CMSE 890-001: Spectral Graph Theory and Related Topics, MSU, Spring 2021

Lecture 24: Graph Sparsification via Random Sampling, Part I

April 13, 2021

Lecturer: Matthew Hirn

30 Graph sparsification via random sampling

30.1 Introduction

Recall that $H = (V, E_H, w_H)$ is an ϵ-approximation of a graph $G = (V, E, w)$ if

$$(1-\epsilon)\boldsymbol{L}_G \preccurlyeq \boldsymbol{L}_H \preccurlyeq (1+\epsilon)\boldsymbol{L}_G\,.$$

In Section 29, Theorem 56, we learned that (d, ϵ)-expander graphs are ϵ-approximations of the re-scaled complete graph $(d/n)K_n$. We also proved in Theorem 59 that

$$\epsilon \geq \frac{2\sqrt{d-1}}{d} - \frac{2\sqrt{d-1}-1}{d(\mathrm{diam}(G)/2-1)}\,,$$

and we observed that the Ramanujan graphs, which are a special type of expander graph, satisfy

$$\epsilon \leq \frac{2\sqrt{d-1}}{d}\,,$$

meaning they are very good approximations of K_n, particularly when n is large. Recall that a (d, ϵ)-expander graph is a d-regular graph and so it has $dn/2$ edges. We can calculate the number of edges in a Ramanujan graph in terms of its approximation error ϵ. Indeed, by the above two inequalities we can take:

$$\epsilon = \frac{2\sqrt{d-1}}{d} \quad \Longrightarrow \quad \epsilon^2 = \frac{4(d-1)}{d^2} = \frac{4}{d} - \frac{4}{d^2} \leq \frac{4}{d} \quad \Longrightarrow \quad \epsilon^{-2} \geq \frac{d}{4} \quad \Longrightarrow \quad d = O(\epsilon^{-2})\,.$$

Therefore, the number of edges in a Ramanujan graph that ϵ approximates $(d/n)K_n$ is equal to

$$\frac{dn}{2} = O(\epsilon^{-2}n)\,.$$

In this section we will prove that any connected graph G, not just the complete graph, can be well approximated by a sparse graph. We will see that by using a careful random sampling of the edges in G, we can obtain an ϵ approximation of G with only $O(\epsilon^{-2}n\log n)$ edges, which is nearly as sparse as the Ramanujan graph (but, again, here we approximate any G, not just K_n). This is a very powerful statement, as it allows us to approximate any

graph with a sparse graph through a relatively simple algorithm. As an example of this power, remember that if H is an ϵ-approximation of G, then by Theorem 15 we have

$$(1-\epsilon)\lambda_i(G) \leq \lambda_i(H) \leq (1+\epsilon)\lambda_i(G)\,, \quad 1 \leq i \leq n\,.$$

Additionally, for all sets $S \subset V$, the number of edges in the boundary of S with respect to H is similar to the number of edges in the boundary of S with respect to G since, recalling a calculation we performed in the proof of Theorem 39, we have:

$$(1-\epsilon)|\partial_G S| = (1-\epsilon)\mathbf{1}_S^T \boldsymbol{L}_G \mathbf{1}_S \leq |\partial_H S| = \mathbf{1}_S^T \boldsymbol{L}_H \mathbf{1}_S \leq (1+\epsilon)\mathbf{1}_S^T \boldsymbol{L}_G \mathbf{1}_S = (1+\epsilon)|\partial_G S|\,.$$

There are other properties for which G and H will be similar as well.

30.2 Overview of the algorithm

Now for the random sampling algorithm that will generate the graph $H = (V, E_H, w_H)$ from the graph $G = (V, E, w)$:

1. Carefully choose a probability $p_{a,b}$ for each edge $(a,b) \in E$, where $p_{a,b}$ gives the probability that we keep the edge in H.

2. Randomly sample the edges from G to include in H using the probabilities $p_{a,b}$.

3. If (a,b) is included in H, we give it weight

$$w_H(a,b) := \frac{w(a,b)}{p_{a,b}}\,.$$

We will show that with properly chosen probabilities $p_{a,b}$, the resulting graph H will have $|E_H| = C\epsilon^{-2} n \log n$ edges with high probability, and H will be an ϵ-approximation of G also with high probability.

