Structure-Aware Spectral Sparsification via Uniform Edge Sampling

Kaiwen He   Petros Drineas   Rajiv Khanna

Purdue University
Department of Computer Science

NeurIPS 2025 Video Presentation

The Scalability Challenge

Spectral Clustering: Powerful method for graph partitioning [Ng 2001, von Luxberg 2007]

- Compute eigenvectors of graph Laplacian
- Embed nodes in low-dimensional space
- Apply k-means clustering

The Scalability Challenge

Spectral Clustering: Powerful method for graph partitioning [Ng 2001, von Luxberg 2007]

- Compute eigenvectors of graph Laplacian
- Embed nodes in low-dimensional space
- Apply k-means clustering

Problem: Clustering massive graphs is expensive! $\mathcal{O}(\#\text{ of edges})$ [Macgregor 2023]

The Scalability Challenge

Spectral Clustering: Powerful method for graph partitioning [Ng 2001, von Luxberg 2007]

- Compute eigenvectors of graph Laplacian
- Embed nodes in low-dimensional space
- Apply k-means clustering

Problem: Clustering massive graphs is expensive! $\mathcal{O}(\#\text{ of edges})$ [Macgregor 2023]

One Solution? Spectral sparsification $\rightarrow$ reduce edges while preserving structure [Spielman 2008]

Classical Approach

Sample edges proportional to *effective resistances*
Requires solving approximate effective resistances for all edges $\rightarrow$ **expensive!**

The Scalability Challenge

Spectral Clustering: Powerful method for graph partitioning [Ng 2001, von Luxberg 2007]

- Compute eigenvectors of graph Laplacian
- Embed nodes in low-dimensional space
- Apply k-means clustering

Problem: Clustering massive graphs is expensive! $\mathcal{O}(\#$ of edges$)$ [Macgregor 2023]

One Solution? Spectral sparsification $\rightarrow$ reduce edges while preserving structure [Spielman 2008]

Classical Approach

Sample edges proportional to *effective resistances*
Requires solving approximate effective resistances for all edges $\rightarrow$ **expensive!**

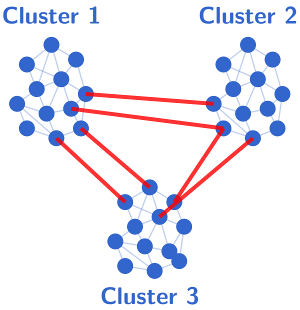
Our Question

Can simple uniform sampling work for well-clustered graphs?

Key Insight: Uniform Sampling IS Structure-Aware

Rough Intuition: In well-separated clusters

- **Intra-cluster edges** lie in the dominant $n-k$ eigenspace of the Laplacian
- **Inter-cluster edges** lie in the bottom k eigenspace of the Laplacian
- Uniform sampling naturally biases towards preserving the dominant n-k eigenspace!



Key issue: Intercluster effective resistances are unbounded! (depend on size of graph)

Structure Ratio Definition

Clusterability constant: $\Upsilon(k) = \frac{\lambda_{k+1}}{\rho_G(k)}$

- λ_{k+1}: $(k+1)$-th eigenvalue of normalized Laplacian
- $\rho_G(k) :=$ min k-way cut

Typical assumption: $\Upsilon(k) = \Omega(k^2)$ [Peng 2015]

Structure Ratio Definition

Clusterability constant: $\Upsilon(k) = \frac{\lambda_{k+1}}{\rho_G(k)}$

- λ_{k+1}: $(k+1)$-th eigenvalue of normalized Laplacian
- $\rho_G(k) :=$ min k-way cut

Typical assumption: $\Upsilon(k) = \Omega(k^2)$ [Peng 2015]

Cluster condition number: $\kappa := \frac{\lambda_n}{\lambda_{k+1}}$.

Structure Ratio Definition

Clusterability constant: $\Upsilon(k) = \frac{\lambda_{k+1}}{\rho_G(k)}$

- λ_{k+1}: $(k+1)$-th eigenvalue of normalized Laplacian
- $\rho_G(k) :=$ min k-way cut

Typical assumption: $\Upsilon(k) = \Omega(k^2)$ [Peng 2015]

Cluster condition number: $\kappa := \frac{\lambda_n}{\lambda_{k+1}}$.

