Algorithms for Big Data (XIV)

Chihao Zhang

Shanghai Jiao Tong University

Dec. 20, 2019

Review

Review

Last week we studied electrical networks using matrices.

Review

Last week we studied electrical networks using matrices.

We defined the graph Laplacian L:

$$L = U^{\mathsf{T}} W U.$$

Review

Last week we studied electrical networks using matrices.

We defined the graph Laplacian L:

$$L = U^{\mathsf{T}} W U.$$

We also defined the notion of effective resistance between two vertices in terms of L:

$$R_{\mathrm{eff}}(u, v) \triangleq (\mathbf{e}_u - \mathbf{e}_v)^{\mathsf{T}} L^{+} (\mathbf{e}_u - \mathbf{e}_v).$$

Sparsification

Sparsification

Given a graph G, the goal of sparsification is to construct a sparse graph H such that

$$(1-\varepsilon)L_G \preccurlyeq L_H \preccurlyeq (1+\varepsilon)L_G.$$

Sparsification

Given a graph G, the goal of sparsification is to construct a sparse graph H such that

$$(1-\varepsilon)L_G \preccurlyeq L_H \preccurlyeq (1+\varepsilon)L_G.$$

Similar Laplacian implies

- similar spectrum;
- similar effective resistance between any two vertices;
- similar clustering;
- ...

The Construction

The Construction

We use $L_{u,v}$ to denote the Laplacian of the unweighted graph containing a single edge $\{u, v\}$.

The Construction

We use $L_{u,v}$ to denote the Laplacian of the unweighted graph containing a single edge $\{u, v\}$.

For a graph $G = (V, E)$, we have

$$L_G = \sum_{\{u,v\} \in E} w_{u,v} \cdot L_{u,v},$$

where $w_{u,v}$ is the weight on the edge $\{u, v\} \in E$.

The Construction

We use $L_{u,v}$ to denote the Laplacian of the unweighted graph containing a single edge $\{u, v\}$.

For a graph $G = (V, E)$, we have

$$L_G = \sum_{\{u,v\} \in E} w_{u,v} \cdot L_{u,v},$$

where $w_{u,v}$ is the weight on the edge $\{u, v\} \in E$.

Let $\{p_{u,v}\}_{\{u,v\} \in E}$ be a collection of probabilities on each pair of vertices.

Let $H = (V, E_H)$ be the sparse graph we are going to construct...

Let $H = (V, E_H)$ be the sparse graph we are going to construct...

H contains the edge $\{u, v\}$ with probability $p_{u,v}$ for every pair $\{u, v\}$ independently.

Let $H = (V, E_H)$ be the sparse graph we are going to construct...

H contains the edge $\{u, v\}$ with probability $p_{u,v}$ for every pair $\{u, v\}$ independently.

If an edge $\{u, v\} \in E_H$, we assign it with weight $w_{u,v}/p_{u,v}$.

Let $H = (V, E_H)$ be the sparse graph we are going to construct...

H contains the edge $\{u, v\}$ with probability $p_{u,v}$ for every pair $\{u, v\}$ independently.

If an edge $\{u, v\} \in E_H$, we assign it with weight $w_{u,v}/p_{u,v}$.

It is easy to verify that

$$\mathbf{E}[L_H] = L_G.$$

Let $H = (V, E_H)$ be the sparse graph we are going to construct...

H contains the edge $\{u, v\}$ with probability $p_{u,v}$ for every pair $\{u, v\}$ independently.

If an edge $\{u, v\} \in E_H$, we assign it with weight $w_{u,v}/p_{u,v}$.

It is easy to verify that

$$\mathbf{E}[L_H] = L_G.$$

We will carefully choose $\{p_{u,v}\}$ to guarantee that

Let $H = (V, E_H)$ be the sparse graph we are going to construct...

H contains the edge $\{u, v\}$ with probability $p_{u,v}$ for every pair $\{u, v\}$ independently.

If an edge $\{u, v\} \in E_H$, we assign it with weight $w_{u,v}/p_{u,v}$.

It is easy to verify that

$$\mathbf{E}\left[L_H\right] = L_G.$$

We will carefully choose $\{p_{u,v}\}$ to guarantee that

- H is sparse with high probability;
- L_H is well-concentrated to its expectation.

A Transformation

A Transformation

Sometimes it is more convenient to work with L_G^+, the pseudo-inverse of L_G.

A Transformation

Sometimes it is more convenient to work with L_G^+, the pseudo-inverse of L_G.

Note that

$$L_H \preccurlyeq (1+\varepsilon)L_G \iff L_G^{+/2} L_H L_G^{+/2} \preccurlyeq (1+\varepsilon) L_G^{+/2} L_G L_G^{+/2}.$$

A Transformation

Sometimes it is more convenient to work with L_G^+, the pseudo-inverse of L_G.