As some initial intuition, the reason for re-weighting the edges is that it will imply the expected matrix of $\boldsymbol{L}_H$ will be equal to $\boldsymbol{L}_G$. To see this, recall that

$$\boldsymbol{L}_G = \sum_{(a,b)\in E} w(a,b)\boldsymbol{L}_{a,b}\,.$$

On the other hand,

$$\boldsymbol{L}_H = \sum_{(a,b)\in E} W_H(a,b)\boldsymbol{L}_{a,b}\,, \tag{73}$$

where $W_H(a,b)$ is the random variable defined as

$$\mathbb{P}\left(W_H(a,b) = \frac{w(a,b)}{p_{a,b}}\right) = p_{a,b} \quad \text{and} \quad \mathbb{P}(W_H(a,b) = 0) = 1 - p_{a,b}\,. \tag{74}$$

It follows that the expected value of $W_H(a,b)$ is:

$$\begin{aligned}\mathbb{E}[W_H(a,b)] &= \mathbb{P}\left(W_H(a,b) = \frac{w(a,b)}{p_{a,b}}\right) \cdot \frac{w(a,b)}{p_{a,b}} + \mathbb{P}(W_H(a,b) = 0) \cdot 0 \\ &= p_{a,b} \cdot \frac{w(a,b)}{p_{a,b}} + (1 - p_{a,b}) \cdot 0 \\ &= w(a,b)\,.\end{aligned}$$

Thus, on average, the weight of each edge is preserved and thus the overall weight of the graph on average is preserved. Furthermore,

$$\mathbb{E}[\boldsymbol{L}_H] = \sum_{(a,b)\in E} \mathbb{E}[W_H(a,b)]\boldsymbol{L}_{a,b} = \sum_{(a,b)\in E} w(a,b)\boldsymbol{L}_{a,b} = \boldsymbol{L}_G\,,$$

indicating that the average graph Laplacian of H over an infinite number of random draws of H will be equal to $\boldsymbol{L}_G$. However, for most random draws of H, $\boldsymbol{L}_H$ will be much sparser than $\boldsymbol{L}_G$ (if we pick the probabilities $p_{a,b}$ correctly).

Remark 30. Since on average the overall weight of G is preserved by H, one can think of the new weights $w_H(a,b) = w(a,b)/p_{a,b}$, which increase the weight of the edge, as aggregating the weights of the edges that are removed from G into the fewer remaining edges left in H. As an additional reference point, if $G = K_n$ with all edge weights equal to one, and if $\widetilde{H}$ is a (d,ϵ)-expander graph also with all edge weights equal to one, then it is $H = (n/d)\widetilde{H}$ that ϵ-approximates K_n. In particular, we see that all the edge weights are blown up in this case too.

30.3 Chernoff bounds

To prove that the algorithm from Section 30.2 works, we will need two types of Chernoff bounds, one for random variables and one for random matrices. All of these results are types of concentration inequalities.

Theorem 60 (Random variable Chernoff bound)**.** *Let $X_1, \ldots, X_m$ be independent random variables taking values in $\{0,1\}$. Define*

$$X := \sum_{i=1}^{m} X_i\,,$$

and let $\mu := \mathbb{E}[X]$. Then for any $0 \le \delta \le 1$,

$$\begin{aligned}\mathbb{P}(X \le (1-\delta)\mu) &\le e^{-\delta^2\mu/2}\,, \\ \mathbb{P}(X \ge (1+\delta)\mu) &\le e^{-\delta^2\mu/3}\,.\end{aligned}$$

Theorem 61 (Random matrix Chernoff bound)**.** *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ be independent random $n \times n$, symmetric, positive semi-definite matrices so that $\|\boldsymbol{X}_i\| \leq R$ almost surely. Define*

$$\boldsymbol{X} := \sum_{i=1}^{m} \boldsymbol{X}_i \,,$$

and let $\mu_{\min}$ and $\mu_{\max}$ be the minimum and maximum eigenvalues of

$$\mathbb{E}[\boldsymbol{X}] = \sum_{i=1}^{m} \mathbb{E}[\boldsymbol{X}_i] \,,$$

respectively. Then for any $0 \leq \epsilon \leq 1$,

$$\mathbb{P}(\lambda_{\min}(\boldsymbol{X}) \leq (1-\epsilon)\mu_{\min}) \leq n e^{-\epsilon^2 \mu_{\min}/2R} \,,$$
$$\mathbb{P}(\lambda_{\max}(\boldsymbol{X}) \geq (1+\epsilon)\mu_{\max}) \leq n e^{-\epsilon^2 \mu_{\max}/3R} \,.$$

Remark 31. It is important to note that the random variables $X_1, \ldots, X_m$ in Theorem 60, and the random matrices $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ in Theorem 61, can have different distributions.

Remark 32. Looking ahead, we will use Theorem 60 to prove that the number of edges in H is $O(\epsilon^{-2} n \log n)$ with high probability, and we will use Theorem 61 to prove that H is an ϵ-approximation of G with high probability.

References

[1] Daniel A. Spielman. Spectral and algebraic graph theory. Book draft, available at: http://cs-www.cs.yale.edu/homes/spielman/sagt/, 2019.

[2] Michael Perlmutter, Feng Gao, Guy Wolf, and Matthew Hirn. Geometric scattering networks on compact Riemannian manifolds. In *Proceedings of The First Mathematical and Scientific Machine Learning Conference, Proceedings of Machine Learning Research*, volume 107, pages 570–604, 2020.

[3] David I. Shuman, Sunil K. Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013.

[4] Stéphane Mallat. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way.* Academic Press, 3rd edition, 2008.

[5] Ronald R. Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21:5–30, 2006.