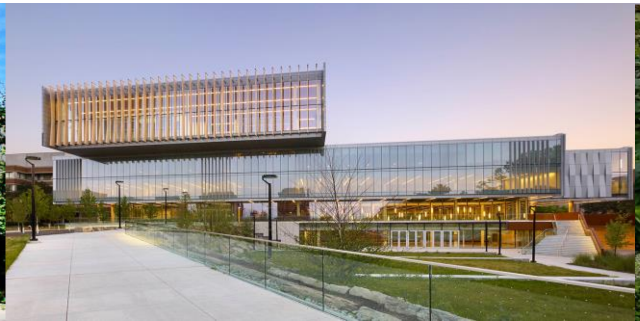
Graph Signal Processing: Theory and Applications to Imaging & Machine Learning

Gene Cheung
York University, Toronto, Canada

June 27, 2023

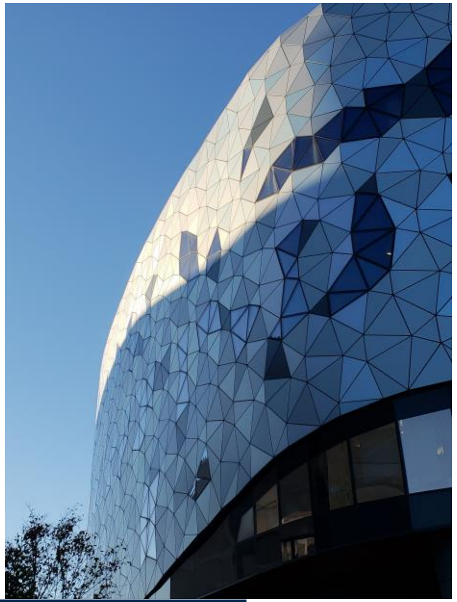
Acknowledgement

- **Graph and Image Signal Processing (GISP) Lab** (York University, Toronto, Canada)
  - Post-docs: Cheng Yang*, Xue Zhang*, Chinthaka Dinesh
  - Grad students: Saghar Bagheri, Tam Thuc Do, Yeganeh Gharedaghi, Niruhan Viswarupan
  - Co-supervised student: Yasaman Parhizkar
  - Visiting researchers: Sadid Sahami (NTHU), Fen Wang* (Xidian), Fei Chen* (Fudan)

- **Collaborators**
  - Michael Brown, Andrew Eckford, Pirathayini Srikantha (York Univ., Canada)
  - Ivan V. Bajic (Simon Fraser Univ., Canada)
  - Antonio Ortega (Univ. of Southern California, USA)
  - Stanley Chan (Purdue Univ., USA)
  - Wai-Tian Tan (Cisco, USA)
  - Phil Chou (Google, USA)
  - Tim Eadie (Growers Edge, USA)
  - Wei Hu (Peking Univ., China), Jin Zeng (Tongji Univ., China)
  - Vicky H. Zhao (Tsinghua Univ., China)
  - Chia-Wen Lin (NTHU, Taiwan)
  - Kazuya Kodama (NII, Japan), Yuichi Tanaka (Osaka Univ. Japan)

cisco

Google GROWERS EDGE



* September 2018 to June 2020.

Outline

- GSP overview
  - Graph frequencies from eigen-pairs
- Graph Learning
  - Positive, signed, directed, Hermitian graphs
- Graph Filtering
- Graph Sampling
- GSP Analysis for GCNs
- Conclusion

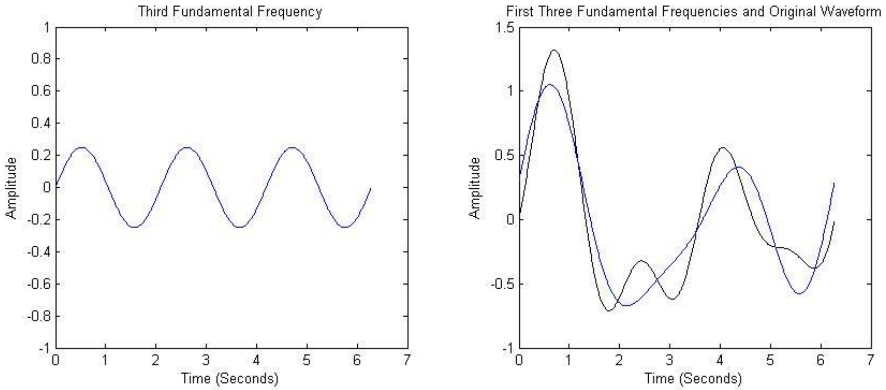
Outline

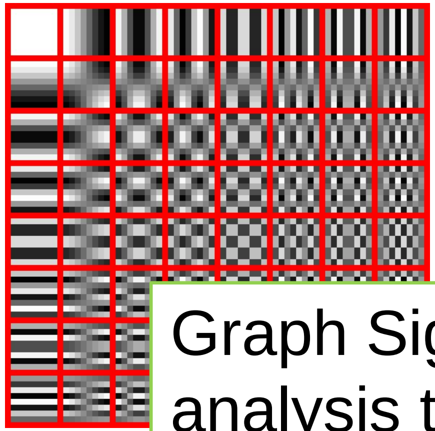
- GSP overview
  - Graph frequencies from eigen-pairs
- Graph Learning
  - Positive, signed, directed, Hermitian graphs
- Graph Filtering
- Graph Sampling
- GSP Analysis for GCNs
- Conclusion

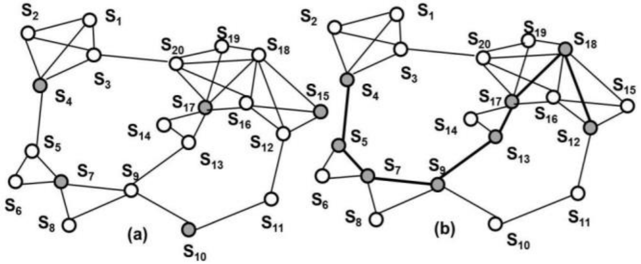
Gene Cheung (genec@yorku.ca)

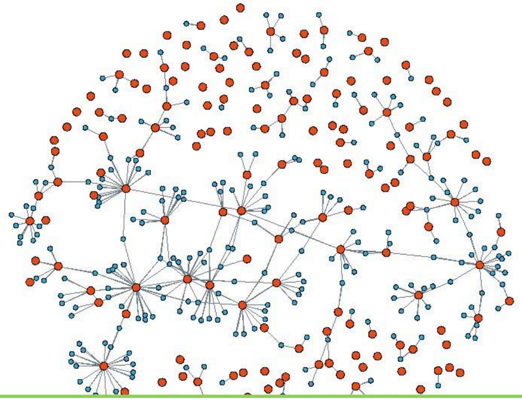
Graph Signal Processing

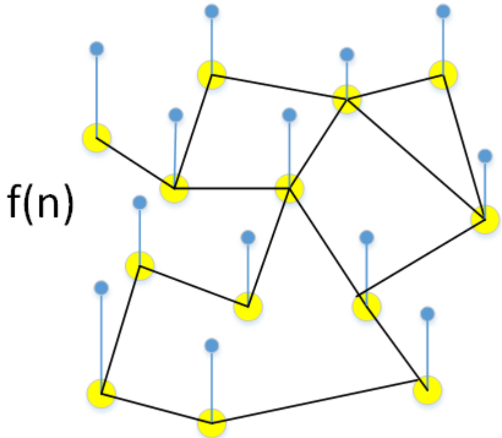
Frequency analysis + Graph kernels = Graph Signal Processing



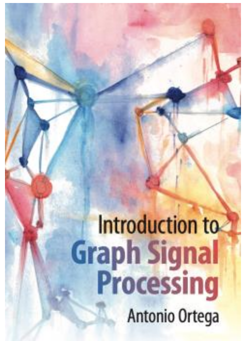


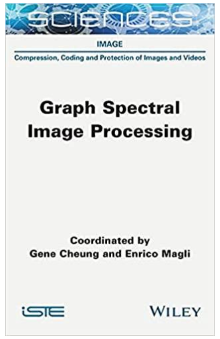






signal on graph kernel





Graph Signal Processing (GSP) studies spectral analysis tools for signals residing on graphs.

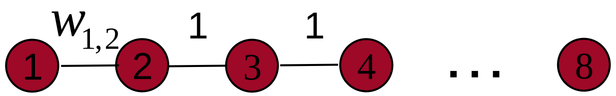
[1] A. Ortega, P. Frossard, J. Kovacevic, J. M. F. Moura, and P. Vandergheynst, “**Graph signal processing: Overview, challenges, and applications**,” *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

Graph Spectrum

Graph Fourier modes: eigenvectors of *graph Laplacian matrix* **L = D - W**.

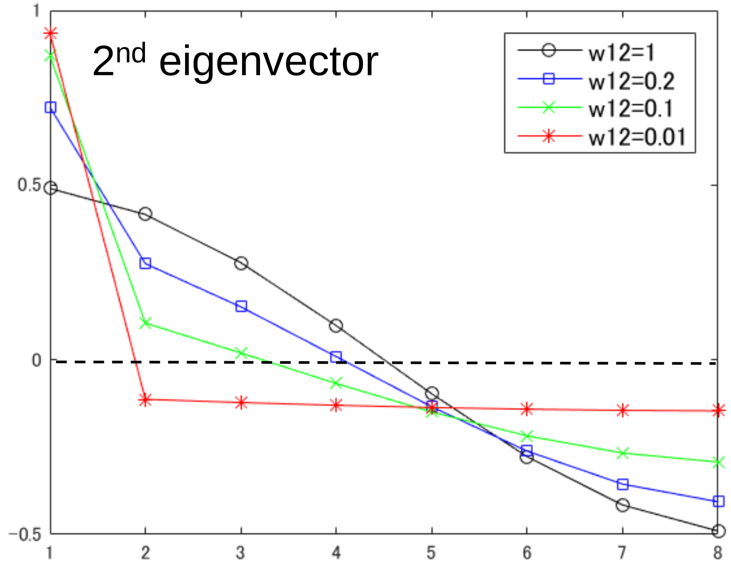
$$L = V \Sigma V^T$$

- Σ: eigenvalues along diagonal
- V: eigenvectors in columns
- V^T: Graph Fourier Transform (GFT)



GFT defaults to ***DCT*** for un-weighted connected line.

GFT defaults to ***DFT*** for un-weighted connected circle.



1. **Eigenvectors** are (***global***) aggregates of (***local***) edge weights.
   - More variations for larger eigenvalues.
2. **Eigenvalues** (≥ 0) as *graph frequencies*.

[1] G. Cheung, E. Magli, Y. Tanaka, M. Ng, "**Graph Spectral Image Processing**," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 907-930, May 2018. .

Graph Frequency Examples (US Temperature)

Weather stations from 100 most populated cities.

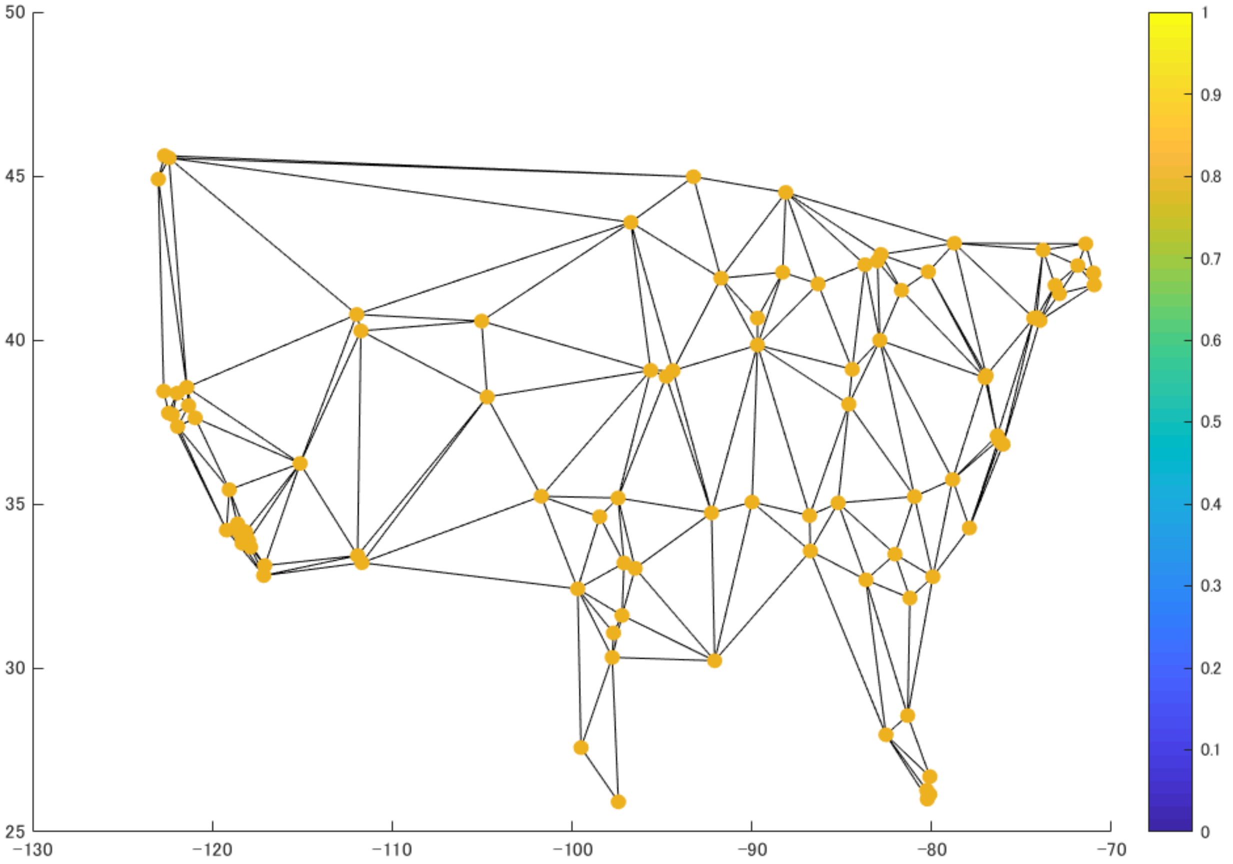
Graph connections from Delaunay Triangulation*.

Edge weights inverse proportional to distance.

location diff.

$$w_{i,j} = \exp\left(\frac{-\|l_i - l_j\|_2^2}{\sigma^2} \right)$$

Edge weights



V1: DC component

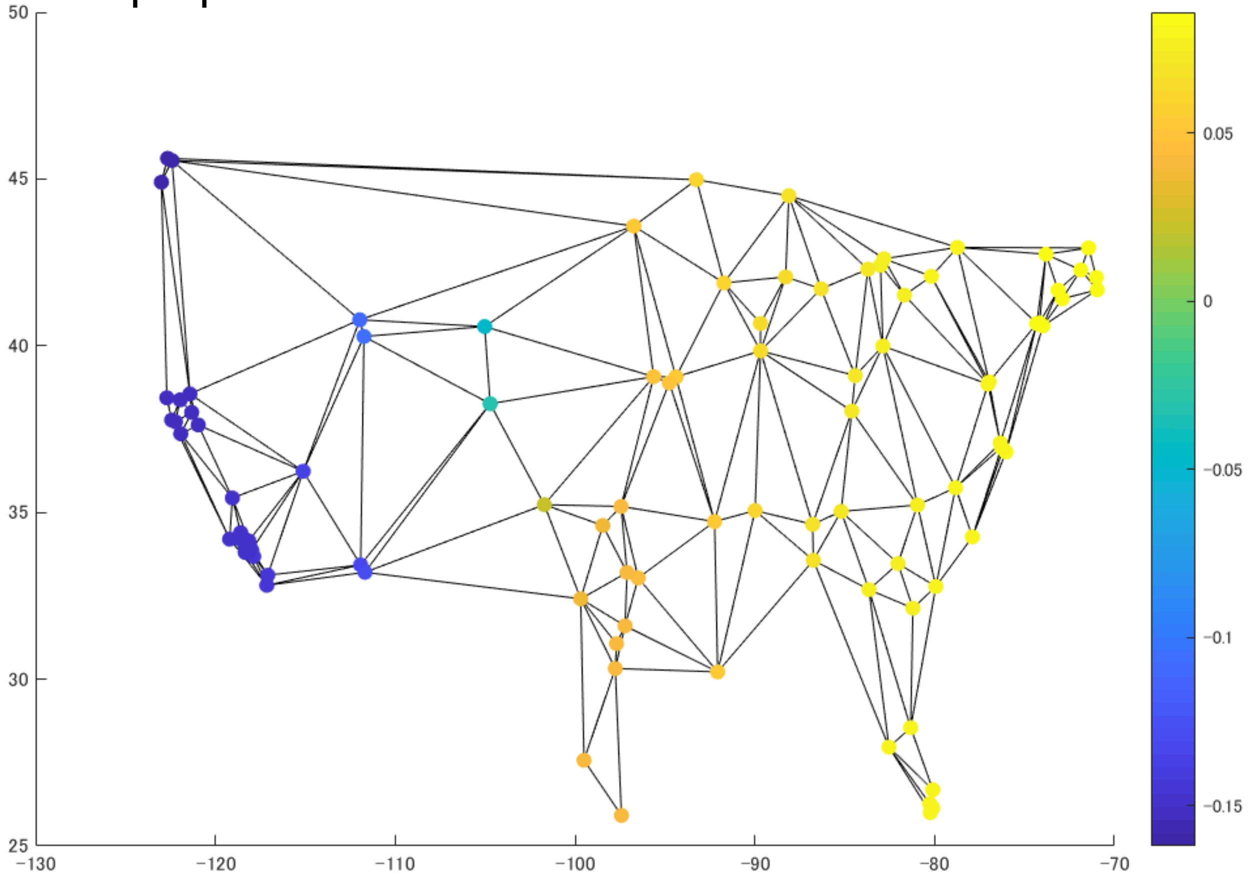
*https://en.wikipedia.org/wiki/Delaunay triangulation

Graph Frequency Examples (US Temperature)

Weather stations from 100 most populated cities.

Graph connections from Delaunay Triangulation*.

