CS369H: Hierarchies of Integer Programming Relaxations Spring 2016-2017

Lecture: SOS Lower bound for Planted Clique

Professor Moses Charikar *Report Author: Kiran Shiragur*

Overview In this lecture we show SOS lower bound for planted clique of roughly $k \approx n^{1/2r}$ for an r round SOS due to [MPW15]. In this lecture we state all of their main results and provide proof sketch.

1 Introduction

Finding Cliques in random graphs is one of the well studied problems in Algorithm design. Let $G(n,p)$ be the Erdös-Renyi random graph on n vertices where between every pair of vertices i and j an edge (i,j) is added with probability p independently. The problem we are interested is planted clique problem introduced by Jerrum [Jer92] and Kucera [Ku95]:

Problem Definition: Planted Clique
Given a graph $G(V,E)$ we are asked to identify which one of the following two distribution the graph is generated:

1. $G(n,1/2)$: Distribution over Erdös-Renyi random graphs with parameter $p = 1/2$.
2. Generate an instance of $G(n,1/2)$ and plant a clique of size k to this graph.

Erdös-Renyi random graph $G(n,1/2)$ has a clique of size at most $(2+o(1))\log n$ with high probability. If you plant a clique of size $k \geq 3\log n$, then it is information theoretically possible to distinguish the distribution input graph was sampled by enumerating over all possible n choose k subsets of vertices. The problem we care about is what one can achieve in polynomial time? and the answer is if $k = \theta(\sqrt{n})$ we have a efficient spectral algorithm which solves the problem due to [AKS98]. The interesting regime is when $3\log n \leq k \leq o(\sqrt{n})$.

In the literature of algorithm design, LP and SDP have been repeatedly used to build algorithms for different problems and are considered important machinery for algorithm design. Here is a general way one uses LP and SDP : For any combinatorial optimization problem, we write the integer linear program (ILP) to capture the problem exactly. Then we relax integral constraints and make the problem convex which can now be solved efficiently. We solve this convex program, take the fractional solutions and round them to integral solutions with good theoretical guarantees. This rounding algorithm works better if the fractional solutions are close to integral solutions. This is captured by the notion of integral gap of a LP/SDP which is defined to be the maximum of the ratio of fractional LP/SDP solution to the optimal integral solution and this captures the strength of LP/SDP (integrality gap). The close this ratio is to 1 the better is the LP/SDP and a way to achieve this integrality gap close to 1 is by adding stronger constraints to the LP/SDP. This is were Sum of Squares plays the role which is an algorithmic way to add constraints to basic LP/SDP. SOS captures most of the interesting algorithms built in the literature. For instance, the Goemans-Williamson SDP for max-cut is captured by the degree 2 SOS. The triangle inequality constraints of ARV SDP is captured by a degree 4 SOS and many others.

Planted clique is an average case problem and we don't have machinery to prove NP-Hardness results. Here we are interested in showing SOS lower bounds which would eliminate the existence of large class of efficient algorithms for planted clique and would serve confidence for the problem being hard.

In this work we show a SOS lower bound for the plated clique. Below is the theorem we prove in this lecture:

Theorem 1.1. *With high probability, for $G \leftarrow G(n, 1/2)$ the natural r-round* SOS *relaxation of the maximum clique problem has an integrality gap of at least $n^{1/2r}/C^r(\log n)^2$.*

As a corollary the following lower bound for the planted clique problem exists,

Corollary 1.1. *With high probability, for $G \leftarrow G(n, 1/2, t)$ the natural r-round* SOS *relaxation of the planted clique problem has an integrality gap of at least $n^{1/2r}/tC^r(\log n)^2$.*

This lower bound implies that any constant round SOS cannot handle a planted clique of size $k = n^{o(1)}$.

Current best result The current best result for planted clique achieves a SOS lower bound of nearly $n^{1/2}$ for any constant rounds of SOS due to [BHK+16].

2 Proof systems and SDP hierarchies

Suppose we are a given a system of axioms or polynomial equations:

$$f_1(x) = 0, f_2(x) = 0, \dots f_m(x) = 0$$

where each $f_i : \mathbb{R}^n \to \mathbb{R}$ is a n-variate polynomial. We wish to decide existence of a solution satisfying all the axioms simultaneously. The problem of refuting a system turns out to be much easier than deciding. The positivstellensatz refutation is an identity of the form:

$$\sum_{i=1}^{m} f_i g_i \equiv 1 + \sum_{i=1}^{N} h_i^2,$$

where g_i's and h_i's are arbitrary polynomials. Clearly if such polynomials exist then no solution exists which would satisfy all the axioms $\{f_1(x) = 0, f_2(x) = 0, \dots f_m(x) = 0\}$ simultaneously because we would have a contradiction $0 \geq 1$. The question is how efficiently can we find such a refutation which motivates the following definition:

Definition 1.1 (Positivstellensatz Refutation)**.** *Let $\mathcal{F} \equiv \{f_1, \dots, f_n : \mathbb{R}^n \to \mathbb{R}\}$, be a system of axioms, where each f_i is a real n-variate polynomial. A positivstellensatz* refutation *of degree r (*$\mathsf{PS}(r)$ *refutation) for $\mathcal{F}$ is an identity of the form*

$$\sum_{i=1}^{m} f_i g_i \equiv 1 + \sum_{i=1}^{N} h_i^2, \tag{2.1}$$

where $g_1, \dots, g_m, h_1, \dots, h_N$ are n-variate polynomials such that $deg(f_i g_i) \leq 2r$ for all $i \in [m]$ and $deg(h_j) \leq r$ for all $j \in [N]$.

