Lecture 5: Local to Global Theorems, SOS and Low Threshold Rank Graphs

Lecturer: Shayan Oveis Gharan *??*

In the last few lectures we introduced expander graphs. Suppose a given G with random walk matrix P (and stationary distribution π_0) is a λ-expander. This means that

$$\lambda = \min_{f \neq \text{const}} \frac{\mathcal{E}(f,f)}{\text{Var}(f)} = \min_{f \neq \text{const}} \frac{\mathbb{E}_{u\sim\pi_0}\mathbb{E}_{\{u,v\}|u}(f(u)-f(v))^2}{\mathbb{E}_{u\sim\pi_0}\mathbb{E}_{v\sim\pi_0}(f(u)-f(v))^2}$$

Suppose we are given a (non-constant) function f that is locally correlated,

$$\mathbb{E}_{u\sim\pi_0}\mathbb{E}_{\{u,v\}|u}(f(u)-f(v))^2 \leq \eta.$$

In other words, we can say that (on average) f assigns almost similar values to the endpoints of every edge. Then, it must also be globally correlated, i.e., we have

$$\mathbb{E}_{u\sim\pi_0}\mathbb{E}_{v\sim\pi_0}(f(u)-f(v))^2 \leq \frac{\eta}{\lambda}.$$

In other words, then (on average) the values that f assigns to any random pair of vertices is almost the same.

Note that this property does not hold if the graph is not an expander graphs.

Lemma 5.1. *For any symmetric matrix $A \in \mathbb{R}^{n\times n}$ and $B \succeq 0$, we have*

$$A \bullet B := \text{Tr}(AB) \leq \sum_{i=1}^n \lambda_i(A)\lambda_i(B).$$

We leave this as an exercise; the main idea of the proof is that for any set of real numbers $a_1, \dots, a_n \in \mathbb{R}$ and $b_1 \geq b_2 \geq \dots \geq \dots \geq b_n \geq 0$ we have

$$\sum_{i=1}^n a_i b_i \leq \sum_{i=1}^n a_{\sigma(i)} b_i,$$

where $\sigma(.)$ is the permutation chosen such that $a_{\sigma(1)} \geq \dots \geq a_{\sigma(n)}$.

In the following lemma we prove a generalization of this fact for low-threshold rank graphs.

Lemma 5.2. *Let $v_1, \dots, v_n \in \mathbb{R}^n$ with $\mathbb{E}_i \|v_i\|^2 = 1$, $\mathbb{E}_{i,j}\langle v_i, v_j\rangle^2 \leq 1/k$ where the expectations are with respect to the uniform distribution. For $C > 0$, any symmetric matrix A with $\|A\| \leq 1$ (i.e., all eigenvalues of A are at most 1), and*

$$\mathbb{E}_i \sum_j A_{i,j}\langle v_i, v_j\rangle \geq 1-\epsilon$$

we have $\lambda_{k(1-1/C)^2}(A) \geq 1 - C\epsilon$ where λ_i is the i-th largest eigenvalue of A.

Proof. First,

$$1 = \mathbb{E}_i \|v_i\|^2 = \frac{1}{n}\operatorname{Tr}(V) = \mathbb{E}_i \lambda_i(V) \tag{5.1}$$

On the other hand, for any integer $1 \leq k' \leq n$, by Cauchy-Schwartz inequality we have,

$$\frac{1}{n}\sum_{i=1}^{k'} \lambda_i(V) \leq \frac{1}{n}\sqrt{k'}\sqrt{\sum_{i=1}^{k'} \lambda_i(V)^2} \leq \frac{\sqrt{k'}}{n}\|V\|_F = \sqrt{k'}\sqrt{\mathbb{E}_{i,j}\langle v_i, v_j\rangle^2} \leq \sqrt{\frac{k'}{k}} \tag{5.2}$$

Let k' be the largest index such that $\lambda_{k'}(A) \geq 1 - C\epsilon$. we need to show $k' \geq k(1-1/C)^2$.