Edge weights inverse proportional to distance.



location diff.

$$w_{i,j} = \exp\left(\frac{-\|l_i - l_j\|_2^2}{\sigma^2}\right)$$

Edge weights

V2: 1st AC component

Graph Frequency Examples (US Temperature)

Weather stations from 100 most populated cities.

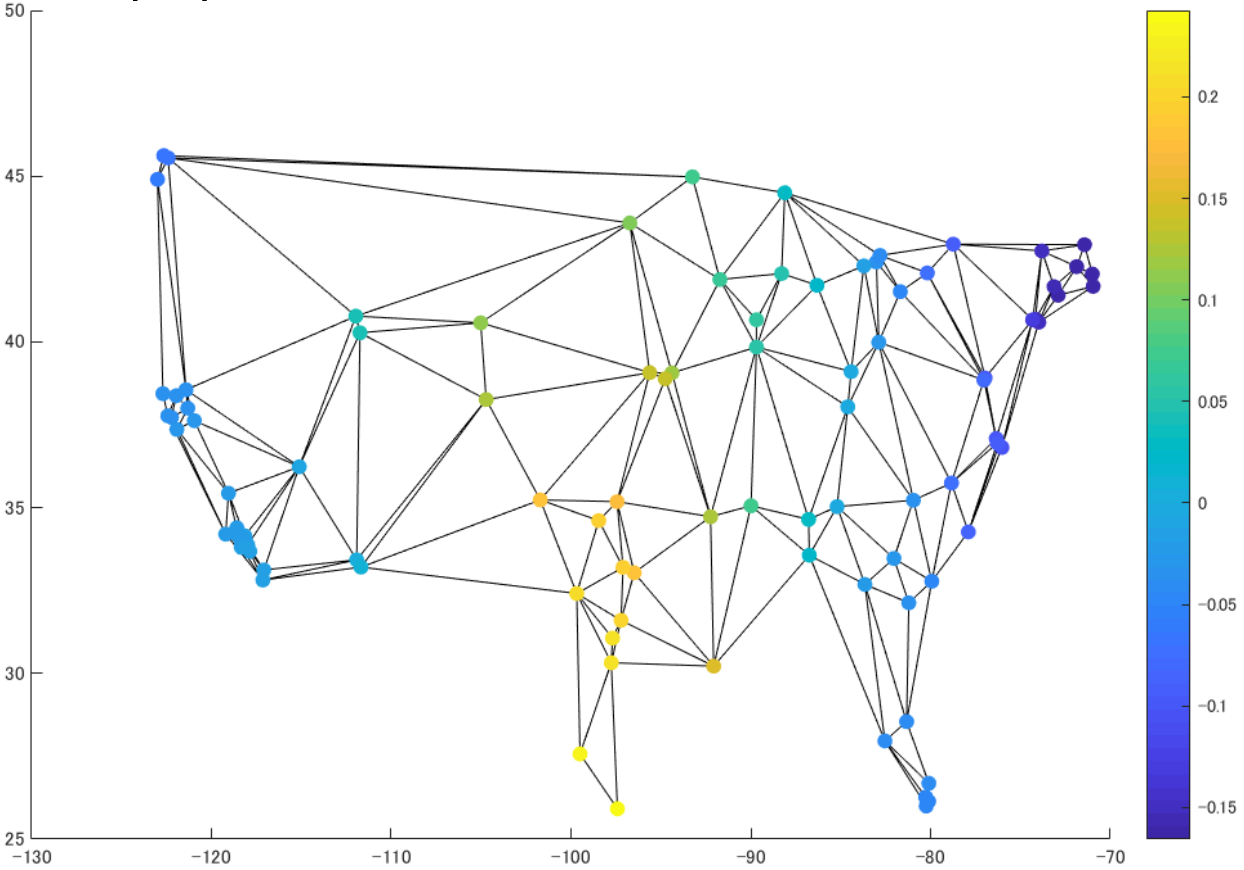
Graph connections from Delaunay Triangulation*.

Edge weights inverse proportional to distance.

location diff.

$$w_{i,j} = \exp\left(\frac{-\|l_i - l_j\|_2^2}{\sigma^2}\right)$$

Edge weights



V3: 2nd AC component

Graph Frequency Examples (US Temperature)

Weather stations from 100 most populated cities.

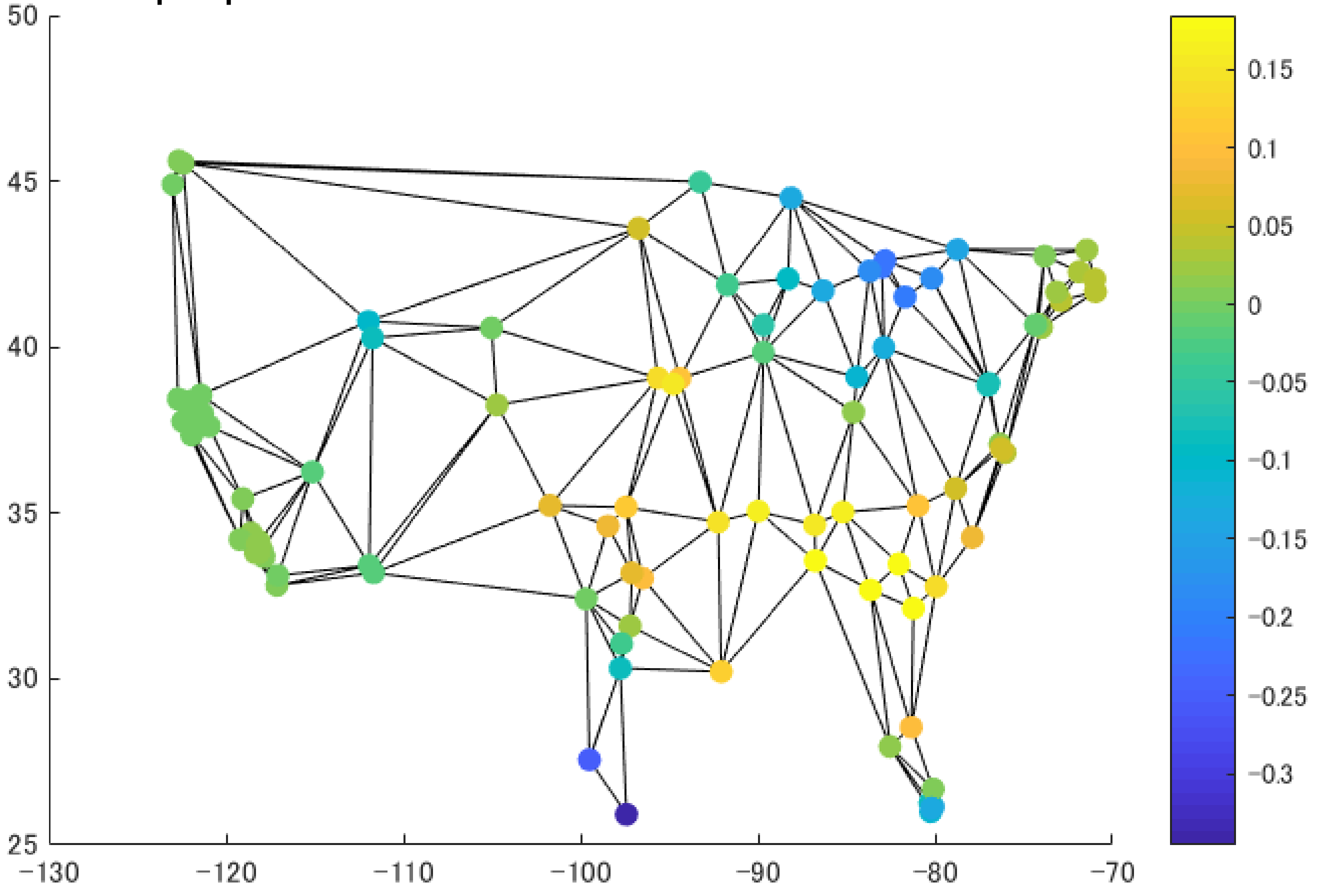
Graph connections from Delaunay Triangulation*.

Edge weights inverse proportional to distance.

location diff.

$$w_{i,j} = \exp\left(\frac{-\left\| l_i - l_j \right\|_2^2}{\sigma^2} \right)$$

Edge weights

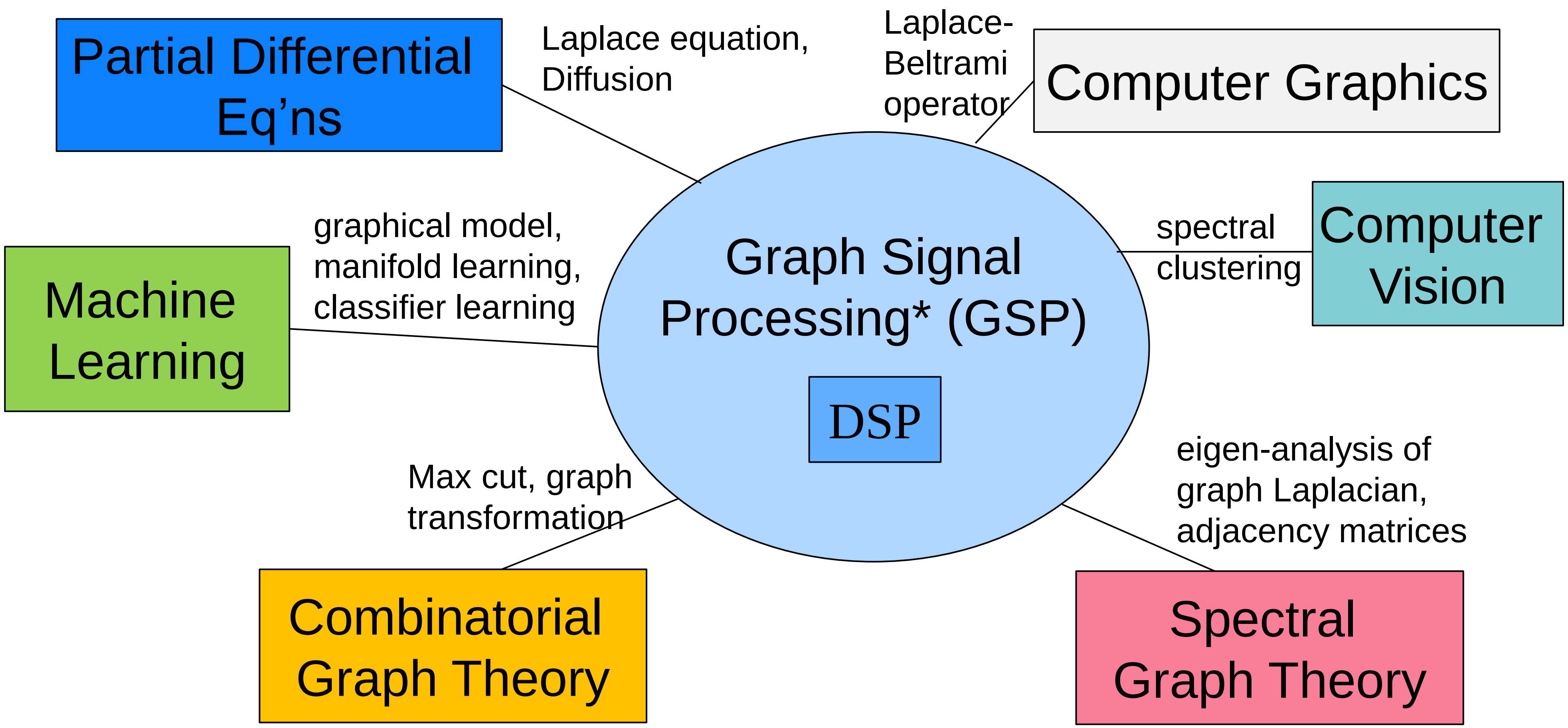


V4: 9th AC component

Gene Cheung (genec@yorku.ca)

GSP and Graph-related Research

GSP: SP framework that unifies concepts from multiple fields.



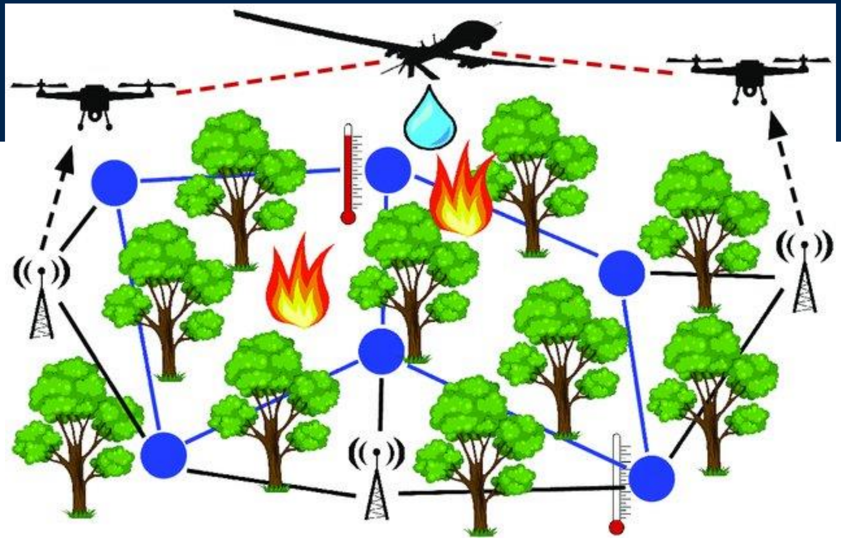
Outline

- GSP overview
  - Graph frequencies from eigen-pairs
- Graph Learning
  - Positive, signed, directed, Hermitian graphs
- Graph Filtering
- Graph Sampling
- GSP Analysis for GCNs
- Conclusion

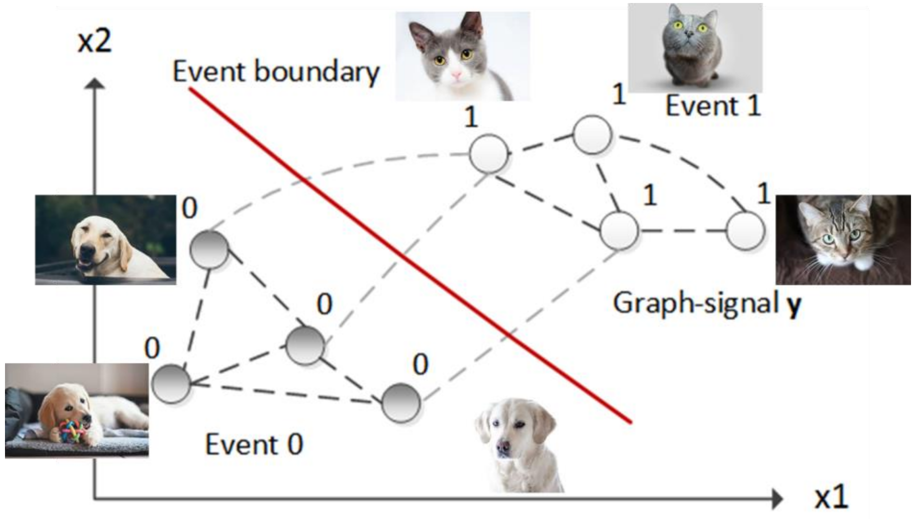
Gene Cheung (genec@yorku.ca)

What is a good graph?

- Graph captures *pairwise relationships*.
  1. Domain knowledge.
  2. Correlations.
  3. Feature distance.



- **Graph Learning from Data**:
  1. Learn sparse **inverse covariance matrix** from observations [1].
     - Graphical Lasso, CLIME.
  2. Learn metric to determine **feature distance** [2].



[1] S. Bagheri, G. Cheung, A. Ortega, F. Wang, "**Learning Sparse Graph Laplacian with *K* Eigenvector Prior via Iterative GLASSO and Projection**," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Toronto, Canada, June 2021 (**best student paper finalist**).

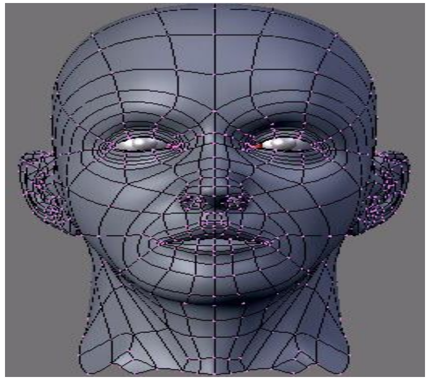
[2] C. Yang, G. Cheung, W. Hu, "**Signed Graph Metric Learning via Gershgorin Disc Alignment**," *IEEE TPAMI*, 2022.

Signed Graphs

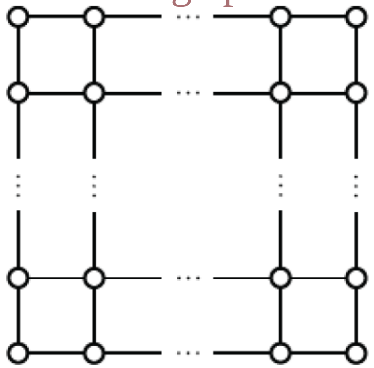
- Previous works designed for **positive graphs**.

