Foundations of Software Technology and Theoretical Computer Science (Bangalore) 2008.
Editors: R. Hariharan, M. Mukund, V. Vinay; pp 187-198

A new approach to the planted clique problem

Alan Frieze[1]*,**Ravi Kannan**[2]

[1] Department of Mathematical Sciences,
Carnegie Mellon University.
Pittsburgh PA15213
USA
alan@random.math.cmu.edu

[2] Microsoft Research Laboratories,
India
kannan100@gmail.com

1 Introduction

It is well known that finding the largest clique in a graph is NP-hard, [8]. Indeed, Hastad [5] has shown that it is NP-hard to approximate the size of the largest clique in an n vertex graph to within a factor $n^{1-\epsilon}$ for any $\epsilon > 0$. Not surprisingly, this has directed some researchers attention to finding the largest clique in a random graph. Let $G_{n,1/2}$ be the random graph with vertex set $[n]$ in which each possible edge is included/excluded independently with probability $1/2$. It is known that **whp** the size of the largest clique is $(2 + o(1)) \log_2 n$, but no known polymomial time algorithm has been proven to find a clique of size more than $(1 + o(1)) \log_2 n$. Karp [9] has even suggested that finding a clique of size $(1 + \epsilon) \log_2 n$ is computationally difficult for any constant $\epsilon > 0$.

Significant attention has also been directed to the problem of finding a *hidden* clique, but with only limited success. Thus let G be the union of $G_{n,1/2}$ and an unknown clique on vertex set P, where $p = |P|$ is given. The problem is to recover P. If $p \geq c(n \log n)^{1/2}$ then, as observed by Kucera [10], with high probability, it is easy to recover P as the p vertices of largest degree. Alon, Krivelevich and Sudakov [1], using spectral analysis, were able to improve this to $p = \Omega(n^{1/2})$. McSherry [11] gives some refinements of this method. In conjunction with a negative result of Jerrum [6] that one possible Markov chain approach fails for $p = o(n^{1/2})$, $p = \Omega(n^{1/2})$ seems like a natural barrier for solving this problem. Feige and Krauthgamer [4] considered finding a planted clique in the context of the semi-random model. Juels and Peinado [7] considered the application of this problem to Cryptographic Security.

Let A_G denote the adjacency matrix of G. The spectral approach of [1] essentially maximizes $x^T A_G x$ over vectors x with $|x| = 1$, expecting that the optimal solution is close to u, defined by $u_i = p^{-1/2} 1_{i \in P}$, ($u$ is the scaled characteristic vector of P) so that we may recover P from the optimal solution.

*Supported in part by NSF grant ccr0200945

FSTTCS 2008
IARCS Annual Conference on
Foundations of Software Technology and Theoretical Computer Science
http://drops.dagstuhl.de/opus/volltexte/2008/1752

In this paper, we define a natural 3-dimensional array A related to the given graph : $A_{i,j,k}$ will be ± 1 depending on whether the parity of the number of edges among the vertices i, j, k is odd or even respectively. Our main result here (Section 2) shows that as long as $p = \Omega(n^{1/3}(\log n)^4)$, the maximum of the cubic form or *tensor* $A(x,x,x) = \sum_{i,j,k} A_{i,j,k}x_ix_jx_k$, $x \in B_n = \{x \in R^n : |x| = 1\}$ is attained close to u. Thus if we can find this maximimum, then we can recover the clique. However, unlike the case of the quadratic form, where the maximization is an eigenvalue computation which is well-known to be solvable in polynomial time, there are in general no known polynomial time algorithms for maxmizing cubic forms. So, our existential result does not automatically lead to an algorithm and this is left as an open question. We make the following conjecture which would yield an algorithm if proved.

