[Spectral] Graph Sparsification

Nikhil Srivastava
Microsoft Research India
Graphs

$G = (V, E, w)$ undirected

$|V| = n$

$w: E \rightarrow R_+$
Sparsification

Approximate any graph $G$ by a sparse graph $H$. 
Sparsification

Approximate any graph $G$ by a sparse graph $H$.

- $H$ is faster to compute with than $G$
- Nontrivial statement about $G$
Cut Approximation [Benczur-Karger’96]

$H$ approximates $G$ if

for every subset $S \subset V$

sum of weights of edges leaving $S$ is preserved

$S \quad \delta S \quad (1 \pm \epsilon) \quad S \quad \delta S$
Example: The Complete Graph

$G = K_n$

$H = \text{random d-regular}$

$|E_G| = O(n^2)$

$|E_H| = O(dn)$
Example: The Complete Graph

$G = K_n$

$H = \text{random d-regular}$

$|E_G| = O(n^2)$

$|E_H| = O(dn)$

$wt_G(\delta S) = |S| \cdot |\overline{S}|$
Example: The Complete Graph

\( G = K_n \)

\( |E_G| = O(n^2) \)

\( \text{wt}_G(\delta S) = |S| \cdot |\overline{S}| \)

\( H = \text{random d-regular} \)

\( |E_H| = O(dn) \)

\( \text{wt}_H(\delta S) \sim (d/n)|S| \cdot |\overline{S}| \)
Example: The Complete Graph

\[ G = K_n \]

\[ \deg H = \text{random d-regular} \]

\[ |E_G| = O(n^2) \]

\[ |E_H| = O(dn) \]

\[ \forall S \subset V, \quad \frac{\omega t_G(\delta S)}{\omega t_H(\delta S)} \approx \left( \frac{n}{d} \right) \]
Example: The Complete Graph

$G = K_n$  \hspace{1cm}  $H = \text{random d-regular } x \ (n/d)$

$|E_G| = O(n^2)$  \hspace{1cm}  $|E_H| = O(dn)$

$\forall S \subset V, \ \frac{wt_G(\delta S)}{wt_{(n/d)}H(\delta S)} \approx 1$
Cut Approximation [Benczur-Karger’96]

$H$ approximates $G$ if for every subset $S \subset V$, sum of weights of edges leaving $S$ is preserved.

[Benczur-Karger’96]: For every $G$ can quickly find $H$ with $O(n \log n / \epsilon^2)$ edges.
Spectral Approximation
The Laplacian Matrix

\[ L_G = D_G - A_G = \sum_{ij \in E} w_{ij} (\delta_i - \delta_j)(\delta_i - \delta_j)^T \]
The Laplacian Matrix

$L_G = D_G - A_G = \sum_{ij \in E} w_{ij} (\delta_i - \delta_j)(\delta_i - \delta_j)^T$

\[
i \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}j\]
The Laplacian Matrix

\[ L_G = D_G - A_G = \sum_{ij \in E} w_{ij} (\delta_i - \delta_j)(\delta_i - \delta_j)^T \]

Quadratic form

\[ x : V \rightarrow \mathbb{R} \]

\[ x^T L_G x = \sum_{ij \in E} w_{ij} (x(i) - x(j))^2 \]
The Laplacian Matrix

\[ L_G = D_G - A_G = \sum_{ij \in E} w_{ij} (\delta_i - \delta_j)(\delta_i - \delta_j)^T \]

Quadratic form

\[ x : V \to \mathbb{R} \]

\[ x^T L_G x = \sum_{ij \in E} w_{ij} (x(i) - x(j))^2 \]

Positive semidefinite

\( \text{Ker}(L_G) = \text{span}(\mathbf{1}) \) if \( G \) is connected
The Laplacian Quadratic Form

An example:
The Laplacian Quadratic Form

An example:

\[ x^T L x = \sum_{i,j} E (x(i) - x(j))^2 \]
The Laplacian Quadratic Form

An example:

$$\mathbf{x}^T \mathbf{L} \mathbf{x} = \sum_{i,j} E \left( \mathbf{x}(i) - \mathbf{x}(j) \right)^2 = 28$$
The Laplacian Quadratic Form

An example:

\[ x^T L x = \sum_{i,j}^E (x(i) - x(j))^2 = 28 \]