Temperature sensor network

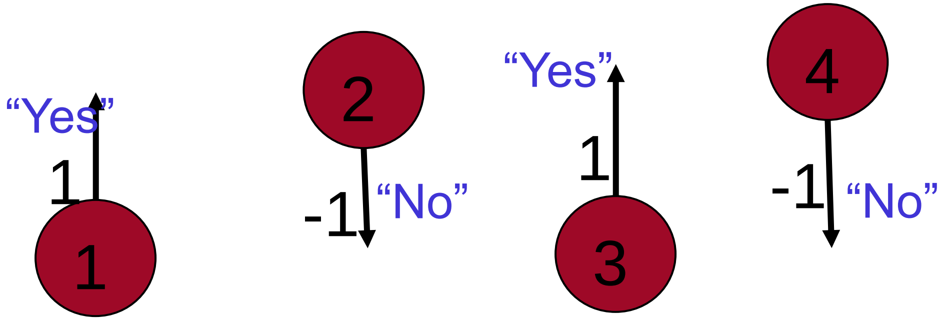
3D geometry

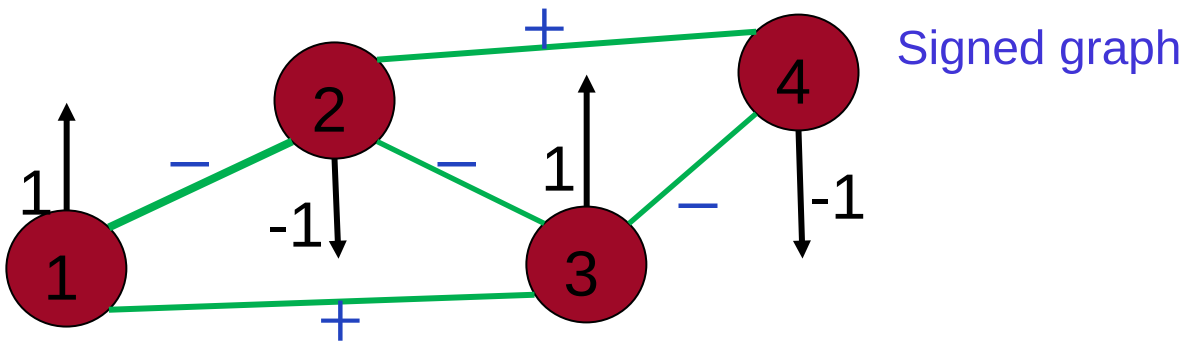


2D image pixels



- Voting records in a Parliament → ***anti-correlation*** represented as **negative edges** [1].



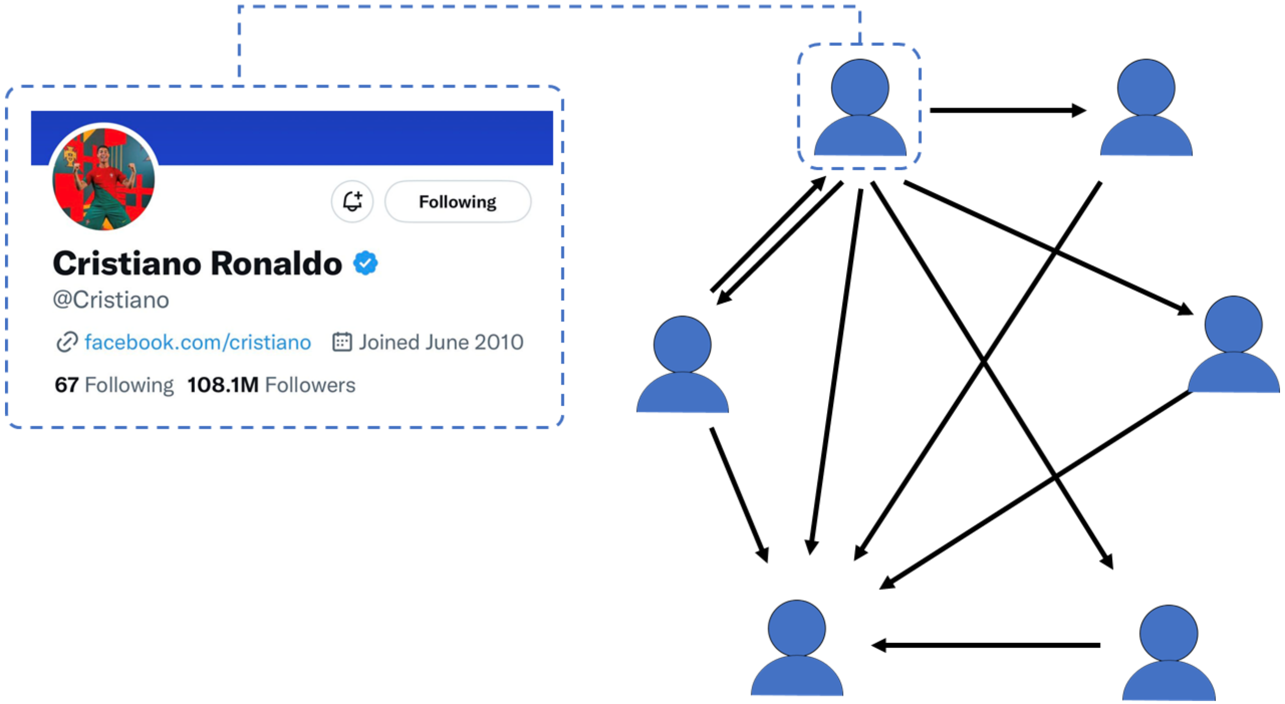


[1] Chinthaka Dinesh, Saghar Bagheri, Gene Cheung, Ivan V. Bajic, "**Linear-time Sampling on Signed Graphs via Gershgorin Disc Perfect Alignment**," *ICASSP'22*, Singapore, May 2022.

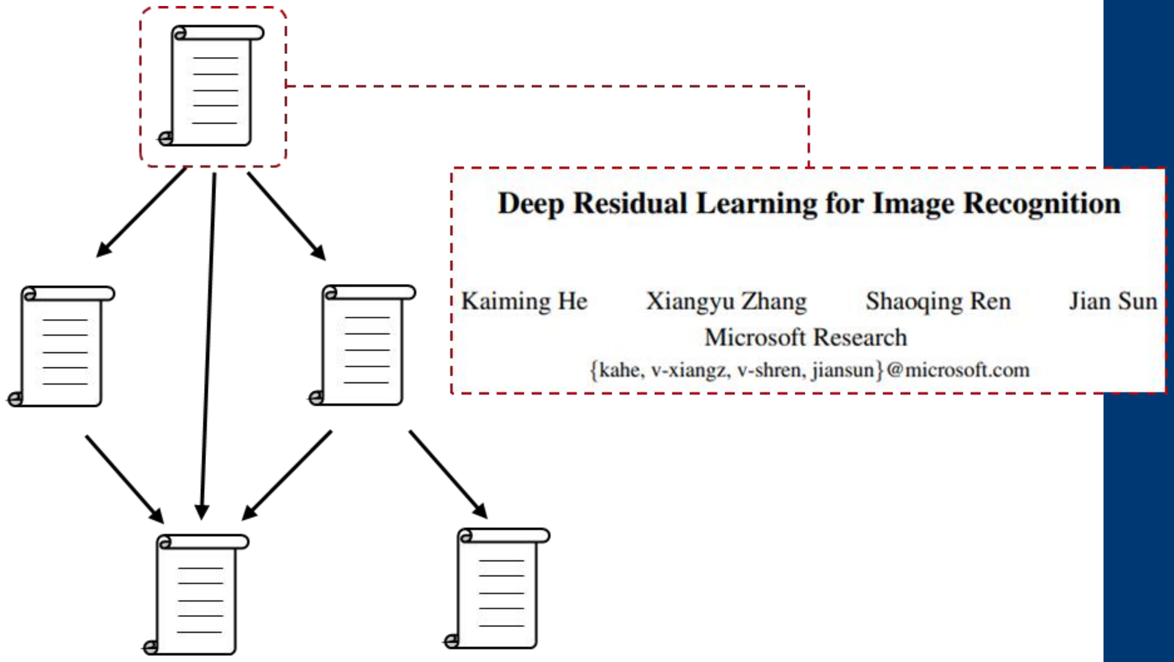
Gene Cheung (genec@yorku.ca)

Directed Graphs

(a) Following Network in Twitter



(b) Paper Citation Network



[1] Yuejiang Li, H. Vicky Zhao, Gene Cheung, "**Eigen-Decomposition-Free Directed Graph Sampling via Gershgorin Disc Alignment**," *ICASSP'23*, Rhodes, Greece, June 2023.

Sparse Precision Matrix Estimation: GLASSO

- Given *empirical covariance matrix* Σ, **Graphical Lasso** computes positive-definite (PD) *precision matrix* Θ:

$$\max_{\Theta} \quad \log\det\Theta - \mathrm{Tr}(\Sigma\Theta) - \rho\,||\Theta||_1$$

- 1^{st} and 2^{nd} terms are ***likelihood***.
- 3^{rd} term promotes **sparsity**.

- Solved via **block coordinate descent** (BCD) algorithm.

[1] Friedman J, Hastie T, Tibshirani R. "**Sparse inverse covariance estimation with the graphical lasso**," *Biostatistics*. 2008; 9(3): 432-441.

Graph Laplacian Estimation

- Assume **precision matrix** is:
  - **Generalized graph Laplacian** (GGLs),
  - **Diagonally dominant generalized graph Laplacian** (DDGLs), or
  - **Combinatorial graph Laplacian** (CGLs).

NOTE: Interpret precision matrix as graph Laplacian

- Given *empirical covariance matrix* S, computes *Laplacian* Θ:

$$\min_{\Theta} \quad \mathrm{Tr}(\Theta \mathbf{K}) - \log\det\Theta \quad \text{subject to} \quad \Theta \in \mathcal{L}_g(A)$$

- **K** = **S** + **H**, **H** is regularization matrix.
- $L_g(A)$ ensures Θ is GGL.
- Solved via **block coordinate descent** (BCD) algorithm.

[1] H. E. Egilmez, E. Pavez and A. Ortega, "**Graph Learning From Data Under Laplacian and Structural Constraints**," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 6, pp. 825-841, Sept. 2017

Graph Laplacian Estimation w/ Eigen-Structure Constraint

- **Key Assumption**: graph Laplacian matrix **L** has *chosen first K eigenvectors*.

1. Side info to derive first *K* e-vectors.
2. Fast computation of first *K* e-vectors.
3. Desire eigen-structure.

- Define **convex cone** $\mathcal{H}_{\mathbf{u}}^{+}$ of PSD matrices with same first *K* eigenvectors.
- Design **projection operator** to $\mathcal{H}_{\mathbf{u}}^{+}$ inspired by **Gram-Schmidt procedure**.
- Given *empirical covariance matrix* $\bar{\mathbf{C}}$, computes *graph Laplacian* $\mathbf{L}$:

$$\min_{\mathbf{L} \in \mathcal{H}_{\mathbf{u}}^{+}} \quad \mathrm{Tr}(\mathbf{L}\bar{\mathbf{C}}) - \log\det\mathbf{L} + \rho\, \|\mathbf{L}\|_1$$

- Solve via alternating BCD and projection algorithm.

[1] S. Bagheri, G. Cheung, A. Ortega, F. Wang, "**Learning Sparse Graph Laplacian with *K* Eigenvector Prior via Iterative GLASSO and Projection**," *IEEE International Conference on Acoustics, Speech and Signal Processing*, Toronto, Canada, June 2021 (**best student paper finalist**).

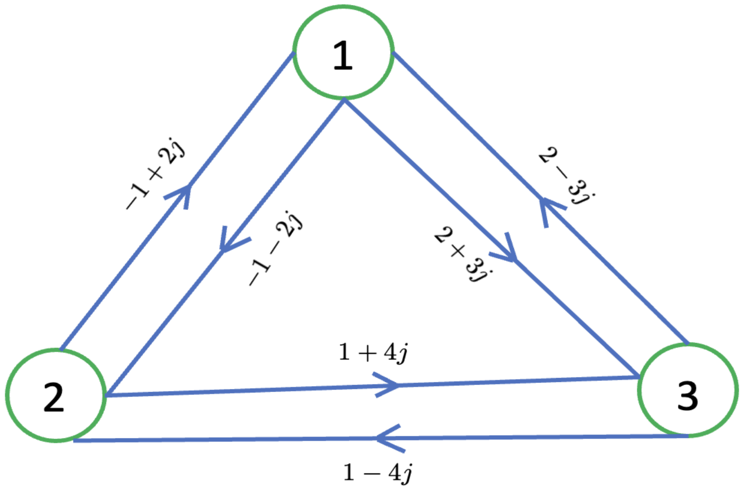
Graph Laplacian Estimation for Complex Graph Signals

- **Complex graph signal**: each node *i* has complex value $x_i \in \mathbb{C}$.
- **Hermitian Graph**: *directed graph* with complex conjugate weights on opposite directed edges between each node-pair.
  - *Hermitian* adjacency, graph Laplacian matrices, w/ REAL eigenvalues (frequencies).

- **Complex graph Laplacian Learning** (generalize CLIME):
  - Linear program to solve:

$$\min_{\mathbf{P}} \quad \|\mathbf{P}\|_1 , \quad \text{s.t. } \|\mathbf{C}\mathbf{P} - \mathbf{I}_N\|_\infty \leq \rho$$

  - Define auxiliary var. to account for real-/imaginary-parts.

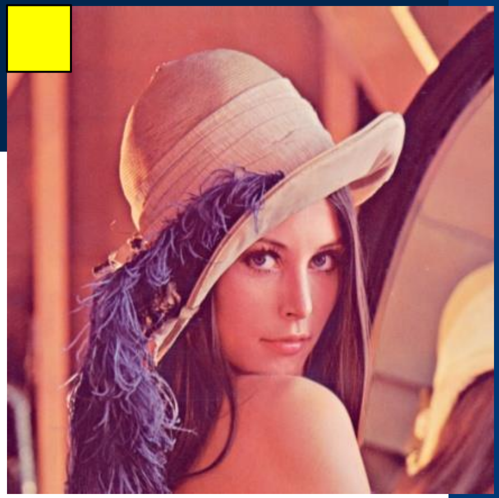


$$\mathcal{L} = \begin{bmatrix} 6 & 1+2j & -2-3j \\ 1-2j & 8 & -1-4j \\ -2+3j & -1+4j & 5 \end{bmatrix}$$

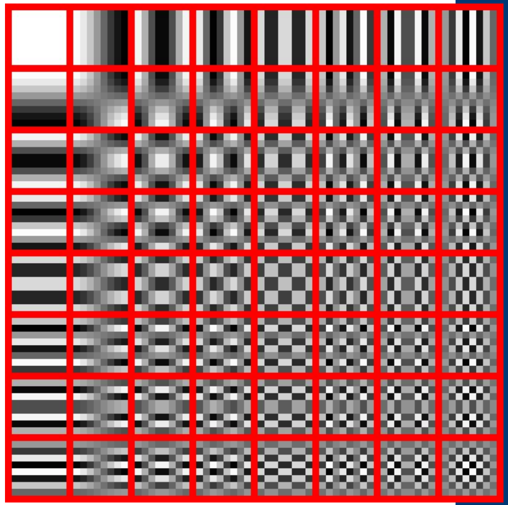
[1] Chinthaka Dinesh, Junfei Wang, Gene Cheung, Pirathayini Srikantha, **"Complex Graph Laplacian Regularizer for Inferencing Grid States**," submitted to *IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids* (SmartGridComm), Glasgow, Scotland, October 2023.

[2] T. Cai, W. Liu, and X. Luo, "**A constrained L1 minimization approach to sparse precision matrix estimation**," *Journal of the American Statistical Association*, vol. 106, no. 494, pp. 594–607, 2011.

Application: Image Coding



- **Transform Coding** is integral component in image compression.
- **Problem**: **DCT** is *fixed* transform and does not adapt locally.
- **Existing Work 1**: **Asymmetric Discrete Sine Transform** (ADST) fits better prediction residuals [1].
- **Existing Work 2**: **Karhunen-Loeve transform** (KLT) adapts well iff ∃ reliable empirical covariance matrix $\bar{\mathbf{C}}$ [2].



[1] J. Han, A. Saxena, V. Melkote, and K. Rose, "**Jointly optimized spatial prediction and block transform for video and image coding**," in *IEEE Transactions on Image Processing*, April 2012, vol. 21, no.4, pp. 1874–1884.

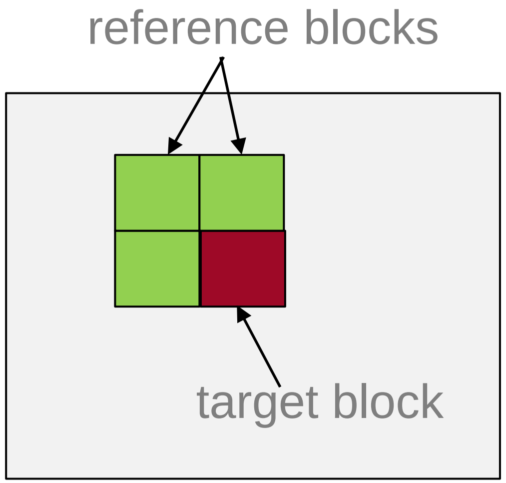
[2] Ian Blanes and Joan Serra-Sagrista, "**Pairwise orthogonal transform for spectral image coding**," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no.3, pp. 961–972, 2011.

Application: Image Coding

- **Key Idea**: derive first K e-vectors from model, compute N-K e-vectors from data.

- **Advantages**:
  1. Reduce degree of freedom when empirical covariance $\bar{\mathbf{C}}$ is unreliable.
  2. Parameter K is tunable depending on covariance reliability.
  3. Reduce computation cost for first K transform coefficients.