Conjecture *Suppose that an $n \times n \times n$ array A is constructed as above from $G_{n,1/2}$ plus a planted clique of size $p \in \Omega(n^{1/3}(\log n)^c)$. Then the function $A(x,x,x)$ has a unique local maximum as x varies over B_n.*

2 The cubic form and the main result

We define the 3-dimensional array :

$$A_{i,j,k} = \begin{cases} 1 & \text{if } i,j,k \text{ are distinct and } G \text{ contains 1 or 3 edges of the triangle } i,j,k. \\ -1 & \text{if } i,j,k \text{ are distinct and } G \text{ contains 0 or 2 edges of the triangle } i,j,k. \\ 0 & \text{if } i,j,k \text{ are not distinct.} \end{cases}$$

We assume that

$$p = C_1 n^{1/3}(\log n)^4.$$

Here $C_1, C_2, \ldots,$ are unspecified positive absolute constants.

For vectors x, y, z, we define

$$A(x,y,z) = \sum_{i,j,k} A_{i,j,k}x_iy_jz_k.$$

x, y, z will denote vectors of length 1 throughout. We will reserve u for the scaled characteristic vector of P defined earlier. The following Theorem (which is the Main Theorem of the paper) will imply (see Corollary 2 below) that if at least one of x, y, z is orthogonal to u, then we have $|A(x,y,z)| \le C_2 n^{1/2}(\log n)^4$. In which case,

$$A(u,u,u) = \frac{p(p-1)(p-2)}{p^{3/2}} \sim p^{3/2} = \omega(A(x,y,z))$$

for all such x, y, z. (We use the notation $a_n = \omega(b_n)$ to mean that $a_n/b_n \to \infty$ as $n \to \infty$).

Let

$$P^{3*} = \{(i,j,k) \in P^3 : i,j,k \text{ are distinct}\}$$

Define the 3-dimensional matrix D by

$$D_{i,j,k} = \begin{cases} 1 & (i,j,k) \in P^{3*}, \\ 0 & \text{otherwise} \end{cases}$$

and let $B = A - D$.

$$B(x,y,z) = A(x,y,z) - \sum_{i,j,k \in P^{3*}} x_i y_j z_k. \tag{1}$$

The entries of A in $P \times P \times P$ contribute $\sum_{(i,j,k)\in P^{3*}} x_i y_j z_k$ to the tensor $A(x,y,z)$; so $B(x,y,z)$ is the contribution due to the random graph alone. The proof of Theorem 1 occupies all of Section 3. We defer the proofs of the corollaries following it to Section 4.

THEOREM 1. *There exists C_3 such that*

$$\mathbf{Pr}\left(\exists x,y,z:\ |B(x,y,z)| \geq C_3 n^{1/2}(\log n)^4\right) = o(1).$$

Let

$$U^* = \{(x,y,z):\ x.u = 0 \text{ or } y.u = 0 \text{ or } z.u = 0\}.$$

COROLLARY 2. *If $(x,y,z) \in U^*$ then*

$$|A(x,y,z)| \leq 2C_3 n^{1/2}(\log n)^4. \tag{2}$$

So, **whp** , we have that

$$A(u,u,u) = \omega\left(\max_{(x,y,z)\in U^*} A(x,y,z)\right). \tag{3}$$

COROLLARY 3. *Suppose the maximum of the multilinear form $A(x,y,z)$ as x,y,z vary over the unit ball is attained at x^*,y^*,z^*. Then,* $\min\{x^*\cdot u, y^*\cdot u, z^*\cdot u\} = 1 - o(1)$.

The above corollary ensures that from x^*,y^*,z^*, we can find the clique P using the Theorem below. (See Section 4.)

THEOREM 4. *There is a polynomial time algorithm which given as input a unit vector v, returns a set P' of cardinality p satisfying the following: If $v\cdot u \geq \frac{C_4 \log n}{p^{1/2}}$, for sufficiently large C_4 then $P' = P$.*

Observe that it is trivial to get a vector v satisfying $v\cdot u \geq 1/p^{1/2}$ by trying out all n unit vectors. Getting a vector v satisfying the hypothesis of the Theorem in polynomial time, however, seems to be non-trivial.

