Notes on Spectural Sparcification of Graphs

PkGu

12/17/2021

Main References

[1]. [arxiv: 0803.0929] Graph Sparsification by Effective Resistance

[2]. [arxiv: 0808.0163] Twice Ramanujan Sparsifiers

Auxiliary References

[3]. [Conbinatorica] Ramanujan Graphs 1988

[4]. [arxiv: cs/0607105]

The general setting

This notes focus on some important advances in the domain about spectral sparsification of graphs. This kind of problem is closely related to the sparsification of matrix, that is, finding a sparse matrix to approximate a given matrix without deviating too much under the operator norm.

Suppose we're given a weighted graph $G = (V, E, w)$ where $|V| = n$ and $|E| = m$, we are seeking for a weighted sparse subgraph $H = (\tilde{V}, \tilde{E}, \tilde{w})$ of G, of which the spectrum of Laplacian matrix $Spec(H)$ is is restricted in $(1 \pm \epsilon)Spec(G)$, to be explicit, for any $x \in \mathbb{R}^V$, $x^\top L_H x = \sum_{u\sim v}(x_u - x_v)^2 \tilde{w}_{u\sim v} \in ((1-\epsilon)x^\top L_G x, (1+\epsilon)x^\top L_G x)$; and the edges are sparse enough, which means, dominated by terms about n and ϵ.

Expanders and Ramanujan graphs

The earliest solved case is the for the complete graphs: It's easy to prove that $Spec(K_n) = \{0, n, ..., n\}$. We're enlighten to approximate a reweighted complete graph $\frac{d}{n}K_n$ by a d-regular graph.

An ϵ-expander is defined to be a d-regular graph H whose non-zero eigenvalues of Laplacian $\{\lambda_i\}$ are close to d: $|\lambda_i - d| < \epsilon d$, equivalently, whose smallest $(n-1)$ eigenvalues $\{\mu_i\}$ are close to 0: $|\mu_i| < \epsilon d$. One intuitive property explains why they're called "expander" is that for a small set S of vertices, it have a uniformly large neighborhood relative to $|S|$.

In [3], a special kind of expanders, called "Ramanujan Graphs", are constructed, whose $\epsilon <= \frac{2\sqrt{d-1}}{d}$ for arbitrarily large n. In this article they considered the Cayley graph of projective linear groups over $\mathbb{F}_q$ acted by a special subset S with $(p+1)$ elements where p and q are prime numbers which $\equiv 1 \mod 4$, and proved that they are Ramanujan graphs.

The Ramanujan graph has approximation accuracy

$$Spec(H) \cdot \frac{d - 2\sqrt{d-1}}{d} \leq Spec(\frac{d}{n} K_n) \leq Spec(H) \cdot \frac{d + 2\sqrt{d-1}}{d}$$

and as an expander, it has $(n-1)d/2$ edges.

Algorithm based on effective resistance

In [1], the authors provided an randomized algorithm for sparsifying common graphs, by sampling according to effective resistance between each edge. Their result is that, a sparsification of precision ratio $(1 \pm \epsilon)$ and total number of edges $O(n \log n/\epsilon^2)$, can be obtained in nearly linear running time.

For a weighted graph with $|V| = n$ and $|E| = m$, the Laplacian an be decomposed as $L = B^\top W B$, where $B \in \mathbb{R}^{m \times n}$ with $B_{ev} = \mathbf{1}_{v=e(head)} - \mathbf{1}_{v=e(tail)}$, and $W \in \mathbb{R}^{m \times m}$ is the diagonal matrix where $W_{ee} = w_e$ represents the weight. The introducing of B gives a orientation on graph.

Here we identify the graph with a electric network, we can introduce the conception "effective resistance": if we let I_{ext} represent the current injected on vertices and I represent the current on oriented edges, U for electric potentials, and the weights w represent the conductance. By the electrical knowledge, we have relations

$$I_{ext} = (B^\top W B)U = LU$$

The effective resistance between an edge is defined to be the amount of potential difference after the implement of a unit current from one vertex to another. if we define the Moore-Penrose pseudoinverse of L as L^+,

$$R_e = (BL^+B^\top)_{e,e}$$

The algorithm states as follows: Given graph $G = (V, E)$ and a parameter q, do q independent sampling from E with probability p_e proportional to $w_e R_e$ and its weight is set to be w_e/qp_e. The weights shall be added if an edge is picked for more than one times.

The authors proved in Section 3 that when q sufficiently large (according to n and ϵ), with high probability, the new graph is $(1 \pm \epsilon)$ close to the original graph. They defined a regularized version of $BL^+B^\top$:

$$\Pi = W^{1/2} B L^+ B^\top W^{1/2}$$

and showed that Π is a projection matrix onto $Im(W^{1/2}B)$ and transferred the problem to showing that if the diagonal random matrix S is denoted to

$$S_{ee} = \frac{\tilde{w}_e}{w_e}$$

with high probability there is

$$\|\Pi S \Pi - \Pi\|_2 \geq \epsilon$$

And in Section 4, they provided a nearly-linear time algorithm for computing the resistance: in fact, any resistance can be inferred from

$$\begin{aligned} R_e &= b_e^\top L^+ b_e \\ &= \|W^{1/2} B L^+ \chi_{e_{head}} - W^{1/2} B L^+ \chi_{e_{tail}}\|_2^2 \end{aligned}$$

and by Johnson-Lindenstrauss Lemma, it's probable to ramdomly project the m-dimensional data onto some smaller space with tiny loss of distances. Set the scaled projection matrix as Q, we only need to approximately compute the $k \times n$ matrix

$$\widetilde{Z} \approx Z = QW^{1/2}BL^{+}$$

which we can refer to [4].

Twice Ramanujan sparsifiers

This algorithm is newly presented in 2018, through which we can find a sparse substitution of any given graph $G = (V, E)$ where $|V| = n$, with at most $d(n-1)$ edges (which is bounded at the doubled bound Ramanujan graphs) and the accuracy is

$$Spec(H) \leq Spec(G) \leq Spec(H) \cdot \frac{d + 2\sqrt{d-1}}{d - 2\sqrt{d-1}}$$

This algorithm is deterministic rather than random, and it removed the "$\log n$" term of edge number in the algorithm before, so it reached a huge improvement.

Noticing that $L = (B^\top W^{1/2})(B^\top W^{1/2})^\top$, though L is not invertible, if we focus only on $im(L)$, the original problem can be reduced to the following matrix version: Given an $m \times n$ matrix V where $VV^\top = \mathbf{I}$, there exists an $m \times m$ diagonal matrix S with at most dn nonzero elements so that

$$\mathbf{I} \leq VSV^\top \leq \frac{d + 2\sqrt{d-1}}{d - 2\sqrt{d-1}} \cdot \mathbf{I}$$

The proof of the matrix version problem is quite technical, and also because I'm too tired to type, so I omit it here.