Electric Potentials

Energy
The Laplacian Quadratic Form

Another example:
The Laplacian Quadratic Form

Another example:

\[ x^T L_G x = 3 \]
Cuts and the Quadratic Form

For characteristic vector \( x_S \in \{0, 1\}^n \) of \( S \subseteq V \)

\[
x_S^T L_G x_S = \sum_{ij \in E} w_{ij} (x(i) - x(j))^2
\]

\[
= \sum_{ij \in (S, \overline{S})} w_{ij}
\]

\[
= wt_G(S, \overline{S})
\]
Cuts and the Quadratic Form

For characteristic vector \( x_S \in \{0, 1\}^n \) of \( S \subseteq V \)

\[
x_S^T L_G x_S = \sum_{ij \in E} w_{ij} (x(i) - x(j))^2
= \sum_{ij \in (S, \overline{S})} w_{ij}
= wt_G(S, \overline{S})
\]

So BK requires:

\[
(1 - \epsilon)x^T L_G x \leq x^T L_H x \leq (1 + \epsilon)x^T L_G x
\]

\[\forall x \in \{0, 1\}^n\]
Spectral Approximation [ST’04]

For characteristic vector \( x_S \in \{0, 1\}^n \) of \( S \subseteq V \)

\[
x_S^T L_G x_S = \sum_{i,j \in E} w_{ij} (x(i) - x(j))^2
= \sum_{i,j \in (S, \overline{S})} w_{ij}
= wt_G(S, \overline{S})
\]

Spielman-Teng’04:

\[
(1 - \epsilon) x^T L_G x \leq x^T L_H x \leq (1 + \epsilon) x^T L_G x
\forall x \in \mathbb{R}^n \quad \forall x \in \{0, 1\}^n
\]
Spectral Approximation [ST’04]

For characteristic vector $x_S \in \{0, 1\}^n$ of $S \subseteq V$

$$x_S^T L_G x_S = \sum_{i,j \in E} w_{ij} (x(i) - x(j))^2$$

Spielman-Teng’04:

$$\left(1 - \epsilon \right) L_G \preceq L_H \preceq L_G \left(1 + \epsilon \right)$$

“Electrically Equivalent”
Spectral Approximation [ST’04]

Thm[ST’04]

Can find $H$ with $O(n \log^8 n/\epsilon^2)$ edges.

ST’04:

$$(1 - \epsilon)L_G \leq L_H \leq L_G(1 + \epsilon)$$

“Electrically Equivalent”
Why?
1. Solving $L_G x = b$ [ST’04]

$x^T L_G x \sim x^T L_H x$ : can solve systems

$\boxed{L_H^{-1} L_G \approx I}$ in $L_G$ by solving systems in $L_H$. 
1. Solving $L_G x = b$ [ST’04]

$x^T L_G x \sim x^T L_H x$ : can solve systems in $L_G$ by solving systems in $L_H$.

\[
L_H^{-1} L_G \approx I
\]

Naïve: $O(n^3)$

Fast Matrix Multiply: $O(n^{2.37})$

ST’04: $O(m\log^{30} n)$

KMP’10: $O(m\log n)$
1. Solving $L_G x = b$ [ST’04]

$$x^T L_G x \sim x^T L_H x : \text{can solve systems}$$

$$L_H^{-1} L_G \approx I$$

in $L_G$ by solving systems in $L_H$.

Naïve: \quad $O(n^3)$

Fast Matrix Multiply: \quad $O(n^{2.37})$

ST’04 \quad $O(m \log^3 0 n)$

KMP’10 \quad $O(m \log n)$

Thm [ST’04] \quad $\forall G$ can find $H$ with $O(n \log^8 n)$ edges.
2. Spectral Graph Theory

Courant-Fischer Thm: $x^T L_G x$ determines $\lambda_i(L_G)$

$$\lambda_{max}(L) = \max \frac{x^T L x}{x^T x} \quad \lambda_{min}(L) = \min \frac{x^T L x}{x^T x}$$

Thus for spectral sparsifier $H$ of $G$:

$$(1 - \epsilon)\lambda_i(G) \leq \lambda_i(H) \leq (1 + \epsilon)\lambda_i(G)$$

Now $H$ inherits many combinatorial properties:

random walks, colorings, spanning trees, etc.
3. Natural Setting

Spectral formulation more tractable: $x^T L x$ better behaved over $\mathbb{R}^n$ than $\{0,1\}^n$. 
Examples
Example: The Complete Graph

\[ G = K_n \quad H = \text{random d-regular } x \ (n/d) \]

\[ |E_G| = O(n^2) \quad |E_H| = O(dn) \]

\[ \forall S \subset V, \quad \frac{wt_G(\delta S)}{wt_{(n/d)}H(\delta S)} \simeq 1 \pm \epsilon \]
Example: The Complete Graph