Interpretation:

- Large $\Upsilon(k) \Rightarrow k$ well-separated clusters exist
- λ_{k+1} large: no $(k+1)$-th cluster
- $\rho_G(k)$ small: sparse cuts between k clusters.
- κ small: clusters have good connectivity.

Main Result: Uniform Sampling Preserves Cluster Structure

Theorem (Informal)

For graphs with well-separated k clusters (large $\Upsilon(k)$), uniformly sampling

$$m = O\left(\frac{\gamma^2 n \log n}{\varepsilon^2}\right) \text{ edges}$$

preserves the spectral subspace for clustering:

$$\left\| V_{n-k} V_{n-k}^T C \right\|_F^2 \leq k \left(\frac{1}{\Upsilon(k)} + \frac{\varepsilon}{1-\varepsilon} \kappa \right)$$

where V_{n-k} = top $(n-k)$ eigenvectors of sparsified Laplacian, C = cluster indicators.

Main Result: Uniform Sampling Preserves Cluster Structure

Theorem (Informal)

For graphs with well-separated k clusters (large $\Upsilon(k)$), uniformly sampling

$$m = O\left(\frac{\gamma^2 n \log n}{\varepsilon^2}\right) \text{ edges}$$

preserves the spectral subspace for clustering:

$$\left\| V_{n-k} V_{n-k}^T C \right\|_F^2 \leq k \left(\frac{1}{\Upsilon(k)} + \frac{\varepsilon}{1-\varepsilon} \kappa \right)$$

where V_{n-k} = top $(n-k)$ eigenvectors of sparsified Laplacian, C = cluster indicators.

Key Points:

- Bottom k-eigenvectors are aligned to the cluster indicator vectors.
- For graphs with good cluster structure γ is a constant **independent** of the size of the graph.

Proof Idea: Three Key Steps

Step 1: Bound rank $(n-k)$-effective resistances
(Intercluster rank $(n-k)$-effective resistances are small!)

Proof Idea: Three Key Steps

Step 1: Bound rank $(n-k)$-effective resistances
(Intercluster rank $(n-k)$-effective resistances are small!)

Step 2: Uniform sampling $\approx$ Importance sampling
(All edges have similar "importance" wrt rank $(n-k)$-effective resistances)

Proof Idea: Three Key Steps

Step 1: Bound rank $(n-k)$-effective resistances
(Intercluster rank $(n-k)$-effective resistances are small!)

↓

Step 2: Uniform sampling $\approx$ Importance sampling
(All edges have similar "importance" wrt rank $(n-k)$-effective resistances)

↓

Step 3: Matrix Chernoff concentration
(Sample $O(\gamma^2 n \log n)$ edges $\Rightarrow$ preserve clustering eigenspace)

Proof Idea: Three Key Steps

Step 1: Bound rank $(n-k)$-effective resistances
(Intercluster rank $(n-k)$-effective resistances are small!)

↓

Step 2: Uniform sampling $\approx$ Importance sampling
(All edges have similar "importance" wrt rank $(n-k)$-effective resistances)

↓

Step 3: Matrix Chernoff concentration
(Sample $O(\gamma^2 n \log n)$ edges $\Rightarrow$ preserve clustering eigenspace)

Conclusion & Impact

Main Contributions:

- **First provable guarantee** that uniform edge sampling preserves spectral clustering structure
- **New theoretical tools**: rank-$(n-k)$ resistances, structure-aware bounds
- **Practical impact**: Enables scalable spectral clustering without expensive preprocessing

Experiments: In our experiments on network models, we show that uniform sampling performs comparably to effective resistance sampling. In problems with well strong clusters, uniform sampling **outperforms** effective resistance sampling.

See our paper for more details and please visit our poster session!

Poster Session # 4: Dec 4th 4:30PM-7:30PM
Poster ID: 117380

Thank you!