$$\begin{aligned}
1 - \epsilon = \frac{1}{n} A \bullet V &\underset{\textit{Lemma 5.1}}{\leq} \mathbb{E}_i \lambda_i(A)\lambda_i(V) \\
&\underset{\lambda_i(A)\leq 1}{\leq} \frac{1}{n}\sum_{i=1}^{k'} \lambda_i(V) + \frac{1}{n}\sum_{i=k'+1}^{n} (1 - C\epsilon)\lambda_i(V) \\
&\underset{(5.1)}{=} 1 - \frac{C\epsilon}{n}\sum_{i=k'+1}^{n} \lambda_i(V)
\end{aligned}$$

Therefore,

$$\frac{1}{C} \geq \frac{1}{n}\sum_{i=k'+1}^{n} \lambda_i(V) = 1 - \frac{1}{n}\sum_{i=1}^{k'} \lambda_i(V) \underset{(5.2)}{\geq} 1 - \sqrt{k'/k}.$$

Therefore, $\sqrt{k'/k} \geq 1 - 1/C$ and $k' \geq k(1-1/C)^2$. □

As a consequence we prove the following statement.

Corollary 5.3 (Local to Global Theorem)**.** *Given a graph $G = (V, E)$ with n vertices and a distribution $\pi_1 : E \to \mathbb{R}_{\geq 0}$ and let π_0 be the corresponding distribution over vertex of G and P be the random walk matrix. Furthermore, suppose we are given a set vectors $v_1, \ldots, v_n$ such that*

$$\mathbb{E}_{i\sim\pi_0}\|v_i\|^2 = 1 \quad \textit{and} \quad \mathbb{E}_{\{i,j\}\sim\pi_1}\langle v_i, v_j\rangle \geq \epsilon.$$

Then,

$$\mathbb{E}_{i,j\sim\pi_0}\langle v_i, v_j\rangle^2 \geq \frac{\epsilon^2}{4\operatorname{rank}_{\epsilon/2}(P)}.$$

Proof. Let $\tilde{P}$ be the normalized Laplacian matrix of G where $\tilde{P}_{i,j} = \frac{\pi_1(\{i,j\})}{2\sqrt{\pi_0(i)\pi_0(j)}}$ and let for any i, let $u_i = \sqrt{n\pi_0(i)}v_i$. then

$$1 - (1-\epsilon) = \epsilon \leq \mathbb{E}_{\{i,j\}\sim\pi_1}\langle v_i, v_j\rangle = \frac{2}{n}\sum_{\{i,j\}\in E} \tilde{P}_{i,j}\langle u_i, u_j\rangle = \mathbb{E}_i \sum_j \tilde{P}_{i,j}\langle u_i, u_j\rangle.$$

Let A be the adjacency matrix of G with $A_{i,j} = \pi_1(\{i,j\})/2$. Then, $P = \Pi_0^{-1}A$ and $\tilde{P} = \Pi_0^{-1/2}A\Pi_0^{-1/2}$ so $P, \tilde{P}$ have the same eigenvalues.

On the other hand,

$$1 \geq \mathbb{E}_{i\sim\pi_0}\|v_i\|^2 = \mathbb{E}_{i\sim\pi_0}\frac{1}{n\pi_0(i)}\|u_i\|^2 = \mathbb{E}_i\|u_i\|^2$$

Suppose,

$$\frac{1}{k} = \mathbb{E}_{i,j\sim\pi_0}\langle v_i, v_j\rangle^2 = \mathbb{E}_{i,j}\langle u_i, u_j\rangle^2.$$

Then, by the above lemma, for $C = (1+\epsilon/2)$ we get

$$1 - C(1-\epsilon) = \epsilon/2 - \epsilon^2/4 \leq 1 - \leq \lambda_{k(1-1/C)^2}(\tilde{P}) \approx \lambda_{k\epsilon^2/4}(P).$$

So, $\text{rank}_{\epsilon/2}(P) \geq \frac{k\epsilon^2}{4} = \frac{\epsilon^2}{4\mathbb{E}_{i,j}\langle v_i, v_j\rangle^2}$ □

5.1 Approximation Algorithms for Max Cut

Let G be a graph with a distribution π_1 over its edges with corresponding random walk operator P and π_0 the stationary distribution of the walk. In this section we describe a $1+\epsilon$ approximation algorithm of [BRS11] for maxcut that runs in time $n^{\text{rank}_{\geq \epsilon^2/2}(P)/\epsilon^4}$.