Remarks: We can assume that $x^* = y^* = z^*$ in Corollary 3. Indeed, for a fixed x, the problem of maximising $A(x,y,z)$ over the unit ball B_n amounts to maximizing $y^T A_x z$ for $y,z \in B_n$. Here A_x is the $n\times n$ matrix defined by $A_x(i,j) = \sum_k A_{i,j,k}x_k$. A_x is a symmetric matrix and so for each x there is a maximum in which $y = z$. Now define a sequence of vector triples $x_k,y_k,z_k, k = 0,1,2,\ldots$, where $x_0,y_0,z_0 = x^*,y^*,z^*$ and $x_1 = x_0$ and $y_1 = z_1$ maximise $y^T A_{x_1} z$ over B_n. Now to obtain $x_2,y_2 = y_1,z_2$ we find $x = z$ to maximise $A(x,y_1,z$ and so on. Any limit point of this sequence $\hat{x},\hat{y},\hat{z}$ must maximise $A(x,y,z)$ and must have $\hat{x} = \hat{y} = \hat{z}$. If for example, $\hat{x} \neq \hat{y}$ then we have the contradiction that there are points of the form ξ,ξ,η arbitrarily close $\hat{x},\hat{y},\hat{z}$.

Remarks: By switching from 2-dimensional matrices to 3-dimensional matrices we have reduced the necessary size of P from $\tilde{O}(n^{1/2})$ to $\tilde{O}(n^{1/3})$. An interesting open question is whether using the natural k-dimensional matrices (whose entries are ± 1 depending on the parity of the number of edges of G in the induced sub-graphs on k vertices) will allow us to go down to $\tilde{O}(n^{1/k})$, for any fixed positive integer k.

Remarks: We note that x^* is a local maximum of the function $A(x,x,x)$ (with respect to first and second order moves) over the unit ball iff

1. x^* is the eigenvector corresponding to the highest eigenvalue of the matrix $A(x^*)$ and
2. the second highest eigenvalue of $A(x^*)$ is at most half the highest.

We can assume that $|x| = 1$. Let $F(x) = A(x,x,x)$ and let h be small and let $x \cdot h = 0$. Then we write $F\left(\frac{x+h}{|x+h|}\right) \leq F(x)$ as

$$F(x) + 3A(x,x,h) + 3A(x,h,h) + O(|h|^3) \leq F(x)(1 + 3|h|^2/2 + O(|h|^4).$$

Then we will need $x \cdot h = 0$ implies $A(x,x,h) = 0$ and $\max_h A(x,h,h) = \lambda_2(A_x)|h|^2$.)

3 Proof of Theorem 1

We will have to make a series of technical modfications. These modifications reduce proving Theorem 1 to Lemma 6 below. In the next Section 3.1, we carry out the central part, namely the proof of Lemma 6.

The first modification is that it is easy to see that if we set to zero all the x_i for which $|x_i| \leq 1/n^2$, as well as similarly for y, z, then the RHS of (1) changes by at most 1. So we will assume that either $x_i = 0$ or $|x_i| \geq 1/n^2$, and similarly for y, z.

Now, here is our second technical modification: Let V_1, V_2, V_3 form an arbitrary partition of V into three subsets, each of size $m = n/3$. Noting that by symmetry, each triangle i, j, k appears in the same number of $V_1 \times V_2 \times V_3$, one can see that

$$\sum_{(i,j,k)} B_{i,j,k} x_i y_j z_k \leq \frac{27}{\binom{n}{m,m,m}} \sum_{V_1,V_2,V_3} \sum_{(i,j,k) \in V_1 \times V_2 \times V_3} B_{i,j,k} x_i y_j z_k$$

So,

$$\left| \sum_{(i,j,k)} B_{i,j,k} x_i y_j z_k \right| \leq \frac{27}{\binom{n}{m,m,m}} \sum_{V_1,V_2,V_3} \left| \sum_{(i,j,k) \in V_1 \times V_2 \times V_3} B_{i,j,k} x_i y_j z_k \right| \tag{4}$$