\( G = K_n \)  \hspace{1cm} \( H = \text{random } d\text{-regular } x (n/d) \)

\[ |E_G| = O(n^2) \hspace{1cm} |E_H| = O(dn) \]

\[ \forall x, \quad \frac{x^T L_G x}{x^T L_H x} \sim 1 \pm \epsilon \]
Example: The Complete Graph

\( G = K_n \) \hspace{1cm} \( H = \text{random } d\text{-regular } x (n/d) \)

\(|E_G| = O(n^2)\) \hspace{1cm} \(|E_H| = O(dn)\)

\( \forall x, \quad \frac{x^T L_G x}{x^T L_H x} \simeq 1 \pm \epsilon \)

\( d = 1/\epsilon^2 \)
Example: Dumbell

\[ G = G_1 + G_2 + G_3 \]

\[ x^T G x = x^T G_1 x + x^T G_2 x + x^T G_3 x \]
Will show how to do this for every graph...
Part 1: Reduction to Linear Algebra
Original Goal

Given $G$

Find sparse $H$

satisfying $L_G \preceq L_H \preceq \kappa \cdot L_G$
Outer Product Expansion

Recall:

\[ L_G = \sum_{ij \in E} (\delta_i - \delta_j)(\delta_i - \delta_j)^T = \sum_{e \in E} b_e b_e^T. \]
Outer Product Expansion

Recall:

\[ L_G = \sum_{ij \in E} (\delta_i - \delta_j)(\delta_i - \delta_j)^T = \sum_{e \in E} b_e b_e^T. \]

For a weighted subgraph \( H \):

\[ L_H = \sum_{e \in E} s_e b_e b_e^T \]

where \( s_e = \text{wt}(e) \) in \( H \).
Original Goal

Given $G$

Find sparse $H$

satisfying

$$L_G \leq L_H \leq \kappa \cdot L_G$$
Original Goal

Given

\[ L_G = \sum_{e \in G} b_e b_e^T \quad b_{ij} = \delta_i - \delta_j \]

Find sparse

\[ s_e \geq 0 \]

satisfying

\[ L_G \preceq L_H = \sum_{e \in G} s_e b_e b_e^T \leq \kappa \cdot L_G \]
More General Goal

Given

\[ V = \sum_{e} v_e v_e^T \]

Find **sparse**

\[ s_e \geq 0 \]

satisfying

\[ V \leq \sum_{e \in G} s_e v_e v_e^T \leq \kappa \cdot V \]
Equivalent General Goal

Given

\[ I = \sum_{e} v_e v_e^T \]

Find \textbf{sparse}

\[ s_e \geq 0 \]

satisfying

\[ I \preceq \sum_{e \in G} s_e v_e v_e^T \preceq \kappa \cdot I \]
Proof of Equivalence

Given

\[ V = \sum_{e} v_e v_e^T \]

Find sparse \( s_e \geq 0 \)

satisfying

\[ V \leq \sum_{e \in G} s_e v_e v_e^T \leq \kappa \cdot V \]

Take \( v'_e = V^{-1/2} v_e \).
Proof of Equivalence

Given

\[ V = \sum_{e} \nu_e \nu_e^T \]

Find sparse \( s_e \geq 0 \)

satisfying

\[ V \preceq \sum_{e \in G} s_e \nu_e \nu_e^T \]

Take

\[ \nu_e' = V^{-1/2} \nu_e. \]

\[ \sum_{e} s_e \nu_e' \nu_e'^T \approx I \]

\[ \sum_{e} \nu_e \nu_e^T \approx V. \]
Core Problem

**Given**

\[ I = \sum_{e} v_e v_e^T \]

**Find** sparse

\[ s_e \geq 0 \]

satisfying

\[ I \preceq \sum_{e \in G} s_e v_e v_e^T \preceq \kappa \cdot I \]
Core Problem

\[ I = \sum_{e} \mathbf{v}_e \mathbf{v}_e^T \]