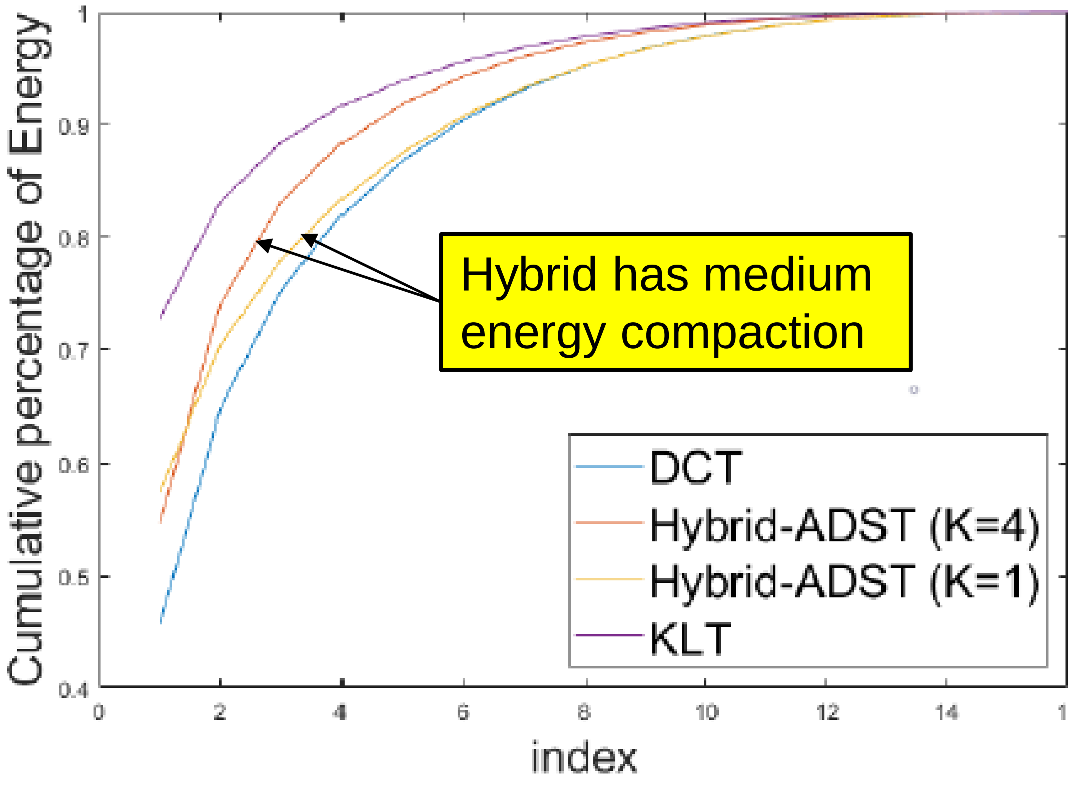
- **Disadvantage**:
  1. Larger computation cost than DCT.



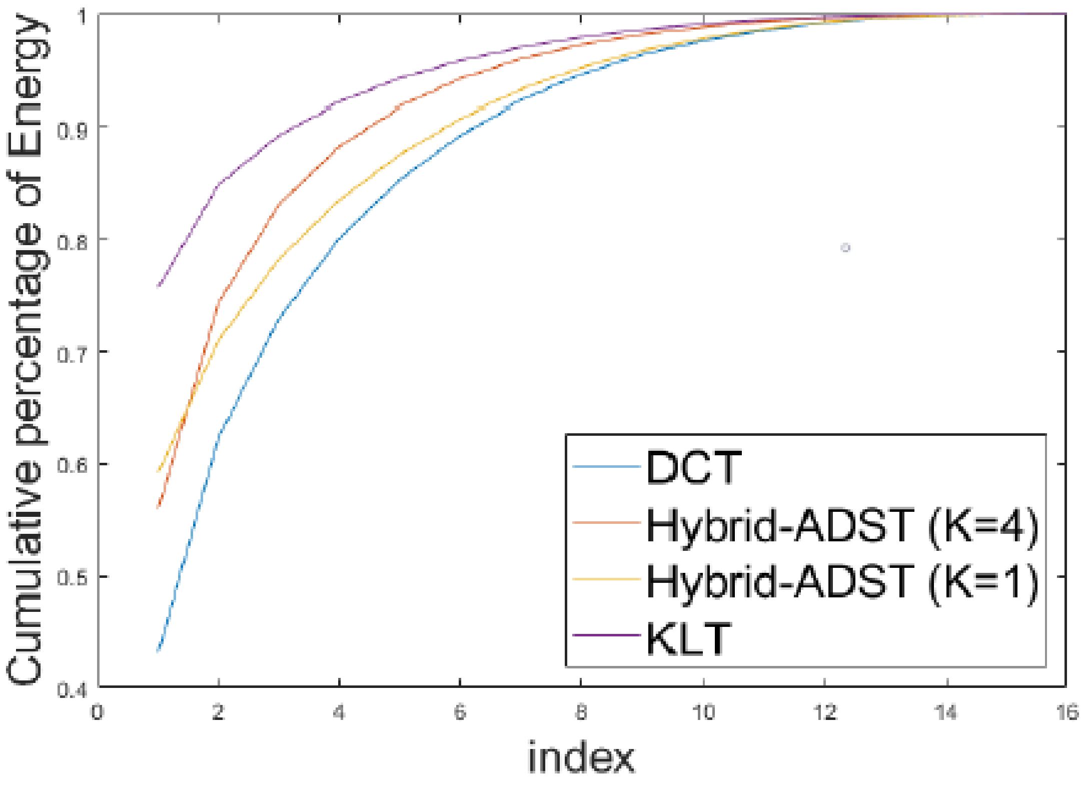
[1] Saghar Bagheri, Tam Thuc Do, Gene Cheung, Antonio Ortega, "**Hybrid Model-based / Data-driven Graph Transform for Image Coding**," *IEEE Conference on Image Processing*, 2022.

Image Coding: results (energy compaction)

- **Setting**: **WebP** image codec [1]. DC4 intra-prediction mode. Improve prediction residual coding of 4x4 block over default DCT.



(a) Airplane



(b) Pepper

[1] https://developers.google.com/speed/webp

Outline

- GSP overview
  - Graph frequencies from eigen-pairs
- Graph Learning
  - Positive, signed, directed, Hermitian graphs
- Graph Filtering
- Graph Sampling
- GSP Analysis for GCNs
- Conclusion

Spectral Graph Filter for Image Denoising

- **Graph Laplacian Regularizer (GLR)** $\mathbf{x}^T\mathbf{L}\mathbf{x}$ is a smoothness measure.

- Denoising has simplest formation model $\mathbf{y} = \mathbf{x} + \mathbf{z}$, thus formulation

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{x}\|_2^2 + \mu \ \mathbf{x}^T\mathbf{L}\mathbf{x}$$

$$\mathbf{x}^T\mathbf{L}\mathbf{x} = \sum_{(i,j)\in E} w_{i,j}\left(x_i - x_j\right)^2 = \sum_k \lambda_k \tilde{x}_k^{\,2}$$

smooth signal low-pass signal

$$(\mathbf{I} + \mu\mathbf{L})\mathbf{x}^* = \mathbf{y}$$

$$\mathbf{x}^* = (\mathbf{I} + \mu\mathbf{L})^{-1}\mathbf{y}$$

$$\mathbf{x}^* = \mathbf{V}\mathrm{diag}(1 + \mu\lambda_1, 1 + \mu\lambda_2, \dots)^{-1}\mathbf{V}^T\mathbf{y}$$

low-pass filter!

[1] J. Pang, G. Cheung, "**Graph Laplacian Regularization for Image Denoising: Analysis in the Continuous Domain**," *IEEE TIP*, vol. 26, no.4, pp.1770-1785, April 2017.
[2] C. Tomasi and R. Manduchi, "**Bilateral filtering for gray and color images**," *IEEE ICCV*, 1998.

Spectral Graph Filter for Image Denoising

- **Graph Laplacian Regularizer (GLR)** $\mathbf{x}^T\mathbf{L}\mathbf{x}$ is a smoothness measure.

- Denoising has simplest formation model $\mathbf{y} = \mathbf{x} + \mathbf{z}$, thus formulation

$$\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{x}||_2^2 + \mu \; \mathbf{x}^T\mathbf{L}\mathbf{x}$$

$$(\mathbf{I} + \mu\mathbf{L})\mathbf{x}^* = \mathbf{y}$$

pixel intensity diff. pixel location diff.

$$w_{i,j} = \exp\left(\frac{-\left\|x_i - x_j\right\|_2^2}{\sigma_1^2} \right) \exp\left(\frac{-\left\|l_i - l_j\right\|_2^2}{\sigma_2^2} \right)$$

Bilateral filter weights

- To promote **Piecewise Smoothness** (PWS), $\mathbf{L}(\mathbf{x})$ is ***signal-dependent***:
  - Fix **L** and solve unconstrained QP each iteration.

$$\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{x}||_2^2 + \mu \; \mathbf{x}^T\mathbf{L}(\mathbf{x})\mathbf{x}$$

← Signal-dependent GLR

[1] J. Pang, G. Cheung, "**Graph Laplacian Regularization for Image Denoising: Analysis in the Continuous Domain**," *IEEE TIP*, April 2017.

[2] C. Tomasi and R. Manduchi, "**Bilateral filtering for gray and color images**," *IEEE ICCV*, 1998.

OGLR Denoising Results: visual comparison

- Subjective comparisons ($\sigma_I = 40$)

Original

Noisy, 16.48 dB

K-SVD, 26.84 dB

BM3D, 27.99 dB

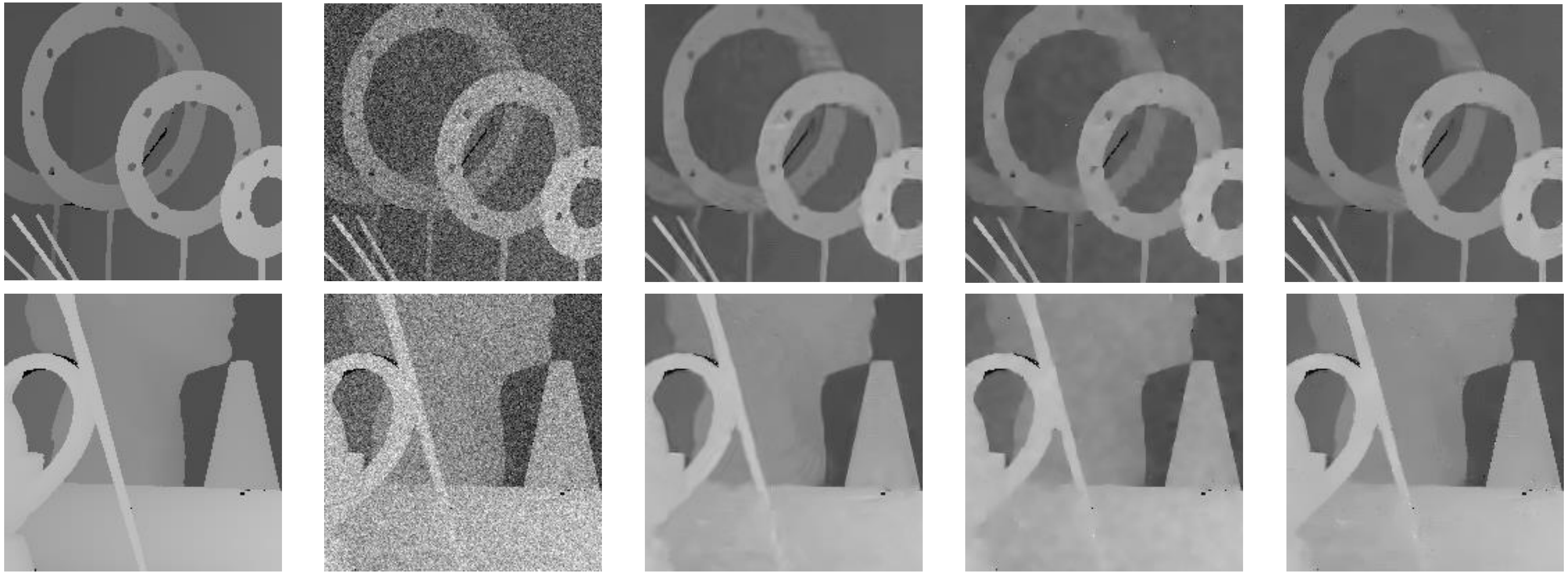
PLOW, 28.11 dB

OGLR, 28.35 dB

[1] J. Pang, G. Cheung, "**Graph Laplacian Regularization for Image Denoising: Analysis in the Continuous Domain**," *IEEE TIP*, vol. 26, no.4, pp.1770-1785, April 2017.

OGLR Denoising Results: visual comparison

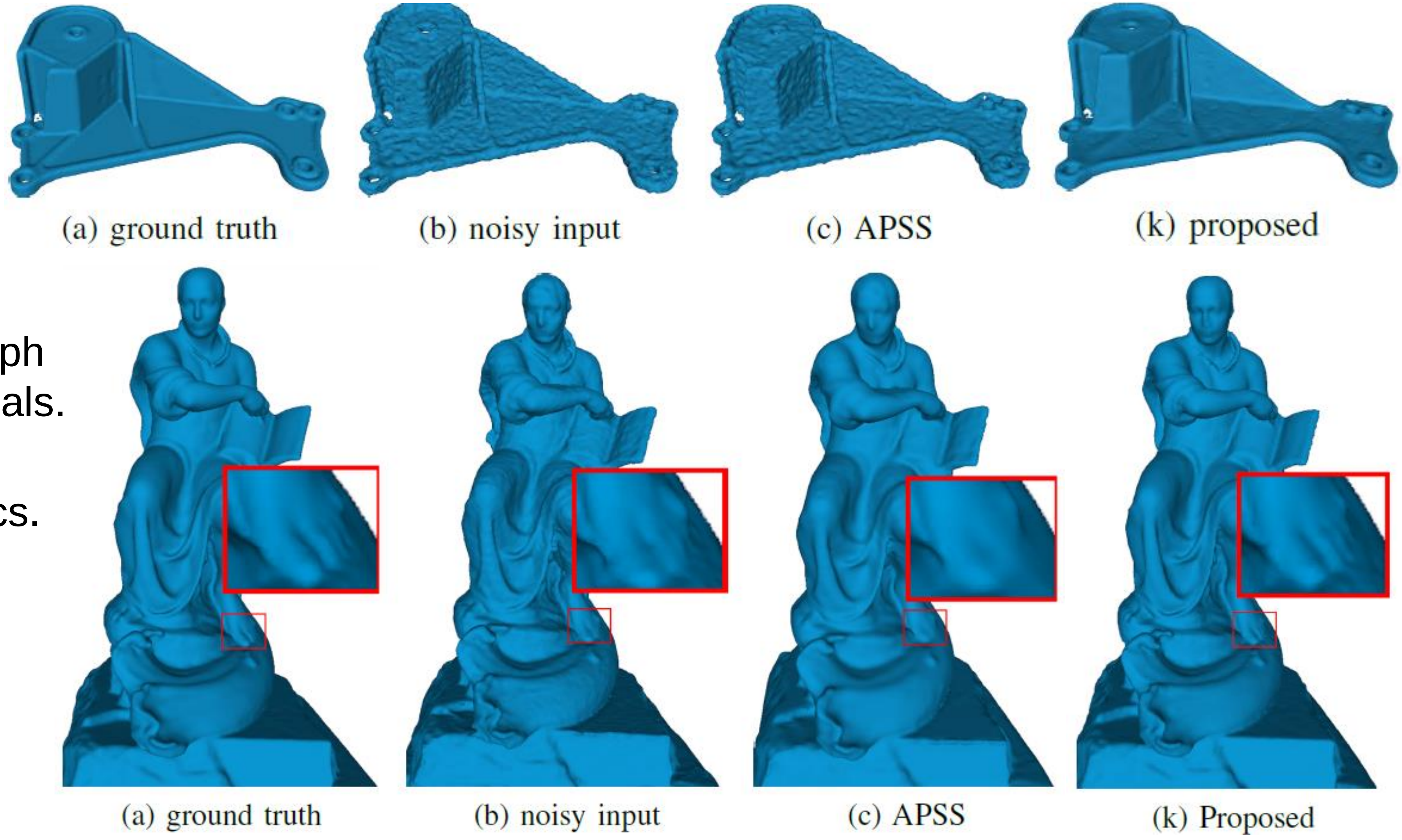
- Subjective comparisons ($\sigma_{\mathrm{I}} = 30$)



Original Noisy, 18.66 dB BM3D, 33.26 dB NLGBT, 33.41dB OGLR, 34.32 dB

[1] J. Pang, G. Cheung, "**Graph Laplacian Regularization for Image Denoising: Analysis in the Continuous Domain**," *IEEE TIP*, vol. 26, no.4, pp.1770-1785, April 2017.

Point Cloud Denoising: Graph filter design



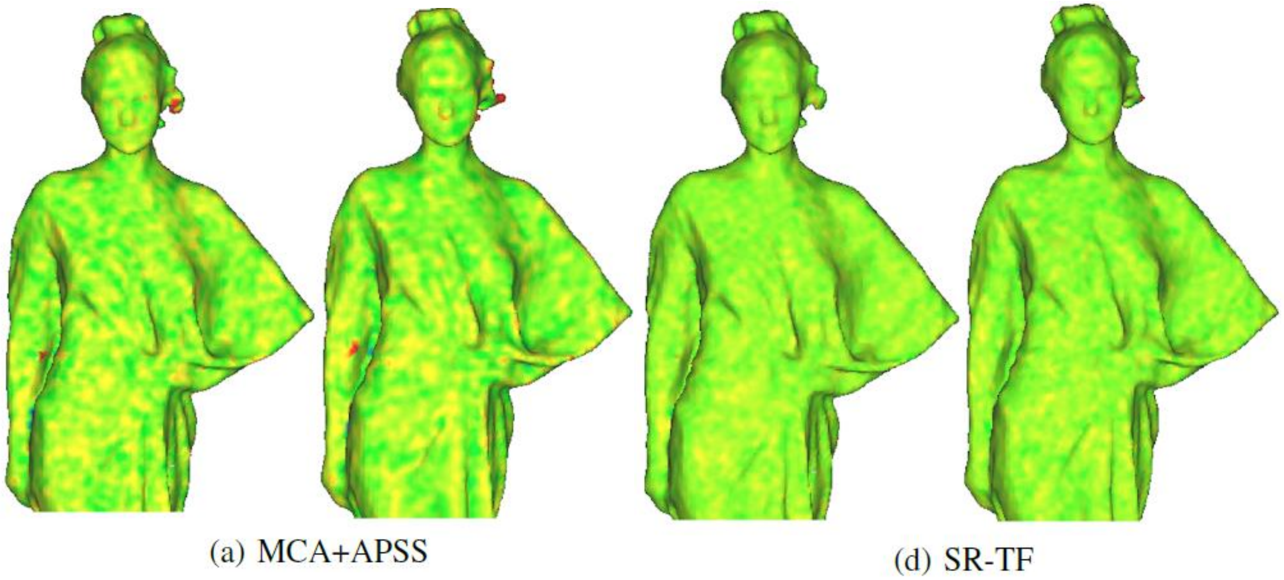
Method:

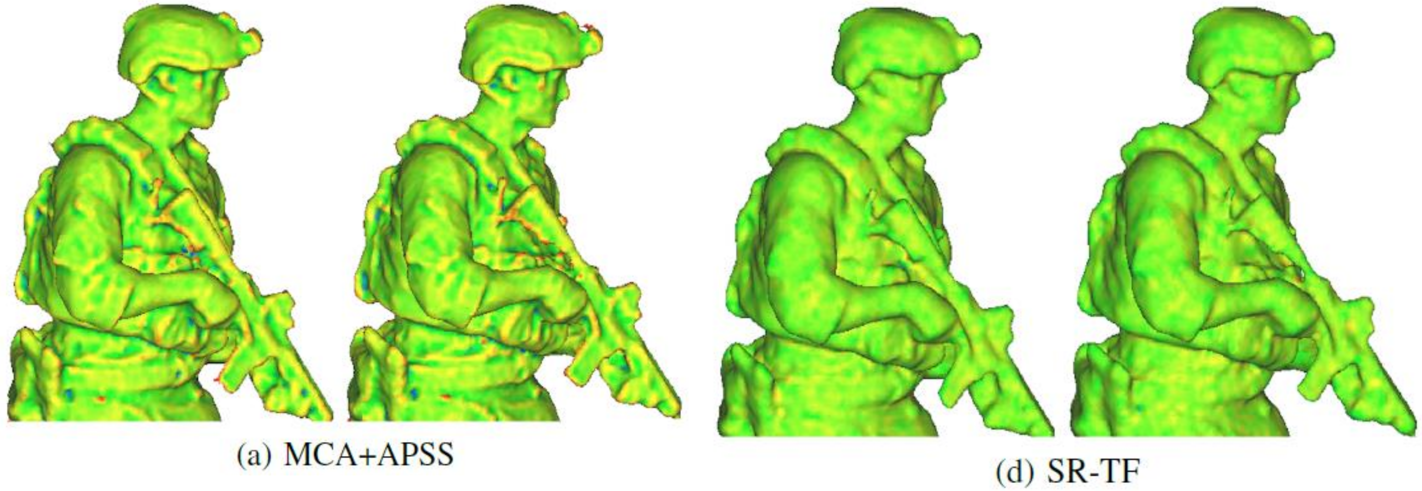
- Construct similarity graph based on surface normals.
- Optimize graph filter based on noise statistics.

