Gene Cheung
Associate Professor, York University
22nd August, 2019

Fast Graph Sampling using Gershgorin Disc Alignment

[1] Yuanchao Bai, Fen Wang, Gene Cheung, Yuji Nakatsukasa, Wen Gao, "**Fast Graph Sampling Set Selection Using Gershgorin Disc Alignment**," submitted to *IEEE Transactions on Signal Processing*, July 2019.

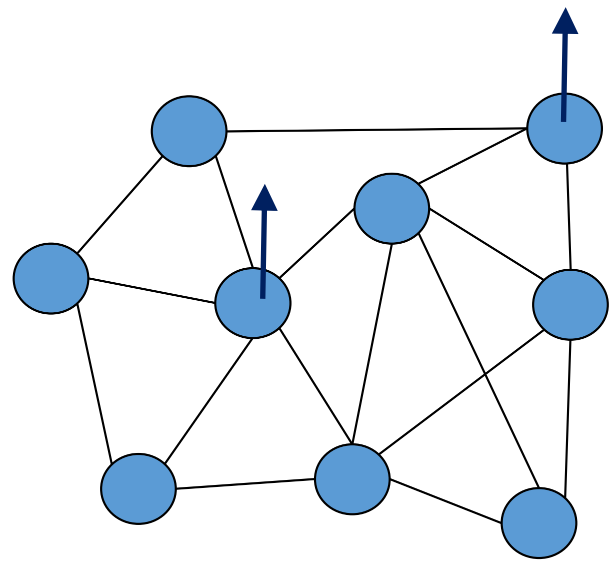
Outline

- What is Graph Sampling?
  - Related Work
- Signal Reconstruction using GLR
- Gershgorin Disc Alignment

Graph Sampling (with and without noise)

- **Q**: How to choose best samples for graph-based reconstruction?

- Existing graph sampling strategies extend **Nyquist sampling** to graph data kernels:
  - Assume ***bandlimited*** signal.
  - Greedily select most "informative" samples by computing extreme eigenvectors of sub-matrix.
  - Computation-expensive.



[1] A. Anis, A. Gadde, and A. Ortega, "**Efficient sampling set selection for bandlimited graph signals using graph spectral proxies**," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3775–3789, 2016.

Related Works

- Graph sampling
  - Aggressive sampling [1]
  - Local measurement [2]
  - Selection sampling
    - Deterministic
      - Bandlimitedness
        - A-optimal (M. Tsitsvero.TSP2016; F. Wang. SPL2018)
        - E-optimal (S. Chen. TSP2015)
        - Spectral proxies (A. Anis. TSP2016; A. Anis. TIT2018)
        - Localized coverage (**A. Sakiyama. TSP2019**)
      - Smoothness: GMRF model (**P. Chen. ICASSP2018**; **Y. Bai. ICASSP 2019**)
    - Random selection [3]

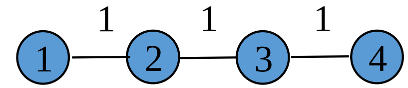
Eigen-decomposition Free

[1] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "**Sampling of graph signals with successive local aggregations**." *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1832–1843, 2016.

[2] X. Wang, J. Chen, and Y. Gu, "**Local measurement and reconstruction for noisy bandlimited graph signals**," *Signal Processing*, vol. 129, pp. 119–129, 2016.

[3] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, "**Random sampling of bandlimited signals on graphs**," *Applied and Computational Harmonic Analysis*, vol. 44, no. 2, pp. 446–475, 2018.

Signal Reconstruction using GLR



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

Sample set {2, 4}

- **Signal Model**:

observation

sampling matrix

desired signal

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

noise

- Signal prior is graph Laplacian regularizer (GLR) [1]:

$$\mathbf{x}^T\mathbf{L}\mathbf{x} = \frac{1}{2}\sum_{i,j} w_{i,j}\left(x_i - x_j\right)^2 = \sum_k \lambda_k \tilde{x}_k^2$$

signal smooth w.r.t. graph

signal contains mostly low graph freq.