\( \mathbf{m} \) vectors in \( \mathbb{R}^n \)
Core Problem

$m$ vectors in $\mathbb{R}^n$

\[ I = \sum_{e} v_e v_e^T \]

\[ \forall u \sum_e \langle u, v_e \rangle^2 = 1 \]

variance is the same in every direction
Core Problem

\[ I = \sum_{e} v_e v_e^T \]

\( m \) vectors in \( \mathbb{R}^n \)

\( O^\sim(n) \) vectors in \( \mathbb{R}^n \)
Core Problem

\[ I = \sum_{e} v_e v_e^T \]

\[ I \leq \sum_{e} s_e v_e v_e^T \leq \kappa I \]
Core Problem

\[ I = \sum_{i \leq m} v_i v_i^T \]

\[ I \preceq \sum_{i} s_e v_i v_i^T \leq \kappa I \]
Examples of the Reduction

Graph

\[ L_G = \sum_e b_e b_e^T \]

\[ I = \sum_e \nu_e \nu_e^T \]

\[ \nu_e = L_G^{-1/2} b_e \]
Examples of the Reduction

Graph

\[ L_G = \sum_e b_e b_e^T \]

\[ I = \sum_e v_e v_e^T \]

\[ v_e = L_G^{-1/2} b_e \]
Examples of the Reduction

Q: Why rescale to identity?
A: All test directions are equally important.

\[ I = \sum_e v_e v_e^T \]

\[ v_e = L_G^{-1/2} b_e \]
Examples of the Reduction

Q: Why rescale to identity?
A: All test directions are equally important.

Rescaling reveals important vectors

\[ I = \sum_e v_e v_e^T \]

\[ v_e = L_G^{-1/2} b_e \]
Part 2: Randomized Solution of LA problem
Part 2: Randomized Solution of LA problem

(implies )
Core Problem

\[
I = \sum_{i \leq m} v_i v_i^T \\
I \leq \sum_i s_v v_i v_i^T \leq \kappa I
\]
Matrix Chernoff Bound [Rudelson’97]

Given \( \sum_i v_i v_i^T = I \)

Sample \( n \log n / \epsilon^2 \) vectors randomly with replacement, by \( p_i \propto \|v_i\|^2 \).

Set \( s_i = 1/p_i \) for chosen vectors.
Matrix Chernoff Bound [Rudelson’97]

Given \[ \sum_i v_i v_i^T = I \]

Sample \[ n \log n/\epsilon^2 \] vectors randomly with replacement, by \[ p_i \propto \|v_i\|^2 \].

Set \[ s_i = 1/p_i \] for chosen vectors.

Rudelson’97: This works whp:

\[ 1 - \epsilon \leq \sum_i s_i v_i v_i^T \leq 1 + \epsilon \]
Effective Resistance View

For a graph $G$, the vectors are $v_e = L_G^{-1/2} b_e$

Lengths of vectors are:

$$\|v_e\|^2 = \|L_G^{-1/2} b_e\|^2 = b_e^T L_G^{-1} b_e = \text{Reff}_G(e)$$
Effective Resistance View

For a graph $G$, the vectors are $v_e = L_G^{-1/2} b_e$

Lengths of vectors are:

$$\|v_e\|^2 = \|L_G^{-1/2} b_e\|^2 = b_e^T L_G^{-1} b_e = \text{Reff}_G(e)$$
Intuition

• Want $G$ and $H$ to be electrically equivalent
• $\text{Reff}(e) = \text{potential difference across } e \text{ when unit current is injected}$
• Edges with higher $\text{Reff}$ are more electrically significant

\[ v_e = L_G^{-1/2} b_e \]
Part 3: Fast Calculation of Sampling Probabilities
Resistances are Distances

Outer product expansion:

\[ L_G = \sum_e b_e b_e^T = B B^T \quad \text{for rows}(B) = \{b_e^T\} \]

Sampling probabilities:

\[
\|\nu_e\|^2 = b_e^T L_G^{-1} b_e \\
= b_e^T L_G^{-1} B B^T L_G^{-1} b_e \\
= \|B L_G^{-1} (\delta_i - \delta_j)\|^2 \quad \text{for } e = ij.
\]
Nearly Linear Time

$$\text{Reff}(ij) = \| BL^{-1}(\delta_i - \delta_j) \|^2$$
Nearly Linear Time

\[ \text{Reff}(ij) = \| BL^{-1} (\delta_i - \delta_j) \|^2 \]

So care about distances between cols. of \( BL^{-1} \)