Now for any x, y, z we have

$$|\sum_{(i,j,k) \in V_1 \times V_2 \times V_3} B_{i,j,k} x_i y_j z_k| \leq (\sum_i |x_i|)(\sum_j |y_j|)(\sum_k |z_k|) \leq n^{3/2}. \tag{5}$$

We will prove below that for each **fixed** partition of V into three equal sized subsets - V_1, V_2, V_3, we have,

$$\mathbf{Pr}\left(\max_{x,y,z} \left| \sum_{(i,j,k) \in V_1 \times V_2 \times V_3} B_{i,j,k} x_i y_j z_k \right| \geq C_5 n^{1/2} (\log n)^4 \right) \leq \frac{1}{n^6}. \tag{6}$$

One can derive Theorem 1 from (4), (5) and (6) by the following simple argument: Say that a partition V_1, V_2, V_3 is bad for A, if $\max_{x,y,z}\left|\sum_{(i,j,k)\in V_1\times V_2\times V_3} B_{i,j,k}x_iy_jz_k\right| \geq C_5n^{1/2}(\log n)^4$ and we let $\mathcal{P}_B$ denote the set of bad partitions. Let

$$g(A) = \frac{|\mathcal{P}_B|}{\binom{n}{m,m,m}}.$$

Then, we know that $\mathbf{E}_A(g(A)) \leq 1/n^6$ from which it follows by Markov inequality that

$$\mathbf{Pr}_A\left(g(A) \geq \frac{100}{n^4}\right) \leq \frac{1}{100n^2}.$$

For any A with $g(A) \leq 100/n^4$, we have from (5)

$$\sum_{V_1,V_2,V_3}\max_{x,y,z}\left|\sum_{(i,j,k)\in V_1\times V_2\times V_3} B_{i,j,k}x_iy_jz_k\right| \leq \left(C_5n^{1/2}(\log n)^4 + \frac{100}{n^4}n^{3/2}\right)\binom{n}{m,m,m}$$

and Theorem 1 follows.

To prove (6), we fix attention from now on on one particular V_1, V_2, V_3. We let

$$X(x,y,z) = \sum_{(i,j,k)\in V_1\times V_2\times V_3} B_{i,j,k}x_iy_jz_k$$

and

$$(x^*,y^*,z^*) = \text{argmax}_{x,y,z}|X(x,y,z)|$$

and suppose that

$$|X(x^*,y^*,z^*)| \geq C_5n^{1/2}(\log n)^4. \tag{7}$$

For sets $R \subseteq V_1, S \subseteq V_2, T \subseteq V_3$ of vertices, we let $\mathbf{B}(R,S,T)$ be the set of triples of vectors (x,y,z) satisfying

$$\begin{aligned}
&|x|,|y|,|z| \leq 1.\\
&R = \{i : x_i \neq 0\}, \quad S = \{j : y_j \neq 0\}, \quad T = \{k : z_k \neq 0\}.\\
&|x_i/x_j| \leq 2,\ \forall i,j \in R,\ |y_i/y_j| \leq 2,\ \forall i,j \in S,\ |z_i/z_j| \leq 2,\ \forall i,j \in T.
\end{aligned}$$

Note that this implies

$$|x_i| \leq \frac{2}{|R|^{1/2}},\ |y_i| \leq \frac{2}{|S|^{1/2}},\ |z_i| \leq \frac{2}{|T|^{1/2}},\ \forall i. \tag{8}$$

Since $\frac{1}{n^2} \leq |x_i^*|, |y_j^*|, |z_k^*| \leq 1$, we can write each of x^*, y^*, z^* as the sum of $\log_2(n^2)$ vectors, each of which has the property that its non-zero components are within a factor of 2 of each other. Thus, (7) implies that there exist R, S, T such that

$$\max_{(x,y,z)\in\mathbf{B}(R,S,T)}|X(x,y,z)| \geq C_6n^{1/2}\log n.$$

So, we see that (7) would lead to the non-occurrence of the event $\mathcal{A}$ in the following Lemma.

