University of Edinburgh

INFR11156: Algorithmic Foundations of Data Science (2019)

Lecture 15: Spectral Sparsification of Graphs (2)

In the last lecture we discussed effective resistance, which is the potential difference between any pair of vertices when a unit current is injected at one vertex and extracted at the other. To derive an algebraic expression for the effective resistance between vertices u and v, we introduced the pseudo-inverse of the Laplacian matrix $L^{\dagger}$, and proved that

$$\mathrm{Reff}(u,v) \triangleq (\delta_u - \delta_v)^{\intercal} L^{\dagger} (\delta_u - \delta_v),$$

where $\delta_u \in \{0,1\}^n$ is the indicator vector of vertex u. We also briefly discussed that, for any graph G with n vertices and m edges, the effective resistances of all the edges can be approximately computed up to a constant factor in $O(m \log^c n)$ time for some constant $c > 0$. In this lecture we will see how effective resistances will be applied to construct spectral sparsifiers. Recall that, for any undirected graph G with n vertices, we call a subgraph H of G, with proper reweighting of the edges, a $(1+\varepsilon)$-spectral sparsifier if

$$(1-\varepsilon) x^{\intercal} L_G x \leq x^{\intercal} L_H x \leq (1+\varepsilon) x^{\intercal} L_G x \tag{1}$$

holds for any $x \in \mathbb{R}^n$, where L_G and L_H are the respective Laplacian matrices of G and H. The main result of this lecture is as follows:

Theorem 1. *Let G be any undirected graph with n vertices and m edges. For any $\varepsilon \in (0,1)$, there is an algorithm that runs in $O(m \log^c n)$ time for some constant $c > 0$ and produces a $(1+\varepsilon)$-spectral sparsifier of G with $O(n \log n / \varepsilon^2)$ edges.*

The algorithm behind Theorem 1 is a very simple procedure that samples edges with probability proportional to their leverage scores, which is defined as $\ell_e \triangleq w_e \cdot \mathrm{Reff}(e)$ for every edge e. Our algorithm is described in Algorithm 1.

Algorithm 1 Algorithm for constructing a spectral sparsifier

---

for every edge e **do**
    let $\ell_e = w_e \cdot \mathrm{Reff}(e)$
    let $p_e = \min\{1, 5 \cdot (\log n) \cdot \ell_e / \varepsilon^2\}$
end for
$H = (V, \emptyset)$
for every edge e **do**
    with probability p_e, add e into H with weight w_e/p_e
end for
return graph H

Notice that in the algorithm every sampled edge e has a new weight w_e/p_e in the resulting graph H, so we first explain the reweighting scheme. Let

$$L_e = b_e b_e^{\intercal}$$

be the Laplacian matrix of the graph consisting of a single edge $e = \{u, v\}$, where

$$b_e \triangleq \delta_u - \delta_v.$$

Then, the Laplacian matrix of graph G can be written as $L_G = \sum_e w_e \cdot L_e$. Since every edge e is sampled with probability p_e and every sampled e has weight w_e/p_e in H, we have

$$\mathbb{E}\left[L_H\right] = \sum_e p_e \cdot \frac{w_e}{p_e} \cdot L_e = \sum_e w_e \cdot L_e = L_G.$$

Hence, the reweighing scheme ensures that the resulting graph H equals to G in expectation. It's also easy to see that, once the algorithm obtains the approximate values of edges' effective resistances that can be computed in $O(m \log^c n)$ time for some $c > 0$, the overall algorithm finishes in $O(m)$ time, which proves the runtime requirement of Theorem 1. Hence, in order to prove Theorem 1, it remains to analyse the number of edges in H, and (1) holds for H.

During the analysis below, we assume $p_e = 5 \cdot (\log n) \cdot \ell_e/\varepsilon^2 \leq 1$ to simplify our analysis. The reason behind this assumption is due to the following trick: for any edge e with $5 \cdot (\log n) \cdot \ell_e/\varepsilon^2 > 1$, we treat e as multiple parallel edges e', each of which is sampled with some probability $p_{e'} < 1$ satisfying $\sum_{e'} p_{e'} = 5 \cdot (\log n) \cdot \ell_e/\varepsilon^2$.

Bounding the number of edges in H. Since we sample edges with probability proportional to their leverage scores, we have

$$\begin{aligned}
\sum_e \ell_e &= \sum_e w_e \mathrm{Reff}(e) \\
&= \sum_e w_e \cdot b_e^\intercal L^\dagger b_e \\
&= \sum_e w_e \cdot \mathrm{tr}\left(L^\dagger b_e b_e^\intercal\right) \\
&= \mathrm{tr}\left(L^\dagger \sum_e w_e b_e b_e^\intercal\right) \\
&= \mathrm{tr}\left(L^\dagger L\right) \\
&= n - 1,
\end{aligned} \tag{2}$$

where the third equality follows by the fact that the trace is invariant under cyclic permutations. The fact $\sum_e \ell_e = n - 1$ can be also explained in a combinatorial way. Notice that ℓ_e is the probability that e appears in a random spanning tree of G when we sample spanning trees with probability proportional to the product of their edge weights. Since every spanning tree has $n - 1$ edges, the sum of these probabilities is $n - 1$.