- **MAP Formulation**:

fidelity term

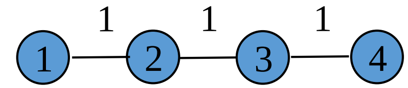
signal prior

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

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

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

Stability of Linear System



- Examine system of linear equations :

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

coefficient matrix **B**

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

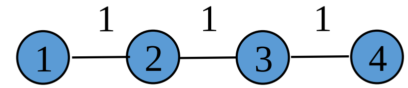
$$\mathbf{H}^{\mathrm{T}}\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Sample set {2, 4}

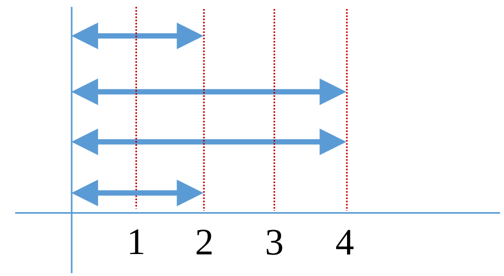
- Stability depends on the **condition number** (λ_{max}/ λ_{min}) of coeff. matrix **B**.
- λ_{max} is upper-bounded by $1+\mu\, 2{*}d_{max}$.
- **Goal**: select samples to maximize λ_{min} (without computing eigen-pairs)!
- Also minimizes worst-case MSE:

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

Gershgorin Circle Theorem



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



- **Gershgorin Circle Theorem**:
  - Row i of **L** maps to a **Gershgorin disc** w/ centre L_{ii} and radius R_i

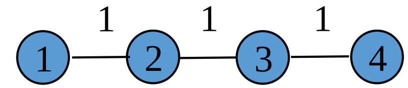
$$R_i = \sum_{j \neq i} |L_{ij}|$$

  - $\lambda_{\min}$ is lower-bounded by smallest left-ends of Gershgorin discs:

$$\min_i \; L_{i,i} - R_i \leq \lambda_{\min}$$

- Graph Laplacian **L** has all Gershgorin disc left-ends at 0 → **L** is psd.

Gershgorin Disc Alignment



- **Main Idea**: Select samples to max smallest disc left-end of coefficient matrix **B**:

$$\mathbf{B} = \mathbf{H}^T \mathbf{H} + \mu \mathbf{L}$$ ← coeff. matrix

- Sample node → shift disc.

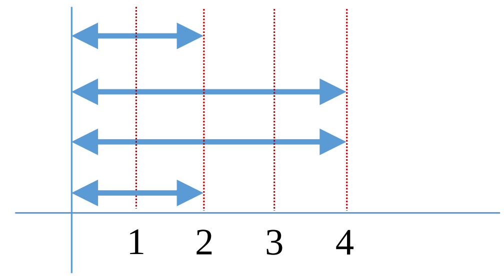
- Consider similar transform of **B**:

$$\mathbf{C} = \mathbf{S}\,\mathbf{B}\,\mathbf{S}^{-1}$$ ← similarity transform

diagonal matrix w/ scale factors

- Scale row → **expand** disc radius.
  → **shrink** neighbors' disc radius.

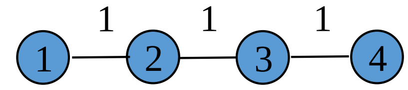
$$\mathbf{B} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$



Sample set { }

Scale factor {1,1,1,1}

Gershgorin Disc Alignment

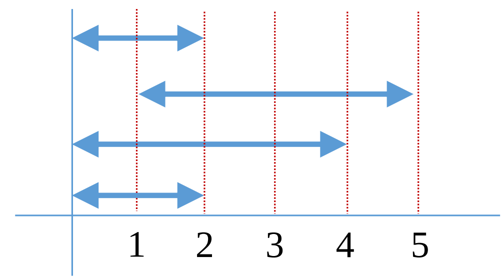


- **Main Idea**: Select samples to max smallest disc left-end of coefficient matrix **B**:

$$\mathbf{B} = \mathbf{H}^T \mathbf{H} + \mu \mathbf{L} \longleftarrow \text{coeff. matrix}$$

$$\mathbf{B} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 3 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

- Sample node → shift disc.



- Consider similar transform of **B**:

$$\mathbf{C} = \mathbf{S}\,\mathbf{B}\,\mathbf{S}^{-1} \longleftarrow \text{similarity transform}$$

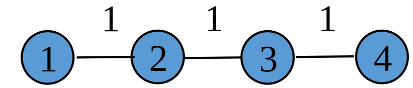
diagonal matrix w/ scale factors

Sample set {2}

Scale factor {1,1,1,1}

- Scale row → **expand** disc radius.
  → **shrink** neighbors' disc radius.