Why is the study of system of polynomial equations important to us?. We could represent the problem of finding clique of size k by the following set of polynomial axioms:

Definition 1.2. *Given a graph G, let* $\mathrm{Clique}(G,k)$ *denote the following set of polynomial axioms:*

$$\begin{aligned} \textit{(Max-Clique):} \quad & x_i^2 - x_i, \ \ \forall i \in [n] \\ & x_i \cdot x_j, \ \ \forall \textit{ pairs } \{i,j\} \notin G \\ & \sum_i x_i - k. \end{aligned} \tag{2.2}$$

The first constraint enforces the boolean condition. The second constraint represents the clique constraint and the last inequality represents the size of clique. Here is the theorem restated in the language of $\mathsf{PS}(r)$ refutation.

Theorem 1.2 (Main)**.** *With high probability over $G \leftarrow G(n,1/2)$, the system* $\mathrm{Clique}(G,k)$ *defined by Equation 2.2 has no* $\mathsf{PS}(r)$ *refutation for $k \leq n^{1/2r}/C^r(\log n)^{1/r}$*

The $\mathsf{PS}(r)$ refutation is related to the $2r$ round SOS and next we describe this connection. Let $\mathcal{P}(n,2r) : \mathbb{R}^n \to \mathbb{R}$ be the set of n-variate real polynomials of total degree at most $2r$.

Definition 1.3 (PSD Mappings)**.** *A linear mapping $\mathcal{M} : \mathcal{P}(n,2r) \to \mathbb{R}$ is said to be positive semi-definite (PSD) if $\mathcal{M}(P^2) \geq 0$ for all n-variate polynomials P of degree at most r.*

Definition 1.4 (Dual Certificates)**.** *Given a set of axioms $f_1, \ldots, f_m$, a dual certificate for the axioms is a PSD mapping $\mathcal{M} : \mathcal{P}(n,2r) \to \mathbb{R}$ such that $\mathcal{M}(f_i g) = 0$ for all $i \in [m]$ and all polynomials g such that $deg(f_i g) \leq 2r$.*

The lemma below builds the connection between degree $2r$ SOS and $\mathsf{PS}(r)$ refutation.

Lemma 1.1 (Dual Certificate)**.** *Given a system of axioms $((f_i))$, there does not exist a $\mathsf{PS}(r)$ refutation of the system if there exists a dual certificate $\mathcal{M} : \mathcal{P}(n,2r) \to \mathbb{R}$ for the axioms.*

With the help of this lemma, all we need to show is the existence of a dual certificate $\mathcal{M} : \mathcal{P}(n,2r) \to \mathbb{R}$ for the clique axioms $\mathrm{Clique}(G,k)$ for $k \geq n^{1/2r}/C^r(\log n)^{1/r}$.

2.1 Proof sketch

To show such a lower bound we do the following:

1. Construct the PSD mapping $\mathcal{M}$ which satisfies all the clique axioms $\mathrm{Clique}(G,k)$.
2. Next show the PSDness of mapping $\mathcal{M}$ which turns out to be the hard part to prove. Here we argue the PSDness of $\mathcal{M}$ by reducing to the PSDness of another PSD mapping M' which would be easier to handle.

3 Dual certificate for $\mathsf{PS}(r)$ refutations of max-clique

In this section we construct the dual certificate for the clique axioms $\mathrm{Clique}(G,k)$. Recall the clique axioms:

$$\begin{aligned} \text{(Max-Clique):} \quad & x_i^2 - x_i, \ \ \forall i \in [n] \\ & x_i \cdot x_j, \ \ \forall \text{ pairs } \{i,j\} \notin G \\ & \sum_i x_i - k. \end{aligned} \tag{3.1}$$

The axioms above suggest any dual certificate $\mathcal{M} \equiv \mathcal{M}_G : \mathcal{P}(n,2r) \to \mathbb{R}$ for clique axioms $\mathrm{Clique}(G,k)$ should satisfy:

$$\begin{aligned} & \mathcal{M}(X_I) = 0, \ \forall I, \ |I| \le 2r, \ I \text{ is not a clique in } G, \\ & \mathcal{M}\left(\left(\sum_{i=1}^n x_i - k\right) X_I\right) = 0, \ \forall I, \ |I| < 2r. \end{aligned} \tag{3.2}$$

where $X_I = \prod_{i \in I} x_i$. The above equations give a set of linear equations that a PSD mapping should satisfy. By looking at the equations we could guess a natural solution to system of equations above: Given a graph G on $[n]$, and $I \subseteq [n]$, $|I| \le 2r$, let

$$deg_G(I) = |\{S \subseteq [n] : I \subseteq S, \ |S| = 2r, \ S \text{ is a clique in } G\}|.$$

which is a generalisation for the degree of a vertex v (for $r = 1$).