Based on (2), the expected number of edges in H equals

$$\sum_e p_e = \sum_e \min\left\{1, 5 \cdot (\log n)\ell_e/\varepsilon^2\right\} \leq \sum_e 5 \cdot (\log n)\ell_e/\varepsilon^2 \leq 5n \log n/\varepsilon^2.$$

By Chernoff bound, it is exponentially unlikely that the number of edges in H is more than a small multiplicity factor of the expected value.

Proving that H is a spectral sparsifier. Our analysis is based on the fact that, for any positive definite matrices A and B, $A \preceq B$ iff

$$B^{-1/2} A B^{-1/2} \preceq I.$$

Similarly, $L_H \preceq L_G$ iff

$$L_G^{\dagger/2} L_H L_G^{\dagger/2} \preceq L_G^{\dagger/2} L_G L_G^{\dagger/2},$$

where $L_G^{\dagger/2}$ is the square root of the pseudo-inverse of L_G. Let

$$\Pi = L_G^{\dagger/2} L_G L_G^{\dagger/2}$$

be the projection onto the range of L_G. Then, by linearity of expectation it holds that

$$\mathbb{E}\left[L_G^{\dagger/2} L_H L_G^{\dagger/2}\right] = L_G^{\dagger/2} \mathbb{E}\left[L_H\right] L_G^{\dagger/2} = L_G^{\dagger/2} L_G L_G^{\dagger/2} = \Pi.$$

We define a random matrix X_e, where

$$X_e = \begin{cases} \frac{w_e}{p_e} \cdot L_G^{\dagger/2} L_e L_G^{\dagger/2} & \text{with probability } p_e \\ 0 & \text{otherwise.} \end{cases}$$

Hence, it holds that

$$\sum_e X_e = L_G^{\dagger/2} L_H L_G^{\dagger/2},$$

and

$$\mathbb{E}\left[\sum_e X_e\right] = \Pi.$$

Notice that proving H is a spectral sparsifier is equivalent to show $\sum_e X_e$ is close to Π with high probability. Our analysis is based on the theorem about the concentration of random matrices, which can be viewed as matrix analog of the Chernoff bound that we saw in Lecture 9.

Lemma 2 (Matrix Chernoff Bound)**.** *Let $X_1, \ldots, X_m \in \mathbb{R}^{n \times n}$ be independent random PSD such that $\lambda_{\max}(X_i) \leq R$ almost surely. Let $X = \sum_{i=1}^m X_i$, and let $\mu_{\min}$ and $\mu_{\max}$ be the minimum and maximum eigenvalues of $\mathbb{E}[X] = \sum_{i=1}^m \mathbb{E}[X_i]$. Then,*

$$\mathbb{P}\left[\lambda_{\min}\left(\sum_{i=1}^m X_i\right) \leq (1-\varepsilon)\mu_{\min}\right] \leq n \left(\frac{\mathrm{e}^{-\varepsilon}}{(1-\varepsilon)^{1-\varepsilon}}\right)^{\mu_{\min}/R}$$

for $0 < \varepsilon < 1$, and

$$\mathbb{P}\left[\lambda_{\max}\left(\sum_{i=1}^m X_i\right) \geq (1+\varepsilon)\mu_{\max}\right] \leq n \left(\frac{\mathrm{e}^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\right)^{\mu_{\max}/R}$$

for $\varepsilon > 0$.

To apply Lemma 2, notice that

$$\begin{aligned} \lambda_{\max}(X_e) &= (w_e/p_e) \cdot \lambda_{\max}\left(L_G^{\dagger/2} L_e L_G^{\dagger/2}\right) \\ &= (w_e/p_e) \cdot \lambda_{\max}\left(L_G^{\dagger/2} b_e b_e^\intercal L_G^{\dagger/2}\right) \\ &\leq (w_e/p_e) \cdot \operatorname{tr}\left(L_G^{\dagger/2} b_e b_e^\intercal L_G^{\dagger/2}\right) \\ &= (w_e/p_e) \cdot \operatorname{tr}\left(b_e^\intercal L_G^{\dagger} b_e\right) \\ &= (w_e/p_e) \cdot \mathrm{Reff}(e) \\ &\leq \varepsilon^2/(5 \cdot \log n), \end{aligned}$$

where the last inequality follows by the definition and our assumption on p_e. Notice that the upper bound above is independent of edge e. Finally, we set $R = \varepsilon^2/(5 \cdot \log n)$ and apply Lemma 2 to obtain that

$$\mathbb{P}\left[\lambda_{\max}\left(\sum_e X_e\right) \geq (1+\varepsilon)\right] \leq n\left(\frac{\mathrm{e}^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\right)^{1/R} \leq n\left(\frac{\mathrm{e}^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\right)^{5\log n/\varepsilon^2} \leq \frac{1}{n^{1.5}}.$$

For the lower bound, remember that we only need to work orthogonal to $\mathbf{1} \in \mathbb{R}^n$, and hence can treat $\lambda_{\min}(\Pi) = 1$. This gives us that

$$\mathbb{P}\left[\lambda_{\min}\left(\sum_e X_e\right) \leq (1-\varepsilon)\right] \leq n\left(\frac{\mathrm{e}^{-\varepsilon}}{(1-\varepsilon)^{1-\varepsilon}}\right)^{5\log n/\varepsilon^2} \leq \frac{1}{n^{1.5}}.$$

By the union bound, we know that with high probability $\sum_e X_e$ is close to Π, which proves (1).