Gershgorin Disc Alignment

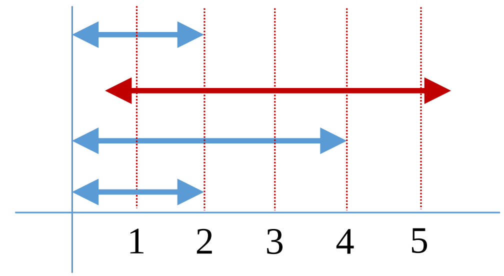


- **Main Idea**: Select samples to max smallest disc left-end of coefficient matrix **B**:

$$\mathbf{B} = \mathbf{H}^T \mathbf{H} + \mu \mathbf{L} \quad \longleftarrow \text{coeff. matrix}$$

$$\mathbf{B} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 3 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

- Sample node → shift disc.



- Consider similar transform of **B**:

$$\mathbf{C} = \mathbf{S}\,\mathbf{B}\,\mathbf{S}^{-1} \quad \longleftarrow \text{similarity transform}$$

diagonal matrix w/ scale factors

Sample set $\{2\}$

Scale factor $\{1, s_2, 1, 1\}$

- Scale row → **expand** disc radius.
  → **shrink** neighbors' disc radius.

Gershgorin Disc Alignment

- **Main Idea**: Select samples to max smallest disc left-end of coefficient matrix **B**:

$$\mathbf{B} = \mathbf{H}^T\,\mathbf{H} + \mu\,\mathbf{L} \;\longleftarrow\; \text{coeff. matrix}$$

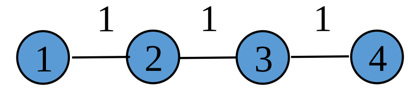
- Sample node → shift disc.

- Consider similar transform of **B**:

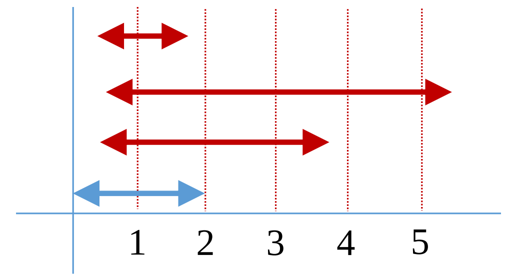
$$\mathbf{C} = \mathbf{S}\,\mathbf{B}\,\mathbf{S}^{-1} \;\longleftarrow\; \text{similarity transform}$$

diagonal matrix w/ scale factors

- Scale row → **expand** disc radius.
  → **shrink** neighbors' disc radius.



$$\mathbf{B} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 3 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$



Sample set {2}

Scale factor $\{1, s_2, 1, 1\}$

Primal Sample Selection Problem

- **Optimization**: Select sample vector **a** and scalars **s**:

$$\max_{\mathbf{a},\mathbf{s}} \min_{i\in\{1,\dots,N\}} \quad c_{ii} - \sum_{j\neq i} |c_{ij}|$$

$$\text{s.t. } \mathbf{C} = \mathbf{S}\left(\mathbf{A} + \mu\mathbf{L}\right)\mathbf{S}^{-1}$$

$$\mathbf{A} = \text{diag}(\mathbf{a}), \quad a_i \in \{0,1\}, \quad \sum_{i=1}^{N} a_i \leq K,$$

$$\mathbf{S} = \text{diag}(\mathbf{s}), \quad s_i > 0.$$

Primal Sample Selection Problem

- **Optimization**: Select sample vector **a** and scalars **s**:

$$\max_{\mathbf{a},\mathbf{s}} \boxed{\min_{i \in \{1,\dots,N\}} \quad c_{ii} - \sum_{j \neq i} |c_{ij}|}$$

smallest disc left-end of **C**

$$\text{s.t. } \mathbf{C} = \mathbf{S}\left(\mathbf{A} + \mu \mathbf{L}\right)\mathbf{S}^{-1}$$

$$\mathbf{A} = \mathrm{diag}(\mathbf{a}), \quad a_i \in \{0,1\}, \quad \sum_{i=1}^{N} a_i \leq K,$$

$$\mathbf{S} = \mathrm{diag}(\mathbf{s}), \quad s_i > 0.$$

Primal Sample Selection Problem

- **Optimization**: Select sample vector **a** and scalars **s**:

$$\max_{\mathbf{a},\mathbf{s}} \boxed{\min_{i \in \{1,\ldots,N\}} \quad c_{ii} - \sum_{j \neq i} |c_{ij}|}$$

smallest disc left-end of **C**

$$\text{s.t.} \ \boxed{\mathbf{C} = \mathbf{S}\left(\mathbf{A} + \mu \mathbf{L}\right)\mathbf{S}^{-1}}$$