LEMMA 5. *For every fixed partition of V into three equal sized sets V_1, V_2, V_3, we have that with probability at least $1 - \frac{1}{n^6}$, the following event $\mathcal{A}$ holds:*

$\mathcal{A}$: For all R, S, T, $R \subseteq V_1, S \subseteq V_2, T \subseteq V_3$,

$$\max_{(x,y,z)\in \mathbf{B}(R,S,T)} |X(x,y,z)| < C_6 n^{1/2} \log n.$$

This in turn will follow from the next lemma:

LEMMA 6. *Suppose R, S, T are fixed pair-wise disjoint subsets of vertices, with $|R| = r, |S| = s, |T| = t$. Then with probability at least $1 - n^{-6(r+s+t)}$, the following event which we will call $\mathcal{A}_{R,S,T}$ happens:*

$$\max_{(x,y,z)\in \mathbf{B}(R,S,T)} |X(x,y,z)| \geq C_6 n^{1/2} \log n.$$

Lemma 5 follows from Lemma 6 by the following argument: For each set of integers r, s, t, the number of subsets (R, S, T) of $\{1, 2, \ldots n\}$ with $|R| = r, |S| = s, |T| = t$ is at most n^{r+s+t}. Thus we will concentrate on proving Lemma 6.

3.1 Proof of Lemma 6

Note that R can be partitioned into two parts - $R \cap P$ and $R \setminus P$, similarly also S, T. So, it suffices to prove that for any fixed R, S, T, each either contained in P or disjoint from P, the following event $\mathcal{B}_{R,S,T}$ happens with probability at least $1 - n^{-6(r+s+t)}$:

$$\mathcal{B}_{R,S,T} : \max_{x,y,z\in \mathbf{B}(R,S,T)} |X(x,y,z)| \leq C_7 n^{1/2} \log n.$$

If $R, S, T \subseteq P$, then $X(x, y, z) = 0$. So, we may assume in what follows that

$$(R \subseteq P \text{ or } R \cap P = \emptyset),\ (S \subseteq P \text{ or } S \cap P = \emptyset),\ (T \subseteq P \text{ or } T \cap P = \emptyset),\ (R \cup S \cup T \not\subseteq P)$$

We consider the following cases, which up to re-naming of R, S, T are exhaustive:

Case 1: $S, T \subseteq P$ and $R \cap P = \emptyset$ and $|R| \leq \max\{|S|, |T|\} \leq |P|$.

In this case we use the Azuma-Hoeffding martingale tail inequality, see for example [3]. We have $\mathbf{E}(X) = 0$ and $X = X(x, y, z)$ is determined by $r(s + t)$ independent random variables (the edges in $R \times (S \cup T)$). Now adding or removing an edge in $R \times S$ (resp. $R \times T$) can change X by at most $\frac{8t}{(rst)^{1/2}}$ (resp. $\frac{8s}{(rst)^{1/2}}$) (recall (8)). Applying the inequality we see that

$$\mathbf{Pr}(|X| \geq C_6 n^{1/2} \log n) \leq 2 \exp\left\{-\frac{C_7 n (\log n)^2}{s+t}\right\} \leq n^{-20(r+s+t)}. \tag{9}$$

(Remember that $r, s, t \leq p = n^{1/3+o(1)}$).

The above deals with one particular $x, y, z \in \mathbf{B}(R, S, T)$.