[1] C. Dinesh, G. Cheung, I. Bajic, "**Point Cloud Denoising via Feature Graph Laplacian Regularization**," vol. 29, pp. 4143-4158, *IEEE TIP*, January 2020..

Point Cloud Video Super-Resolution

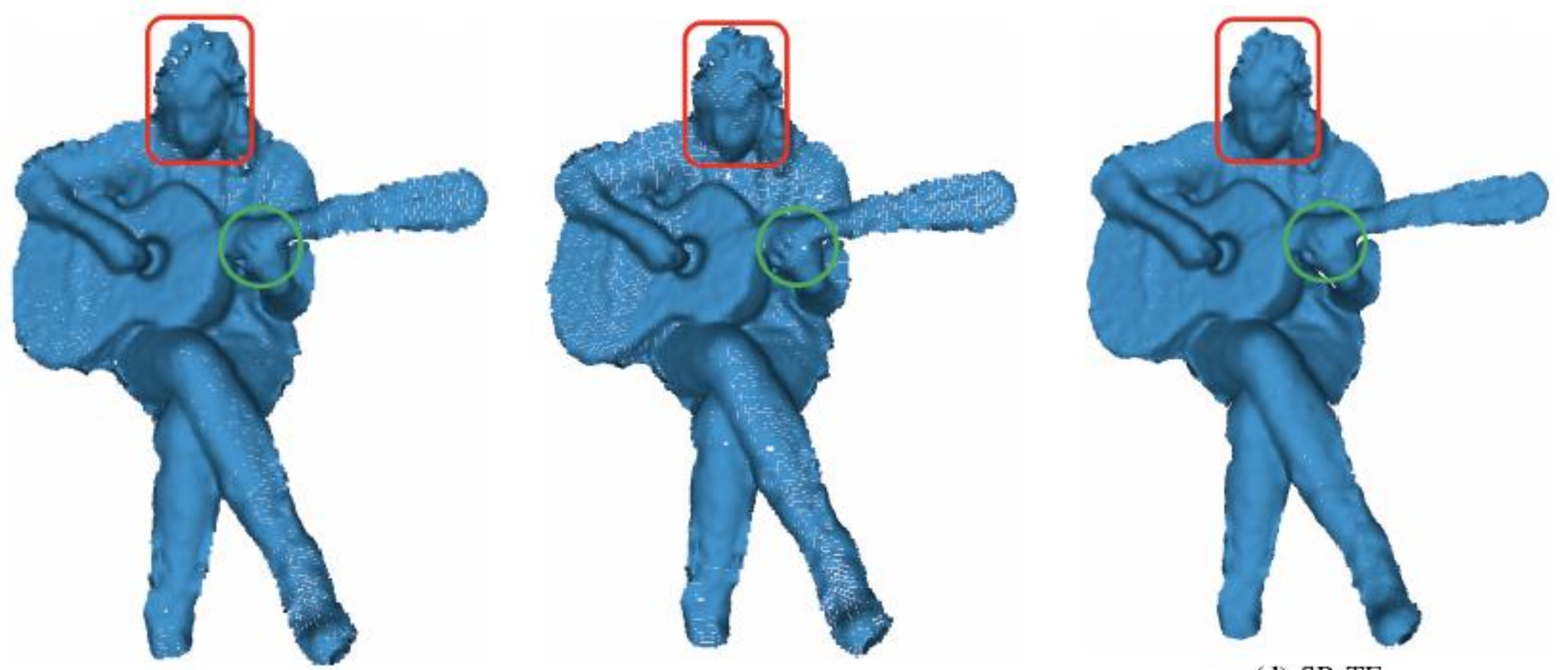
- **Problem**: Acquired point cloud video has low point density.
- **Solution**:
  - Find similar 3D patches in consecutive frames.
  - Super-resolve patches in consecutive frames via GLR.





[1] Chinthaka Dinesh, Gene Cheung, Ivan V. Bajic, "**Point Cloud Video Super-Resolution via Partial Point Coupling and Graph Smoothness**," *IEEE TIP*, June 2022.

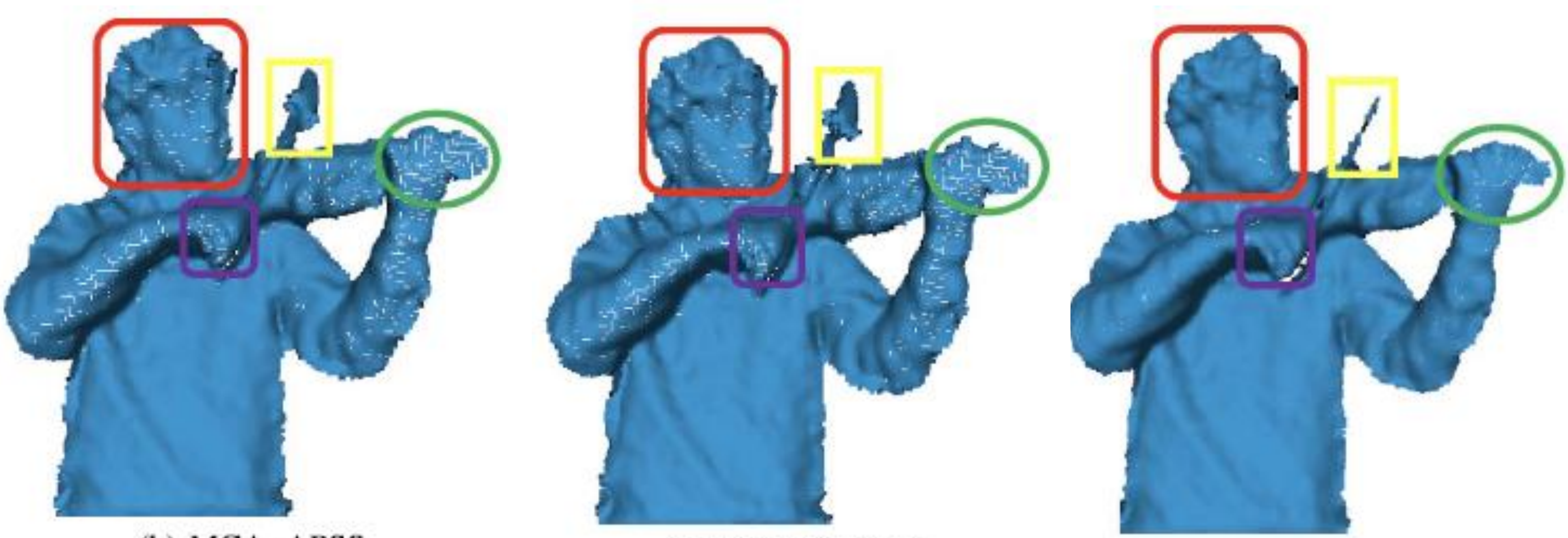
Point Cloud Video Super-Resolution



(b) MCA+APSS (c) MCA+RIMLS (d) SR-TF

[1] Chinthaka Dinesh, Gene Cheung, Ivan V. Bajic, "**Point Cloud Video Super-Resolution via Partial Point Coupling and Graph Smoothness**," *IEEE Transactions on Image Processing*, vol. 31, pp.4117-4132, June 2022.

Point Cloud Video Super-Resolution



(b) MCA+APSS   (c) MCA+RIMLS   (d) SR-LF

[1] Chinthaka Dinesh, Gene Cheung, Ivan V. Bajic, "**Point Cloud Video Super-Resolution via Partial Point Coupling and Graph Smoothness**," *IEEE Transactions on Image Processing*, vol. 31, pp.4117-4132, June 2022.

Outline

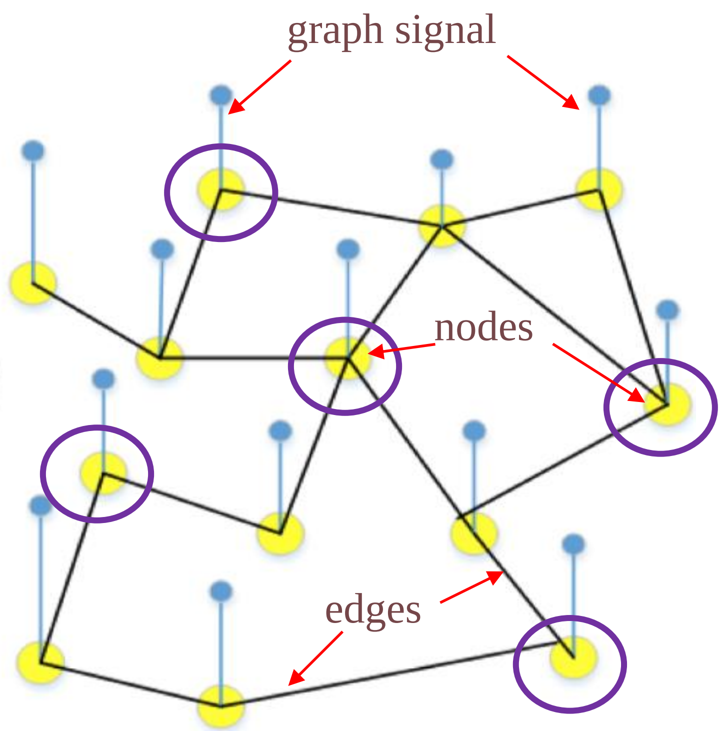
- GSP overview
  - Graph frequencies from eigen-pairs
- Graph Learning
  - Positive, signed, directed, Hermitian graphs
- Graph Filtering
- Graph Sampling
- GSP Analysis for GCNs
- Conclusion

Gene Cheung (genec@yorku.ca)

Motivation: graph sampling

Graph sampling [1]: Choose a node subset, so that the entire signal can be reconstructed.

- Graph sampling strategies extend **Nyquist sampling** to graph data kernel.
- Bandlimited or smooth signal assumption.



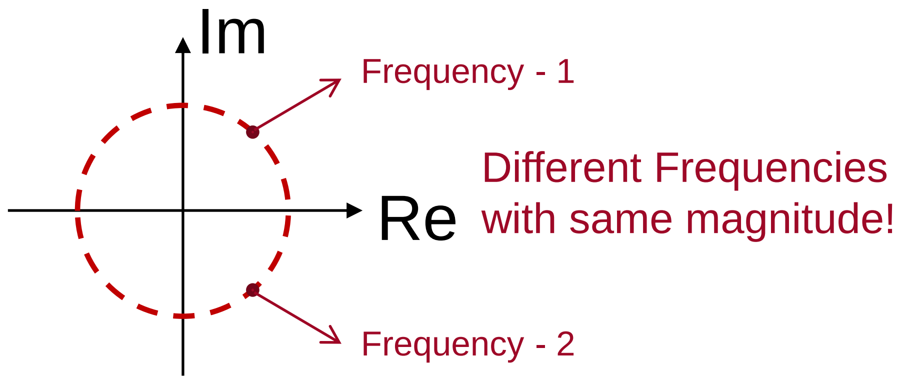
[1] Y. Tanaka et al., "**Sampling signals on graphs: From theory to applications**," *IEEE Signal Process. Mag.*, vol. 37, no. 6, pp. 14–30, 2020.

Existing Work: graph frequencies

- Mainly focus on undirected graph.
  - Symmetric Graph Laplacian $\mathbf{L} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$
- Categories
  - Bandlimited prior[1-3]: $\mathbf{x} = \sum_{i=1}^{M} \widetilde{x}_i \mathbf{u}_i$
  - Smoothness prior[4-5]: smaller $\mathbf{x}^\top \mathbf{L}\mathbf{x}$

Not obvious for directed graphs!

- Directed graph → **Asymmetric $\mathbf{L}$**
- Complex graph frequency



- Meaningless quadratic form $\mathbf{x}^\top \mathbf{L}\mathbf{x}$

[1] S. Chen, R. Varma, A. Sandryhaila, and J. Kovacevic, "Discrete signal processing on graphs: Sampling theory," TSP, vol. 63, no. 24, pp. 6510– 6523, 2015.
[2] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," TSP, vol. 64, no. 14, pp. 3775–3789, 2016.
[3] A. Sakiyama, Y. Tanaka, T. Tanaka, and A. Ortega, "Eigendecomposition-free sampling set selection for graph signals," TSP, vol. 67, no. 10, pp. 2679–2692, 2019.
[4] Y. Tanaka and Y. C. Eldar, "Generalized sampling on graphs with subspace and smoothness priors," TSP, vol. 68, pp. 2272–2286, 2020.
[5] Y. Bai, F. Wang, G. Cheung, Y. Nakatsukasa, and W. Gao, "Fast graph sampling set selection using Gershgorin disc alignment," TSP, vol. 68, pp. 2419–2434, 2020.

Existing Work: computation

Existing graph sampling methods

Eigen-decomposition-based methods [1,2]

Computational expensive

eigen-decomposition-free methods

Spectral proxies (SP) [3]

Neumann series (NS) [4]

Localization operator (LO) [5]

Gershgorin disc alignment (GDA) [6]

Fast method for graph sampling!

[1] M. Tsitsvero, S. Barbarossa, and P. Di Lorenzo, “Signals on graphs: Uncertainty principle and sampling,” *IEEE TSP*, vol. 64, no. 18, pp. 4845–4860, 2016.
[2] S. Chen, R. Varma, A. Sandryhaila, and J. Kovavcevic, “Discrete signal processing on graphs: Sampling theory,” *IEEE TSP*, vol. 63, no. 24, pp. 6510–6523, 2015.
[3] A. Anis et al., “Efficient sampling set selection for bandlimited graph signals using graph spectral proxies,” *IEEE TSP*, vol. 64, no. 14, pp.3775–3789, 2016.
[4] F. Wang et al., “Low-complexity graph sampling with noise and signal reconstruction via Neumann series,” *IEEE TSP*, vol. 67, no. 21, pp. 5511–5526, 2019.
[5] A. Sakiyama, Y. Tanaka, T. Tanaka, and A. Ortega, “Eigendecomposition-free sampling set selection for graph signals,” *IEEE TSP*, vol. 67, no. 10, pp. 2679–2692, 2019.
[6] Y. Bai, F. Wang, G. Cheung, Y. Nakatsukasa, and W. Gao, Fast graph sampling set selection using Gershgorin disc alignment,” *IEEE TSP*, vol. 68, pp. 2419–2434, 2020.

Signal Reconstruction from Samples

- **Signal Model**:

observation; sampling matrix; desired signal; noise

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}$$

- Signal prior is **graph Laplacian regularizer** (GLR):

$$\mathbf{x}^T\mathbf{L}\mathbf{x} = \sum_{i,j} w_{i,j}(x_i - x_j)^2 = \sum_k \lambda_k \tilde{x}_k^2$$

signal smooth w.r.t. graph

signal contains mostly low graph freq.

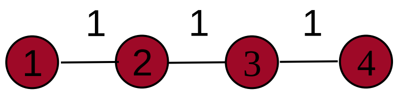
- **MAP Formulation**:

fidelity term; signal prior

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \mu\ \mathbf{x}^T\mathbf{L}\mathbf{x}$$

$$(\mathbf{H}^T\mathbf{H} + \mu\mathbf{L})\mathbf{x}^* = \mathbf{y}$$

linear system of eqn's solved using *conjugate gradient*



Sample set {2, 4}

$$\mathbf{H} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{H}^{\mathrm{T}}\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

[1] Y. Bai, F. Wang, G. Cheung, Y. Nakatsukasa, W. Gao, "**Fast Graph Sampling Set Selection Using Gershgorin Disc Alignment**," vol. 68, pp. 2419-2434, *IEEE Transactions on Signal Processing*, March 2020.

Stability of Linear System

- Examine solution's linear system:

$$(\mathbf{H}^T\mathbf{H} + \mu\mathbf{L})\mathbf{x}^* = \mathbf{y}$$

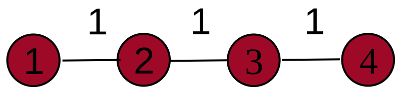
coefficient matrix **B**

- Stability depends on **condition number** (λ_{max}/ λ_{min}) of **B**.
- λ_{max} is upper-bounded by 1+μ2*d_{max}.