Note that

$$L_H \preccurlyeq (1+\varepsilon)L_G \iff L_G^{+/2} L_H L_G^{+/2} \preccurlyeq (1+\varepsilon) L_G^{+/2} L_G L_G^{+/2}.$$

The matrix $L_G^{+/2} L_G L_G^{+/2}$ is the projection onto the column space of L_G.

A Transformation

Sometimes it is more convenient to work with L_G^+, the pseudo-inverse of L_G.

Note that

$$L_H \preccurlyeq (1+\varepsilon)L_G \iff L_G^{+/2} L_H L_G^{+/2} \preccurlyeq (1+\varepsilon) L_G^{+/2} L_G L_G^{+/2}.$$

The matrix $L_G^{+/2} L_G L_G^{+/2}$ is the projection onto the column space of L_G.

We will now study $L_G^{+/2} L_H L_G^{+/2}$.

Chernoff Bound for Matrices

Chernoff Bound for Matrices

The main tool to establish concentration is the following analogue of Chernoff bound for matrices.

Chernoff Bound for Matrices

The main tool to establish concentration is the following analogue of Chernoff bound for matrices.

Theorem

Let $X_1, \ldots, X_n \in \mathbb{R}^{n \times n}$ be independent random positive semi-definite matrices such that $\lambda_{max}(X_i) \le R$ almost surely. Let $X = \sum_{i=1}^{n} X_i$. Let μ_{min} and μ_{max} be the minimum and maximum eigenvalues of $\mathbf{E}[X]$ respectively. Then

- $\mathbf{Pr}\left[\lambda_{min}(X) \le (1-\varepsilon)\mu_{min}\right] \le n \left(\frac{e^{-\varepsilon}}{(1-\varepsilon)^{1-\varepsilon}}\right)^{\mu_{min}/R}$, for $0 < \varepsilon < 1$, and
- $\mathbf{Pr}\left[\lambda_{max}(X) \ge (1+\varepsilon)\mu_{max}\right] \le n \left(\frac{e^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\right)^{\mu_{max}/R}$, for $\varepsilon > 0$.

SETTING $p_{u,v}$

Setting $p_{u,v}$

For every pair of vertices u and v, we define

$$p_{u,v} \triangleq \frac{1}{R} w_{u,v} \| L_G^{+/2} L_{u,v} L_G^{+/2} \|.$$

Setting $p_{u,v}$

For every pair of vertices u and v, we define

$$p_{u,v} \triangleq \frac{1}{R} w_{u,v} \| L_G^{+/2} L_{u,v} L_G^{+/2} \|.$$

Following our construction of H, for every $\{u, v\}$, define a random variable

$$X_{u,v} = \begin{cases} (w_{u,v}/p_{u,v}) L_G^{+/2} L_{u,v} L_G^{+/2}, & \text{w.p. } p_{u,v} \\ 0, & \text{otherwise.} \end{cases}$$

Setting $p_{u,v}$

For every pair of vertices u and v, we define

$$p_{u,v} \triangleq \frac{1}{R} w_{u,v} \| L_G^{+/2} L_{u,v} L_G^{+/2} \|.$$

Following our construction of H, for every $\{u, v\}$, define a random variable

$$X_{u,v} = \begin{cases} (w_{u,v}/p_{u,v}) L_G^{+/2} L_{u,v} L_G^{+/2}, & \text{w.p. } p_{u,v} \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$L_G^{+/2} L_H L_G^{+/2} = \sum_{\{u,v\} \in E} X_{u,v}, \text{ and}$$

Setting $p_{u,v}$

For every pair of vertices u and v, we define

$$p_{u,v} \triangleq \frac{1}{R} w_{u,v} \| L_G^{+/2} L_{u,v} L_G^{+/2} \|.$$

Following our construction of H, for every $\{u, v\}$, define a random variable

$$X_{u,v} = \begin{cases} (w_{u,v}/p_{u,v}) L_G^{+/2} L_{u,v} L_G^{+/2}, & \text{w.p. } p_{u,v} \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$L_G^{+/2} L_H L_G^{+/2} = \sum_{\{u,v\} \in E} X_{u,v}, \text{ and}$$

$$\lambda_{\max}\left(X_{u,v}\right) \le R.$$

Relation to Resistance

Relation to Resistance

It remains to compute $p_{u,v}$.

Relation to Resistance

It remains to compute $p_{u,v}$.

It is easy to verify that

$$L_G^{+/2} L_{u,v} L_G^{+/2} = L_G^{+/2} (\mathbf{e}_u - \mathbf{e}_v)(\mathbf{e}_u - \mathbf{e}_v)^T L_G^{+/2}$$

is a rank-1 matrix.