Note next that there is a $1/(r + s + t)^2$-net $\mathcal{L}$ of $\mathbf{B}(R, S, T)$ of size at most $O((r + s + t)^{6(r+s+t)})$. (I.e., there is a set $\mathcal{L}$ of $O((r + s + t)^{6(r+s+t)})$ elements of $\mathbf{B}(R, S, T)$ so that for

each element (x,y,z) of $\mathbf{B}(R,S,T)$, there is some element (x',y',z') of $\mathcal{L}$ such that $|(x-x',y-y',z-z')| \le 1/(r+s+t)^2)$. Now, (9) implies that

$$\Pr\left(\exists (x',y',z') \in \mathcal{L} : |X(x',y',z')| \ge C_6 n^{1/2}\log n\right) \le n^{-12(r+s+t)}.$$

Lemma 6 follows from this and

$$\begin{aligned}|A(x,y,z)-A(x',y',z')| \le \\ |A(x,y,z)-A(x',y,z)|+|A(x',y,z)-A(x',y',z)|+|A(x',y',z)-A(x',y',z')| \\ \le \frac{4rst}{(r+s+t)^2}\left(\frac{1}{(st)^{1/2}}+\frac{1}{(rt)^{1/2}}+\frac{1}{(rs)^{1/2}}\right).\end{aligned}$$

Case 2 $|R| \ge |S|,|T|$ and either (i) $R \subseteq P$ and $S \cap P = T \cap P = \emptyset$ or (ii) $R \cap P = \emptyset$.

In either of the two sub-cases (i) and (ii), all the edges in G from $R \times (S \cup T)$ are from the random graph, not from the planted clique. Also, fix attention on one particular $(x,y,z) \in \mathbf{B}(R,S,T)$.

In this case, to prove an upper bound on $|X(x,y,z)|$, we bound its ℓth moment, where ℓ is an even integer to be chosen later.

Let I be the set of triples (i,j,k), where i,j,k are distinct and at most 2 of them are in P. Let Ω_ℓ denote the set of ordered sequences of ℓ triangles $T_1,T_2,\ldots,T_\ell$ where $T_i \in I \cap (R \times S \times T)$ for $i = 1,2,\ldots,\ell$. Let $X = X(x,y,z)$. We have

$$\mathbf{E}(X^\ell) = \sum_{\mathcal{T}\in\Omega_\ell} \mathbf{E}\left(\prod_{i=1}^{\ell} A(T_i)\right)\prod_{i=1}^{\ell} Z(T_i). \tag{10}$$

where if $T_i = (\alpha,\beta,\gamma)$ then $A(T_i) = A_{\alpha,\beta,\gamma}$ and $Z(T_i) = x_\alpha y_\beta z_\gamma$.

Consider an edge $e \in R \times (S \cup T)$ such that e appears in an odd number of triangles in $\mathcal{T}$. If we consider the measure preserving map f_e which deletes e if it appears in G and adds it otherwise then we see that

$$\prod_{i=1}^{\ell} A(f_e(T_i)) = -\prod_{i=1}^{\ell} A(T_i)$$

and so $\mathbf{E}\left(\prod_{r=1}^{\ell} A(T_r)\right) = 0$. This implies that it is sufficient to sum over those $\mathcal{T}$ in which each edge of $R \times (S \cup T)$ appears an even number of times. Let $\Omega^*_\ell(R,S,T)$ denote the set of ordered sequences $(i_1,j_1,k_1),\ldots,(i_\ell,j_\ell,k_\ell) \in (I \cap (R \times S \times T))^\ell$ such that each pair $(i,j) \in R \times S$ and each pair $(i,k) \in R \times T$ appears an even number of times.

