Lecture 6: Intro to Graph and Matrix Sparsification

Instructor: Lorenzo Orecchia Scribe: Erasmo Tani

A simple bound on L

Lemma 1. *For any graph G:*

$$0 \preceq L_G \preceq 2D$$

Proof. we have:

$$x^T L_G x = \sum_{\{i,j\}\in E} (x_i - x_j)^2 \leq \sum_{\{i,j\}\in E} 2(x_i^2 + x_j^2) = 2\sum_{i\in V} d_i x_i^2 = 2x^T Dx$$

□

Lemma 2. *If $\exists x \neq 0 : x^T L_G x = 2x^T Dx$*

Lemma 3 (From last time)**.** $\lambda_2 = 0 \iff$ *G is disconnected.*

Lemma 4. $\lambda_n(\mathcal{L}) = 2 \iff$ *G is bipartite.*

There are robust versions of both the above lemmas and we will see them in future classes.

Sparsification

Given a graph $G = (V, E)$ then:

$$L_G = \sum_{e\in E} \chi_e \chi_e^T$$

where:

$$\chi_{u,v} := e_u - e_v$$

note that the sign doesn't matter in general. Our goal is to construct a weighted graph $H = (V, E_H, w_h)$ s.t. (ε-sparsifier)

1. $||L_G - L_H|| \leq \varepsilon$
2. H is sparse i.e. $|E_H|$ is small $\tilde{O}(n) = O(n \text{ polylog } n)$.

There are two different notions of error we are going to discuss:

Aboslute Error Here we want:

$$|x^T L_G x - x^T L_H x| \leq \varepsilon \cdot x^T Dx$$

This is an easy definition to work with however in some cases it does not represent our notion of similarity in a combinatorial sense. Consider the example given in Figure 1. In some sense in this setting we are getting the small eigenvalues wrong.

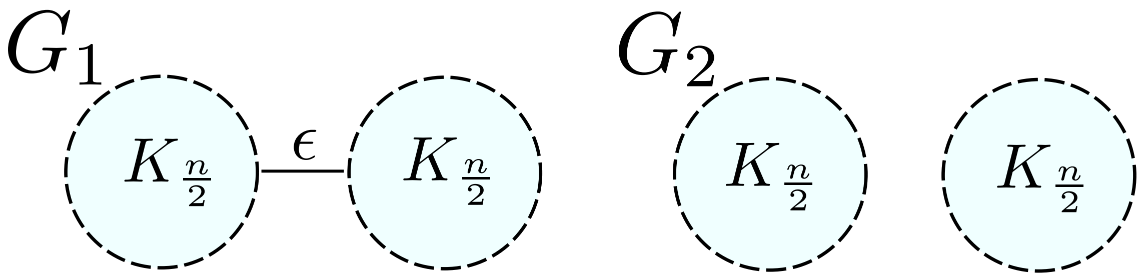

Figure 1: Two graphs that are ϵ-similar in additive error. Here G_1 is connected while G_2 is not.

Relative Error Here the measure of error is:

$$|x^T L_G x - x^T L_H x| \leq \varepsilon \cdot x^T L_G x$$

One thing worth nothing is that sparsifiers will approximately preserve the degree of every vertex (check by picking x to be the indicator random vector of a vertex). This holds independently of which of the two above notions of error is used. All the sparsifier we will look at have the additional property that $E_H \subseteq E_G$.

PSD Sparsification

We are given a symmetric matrix $A \succeq 0$, $A \in \mathbb{R}^{n \times n}$ and it is given as the sum of rank 1 matrices:

$$A = \sum_{i=1}^{m} u_i u_i^T \qquad\qquad m >> n$$

A is telling me how the mass of the u_i vectors is distributed. Look for instance at:

$$x^T A x = \sum_{i=1}^{n} (u_i^T x)^2$$

we may want to construct a matrix sparsifier:

$$\tilde{A} = \sum_{i=1}^{m} s_i u_i u_i^T$$

where we want:

$$\mathbb{E}[s_i] = 1$$

Choosing the distribution of the s_is amounts to deciding the relative importance of each vector in the decomposition of A. (Statistical Leverage).

$$\forall x : \qquad |x^T A x - x^T \tilde{A} x| \leq \varepsilon \cdot x^T A x$$

we do a change of matrix: $y = A^{1/2} x$

$$|y^T I y - y^T A^{-1/2} \tilde{A} A^{-1/2} y| \leq \varepsilon \cdot ||y||^2$$

$$\tilde{A} = \sum_{i=1}^{T} s_i A^{-1/2} u_i u_i^T A^{-1/2}$$

Suppose you have a bunch of vectors $\{\vec{a}_i\}$ and you want to estimate the sum by sampling.

$$A = \sum_{i=1}^{m} a_i \qquad\qquad \vec{a}_i \in \mathbb{R}^n, ||\vec{a}_i||^2$$

we can do:

$$s_i = \begin{cases} \frac{1}{p_i} & \text{w.p. } \; p_i \propto ||a_i||^2 \\ 0 \end{cases}$$

Matrix Chernoff Bounds

Scalar Chernoff Bound

Given a sum of independent random variabels:

$$X = \sum X_i$$

we can look at:

$$g(\theta) = \mathbb{E}[e^{\theta X}] = \mathbb{E}[e^{\theta \Sigma X_i}] = \prod \mathbb{E}[e^{\theta X_i}]$$

We then have:

$$Pr[X > \lambda] = Pr[e^{\theta X} > e^{\theta\lambda}] \leq \frac{\mathbb{E}[e^{\theta x}]}{e^{\theta\lambda}}$$

Matrix Chernoff Bound

Given a matrix random variable $Y \in S^{n\times n}$ its mgf is:

$$\mathbb{E}[e^{\theta Y}]$$

where $e^{\theta Y}$ is a matrix exponential.