Now define $\mathcal{M} \equiv \mathcal{M}_G : \mathcal{P}(n,2r) \to \mathbb{R}$ as follows: for $I \subseteq [n], |I| \le 2r$, let

$$\mathcal{M}\left(\prod_{i \in I} x_i\right) = deg_G(I) \cdot \frac{k(k-1)\cdots(k-|I|+1)}{2r(2r-1)\cdots(2r-|I|+1)} = \deg_G(I) \cdot \frac{\binom{k}{|I|}}{\binom{2r}{|I|}}. \tag{3.3}$$

Claim 1.1. *For any graph G, $\mathcal{M} \equiv \mathcal{M}_G$ defined by Equation 3.3 satisfies Equations 3.2.*

Proof. 1. The first set of constraints are:

$$\mathcal{M}(X_I) = 0, \ \forall I, \ |I| \le 2r, \ I \text{ is not a clique in } G$$

For any $I \subseteq [n], |I| \le 2r$, and I is not a clique in G. Any superset S of I ($I \subseteq S$) of size $2r$ would note be a clique in G and $deg_G(I) = 0$ which gives us $\mathcal{M}(X_I) = \deg_G(I) \cdot \frac{\binom{k}{|I|}}{\binom{2r}{|I|}} = 0$

2. The second set of constraints are:

$$\mathcal{M}\left(\left(\sum_{i=1}^n x_i - k\right) X_I\right) = 0, \ \forall I, \ |I| < 2r.$$

$$\mathcal{M}\left(\left(\sum_{i=1}^n x_i - k\right) X_I\right) = \mathcal{M}\left(\sum_{i=1}^n x_i X_I\right) - \mathcal{M}(kX_I) = \mathcal{M}\left(\sum_{i \notin I} X_{I \cup \{i\}} + |I| X_I\right) - \mathcal{M}(kX_I)$$

$$= \mathcal{M}\left(\sum_{i\notin I} X_{I\cup\{i\}}\right) + (|I|-k)\,\mathcal{M}\left(X_I\right) = (|I|-k)\cdot deg_G(I)\cdot\frac{\binom{k}{|I|}}{\binom{2r}{|I|}} + \sum_{i\notin I} deg_G(I\cup\{i\})\cdot\frac{\binom{k}{|I|+1}}{\binom{2r}{|I|+1}} = 0$$

The last equality follows by the following observation:

$$deg_G(I) = \frac{1}{2r-|I|}\sum_{i\notin I} deg_G(I\cup\{i\})$$

□

To make the analysis simpler we use the following lemma.

Lemma 1.2. *For any P of degree at most r we may write $P = P_1 + \sum_i P_{2i}(x_i^2 - x_i) + P_3(\sum_i x_i - k)$ where P_1 is multilinear and homogeneous of degree r, P_3 has degree at most $r-1$, and all P_{2i} have degree at most $r-2$.*

We leave the proof of this lemma as an exercise. One of the implications of this lemma is the following corollary.

Corollary 1.2. *If $\mathcal{M}(P_1^2) \geq 0$ for all multilinear homogeneous P_1 of degree r then $\mathcal{M}$ is PSD.*

Proof. For any P of degree at most r we may write

$$P = P_1 + \sum_i P_{2i}(x_i^2 - x_i) + P_3(\sum_i x_i - k)$$

$$\mathcal{M}(P^2) = \mathcal{M}\left(\left(P_1 + \sum_i P_{2i}(x_i^2 - x_i) + P_3(\sum_i x_i - k)\right)^2\right) = \mathcal{M}(P_1^2) \geq 0$$

The second equality is true because the following square and cross terms are mapped to zero.

$$P_1\cdot\sum_i P_{2i}(x_i^2-x_i) + P_1\cdot P_3(\sum_i x_i - k) + \sum_i P_{2i}(x_i^2-x_i)\cdot P_3(\sum_i x_i - k) + (\sum_i P_{2i}(x_i^2-x_i))^2 + (P_3(\sum_i x_i - k))^2$$

Because they are either of the form $\mathcal{M}\left(\left(\sum_{i=1}^n x_i - k\right)X_I\right)$ for some I ($|I| < 2r$) or $\mathcal{M}\left(\left(x_i^2 - x_i\right)X_I\right)$ for some I ($|I| < 2r-1$) which are mapped to zero by the constraints enforced by the clique axioms $\mathrm{Clique}(G,k)$. □

This corollary simplifies our analysis to showing the following moment matrix $M \equiv M_G \in \mathbb{R}^{\binom{[n]}{r}\times\binom{[n]}{r}}$ is PSD with high probability for $G \leftarrow G(n,1/2)$: for $I, J \in \binom{[n]}{r}$,

$$M(I,J) = deg_G(I\cup J)\cdot\frac{\binom{k}{|I\cup J|}}{\binom{2r}{|I\cup J|}}. \tag{3.4}$$

Note the moment matrix M is defined only for the sets $I, J \in \binom{[n]}{r}$. In the remaining part of the lecture, we show that M is PSD with high probability for $k \leq \Omega_r(n^{1/2r}/(\log n)^{1/r})$.