C is similar transform of coeff. matrix

$$\mathbf{A} = \text{diag}(\mathbf{a}), \quad a_i \in \{0,1\}, \quad \sum_{i=1}^{N} a_i \leq K,$$

$$\mathbf{S} = \text{diag}(\mathbf{s}), \quad s_i > 0.$$

Primal Sample Selection Problem

- **Optimization**: Select sample vector **a** and scalars **s**:

$$\max_{\mathbf{a},\mathbf{s}} \boxed{\min_{i \in \{1,\ldots,N\}} \quad c_{ii} - \sum_{j \neq i} |c_{ij}|}$$

smallest disc left-end of **C**

$$\text{s.t.} \ \boxed{\mathbf{C} = \mathbf{S}\left(\mathbf{A} + \mu \mathbf{L}\right)\mathbf{S}^{-1}}$$

C is similar transform of coeff. matrix

$$\mathbf{A} = \mathrm{diag}(\mathbf{a}), \quad \boxed{a_i \in \{0, 1\}}, \quad \boxed{\sum_{i=1}^{N} a_i \leq K,}$$

sample vector **a** is binary and within budget *K*

$$\mathbf{S} = \mathrm{diag}(\mathbf{s}), \quad s_i > 0.$$

Primal Sample Selection Problem

- **Optimization**: Select sample vector **a** and scalars **s**:

$$\max_{\mathbf{a},\mathbf{s}} \boxed{\min_{i\in\{1,\ldots,N\}} \quad c_{ii} - \sum_{j\neq i} |c_{ij}|}$$

smallest disc left-end of **C**

$$\text{s.t.} \boxed{\mathbf{C} = \mathbf{S}\left(\mathbf{A} + \mu\mathbf{L}\right)\mathbf{S}^{-1}}$$

C is similar transform of coeff. matrix

$$\mathbf{A} = \mathrm{diag}(\mathbf{a}), \quad \boxed{a_i \in \{0,1\}}, \quad \boxed{\sum_{i=1}^{N} a_i \leq K,}$$

sample vector **a** is binary and within budget K

$$\mathbf{S} = \mathrm{diag}(\mathbf{s}), \quad \boxed{s_i > 0.}$$

scalars **s** are positive

Primal Sample Selection Problem

- **Optimization**: Select sample vector **a** and scalars **s**:

$$\max_{\mathbf{a},\mathbf{s}} \boxed{\min_{i \in \{1,\ldots,N\}} \quad c_{ii} - \sum_{j \neq i} |c_{ij}|}$$

smallest disc left-end of **C**

$$\text{s.t.} \boxed{\mathbf{C} = \mathbf{S}\left(\mathbf{A} + \mu \mathbf{L}\right)\mathbf{S}^{-1}}$$

C is similar transform of coeff. matrix

$$\mathbf{A} = \mathrm{diag}(\mathbf{a}), \quad \boxed{a_i \in \{0,1\}}, \quad \boxed{\sum_{i=1}^{N} a_i \leq K,}$$

sample vector **a** is binary and within budget K

$$\mathbf{S} = \mathrm{diag}(\mathbf{s}), \quad \boxed{s_i > 0.}$$

scalars **s** are positive

- **Difficulty**: max-min objective is hard to optimize.

Dual Sample Selection Problem

- **Dual Formulation**: Select sample vector **a** and scalars **s**:

$$\min_{\mathbf{a},\mathbf{s}} \sum_{i=1}^{N} a_i$$

$$\text{s.t. } \mathbf{C} = \mathbf{S}\left(\mathbf{A} + \mu\mathbf{L}\right)\mathbf{S}^{-1}, \quad c_{ii} - \sum_{j \neq i} |c_{ij}| \geq T, \quad \forall i$$

$$\mathbf{A} = \mathrm{diag}(\mathbf{a}), \qquad a_i \in \{0, 1\},$$

$$\mathbf{S} = \mathrm{diag}(\mathbf{s}), \qquad s_i > 0.$$

Dual Sample Selection Problem

- **Dual Formulation**: Select sample vector **a** and scalars **s**:

$$\min_{\mathbf{a},\mathbf{s}} \boxed{\sum_{i=1}^{N} a_i}$$

total number of samples

$$\text{s.t. } \mathbf{C} = \mathbf{S}\left(\mathbf{A} + \mu\mathbf{L}\right)\mathbf{S}^{-1}, \quad c_{ii} - \sum_{j \neq i} |c_{ij}| \geq T, \quad \forall i$$