Relation to Resistance

It remains to compute $p_{u,v}$.

It is easy to verify that

$$L_G^{+/2} L_{u,v} L_G^{+/2} = L_G^{+/2} (\mathbf{e}_u - \mathbf{e}_v)(\mathbf{e}_u - \mathbf{e}_v)^T L_G^{+/2}$$

is a rank-1 matrix.

Therefore

$$\|L_G^{+/2} L_{u,v} L_G^{+/2}\| = \mathrm{Tr}(L_G^{+/2} L_{u,v} L_G^{+/2}) = (\mathbf{e}_u - \mathbf{e}_v)^T L_G^{+} (\mathbf{e}_u - \mathbf{e}_v) = R_{\mathrm{eff}}(u, v).$$

Relation to Resistance

It remains to compute $p_{u,v}$.

It is easy to verify that

$$L_G^{+/2} L_{u,v} L_G^{+/2} = L_G^{+/2} (\mathbf{e}_u - \mathbf{e}_v)(\mathbf{e}_u - \mathbf{e}_v)^T L_G^{+/2}$$

is a rank-1 matrix.

Therefore

$$\|L_G^{+/2} L_{u,v} L_G^{+/2}\| = \mathrm{Tr}(L_G^{+/2} L_{u,v} L_G^{+/2}) = (\mathbf{e}_u - \mathbf{e}_v)^T L_G^{+} (\mathbf{e}_u - \mathbf{e}_v) = R_{\mathrm{eff}}(u, v).$$

We can then use the algorithm learnt in the last lecture to approximate $R_{\mathrm{eff}}(u, v)$.

Analysis

Analysis

We now compute $\mathbf{E}\left[|E_H|\right]$.

Analysis

We now compute $\mathbf{E}\left[|E_H|\right]$. It holds that

$$\mathbf{E}\left[|E_H|\right] = \sum_{\{u,v\}\in E} p_{u,v} = \frac{\sum_{\{u,v\}\in E} w_{u,v} \cdot R_{\mathrm{eff}}(u,v)}{R}.$$

Analysis

We now compute $\mathbf{E}[|E_H|]$. It holds that

$$\mathbf{E}[|E_H|] = \sum_{\{u,v\}\in E} p_{u,v} = \frac{\sum_{\{u,v\}\in E} w_{u,v} \cdot R_{\text{eff}}(u,v)}{R}.$$

We can also directly compute

$$\begin{aligned}
\sum_{\{u,v\}\in E} w_{u,v} R_{\text{eff}}(u,v) &= \sum_{\{u,v\}\in E} w_{u,v} (\mathbf{e}_u - \mathbf{e}_v)^T L_G^{+} (\mathbf{e}_u - \mathbf{e}_v) \\
&= \sum_{\{u,v\}\in E} w_{u,v} \mathrm{Tr}(L_G^{+} (\mathbf{e}_u - \mathbf{e}_v)(\mathbf{e}_u - \mathbf{e}_v)^T) \\
&= \mathrm{Tr}\left(\sum_{\{u,v\}\in E} L_G^{+} w_{u,v} (\mathbf{e}_u - \mathbf{e}_v)(\mathbf{e}_u - \mathbf{e}_v)^T \right) \\
&= \mathrm{Tr}\left(L_G^{+} L_G\right) = n - 1.
\end{aligned}$$

Therefore, $\mathbf{E}[|E_H|] = \frac{n-1}{R}$.

Therefore, $\mathbf{E}[|E_H|] = \frac{n-1}{R}$.

Note that $\mathbf{E}[|E_H|]$ is the sum of m independent Bernoulli trials, therefore, for suitable R, we can control its concentration using the standard Chernoff bound.

Therefore, $\mathbf{E}[|E_H|] = \frac{n-1}{R}$.

Note that $\mathbf{E}[|E_H|]$ is the sum of m independent Bernoulli trials, therefore, for suitable R, we can control its concentration using the standard Chernoff bound.

We choose $R = \frac{\varepsilon^2}{3.5 \log n}$, then $|E_H| \leq 4\varepsilon^{-2} n \log n$ with high probability.

Therefore, $\mathbf{E}[|E_H|] = \frac{n-1}{R}$.

Note that $\mathbf{E}[|E_H|]$ is the sum of m independent Bernoulli trials, therefore, for suitable R, we can control its concentration using the standard Chernoff bound.

We choose $R = \frac{\varepsilon^2}{3.5 \log n}$, then $|E_H| \leq 4\varepsilon^{-2} n \log n$ with high probability.

Now we can apply Matrix Chernoff bound to obtain the concentration bound needed.

$p_{u,v} > 1$?

$p_{u,v} > 1$?