Goal: select **H** to maximize $\lambda_{\min}(\mathbf{B})$ (E-optimality criterion)

Also minimizes **worst-case MSE**:

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \mu \left\| \frac{1}{\lambda_{min}(\mathbf{B})} \right\|_2 \|\mathbf{L}(\mathbf{x} + \tilde{\mathbf{n}})\|_2 + \|\tilde{\mathbf{n}}\|_2$$



$$\mathbf{L} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

$$\mathbf{H}^{\mathrm{T}}\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Sample set {2, 4}

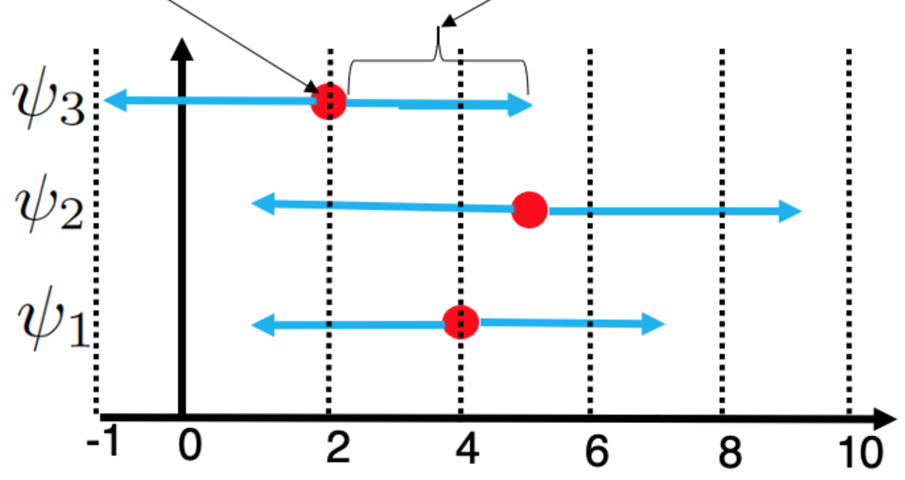
[1] Y. Bai, F. Wang, G. Cheung, Y. Nakatsukasa, W. Gao, "**Fast Graph Sampling Set Selection Using Gershgorin Disc Alignment**," vol. 68, pp. 2419-2434, *IEEE Transactions on Signal Processing*, March 2020.

Gershgorin Circle Theorem

- GCT relates matrix entries to bounds of eigenvalues.

Row i of matrix $\mathbf{F}$ maps to a Gershgorin disc with centre $c_i = \mathbf{F}(i,i)$ and radious $r_i = \sum_{j \neq i} |\mathbf{F}(i,j)|$

$$\mathbf{F} = \begin{bmatrix} 2 & -2 & -1 \\ -2 & 5 & -2 \\ -1 & -2 & 4 \end{bmatrix}$$



$\lambda_{\min}(\mathbf{F})$ is lower-bounded by the *smallest left-end of Gershgorin discs*:

$$\lambda^{-}_{\min}(\mathbf{F}) \triangleq \min_i c_i - r_i \leq \lambda_{\min}(\mathbf{F})$$

[1] R. S. Varga, *Gershgorin and his circles*, Springer, 2004.

GDA Sampling for Positive Graphs

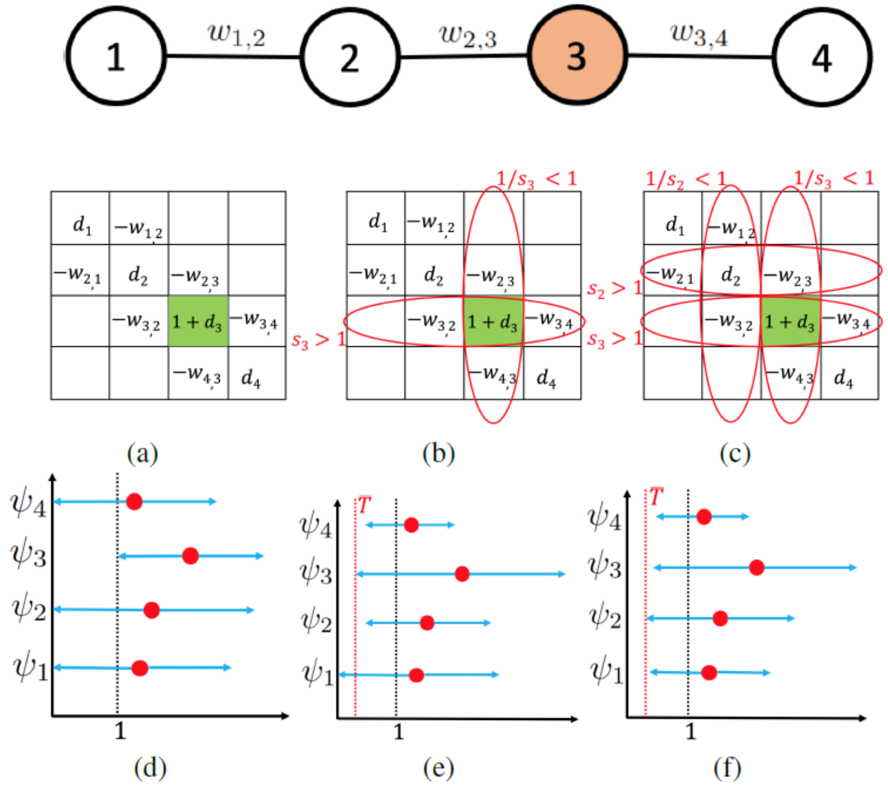
- We focus on maximizing $\lambda^-_{\min}(\mathbf{B})$:

$\mathbf{H}^\top\mathbf{H} + \mu\mathcal{L}$

$$\max_{\mathbf{H},\mathbf{S}\,|\,\mathrm{Tr}(\mathbf{H}^\top\mathbf{H})\leq K} \lambda^-_{\min}\left(\mathbf{S}\mathbf{B}\mathbf{S}^{-1}\right)$$

diagonal *scaling* matrix

- Given Gershgorin disc left-ends of $\mathcal{L}$ is at the same exact value, GDA graph sampling [1]: Select samples to max smallest disc left-end $\lambda^-_{\min}(\mathbf{B})$ of coefficient matrix **B** via:
  - *Disc shifting* (choosing sample *i*).
  - *Disc scaling* (estimating influence on neighbors given sample *i*).



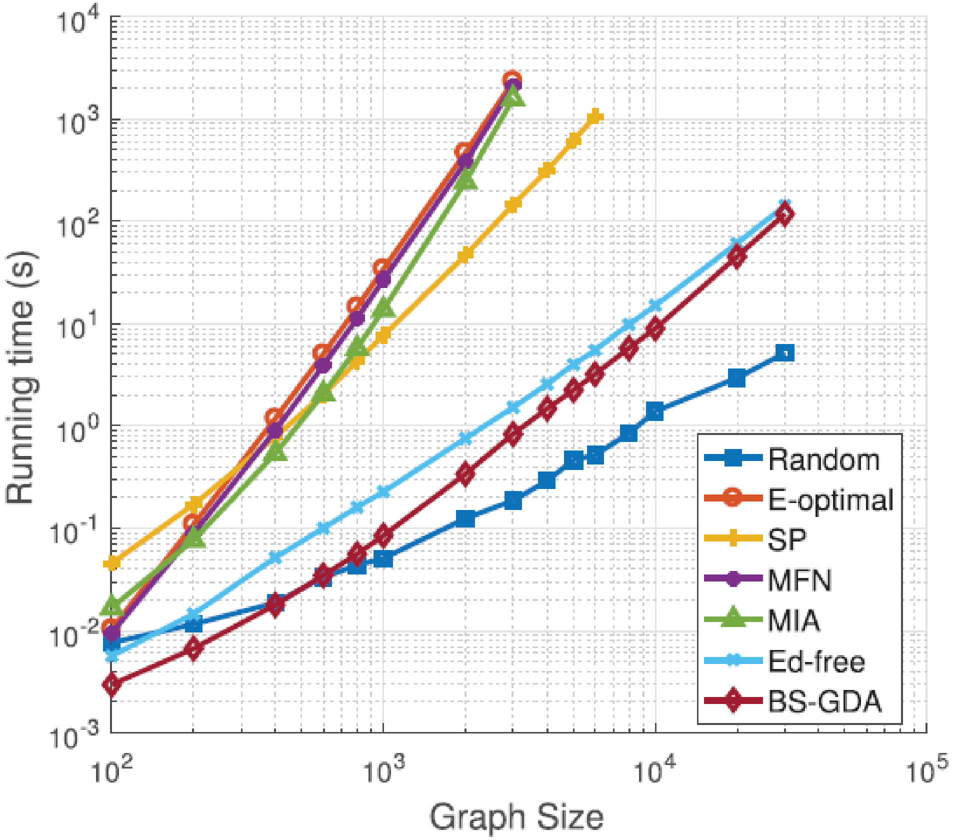
[1] Y. Bai, F.Wang, G. Cheung, Y. Nakatsukasa, and W. Gao, "**Fast graph sampling set selection using gershgorin disc alignment**," *IEEE TSP*, 2020.

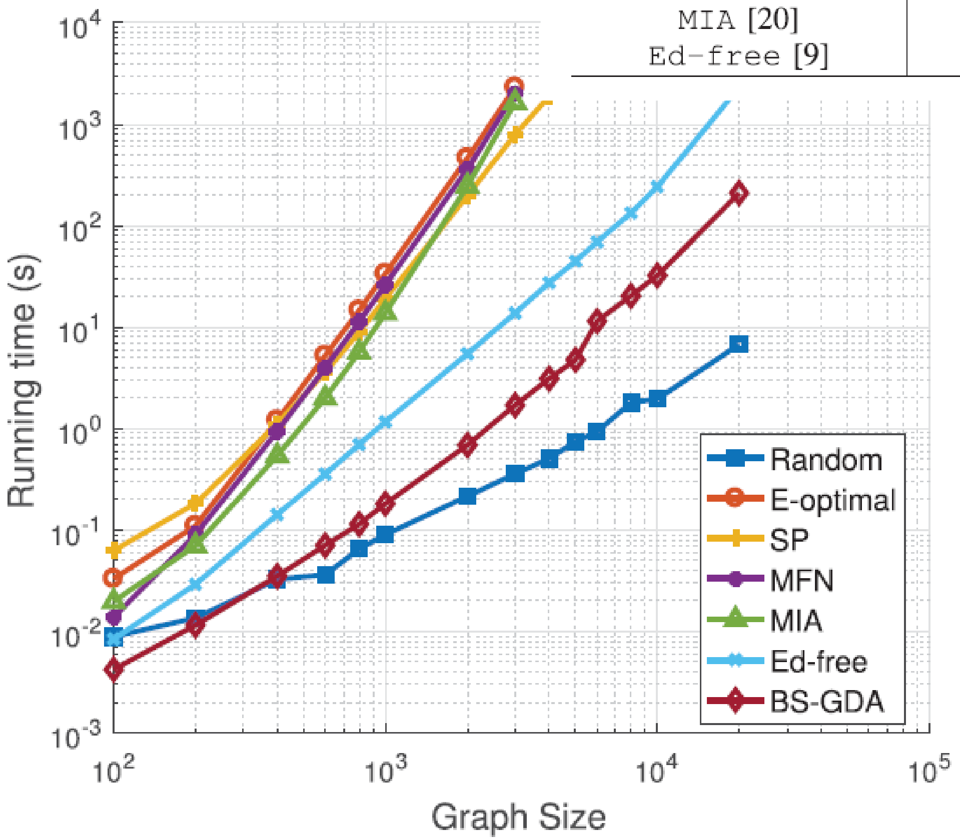
Results: positive graph sampling (speed)

- Running time comparisons on two different graphs. (a) Random sensor raph. (b) Community graph.

TABLE II
SPEEDUP FACTORS OF OUR ALGORITHM WITH RESPECT TO OTHER SAMPLING ALGORITHMS FOR $N = 3000$

Sampling Algorithms	Sensor	Community
Random [27]	0.22	0.21
E-optimal [24]	2812.77	1360.76
SP [16]	174.09	466.18
MFN [22]	2532.91	1184.23
MIA [20]	1896.19	964.65
Ed-free [9]	1.82	8.11

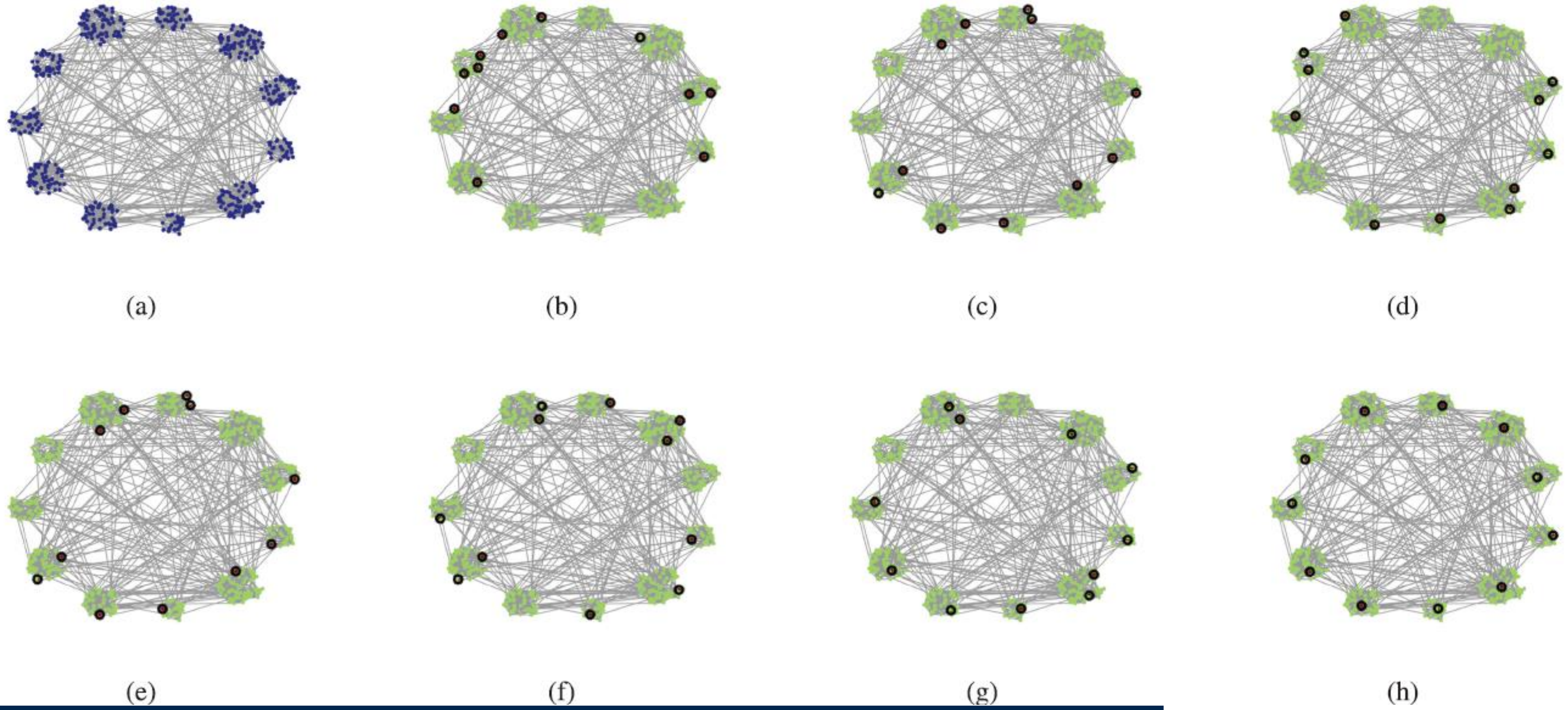




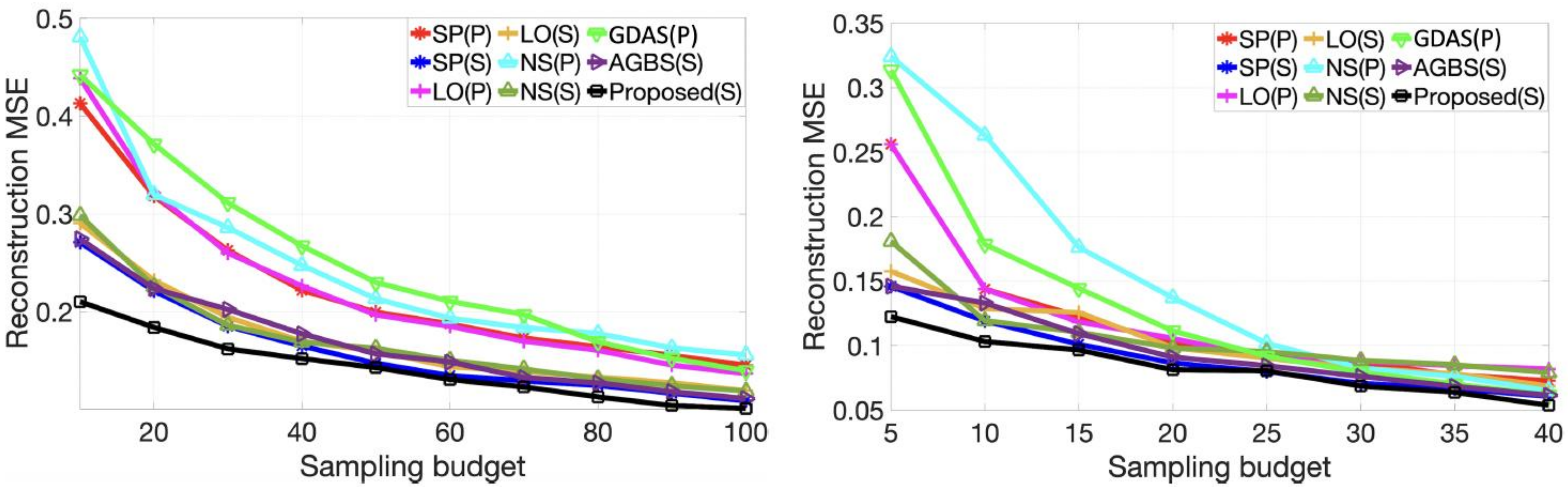
Gene Cheung (genec@yorku.ca)

Results: community graph sampling

- Visualization of selected nodes on the community graph ($N = 500, K = 11$). Black circles denote sampled nodes. (a) Original graph. (b) Random [28].(c) E-optimal [25]. (d) SP [16]. (e) MFN [23]. (f) MIA [20]. (g) Ed-free [9]. (h) The proposed BS-GDA.