$$\mathbf{A} = \mathrm{diag}(\mathbf{a}), \qquad a_i \in \{0, 1\},$$

$$\mathbf{S} = \mathrm{diag}(\mathbf{s}), \qquad s_i > 0.$$

Dual Sample Selection Problem

- **Dual Formulation**: Select sample vector **a** and scalars **s**:

$$\min_{\mathbf{a},\mathbf{s}} \boxed{\sum_{i=1}^{N} a_i}$$

total number of samples

$$\text{s.t. } \mathbf{C} = \mathbf{S}\left(\mathbf{A} + \mu\mathbf{L}\right)\mathbf{S}^{-1}, \quad \boxed{c_{ii} - \sum_{j \neq i} |c_{ij}| \geq T, \quad \forall i}$$

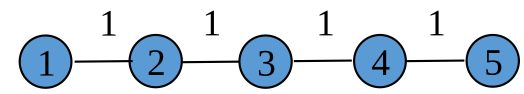
$$\mathbf{A} = \text{diag}(\mathbf{a}), \qquad a_i \in \{0, 1\},$$

$$\mathbf{S} = \text{diag}(\mathbf{s}), \qquad s_i > 0.$$

all disc left-ends are at least T

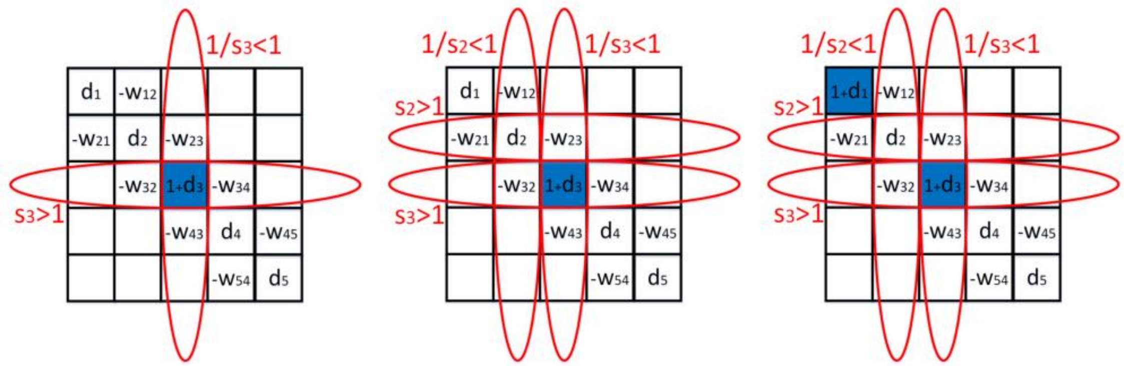
- **Proposition**: If there exists threshold T s.t. optimal sol'n (**a**,**s**) to dual satisfies $\Sigma\, a_i = K$, one dual sol'n is also optimal to primal.