Theorem 1.3 (Main Technical Theorem)**.** *There exists a constant $c > 0$ such that, with high probability over $G \leftarrow G(n,1/2)$, the matrix M_G defined by Equation 3.4 is PSD for $k \leq 2^{-cr}\cdot(\sqrt{n}/\log n)^{1/r}$.*

4 Notations

1. For any set $I \subseteq [n]$, let $\mathcal{E}(I) = \{\{i, j\} : i \neq j \in I\}$.
2. For $0 \leq i \leq r$, let
$$\alpha(i) = \frac{\binom{k}{2r-i}}{\binom{2r}{2r-i}} \cdot \binom{n-2r+i}{i} \cdot 2^{-r^2-\binom{i}{2}}. \tag{4.1}$$
3. For $0 \leq i \leq r$, let
$$\beta(i) = \binom{k}{2r-i} / \binom{2r}{2r-i}. \tag{4.2}$$
4. For $0 \leq i \leq r$, let $p(i) = 2^{-(r-i)^2}$. Then, for $I, J \in \binom{[n]}{r}$ with $|I \cap J| = i$, $p(i)$ is the probability that $\mathcal{E}(I \cup J) \setminus (\mathcal{E}(I) \cup \mathcal{E}(J)) \subseteq G$.

5 Reduction to PSDness of M'

The moment matrix M constructed above contains many zero rows and columns which are difficult to analyse. We define the matrix M' such that if M' is PSD then M is PSD and is much easier to work with.

For every $T \subseteq [n]$, let $M_T \in \mathbb{R}^{\binom{[n]}{r} \times \binom{[n]}{r}}$, with $M_T(I, J) = \beta(|I \cap J|)$ if $I \cup J \subseteq T$, and G contains every edge in $\mathcal{E}(T) \setminus \mathcal{E}(I) \cup \mathcal{E}(J)$. Define M' as:

$$M' = \sum_{T:|T|=2r} M_T. \tag{5.1}$$

The matrix M has a non-zero entry at $M(I, J)$ if $I \cup J$ is a clique in G and from the definition of M' one could work out that $M'(I, J) = M(I, J)$ and M' might fill the remaining zero entries of M with some value ≥ 0.

Lemma 1.3. *If M' is PSD then M is PSD.*

This is true because the non-zero part of M is a principal sub matrix of M' and the non-zero eigenvalues of M are greater than the minimum eigenvalue of M'.

To show PSDness of M' we need the machinery called Johnson Scheme which we describe next.

6 Johnson Scheme

We need the following definitions and results about Johnson Scheme which would be used to prove the PSDness of M'.

Definition 1.5 (Set-Symmetry)**.** *A matrix $M \in \mathbb{R}^{\binom{[n]}{r} \times \binom{[n]}{r}}$ is set-symmetric if for every $I, J \in \binom{[n]}{r}$, $M(I, J)$ depends only on the size of $|I \cap J|$.*

Definition 1.6 (Johnson Scheme)**.** *For $n, r \leq n/2$, let $\mathcal{J} \equiv \mathcal{J}_{n,r} \subseteq \mathbb{R}^{\binom{[n]}{r} \times \binom{[n]}{r}}$ be the subspace of all set-symmetric matrices. $\mathcal{J}$ is called the Johnson scheme.*

Johnson scheme is interesting because it exhibits two interesting Basis.

Definition 1.7 (D-Basis)**.** *For $0 \leq \ell \leq r \leq n$, let $D_\ell \equiv D_{n,r,\ell} \in \mathbb{R}^{\binom{[n]}{r} \times \binom{[n]}{r}}$ be defined by*

$$D_\ell(I, J) = \begin{cases} 1 & |I \cap J| = \ell \\ 0 & \text{otherwise.} \end{cases} \tag{6.1}$$