Results: signed graph sampling



(a) Canadian parliament voting dataset

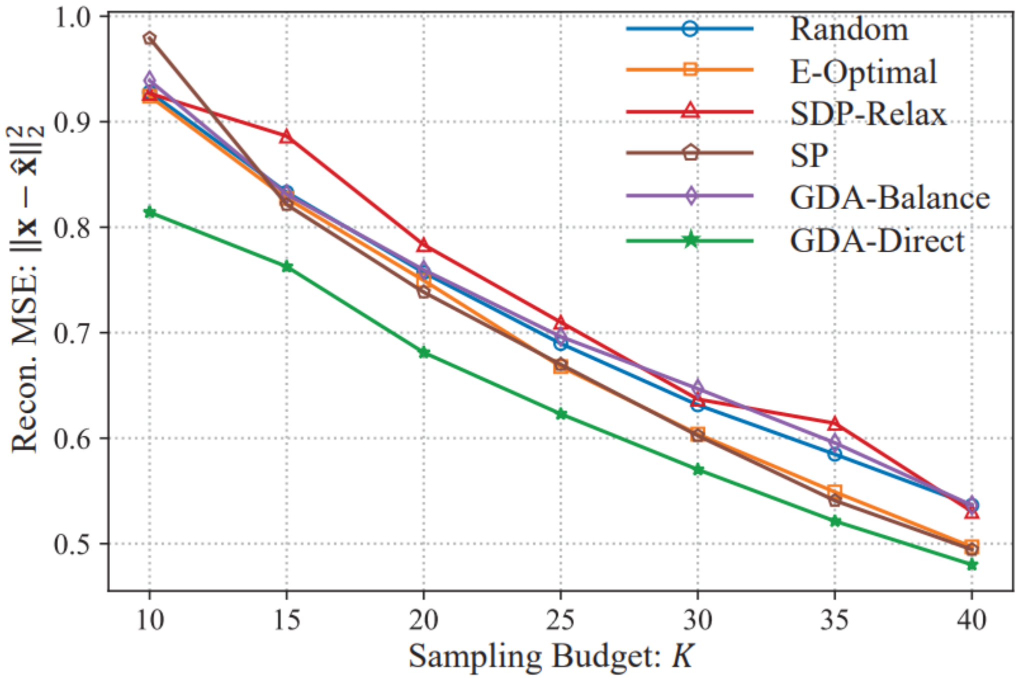
(b) US senate voting dataset

For the Canadian dataset, our scheme reduced the lowest MSE among competitor schemes by 22.2%, 18.2%, 13.5%, 10.4%, for sampling budget 10, 20, 30, 40.

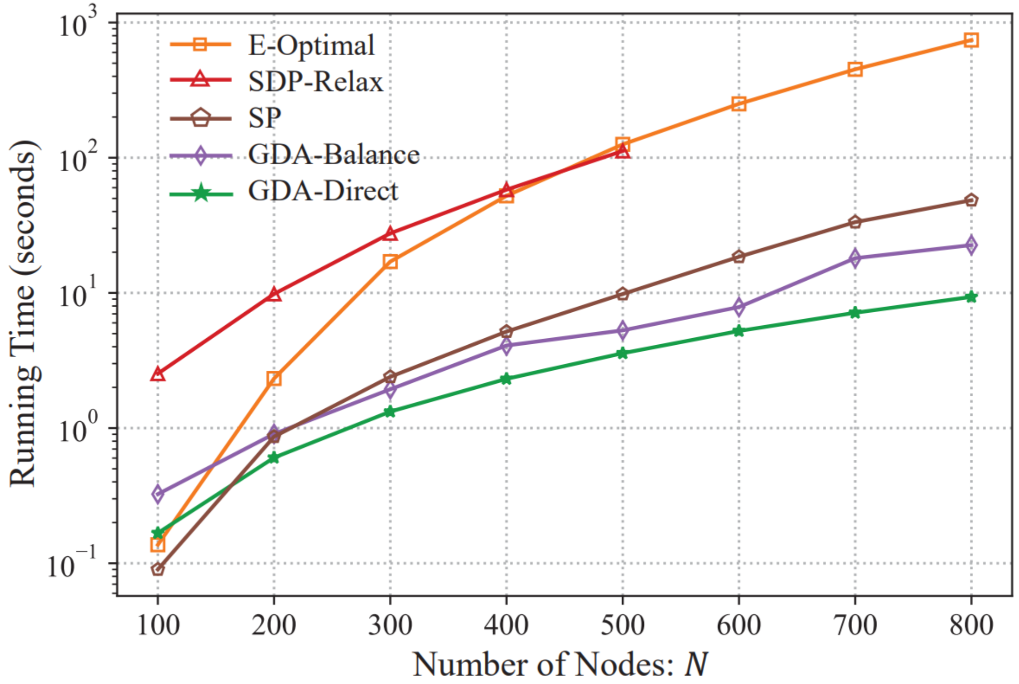
[1] Chinthaka Dinesh, Saghar Bagheri, Gene Cheung, Ivan V. Bajic, "**Linear-time Sampling on Signed Graphs via Gershgorin Disc Perfect Alignment**," *ICASSP*, 2022.

Results: directed graph sampling

(a) Reconstruction MSE on Erdös-Rényi Random Graph. GS = Diffusion Signal



(b) Running Time on Erdös-Rényi Random Graph



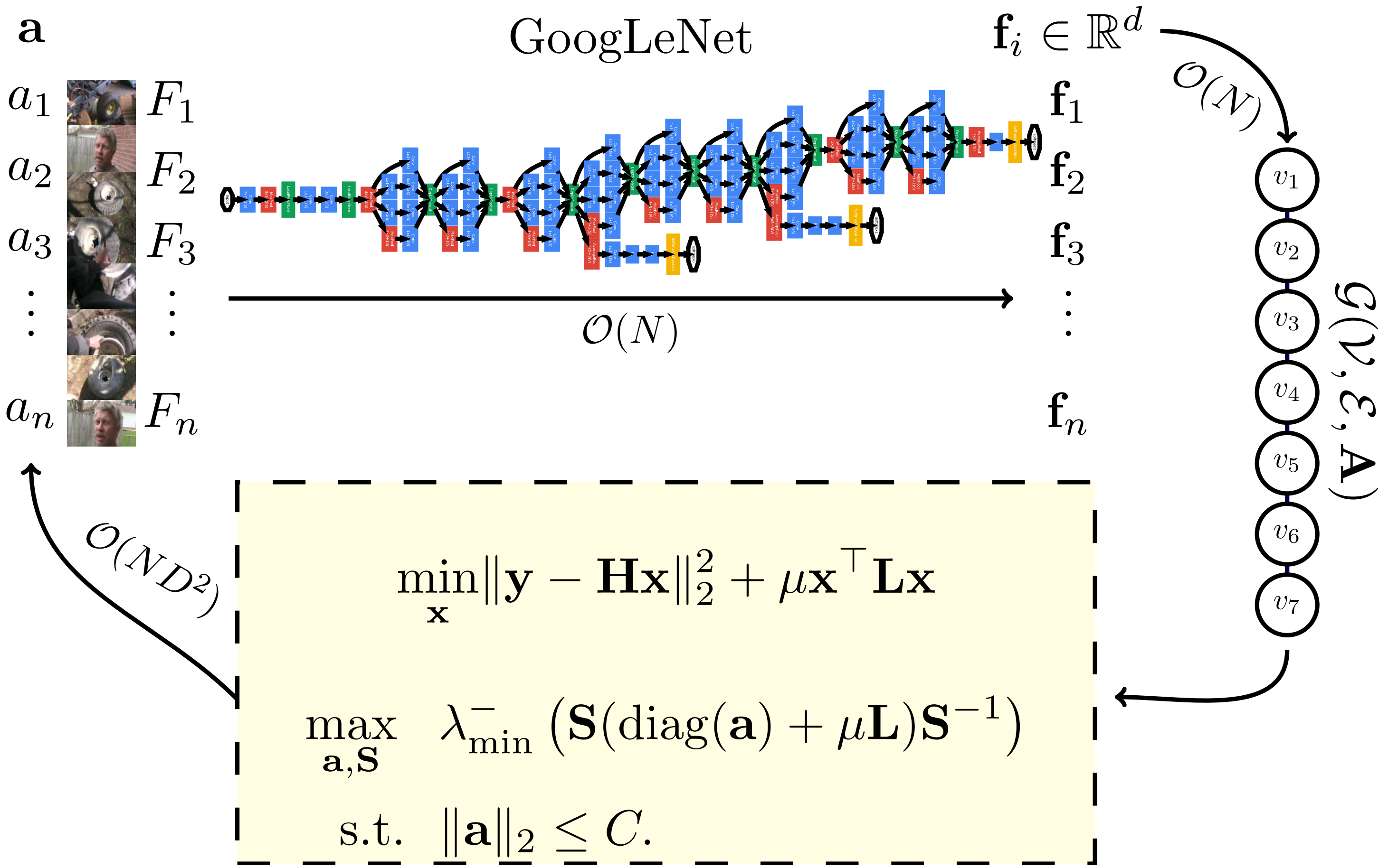
- GDA-Direct achieves lowest Recon. MSE
- Decreases MSE by 11.9%

- GDA-Direct is the fastest method
- Speeds up by 1.4 times.

[1] Yuejiang Li, H. Vicky Zhao, Gene Cheung, "**Eigen-Decomposition-Free Directed Graph Sampling via Gershgorin Disc Alignment**," *ICASSP'23*, Rhodes, Greece, June 2023.

Graph Sampling Application 1: video summarization

- **Problem**: Select ***Keyframes*** to summarize short video.
- **Solution**: Construct path graph from GoogLeNet features, choose graph samples.



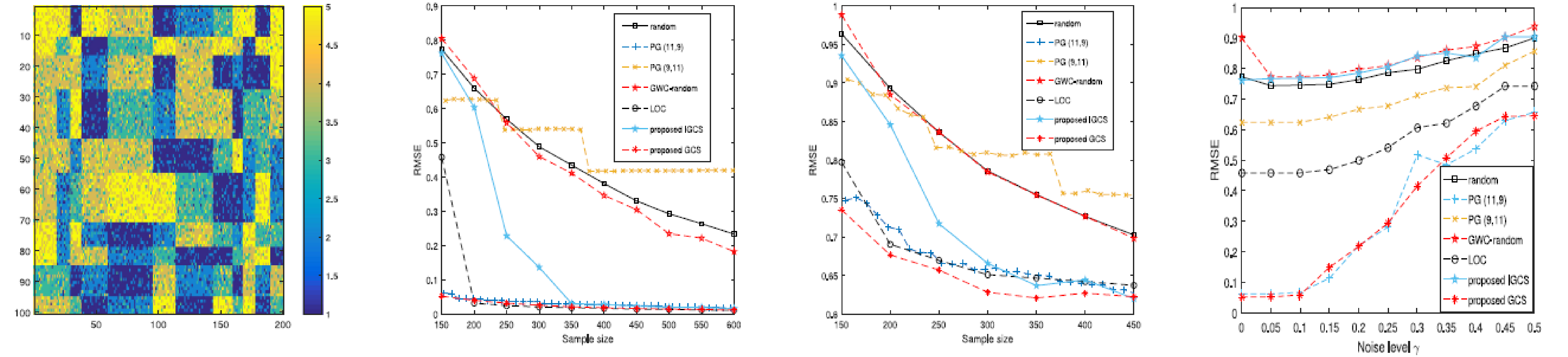
[1] S. Sahami, G. Cheung, C.-W. Lin, "**Fast Graph Sampling for Short Video Summarization using Gershgorin Disc Alignment**," *IEEE ICASSP*, May 2022.

Graph Sampling Application 2: matrix completion

- Pre-select a subset of matrix entries for sampling to maximize **matrix completion** fidelity.
- **Challenge**: select sampling set Ω to maximize λ_{min} of $\tilde{\mathbf{A}}_\Omega + \alpha \mathbf{I}_n \otimes \mathbf{L}_r + \beta \mathbf{L}_c \otimes \mathbf{I}_m$

graph Laplacians for row / column graphs

- RMSE of different sampling methods for MC on Synthetic Netflix. The matrix was completed using the *double graph smoothness* based method.



(a) Noisy synthetic Netflix signal (b) RMSE on noiseless signal (c) RMSE on noisy signal with $\gamma = 0.6$ (d) RMSE on different noise level γ

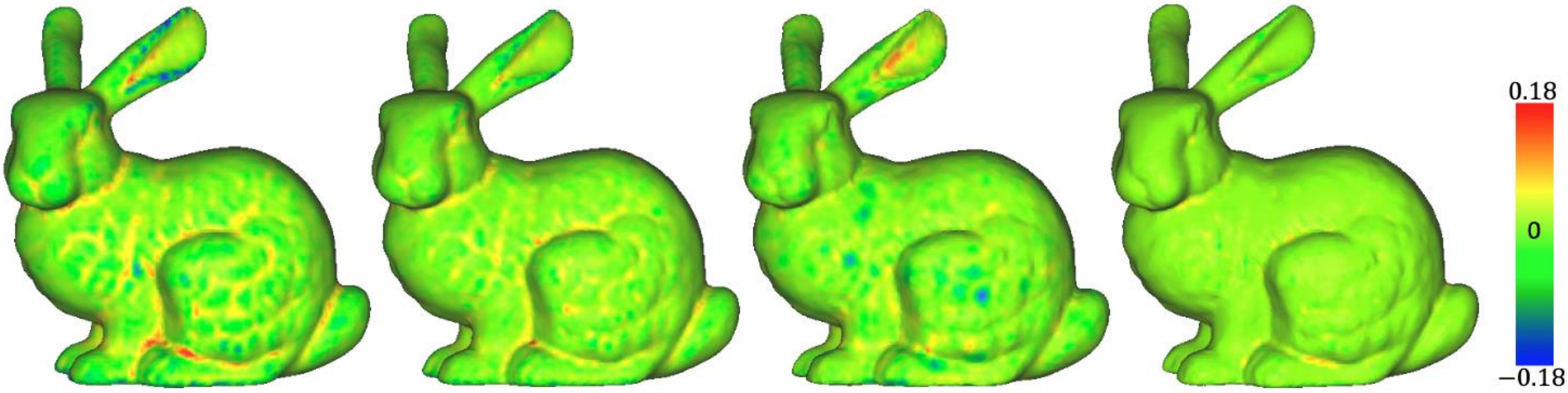
[1] F. Wang, Y. Wang, G. Cheung, C. Yang, "**Graph Sampling for Matrix Completion Using Recurrent Gershgorin Disc Shift**," vol. 68, pp. 1814-2829, *IEEE Transactions on Signal Processing*, April 2020.

Graph Sampling Application 3: 3D point cloud sub-sampling

- Reduce 3D point cloud size by sub-sampling while preserving the overall object shape.
- **Challenge**: select sampling matrix **H** to maximize λ_{min} of $\mathbf{H}^\top\mathbf{H} + \mu\mathcal{L}$

generalized graph Laplacian

- SR reconstruction results from diff. methods of sub-sampled Bunny under 0.2 sub-sampling ratio.



(a) BGFS (b) PDS (c) FPS (d) proposed

[1] C. Dinesh, G. Cheung, I. V. Bajic, "**Point Cloud Sampling via Graph Balancing and Gershgorin Disc Alignment**," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, January 2023.

Outline

- GSP overview
  - Graph frequencies from eigen-pairs
- Graph Learning
  - Positive, signed, directed, Hermitian graphs
- Graph Filtering
- Graph Sampling
- GSP Analysis for GCNs
- Conclusion

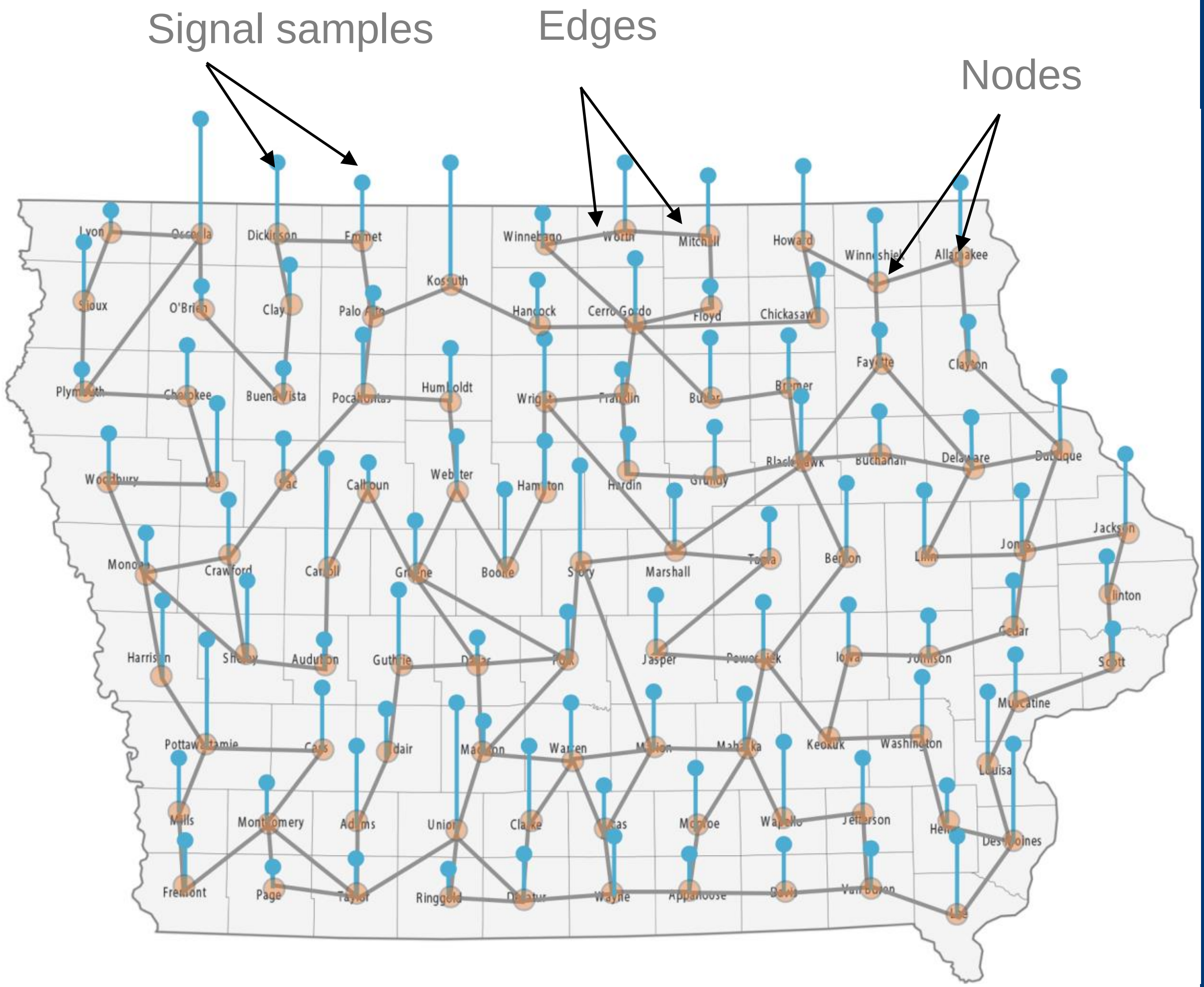
Motivation: graph sparsification

Crop Yield Prediction:

- Reduce price fluctuation.
- Stabilize food supply.
- Minimize uncertainty for farmers.