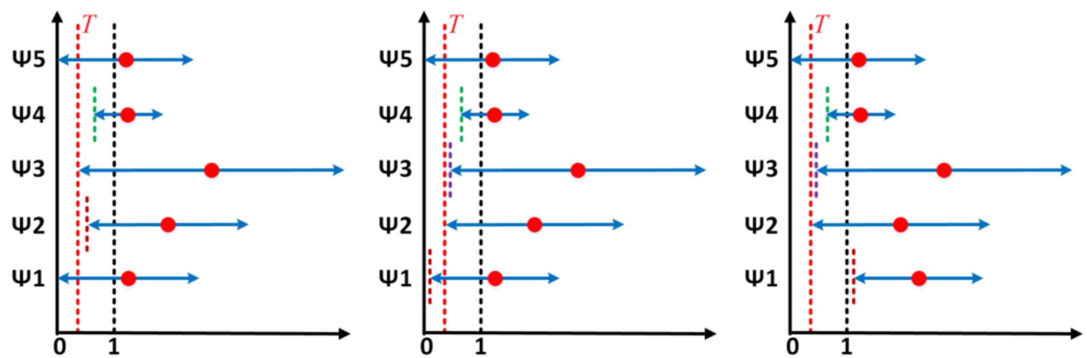
Solving the Dual: align disc at T



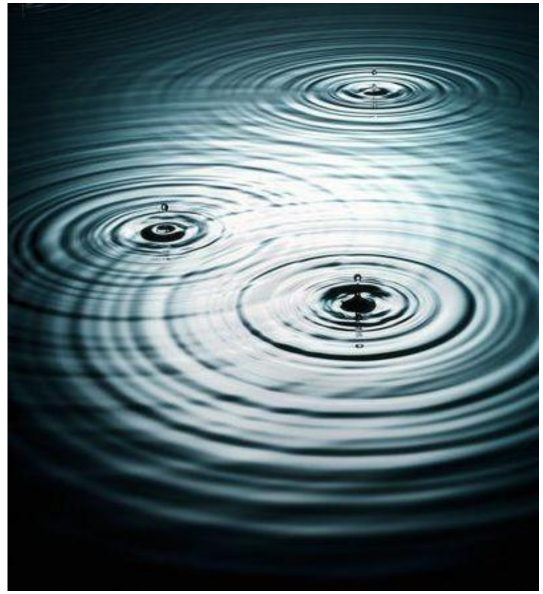
- **Breadth First Iterative Sampling (BFIS)**:
  - Given initial node set, threshold T.

1. Sample chosen node i
   (shift disc)
2. Scale row i
   (expand disc radius i to T)
3. If disc left-end of connected node $j > T$,
   Scale row j
   (expand disc radius j to T)
   Else,
   Add node j to node set.
4. Goto 1 if node set not empty.
5. Output sample set and count K.

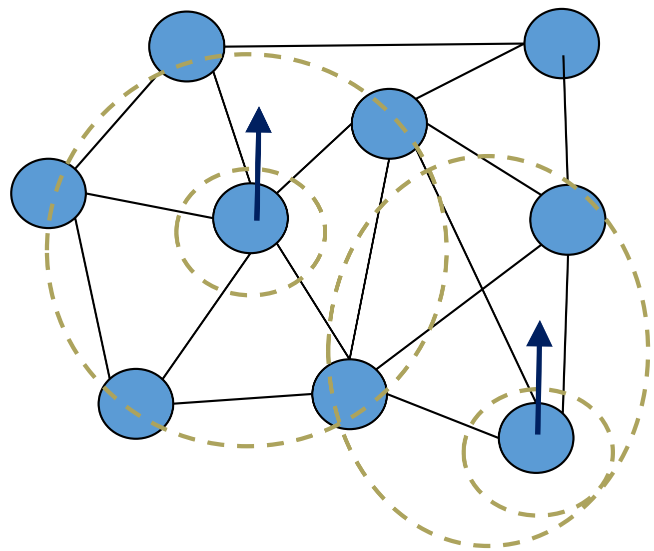




Disc-based Sampling (intuition)



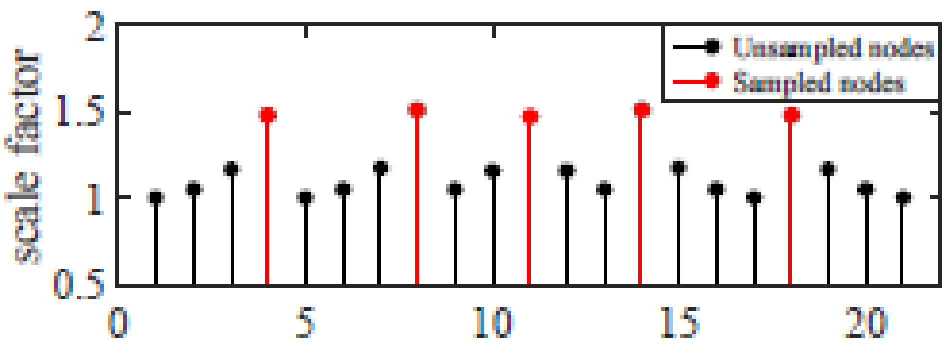
- **Analogy**: throw pebbles into a pond.
- **Disc Shifting**: throw pebble at sample node *i*.
- **Disc Scaling**: ripple to neighbors of node *i*.
- **Goal**: Select min # of samples so ripple at each node is at least *T*.



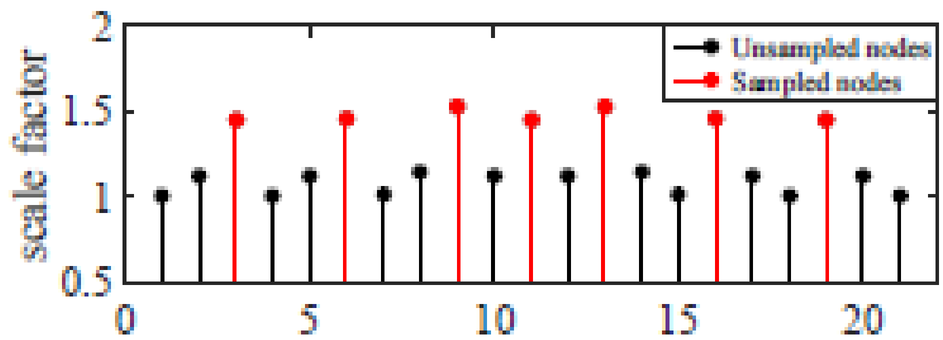
Gershgorin Disc Alignment (math)

- **Binary Search with BFIS**:
  - Sample count K inverse proportional to threshold T.
  - Binary search on T to drive count K to budget.