Diagram of vectors \( BL^{-1} \delta_i \) and \( BL^{-1} \delta_j \)
Nearly Linear Time

\[ \text{Reff}(ij) = \| BL^{-1}(\delta_i - \delta_j) \|^2 \]

So care about distances between cols. of \( BL^{-1} \)
Nearly Linear Time

\[ \text{Reff}(ij) = \| BL^{-1}(\delta_i - \delta_j) \|^2 \]

So care about distances between cols. of \( BL^{-1} \)
Johnson-Lindenstrauss: Take random \( Q_{\log n \times m} \)

Set \( Z = QBL^{-1} \)

\[
\begin{bmatrix}
Q_{(\log n \times m)} & (m \times n) & Z_{(\log n \times n)}
\end{bmatrix}
\]

\[ = \]

\[
\begin{bmatrix}
BL^{-1}
\end{bmatrix}
\]
Nearly Linear Time

\[ \text{Reff}(ij) \sim \| Z(\delta_i - \delta_j) \|^2 \]

\[ R^{\log n} \]
Nearly Linear Time

Find rows of $Z_{\log n \times n}$ by

$Z = QBL^{-1}$

$ZL = QB$

$z_iL = (QB)_i$

$\text{Reff}(ij) \sim \|Z(\delta_i - \delta_j)\|^2$
Nearly Linear Time

Find rows of $Z_{\log n \times n}$ by

$$Z = QBL^{-1}$$

$$ZL = QB$$

$$z_iL = (QB)_i$$

Solve $O(\log n)$ linear systems in $L$ using fast Laplacian solver solver which uses combinatorial $O(n\log^c n)$ sparsifier.

Can show approximate $R_{\text{eff}}$ suffice.
Is this circular?

- Wanted sparsifier to solve $Lx=b$
- Need to solve $Lx=b$ to compute sampling probabilities

- [Koutis-Miller-Peng’10/11]: there is a way to get around this. Gives $O^*(m\log n)$ solver.
Connection to Low Rank Appx.

Norms $\|v_e\|^2$ are leverage scores of rows of $B$

The sampling algorithm is identical to:
[Rudelson-Vershynin’07]
‘low rank’ $k=n$ approximation of $B$
by sampling $O(n \log n)$ rows

But can compute $\|v_e\|^2$ very fast using Laplacian solvers
Can we do better?

Given \( \sum_i v_i v_i^T = I \)

Sample \( n \log n / \epsilon^2 \) vectors randomly with replacement, by \( p_i \propto \|v_i\|^2 \).
Set \( s_i = 1 / p_i \) for chosen vectors.

Rudelson’97: This works whp:

\[
1 - \epsilon \leq \sum_i s_i v_i v_i^T \leq 1 + \epsilon
\]
Deterministic Solution
[Batson-Spielman-S’09]

Spectral Sparsification Theorem:

Given $\sum_{i \leq m} v_i v_i^T = I_n$ there are $s_i \geq 0$ with:

- $(1 - \epsilon)I \leq \sum_i s_i v_i v_i^T \leq (1 + \epsilon)I$
- $\text{supp}(s) \leq 4n/\epsilon^2$. 
Deterministic Solution [BSS’09]

Spectral Sparsification Theorem:

Given \( \sum_{i \leq m} b_e b_e^T = L_G \) there are \( s_e \geq 0 \) with:

- \( (1 - \epsilon)L_G \leq \sum_i s_i v_i v_i^T \leq (1 + \epsilon)L_G \)
- \( \text{supp}(s) \leq 4(n - 1)/\epsilon^2 \).
Deterministic Solution [BSS’09]

Spectral Sparsification Theorem:

Given $\sum_{i \leq m} b_e b_e^T = L_G$ there are $s_e \geq 0$ with:

- $(1 - \epsilon)L_G \preceq \sum_i s_i v_i v_i^T \preceq (1 + \epsilon)L_G$
- $\text{ran}(s) \leq 4(n - 1)/\epsilon^2$.

Open: Fast Algorithm?
Other Connections

• Column Subset Selection
  – [Spielman-S’10] Restricted Invertibility
  – [Boutsidis-Drineas-MagdonIsmail’10]

• Improved Matrix Chernoff Bounds
  – [Vershynin-S’11] Can remove log\(n\) in Rudelson for smooth distributions

• Higher Rank Spectral Sparsification
  – [deCarli-Silva-Harvey-Sato’11]