SOTA methods use **GCN**s:

- Dense underlying graph kernel can be used.
- Requires long training and execution time.



[1] Saghar Bagheri, Gene Cheung, Timothy Eadie, "**Graph Sparsification for GCN towards Optimal Crop Yield Prediction**," *International Geoscience and Remote Sensing Symposium* (IGARSS), Pasadena, CA, July 2023.

Graph Sparsification: Fiedler number

Fiedler number:

- Second smallest eigenvalue λ_2 of the Laplacian matrix **L**.
- Known to quantify "connectedness" of underlying graph.

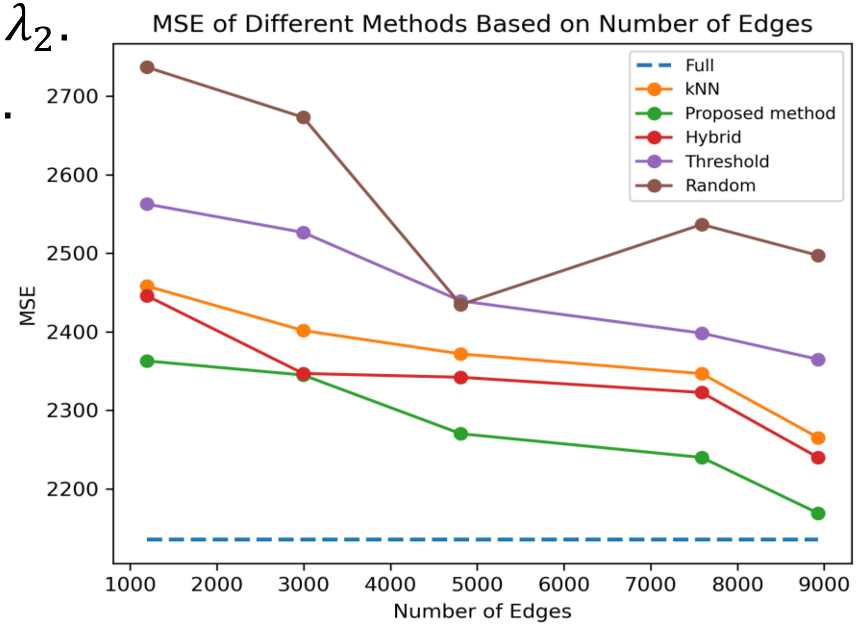
Fast method:

- In each iteration, remove edge that induces min change in λ_2.
- Modified Laplacian $\tilde{\mathbf{L}} = \mathbf{L} + \mathbf{E}^{m,n}$ with edge *(m, n)* removed.

$$\tilde{\lambda} \in [\, \lambda_2 - \|\mathbf{E}^{m,n}\mathbf{v}_2\|_2,\ \lambda_2 + \|\mathbf{E}^{m,n}\mathbf{v}_2\|_2 \,].$$

- Choose edge such that

$$(m^*, n^*) = \arg \min_{(m,n)\in\mathcal{E}} \|\mathbf{E}^{m,n}\mathbf{v}_2\|_2 .$$

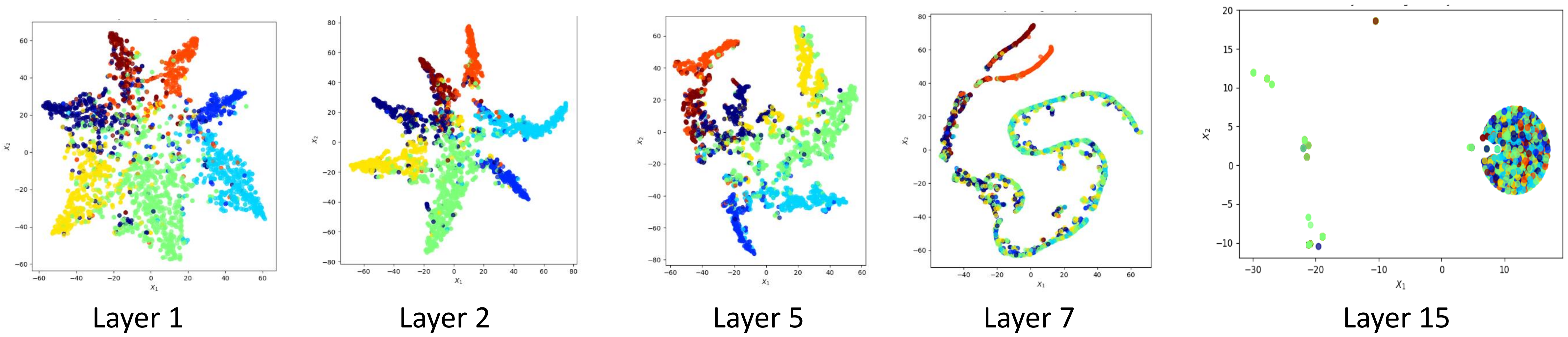


[1] Saghar Bagheri, Gene Cheung, Timothy Eadie, "**Graph Sparsification for GCN towards Optimal Crop Yield Prediction**," *International Geoscience and Remote Sensing Symposium* (IGARSS), Pasadena, CA, July 2023.

Motivation: graph learning with spectrum prior

Stacking more GCN layers → node representations become indistinguishable [1]

Q: How to alleviate over-smoothing?



t-SNE visualization of output features from GCNs[2] w/ different layers for Cora dataset

[1] Qimai Li, Zhichao Han, and Xiao-Ming Wu, “Deeper insights into graph convolutional networks for semi-supervised learning,” AAAI, 2018.
[2] Thomas N. Kipf and Max Welling, “Semi-supervised lassification with graph convolutional networks,” ICLR, 2017.
[3] Jin Zeng, Yang Liu, Gene Cheung, Wei Hu, "**Sparse Graph Learning with Spectrum Prior for Deep Graph Convolutional Networks**," ICASSP, June 2023.

Oversmoothing in GCNs

Multilayer GCN

- $\mathcal{G}$: Undirected graph with N nodes
- Output from GCN associated with $\mathcal{G}$ with L layers: $f = f_L \circ \cdots \circ f_1$
- l-th layer output with input $\mathbf{X}$: $f_l(\mathbf{X}) \triangleq \sigma(\, \mathbf{P}\, \mathbf{X}\, \mathbf{\Theta}^{(l)}\,)$

Activation operator — Graph operator — Filter weight

Over-smoothing in GCN

- $\mathcal{M}$: subspace spanned by 1st eigenvectors of ***normalized graph Laplacian*** $\tilde{\mathbf{L}} = \mathbf{I} - \mathbf{P}$
- More GCN layers → $f(\mathbf{X})$ converges to $\mathcal{M}$ [1]

Uninformative signals

$$\mathrm{d}_{\mathcal{M}}(f_l(\mathbf{X})) \leq r * \mathrm{d}_{\mathcal{M}}(\mathbf{X})$$

L2 distance to $\mathcal{M}$

Convergence rate characterized by graph spectrum

[1] Kenta Oono and Taiji Suzuki, “Graph neural networks exponentially lose expressive power for node classification,” ICLR, 2020.

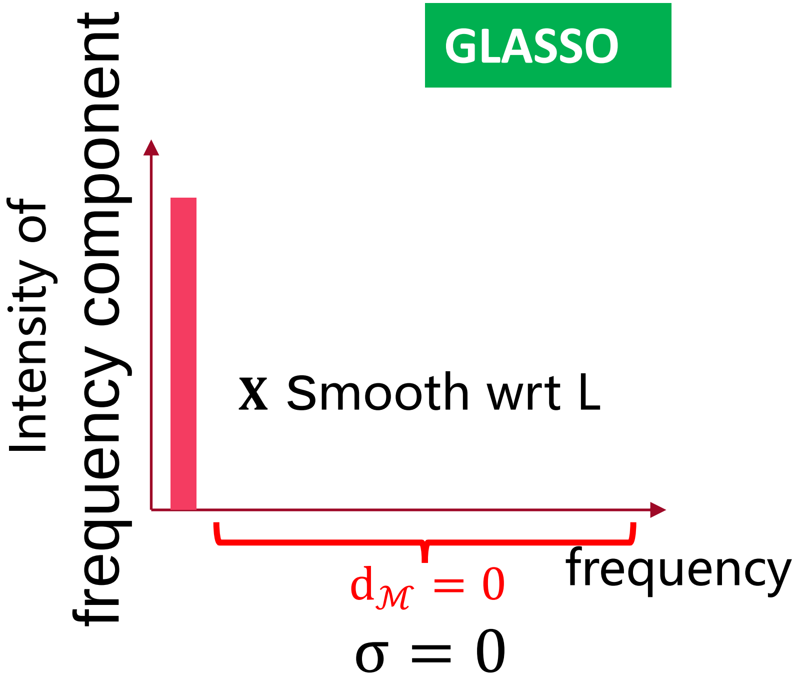
Sparse Graph Learning with Spectrum Prior

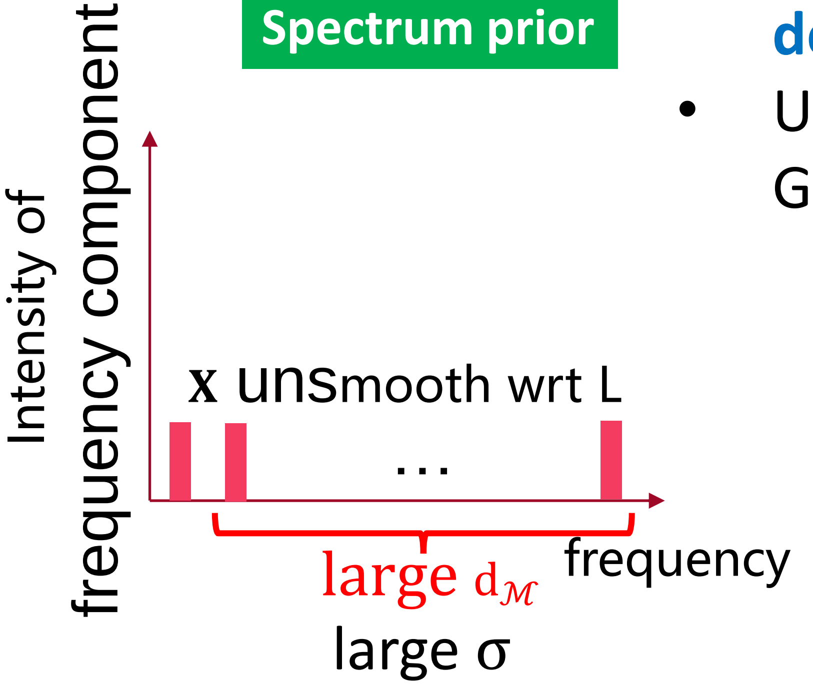
SGL-GCN: Sparse Graph Learning with Spectrum Prior for GCN

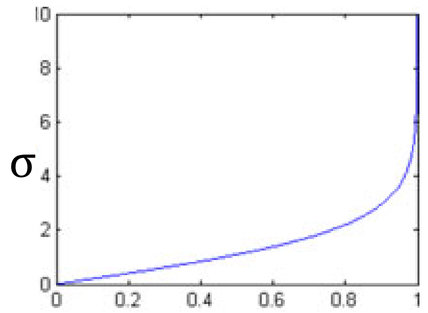
$$\min_{\mathbf{L} \succeq \mathbf{0}} \underbrace{\mathrm{Tr}(\mathbf{LC}) - \log\det(\mathbf{L}) + \rho||\mathbf{L}||_1}_{\text{GLASSO}} + \underbrace{\sigma \mathrm{Tr}(\mathbf{L})}_{\text{Spectrum prior}}$$

- Solve via **block coordinate descent** (BCD)
- Use output L for GCN

Weight σ:







$\sigma = \ln\left(\frac{1+M(S)}{1-M(S)}\right)$ where $M(S)$ is smoothness, S is given data

Gene Cheung (genec@yorku.ca)

LASSONDE SCHOOL OF ENGINEERING | YORK UNIVERSITÉ UNIVERSITY

Results: robust graph learning

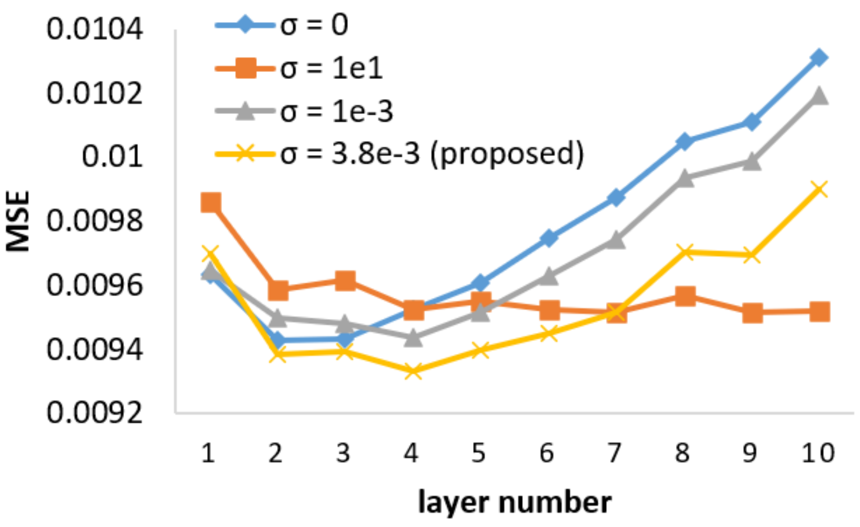
METR-LA Dataset

- Task: predict traffic speed given past 50 mins data (sampled every 5 mins)
- C: observations in training data

Proposed SGL-GCN

- Validated weight σ
- Higher acc w/ deeper GCN

Validation of weight computation



9.43e-3@Layer2
✗ Over-smoothing
✗ Acc

9.33e-3@Layer4
✔ Over-smoothing
✔ Acc

9. 51e-3@Layer7
✔ Over-smoothing
✗ Acc

Methods	2 layers	4 layers	8 layers
GCN	10.76	11.81	17.38
DropEdge[1]	10.79	12.25	17.72
Oono's [2]	10.79	11.89	17.32
SGL-GCN w/o spectrum	9.43	9.52	10.05
SGL-GCN	**9.38**	**9.33**	**9.70**

[1] Yu Rong, Wenbing Huang, et al.,"Dropedge: Towards deep graph convolutional networks on node classification," ICLR, 2020.
[2] Kenta Oono and Taiji Suzuki, "Graph neural networks exponentially lose expressive power for node classification," ICLR, 2020.

Outline

- GSP overview
  - Graph frequencies from eigen-pairs
- Graph Learning
  - Positive, signed, directed, Hermitian graphs
- Graph Filtering
- Graph Sampling
- GSP Analysis for GCNs
- Conclusion

Gene Cheung (genec@yorku.ca)

Conclusion

- GSP analyzes and processes discrete signals on graphs.
- Graphs captures pairwise relationships:
  - Positive, signed, directed, Hermitian graphs.

- Graph Learning
  - Statistical sparse graph learning, metric learning
- Graph Filtering
  - Spectral filters in graph spectrum
- Graph Sampling
  - Fast algorithms based on Gershgorin circle theorem
- GSP analysis for GCNs
  - Graph spectrum affects performance, oversmoothing

Applications:
Image coding, denoising, deblurring, interpolation, contrast enhancement, light field image denoising, 3D point cloud denoising, sub-sampling, super-resolution, matrix completion, semi-supervised classifier learning, video summarization, crop yield prediction

Gene Cheung (genec@yorku.ca)

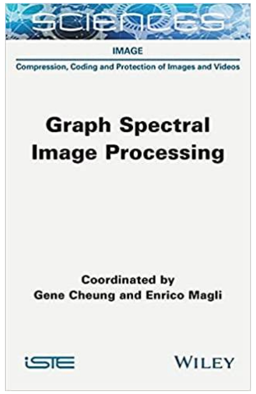
Contact Info

- **Homepage**:

https://www.eecs.yorku.ca/~genec/index.html

- **E-mail**:

genec@yorku.ca



- **New book**:

G. Cheung, E. Magli, (edited) ***Graph Spectral Image Processing***, ISTE/Wiley, August 2021.