Definition 1.8 (P-Basis)**.** *For $0 \leq t \leq r$, let $P_t \equiv P_{n,r,t} \in \mathbb{R}^{\binom{[n]}{r} \times \binom{[n]}{r}}$ be defined by*

$$P_t(I, J) = \binom{|I \cap J|}{t}.$$

Equivalently, for $T \subseteq [n]$, if we let P_T be the PSD rank one matrix

$$P_T = 1\left(\{I : I \subseteq [n], I \supseteq T\}\right) \cdot 1\left(\{I : I \supseteq [n], I \subseteq T\}\right)^\dagger,$$

then

$$P_t = \sum_{T : T \subseteq [n], |T| = t} P_T. \tag{6.2}$$

Claim 1.2. *For fixed n, r, the following relations hold:*

1. *For $0 \leq t \leq r$, $P_t = \sum_{\ell=t}^{r} \binom{\ell}{t} D_\ell$.*
2. *For $0 \leq \ell \leq r$, $D_\ell = \sum_{t=\ell}^{r} (-1)^{t-\ell} \binom{t}{\ell} P_t$.*

Lemma 1.4. *Fix $n, r \leq n/2$ and let $\mathcal{J} \equiv \mathcal{J}(n, r)$ be the Johnson scheme. Then, for P_t as defined by Equation 6.2, there exist subspaces $V_0, V_1, \ldots, V_r \in \mathbb{R}^{\binom{[n]}{r}}$ that are orthogonal to one another such that:*

1. *$V_0, \ldots, V_r$ are eigenspaces for $\{P_t : 0 \leq t \leq r\}$ and consequently for all matrices in $\mathcal{J}$.*
2. *For $0 \leq j \leq r$, $dim(V_j) = \binom{n}{j} - \binom{n}{j-1}$.*
3. *For any matrix $Q \in \mathcal{J}$, let $\lambda_j(Q)$ denote the eigenvalue of Q within the eigenspace V_j. Then,*

$$\lambda_j(P_t) = \begin{cases} \binom{n-t-j}{r-t} \cdot \binom{r-j}{t-j} & j \leq t \\ 0 & j > t \end{cases}. \tag{6.3}$$

7 PSDness of M'

The prove the PSDness of M' we break the matrix M' into three matrices:

$$M' = E + L + \Delta$$

where matrix E is the expected matrix with a large minimum eigenvalue (roughly $\Omega_r(k^r n^r)$), matrix L is locally random matrix with bounded spectral norm $\|L\| < Ck^{2r} n^{r-1/2} \log n$ and matrix Δ is a global noise matrix with each entry being small and $\|\Delta\| < Ck^{2r} n^{r-1/2} \log n$.

7.1 Matrix E

In this subsection we define the matrix E and show it has a large minimum eigenvalue. The matrix E is just the expected matrix $E = \mathbb{E}[M']$. One could write down the expected matrix E as follows:

Claim 1.3. *For $I, J \in \binom{n}{r}$, and $E = \mathbb{E}[M']$,*

$$E(I,J) = \binom{n - |I \cup J|}{2r - |I \cup J|} \cdot \frac{\binom{k}{|I \cup J|}}{\binom{2r}{|I \cup J|}} \cdot 2^{-r^2 - \binom{|I \cap J|}{2}}. \tag{7.1}$$

The proof follows by arguing the expected value of $\mathbb{E}[deg_G(I \cup J)]$ for a random $G(n, 1/2)$ and is left as an exercise. Next we show that expected matrix E exhibits a large minimum eigenvalue. To prove this statement we use the machinery of Johnson Scheme defined in previous section.

The first observation is that the expected matrix E is a set symmetric matrix and we can use the machinery of Johnson Scheme. The matrix E can be easily written as the linear combination in D-Basis. To be precise:

$$E = \sum_{\ell} e_\ell D_\ell$$

where $e_\ell = \binom{n - 2r + \ell}{\ell} \cdot \frac{\binom{k}{2r-\ell}}{\binom{2r}{2r-\ell}} \cdot 2^{-r^2 - \binom{\ell}{2}}$ (For matrix D_ℓ, $|I \cup J| = 2r - \ell$ and $|I \cap J| = \ell$).

We understand the eigenvalues of P_t pretty well and we write E as a linear combinations in P-Basis. We already know from above that $D_\ell = \sum_{t=\ell}^{r} (-1)^{t-\ell} \binom{t}{\ell} P_t$ and we can write E as follows:

$$E = \sum \alpha_t P_t$$

One could work out the details to show $\alpha_t = \sum_{\ell=0}^{t} (-1)^{t-\ell} \binom{t}{\ell} e_\ell$. Now write e_ℓ recursively in terms of $e_{\ell-1}$:

$$e_\ell = \frac{n - 2r + \ell}{2^{1-l}(k - 2r + \ell)} \cdot e_{\ell-1}$$

If $k < \frac{n - 2r}{3r \cdot 2^{r-1}}$, then the terms in the sum of α_t increase geometrically by a constant factor and the sum will be dominated by the last term e_t. In particular one could show that $\alpha_t \geq \frac{e_t}{2} > 0$ and

$$\alpha_r \geq \frac{e_r}{2} = 2^{-O(r^2)} k^r n^r$$

Since P_t's are PSD and $P_r = I$ is just the identity matrix, we have that E is PSD with minimum eigenvalue of at least $2^{-O(r^2)} k^r n^r$ which proves the following lemma.