Let $X_1, \ldots, X_n$ be random variables in $\{-1, +1\}$. We can think of the maximum-cut as the problem of finding a joint distribution over $X_1, \ldots, X_n$ that maximizes.

$$\max \mathbb{E}_{\{i,j\}\sim\pi_1} \mathbb{P}\left[X_i \neq X_j\right]$$

Now, how can we optimize over such high dimensional family of distributions? The idea is that for each vertex i and a side $s \in \{-1, +1\}$ we have a vector $v_{i,s}$ with $\|v_{i,+1}\|^2 = \mathbb{P}\left[X_i = +1\right]$ and $\|v_{i,-1}\|^2 = \mathbb{P}\left[X_i = -1\right]$, and, for each pair of vertices $\{i,j\}$ and $s_1, s_2 \in \{-1, +1\}$ we have a vector v_{i,j,s_1,s_2} with property that

$$\|v_{i,j,s_1,s_2}\|^2 = \langle v_{i,s_1}, v_{j,s_2}\rangle = \mathbb{P}\left[X_i = s_1, X_j = s_2\right].$$

Now, we can obtain all these vectors by writing a SDP relaxation. Note that the SDP solution does not give us an actual distribution over all n variables $X_1, \ldots, X_n$; it just gives a locally consistent distribution.

Because we don't have a joint distribution over all variables, perhaps the simplest idea to round is to run an independent rounding: For any vertex i, put i at the -1-side with probability $\|v_{i,-1}\|^2$ and put it on the $+1$-side otherwise. How well does this algorithm do with respect to the objective function? The loss is at most

$$\mathbb{E}_{\{i,j\}\sim\pi_1}\mathbb{P}_{sdp}\left[X_i \neq X_j\right] - \mathbb{P}_{indep}\left[X_i \neq X_j\right],$$

i.e., for every edge $\{i,j\}$ the probability that i,j map to the opposite sides of the cut in the SDP solution minus the probability that they map to opposite sides in the independent rounding solution.

Further notice,

$$\mathbb{E}\left[X_i X_j\right] = \mathbb{P}\left[X_i = X_j\right] - \mathbb{P}\left[X_i \neq X_j\right] = 1 - 2\mathbb{P}\left[X_i \neq X_j\right].$$

So, we can write

$$\begin{aligned} ALG &= SDP - \mathbb{E}_{\{i,j\}\sim\pi_1}(\mathbb{P}_{sdp}\left[X_i \neq X_j\right] - \mathbb{P}_{indep}\left[X_i \neq X_j\right]) \\ &= SDP - \mathbb{E}_{\{i,j\}\sim\pi_1}(\frac{1}{2}(1 - \mathbb{E}\left[X_i X_j\right]) - (\mathbb{P}\left[X_i = -1\right]\mathbb{P}\left[X_j = +1\right] + \mathbb{P}\left[X_i = +1\right]\mathbb{P}\left[X_j = -1\right]) \end{aligned}$$

Using $\frac{1}{2} = \frac{1}{2}(\mathbb{P}\left[X_i = +1\right] + \mathbb{P}\left[X_i = -1\right])(\mathbb{P}\left[X_j = +1\right] + \mathbb{P}\left[X_j = -1\right])$

$$\begin{aligned} &= SDP - \frac{1}{2}\mathbb{E}_{\{i,j\}\sim\pi_1}(-\mathbb{E}\left[X_i X_j\right] + (\mathbb{P}\left[X_i = +1\right] - \mathbb{P}\left[X_i = -1\right])(\mathbb{P}\left[X_j = +1\right] - \mathbb{P}\left[X_j = -1\right])) \\ &= SDP + \frac{1}{2}\mathbb{E}_{\{i,j\}\sim\pi_1}\operatorname{Cov}(X_i, X_j) \end{aligned}$$

where recall that

$$\mathrm{Cov}(X_i, X_j) = \mathbb{E}\left[X_i X_j\right] - \mathbb{E}\left[X_i\right]\mathbb{E}\left[X_j\right].$$