- **Example**: line graph with equal edge weight.
  - Uniform sampling.



(a)



(b)

Results: Graph Sampling

- GDA is 100x to 1000x faster than state-of-art methods computing e-vectors.
- GDA is “comparable” in complexity to Random [23] and Ed-free [8].

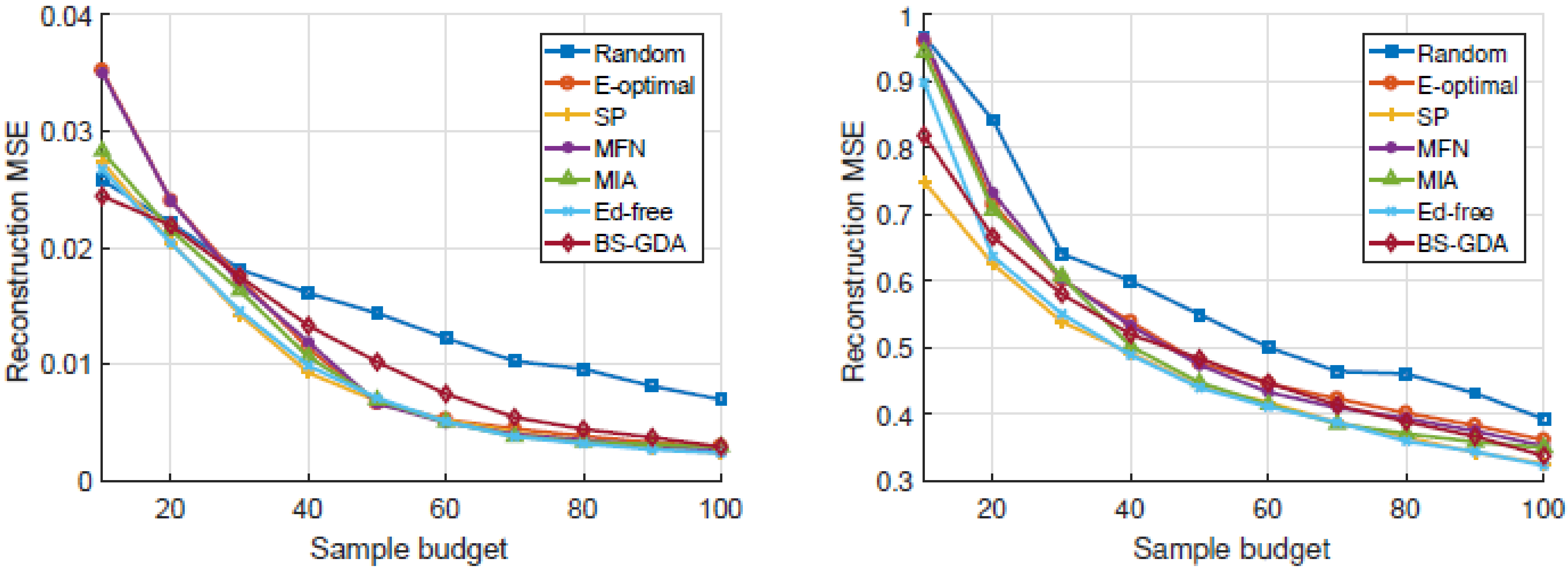
TABLE II
SPEEDUP FACTORS OF OUR ALGORITHM WITH RESPECT TO OTHER SAMPLING ALGORITHMS FOR $N = 3000$

Sampling Methods	Sensor	Community
Random [23]	0.22	0.21
E-optimal [20]	2812.77	1360.76
SP [12]	174.09	466.18
MFN [18]	2532.91	1184.23
MIA [16]	1896.19	964.65
Ed-free [8]	1.82	8.11

[1] Yuanchao Bai, Fen Wang, Gene Cheung, Yuji Nakatsukasa, Wen Gao, "**Fast Graph Sampling Set Selection Using Gershgorin Disc Alignment**," submitted to *IEEE Transactions on Signal Processing*, July 2019.