Lemma 1.5. *If $k < \frac{n - 2r}{3r \cdot 2^{r-1}}$ and $r \leq \frac{k}{2}$ then E is PSD with minimal eigenvalue $2^{-O(r^2)} k^r n^r$*

7.2 Bounding the norm of locally random matrix L

Define $L \in \mathbb{R}^{\binom{[n]}{r} \times \binom{[n]}{r}}$ as follows: for $I, J \in \binom{[n]}{r}$,

$$L(I,J) = \begin{cases} \alpha(|I \cap J|) \cdot \dfrac{1 - p(|I \cap J|)}{p(|I \cap J|)} & \text{if } \mathcal{E}(I \cup J) \setminus (\mathcal{E}(I) \cup \mathcal{E}(J)) \subseteq G \\ -\alpha(|I \cap J|) & \text{otherwise} \end{cases}. \tag{7.2}$$

In this section we prove the following main lemma:

Lemma 1.6. *For some constant $C > 0$, with probability at least $1 - 1/n$ over the random graph G,*

$$\|L\| \leq O(1) \cdot 2^{Cr^2} \cdot k^{2r} \cdot n^r \cdot \frac{\log n}{\sqrt{n}}.$$

To prove this lemma we define the following special matrices: for $V, W \in \binom{[n]}{a}$,

$$R_a(V,W) = \begin{cases} 2^{a^2} - 1 & \text{if } V \cap W = \emptyset \text{ and } \{\{v,w\} : v \in V, w \in W\} \subseteq G \\ -1 & \text{if } V \cap W = \emptyset \text{ and } \{\{v,w\} : v \in V, w \in W\} \not\subseteq G \\ 0 & \text{if } V \cap W \neq \emptyset \end{cases}. \tag{7.3}$$

Claim 1.4. *If $n \geq 100$, for all $\varepsilon \in (0,1)$, $\mathsf{Pr}\left[||R_a|| > 2^{a^2+2a+2} \ln(\frac{n}{\varepsilon}) n^{a-\frac{1}{2}}\right] < \varepsilon$.*

The proof relies on the trace of powers of matrix R_a and we ask the reader to refer the paper for the proof. Let us introduce some notations before we proceed:

1. For a matrix $X \in \mathbb{R}^{\binom{[n]}{r_1} \times \binom{[n]}{r_2}}$, and $0 \leq i \leq \min\{r_1, r_2\}$, let $X^i \in \mathbb{R}^{\binom{[n]}{r_1} \times \binom{[n]}{r_2}}$ be the matrix such that $X^i(I,J) = X(I,J)$ if $|I \cap J| = i$ and 0 otherwise
2. For a matrix $X \in \mathbb{R}^{\binom{[n]}{r_1 - i} \times \binom{[n]}{r_2 - i}}$, let $X^{(i)} \in \mathbb{R}^{\binom{[n]}{r_1} \times \binom{[n]}{r_2}}$, be defined as follows:

$$X^{(i)}(I,J) = \begin{cases} X(I \setminus (I \cap J), J \setminus (I \cap J)) & \text{if } |I \cap J| = i \\ 0 & \text{otherwise} \end{cases}. \tag{7.4}$$

The matrices R_a would help us bound the spectral norm of L because we can write L as:

$$L = \sum_{i=0}^{r} L^i$$

$$L^i = \alpha_i \cdot R^{(i)}_{r-i}. \tag{7.5}$$

Recall matrix L is defined as: for $I, J \in \binom{[n]}{r}$,

$$L(I,J) = \begin{cases} \alpha(|I \cap J|) \cdot \dfrac{1 - p(|I \cap J|)}{p(|I \cap J|)} & \text{if } \mathcal{E}(I \cup J) \setminus (\mathcal{E}(I) \cup \mathcal{E}(J)) \subseteq G \\ -\alpha(|I \cap J|) & \text{otherwise} \end{cases}. \tag{7.6}$$

$$L^i(I,J) = L(I,J) \text{ if } |I \cap J| = i \text{ and } 0 \text{ otherwise}$$

Let us prove equation 7.5: for any $I, J \in \binom{[n]}{r}$ and $|I \cap J| = i$,

1. If $\mathcal{E}(I \cup J) \setminus (\mathcal{E}(I) \cup \mathcal{E}(J)) \subseteq G$, then define $V := I \setminus (I \cap J)$, and $W := J \setminus (I \cap J)$ and observe that $V \cap W = \emptyset$ and $\{\{v, w\} : v \in V, w \in W\} \subseteq G$ and

$$R^{(i)}_{r-i}(I,J) = 2^{(r-i)^2-1} = \frac{1-p(i)}{p(i)} = \frac{1-p(|I \cap J|)}{p(|I \cap J|)}$$

   and hence $L_i(I,J) = \alpha(i) R^{(i)}_{r-i}(I,J)$ if $\mathcal{E}(I \cup J) \setminus (\mathcal{E}(I) \cup \mathcal{E}(J)) \subseteq G$.