So, if $\mathbb{E}_{\{i,j\}\sim\pi_1} \mathrm{Cov}(X_i, X_j) \geq -\epsilon$ we obtain a multiplicative $1-\epsilon$ approximation to max-cut. But what if the covariance is too small? The main idea of [BRS11] is the following fundamental inequality proved in the following lemma:

$$\begin{aligned}\mathbb{E}_{i\sim\pi_0}\mathbb{E}_{j\sim\sim\pi_0}(\mathbb{E}\,\mathrm{Var}[X_j] - \mathbb{E}\,\mathrm{Var}[X_j|X_i]) &\geq \mathbb{E}_{i,j\sim\pi_0} \mathrm{Cov}(X_i, X_j)^2 \frac{1}{2}\left(\frac{1}{\mathrm{Var}[X_i]} + \frac{1}{\mathrm{Var}[X_j]}\right)\\ &\geq \mathbb{E}_{i,j\sim\pi_0}\mathrm{Cov}(X_i, X_j)^2 \qquad (5.3)\end{aligned}$$

where the inequality uses that $\mathrm{Var}[X_i] \leq \mathbb{E}X_i^2 = 1$ for any i.

Lemma 5.4. *For any two i, j we have*

$$\mathbb{E}\,\mathrm{Var}[X_j|X_i] \leq \mathrm{Var}[X_j] - \frac{\mathrm{Cov}(X_i, X_j)^2}{\mathrm{Var}[X_i]}$$

Proof. First, recall law of total variance: For any two random variables X, Y,

$$\mathrm{Var}[Y] = \mathbb{E}\,\mathrm{Var}[Y|X] + \mathrm{Var}[\mathbb{E}(Y|X)]$$

Having that it is enough to show

$$\mathrm{Var}[\mathbb{E}(X_j|X_i)] \geq \frac{\mathrm{Cov}(X_i, X_j)^2}{\mathrm{Var}[X_i]}$$

The above identity holds in general as long as X_i only takes two different values. First, notice that Variance and Covariance are shift-invariant, so we can assume $\mathbb{E}X_i = \mathbb{E}X_j = 0$. So, it is enough to show

$$\mathbb{E}[\mathbb{E}(X_j|X_i)^2] = \frac{\mathbb{E}[X_i X_j]^2}{\mathbb{E}[X_i^2]}$$

Now, the above basically follows by Cauchy-Schwartz inequality:

$$\mathbb{E}\left[X_i X_j\right]^2 = \mathbb{E}\left[\mathbb{E}\left[X_i X_j|X_i\right]\right]^2 = \mathbb{E}\left[X_i \mathbb{E}\left[X_j|X_i\right]\right]^2 \underset{Cauchy}{\leq} \mathbb{E}\left[X_i^2\right]\mathbb{E}\left[\mathbb{E}\left[X_j|X_i\right]^2\right]$$

as desired. □

Having that, one can imagine a win-win strategy: Either $\mathbb{E}_{\{i,j\}\sim\pi_1}\mathrm{Cov}(X_i, X_j) \geq -\epsilon$ and we get a $1-\epsilon$ approximation to maxcut or, $\mathbb{E}_{\{i,j\}\sim\pi_1}\mathrm{Cov}(X_i, X_j) < -\epsilon$. In the latter case, by the following fact, the Covariance matrix is PSD. So, there are vectors $u_1, \ldots, u_n$ such that $\langle u_i, u_j\rangle = \mathrm{Cov}(X_i, X_j)$ for all i, j. Let $w_i = u_i^{\otimes 2}$. So, by Cauchy-Schwartz inequality the aforementioned assumption implies that

$$\epsilon^2 \leq \mathbb{E}_{\{i,j\}\sim\pi_1}\mathrm{Cov}(X_i, X_j)^2 = \mathbb{E}_{\{i,j\}\sim\pi_1}\langle u_i, u_j\rangle^2 = \mathbb{E}_{\{i,j\}\sim\pi_1}\langle w_i, w_j\rangle.$$

