locally linear embedding", in Advances in neural information processing systems (NIPS), 2001, pp. 1255-1262.
- A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in Advances in neural information processing systems (NIPS), 2002, pp. 849856.
- P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," Social Networks, vol. 5, no. 2, pp. 109137, 1983.
- P.-Y. Chen and A. Hero, "Phase transitions in spectral community detection," IEEE Trans. Signal Process., vol. 63, no. 16, pp. 43394347, Aug 2015.
- P.-Y. Chen and A. O. Hero, "Phase transitions and a model order selection criterion for spectral graph clustering," arXiv preprint arXiv:1604.03159, 2016.
- P.-Y. Chen, B. Zhang, M. Hasan and A. O. Hero, "Incremental Method for Spectral Clustering of Increasing Orders," ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD) Workshop on Mining and Learning with Graphs (MLG), Aug. 2016
- P.-Y. Chen and A. O. Hero, "Multilayer Spectral Graph Clustering via Convex Layer Aggregation," IEEE GlobalSIP, 2016.
- V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," Journal of Statistical Mechanics: Theory and Experiment, no. 10, 2008.
- F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborova, and P. Zhang, "Spectral redemption in clustering sparse networks," Proc. National Academy of Sciences, vol. 110, pp. 20 93520 940, 2013.
- L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," in Advances in neural information processing systems (NIPS), 2004, pp. 16011608.