2. If $\mathcal{E}(I \cup J) \setminus (\mathcal{E}(I) \cup \mathcal{E}(J)) \not\subseteq G$, then as before define $V := I \setminus (I \cap J)$, and $W := J \setminus (I \cap J)$ and observe that $V \cap W = \emptyset$ and $\{\{v, w\} : v \in V, w \in W\} \not\subseteq G$ and

$$R^{(i)}_{r-i}(I,J) = -1$$

   and hence $L_i(I,J) = -\alpha(i) = \alpha(i) \cdot R^{(i)}_{r-i}(I,J)$ if $\mathcal{E}(I \cup J) \setminus (\mathcal{E}(I) \cup \mathcal{E}(J)) \not\subseteq G$

For any other $I, J \in \binom{[n]}{r}$ and $|I \cap J| \neq i$ both $L_i(I,J)$ and $R^{(i)}_{r-i}(I,J)$ are equal to zero and the equation 7.5 is satisfied. All we need now is to bound the spectral norm of matrices $R^{(i)}_{r-i}$ for all i.

We already know from the claim 1.4, the spectral norm of the matrix R_{r-i} is bounded above by $2^{a^2+2a+2} \ln(\frac{n}{\varepsilon}) n^{a-\frac{1}{2}}$ where $a = r - i$, which evaluates to $\leq O(1) 2^{r^2} \frac{\log n \cdot n^r}{\sqrt{n}}$. We next use the following technical lemma to bound the spectral norm of $R^{(i)}_{r-i}$ and we omit the proof of this lemma here.

Lemma 1.7. *For* $0 \leq i \leq \min\{r_1, r_2\}$ *and* $R \in \mathbb{R}^{\binom{[n]}{r_1-i} \times \binom{[n]}{r_2-i}}$, *if* $R = R^0$ *then* $\|R^{(i)}\| \leq \binom{r_1}{i}\binom{r_2}{i} \cdot \|R\|$.

With the help of Lemma 1.7 we have: $\|R^{(i)}_{r-i}\| \leq \binom{r}{i}\binom{r}{i} \cdot \|R_{r-i}\| \leq O(1)\binom{r}{i}\binom{r}{i} 2^{r^2} \frac{\log n \cdot n^r}{\sqrt{n}}$ and $\alpha(i) = \frac{\binom{k}{2r-i}}{\binom{2r}{2r-i}} \cdot \binom{n-2r+i}{i} \cdot 2^{-r^2-\binom{i}{2}}$. Which gives us a bound on the spectral norm of L_i:

$$\|L_i\| \leq \alpha(i) \|R^{(i)}_{r-i}\| \leq O(1) \cdot 2^{Cr^2} \cdot k^{2r-i} \cdot n^r \cdot \frac{\log n}{\sqrt{n}}$$

Which proves our Lemma 1.6 because

$$\|L = \|\sum_{i=0}^{r} L^i\| \leq r \max_i \|L_i\|$$

.

7.3 Bounding the norm of the global error matrix Δ

The global error matrix $\Delta = M' - E - L$. In this section we show the spectral norm of Δ matrix is small. This statement turns out to be true because each entry of the Δ matrix is small with high probability which leads to a smaller spectral norm. We will make these statements more formal in this section.

Let $\mathcal{A}$ be the event that $\mathcal{E}(I \cup J) \setminus (\mathcal{E}(I) \cup \mathcal{E}(J)) \subseteq G$. Conditioned on $\neg\mathcal{A}$, the value of $\Delta(I,J) = M'(I,J) - E(I,J) - L(I,J) = 0 - \alpha(I \cap J) - (-\alpha(|I \cap J|)) = 0$, here $M'(I,J) = 0$ because there exist no $T \supseteq I \cup J$ such that G contains all the edges in $\mathcal{E}(T) \setminus \mathcal{E}(I) \cup \mathcal{E}(J)$ because $\mathcal{E}(I \cup J) \setminus \mathcal{E}(I) \cup \mathcal{E}(J) \subseteq \mathcal{E}(T) \setminus \mathcal{E}(I) \cup \mathcal{E}(J)$ and we have conditioned on $\neg\mathcal{A}$.