Results: Graph Sampling

- **Small graphs**: GDA has roughly the same reconstruction MSE.
  - Random sensor graph of size 500 for two signal types.

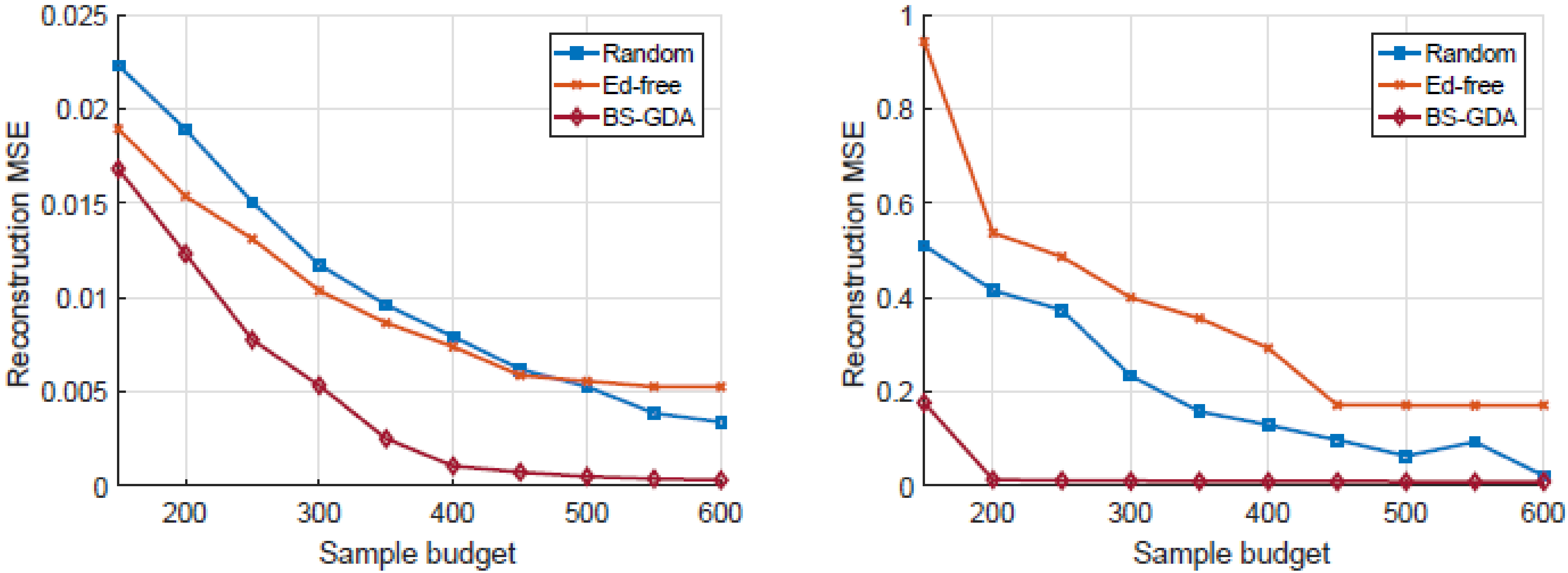


(a) (b)

[1] Yuanchao Bai, Fen Wang, Gene Cheung, Yuji Nakatsukasa, Wen Gao, "**Fast Graph Sampling Set Selection Using Gershgorin Disc Alignment**," submitted to *IEEE Transactions on Signal Processing*, July 2019.

Results: Graph Sampling

- **Large graphs**: GDA has smallest reconstruction MSE.
  - Minnesota road graph of size 2642 and for two signal types.



(e) (f)

[1] Yuanchao Bai, Fen Wang, Gene Cheung, Yuji Nakatsukasa, Wen Gao, "**Fast Graph Sampling Set Selection Using Gershgorin Disc Alignment**," submitted to *IEEE Transactions on Signal Processing*, July 2019.

Summary

- Graph Sampling
  - Generalization of Nyquist sampling to graph domain.
  - Existing works require computation of extreme eigenvectors.

- Disc-based graph sampling
  - Each eigenvalue is contained in a Gershgorin disc.
  - Maximize smallest disc left-ends to maximize $\lambda_{\min}$.
  - Roughly linear time, 100x to 1000x faster than e-vector schemes.

Q&A

- Email: genec@yorku.ca
- Homepage: https://www.eecs.yorku.ca/~genec/index.html