Then, by Corollary 5.3, we get that

$$\mathbb{E}_{i,j\sim\pi_0}\mathrm{Cov}(X_i, X_j)^2 \geq \mathbb{E}_{i,j\sim\pi_0}\langle u_i, u_j\rangle^4 = \mathbb{E}_{i,j\sim\pi_0}\langle w_i, w_j\rangle^2 \geq \frac{\epsilon^4}{4\,\mathrm{rank}_{\epsilon^2/2}(P)}.$$

Plugging this in to (5.3) if we sample $i \sim \pi_0$, round X_i (to $+1$ or -1 with respect to its marginals) then the expected variance of all remaining variables decrease by $\frac{\epsilon^4}{4\,\mathrm{rank}_{\epsilon^2/2}(P)}$. But in this case we need that every X_j to be well-defined after conditioning $X_i = +1/-1$. This leads to the Lasserre/Sum-Squares hierarchy.

Fact 5.5. *Given a set of vectors v_{i,j,s_1,s_2} as defined above the covariance matrix is PSD.*

Proof. We just sketch the proof: The idea is to define a vector $u_i = v_{i,+1} - v_{i,-1} - \mathbb{E}\left[X_i\right]$ where $\mathbb{E}\left[X_i\right] = \|v_{i,+1}\|^2 - \|v_{i,-1}\|^2$ and show that the covariance matrix is just the Gram-matrix of $u_1, \ldots, u_n$. □

Sum of Squares Hierarchy. In the $m+2$ rounds of the sum of squares hierarchy for every set $S \subseteq V$ of size $|S| \leq m+2$ and any $\alpha \in \{+1,-1\}^{|S|}$ we have a vector $v_{S,\alpha}$. For any two sets S, T with $|S \cup T| \leq m+2$ and $\alpha \in \{+1,-1\}^{|S|}, \beta \in \{+1,-1\}^{|T|}$ we have

$$\langle v_{S,\alpha}, v_{T,\beta} \rangle = \mathbb{P}\left[X_S = \alpha, X_T = \beta\right]$$

Now, this allows us to follow the same line of reasoning for m many steps, each time conditioning a variable to be $+1$ or -1. Now, after $O(\text{rank}_{\epsilon^2/2}(P)/\epsilon^4)$ many steps either we get to (conditional) variables $X_1, \ldots, X_n$ with $\mathbb{E}_{i \sim \pi_0} \text{Var}[X_i] < 0$ but this is impossible. So, the procedure should end in $O(\text{rank}_{\epsilon^2/2}(P)/\epsilon^4)$ many conditiong.

This finishes the proof of [BRS11].

Remark 5.6. *It remains a fascinating open problem to design a sub-exponential time algorithm for the max-cut problem. Based on this discussion, all we need to do is to design a (sub-exponential) time algorithm for high threshold rank graphs, i.e., graphs where* $\text{rank}_\epsilon(P) \geq n^{\Omega(1)}$. *The byproduct of the above method is that we only need to focus on the case where* $\mathbb{E}_{i,j \sim \pi_0} \text{Cov}(X_i, X_j)^2 \leq n^{-c}$ *for some constant* $c > 0$. *Or, in other words, that there are vectors* $v_1, \ldots, v_n$ *with* $\|v_i\|^2 \approx 1$ *and* $\mathbb{E}_{i,j \sim \pi_0} \langle v_i, v_j \rangle^2 < n^{-c}$. *If in such a case one can get a tight approximation for max-cut efficiently, similar to the algorithm for unique games in the last lecture, it would lead to a sub-exponential time* $1+\epsilon$ *approximation algorithm.*

References

[BRS11] B. Barak, P. Raghavendra, and D. Steurer. "Rounding Semidefinite Programming Hierarchies via Global Correlation". In: *FOCS*. Ed. by R. Ostrovsky. IEEE Computer Society, 2011, pp. 472–481 (cit. on pp. 5-3, 5-4, 5-5).