All we care about is the following quantity:

$$\mathbb{E}[\Delta(I,J) \mid \mathcal{A}]$$

We want to show that $\Delta(I,J)$ conditioned on $\mathcal{A}$ is very small. First let us explicitly write down the matrix:

$$\Delta(I,J) = \begin{cases} M'(I,J) - \alpha(|I \cap J|)/p(|I \cap J|) & \text{if } \mathcal{E}(I \cup J) \setminus (\mathcal{E}(I) \cup \mathcal{E}(J)) \subseteq G \\ 0 & \text{otherwise} \end{cases}. \tag{7.7}$$

We already know the expected value $\mathbb{E}\deg_G(I \cup J) = 2^{-\binom{2r}{2}+\binom{2r-i}{2}} \cdot \binom{n-2r+i}{i}$. Suppose $\deg_G(I \cup J)$ is close to its expected value:

$$\deg_G(I \cup J) \approx 2^{-\binom{2r}{2}+\binom{2r-i}{2}} \cdot \binom{n-2r+i}{i}$$

$$M'(I,J) = \beta(i) \cdot \deg_G(I \cup J) \approx \beta(i) 2^{-\binom{2r}{2}+\binom{2r-i}{2}} \cdot \binom{n-2r+i}{i} = \alpha(|I \cap J|)/p(|I \cap J|)$$

$$M'(I,J) \approx \alpha(|I \cap J|)/p(|I \cap J|) = \alpha(|I \cap J|)/p(|I \cap J|) + noise$$
$$\Delta(I,J) = M'(I,J) - \alpha(|I \cap J|)/p(|I \cap J|) = noise$$

We need to make each of these statements rigorous and we use concentration inequalities to claim such results and also show that the noise part is small. Below are the precise statements and we omit the proof here:

Claim 1.5. *For some constant $C > 0$,*

$$\Pr\left[\left|\deg_G(I \cup J) - 2^{-\binom{2r}{2}+\binom{2r-i}{2}} \cdot \binom{n-2r+i}{i}\right| > 2(\ln(C/\varepsilon))^2 n^{i-1/2} \mid (I \cup J \textit{ a clique})\right] < \varepsilon.$$

Observe the claim works with conditioning on $\mathcal{A}$ to show concentration bounds. Choosing $\varepsilon = 1/n^{2r+1}$ and applying a union bound over all sets I, J we have with high probability:

$$|\Delta(I,J)| = \left|M'(I,J) - \beta(i) 2^{-\binom{2r}{2}+\binom{2r-i}{2}} \cdot \binom{n-2r+i}{i}\right| < Cr2^{2r^2} \cdot k^{2r-i} \cdot (\log n) \cdot n^{i-1/2}.$$

is small conditioned on $\mathcal{A}$. This concentration implies the lemma below:

Lemma 1.8. *For some universal constant C, and $n > C2^{4r^2}$, with probability at least $1 - 1/n$ over the random graph G, for all $I, J \in \binom{[n]}{r}$, with $i = |I \cap J|$,*

$$|\Delta(I, J)| \leq 2^{Cr^2} \cdot k^{2r-i} \cdot n^i \cdot \frac{\log n}{\sqrt{n}}.$$

Since each entry is small, it is now easy to argue the spectral norm of the matrix Δ and we omit the proof of the following Lemma:

Lemma 1.9. *For $n > C2^{4r^2}$, with probability at least $1 - 1/n$ over the random graph G,*

$$\|\Delta\| \leq 2^{Cr^2} \cdot k^{2r} \cdot n^r \cdot \frac{\log n}{\sqrt{n}}.$$

7.4 Combining all the results

We have

$$M' = E + L + \Delta$$

Assuming all the necessary conditions used while proving the lemmas. We have with high probability (at least $1 - 1/n$):

1. The minimum eigenvalue of E is at least $2^{-O(r^2)}k^r n^r$.
2. $$\|L\| \leq O(1) \cdot 2^{Cr^2} \cdot k^{2r} \cdot n^r \cdot \frac{\log n}{\sqrt{n}}.$$
3. $$\|\Delta\| \leq 2^{Cr^2} \cdot k^{2r} \cdot n^r \cdot \frac{\log n}{\sqrt{n}}.$$

which implies $\|M'\| \geq 2^{-O(r^2)}k^r n^r - 2^{Cr^2} \cdot k^{2r} \cdot n^r \cdot \frac{\log n}{\sqrt{n}} - 2^{Cr^2} \cdot k^{2r} \cdot n^r \cdot \frac{\log n}{\sqrt{n}}$, just noting down the higher order terms:

$$\|M'\| \geq k^r n^r - k^{2r} n^{r-1/2}$$

For M' to be PSD we need $k^r n^r - k^{2r} n^{r-1/2} \geq 0$ which is true when $k \leq n^{1/2r}$ and PSDness of M' implies the PSDness of M.

References

[AKS98] Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. *Random Structures Algorithms*, 13(3-4):457–466, 1998.

[BHK+16] Boaz Barak, Samuel B. Hopkins, Jonathan A. Kelner, Pravesh Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *CoRR*, abs/1604.03084, 2016.

[Jer92] Mark Jerrum. Large cliques elude the metropolis process. *Random Structures and Algorithms*, 3(4):347–359, 1992.

[Ku95] Ludk Kuera. Expected complexity of graph partitioning problems. *Discrete Applied Mathematics*, 57(2):193 – 212, 1995. Combinatorial optimization 1992.

[MPW15] Raghu Meka, Aaron Potechin, and Avi Wigderson. Sum-of-squares lower bounds for planted clique. *CoRR*, abs/1503.06447, 2015.