LEMMA 7.

$$|\Omega^*_\ell(R,S,T)| \le \ell!\binom{\ell+r-1}{r-1}(4st)^{\ell/2}.$$

Proof Fix $d_i \ge 0, i \in R$ and let us first count the sequences in $\Omega'_\ell(R,S,T)$ in which $i \in R$ appears d_i times. Note that $\sum_{i\in R} d_i = \ell$. Now fix $i \in R$ and consider the d_i triangles $(i,s_1,t_1),\ldots(i,s_{d_i},t_{d_i})$ which contain i. Then consider the bipartite multigraph Γ on $S \cup T$

with edges $(s_1, t_1), \ldots, (s_{d_i}, t_{d_i})$. By assumption, each vertex of Γ is of even degree and so by Lemma 8 (below) there are at most $(4st)^{d_i/2}$ choices for Γ. Multiplying over i we see that there are at most $(4st)^{\ell/2}$ choices for any given sequence $d_1, \ldots, d_r$. The number of choices for $d_1, \ldots, d_r$ is at most $\binom{\ell+r-1}{r-1}$ and the lemma follows by multiplying by $\ell!$ to get an ordered sequence. □

Let $N(s, t, \mu)$ denote the number of bipartite multigraphs with vertex sets S, T on the two sides, with μ edges and such that each vertex has even degree.

LEMMA 8.

$$N(s, t, \mu) \leq (4st)^{\mu/2}.$$

Proof First note that for $f \geq 1$

$$\frac{2^{2f}}{2f^{1/2}} \leq \frac{(2f)!}{(f!)^2} \leq 2^{2f}.$$

Let $2e_1, 2e_2, \ldots, 2e_s$ and $2f_1, 2f_2, \ldots, 2f_t$ denote the degrees of vertices in S, T respectively. Then

$$\begin{aligned}
N(s,t,\mu) &\leq \sum_{\substack{2e_1+\cdots+2e_s=\mu \\ 2f_1+\cdots+2f_t=\mu}} \mu! \min\left\{\prod_{i\in S}\frac{1}{(2e_i)!}, \prod_{j\in T}\frac{1}{(2f_j)!}\right\} \\
&\leq \sum_{\substack{2e_1+\cdots+2e_s=\mu \\ 2f_1+\cdots+2f_t=\mu}} \mu! \left(\prod_{i\in S}\frac{1}{(2e_i)!}\prod_{j\in T}\frac{1}{(2f_j)!}\right)^{1/2} \\
&\leq \sum_{\substack{2e_1+\cdots+2e_s=\mu \\ 2f_1+\cdots+2f_t=\mu}} (\mu/2)!^2 2^\mu \prod_{i\in S}\frac{2^{1/2}e_i^{1/4}}{2^{e_i}e_i!}\prod_{j\in T}\frac{2^{1/2}f_j^{1/4}}{2^{f_j}f_j!} \\
&\leq 2^\mu \left(\sum_{e_1+\cdots+e_s=\mu/2}(\mu/2)!\prod_{i\in S}\frac{1}{e_i!}\right)\left(\sum_{f_1+\cdots+f_t=\mu/2}(\mu/2)!\prod_{j\in T}\frac{1}{f_j!}\right) \\
&= 2^\mu s^{\mu/2}t^{\mu/2},
\end{aligned}$$

the last because $\left(\sum_{e_1+\cdots+e_t=\mu/2}(\mu/2)!\prod_{j\in T}\frac{1}{e_j!}\right)$ is the number of ways of parititioning the set $\{1, 2, \ldots \mu/2\}$ into t subsets and this number also equals $t^{\mu/2}$. □

Thus,

$$\begin{aligned}
\mathbf{E}(X^\ell) &= \sum_{\mathcal{T}\in\Omega_\ell^*}\mathbf{E}\left(\prod_{r=1}^{\ell}A(T_r)\right)\prod_{r=1}^{\ell}Z(T_r) \\
&\leq |\Omega_\ell^*|\cdot\frac{8}{(rst)^{\ell/2}} \\
&\leq \binom{\ell+r-1}{r-1}\cdot\frac{2^{\ell+3}\ell!}{r^{\ell/2}} \\
&\leq \frac{2^{\ell+4}\ell^{\ell+1/2}e^r}{r^{\ell/2}}.
\end{aligned}$$

Now ℓ even implies that $X^\ell \geq 0$ and so applying the Markov inequality, we see that for any $\xi > 0$,

$$\mathbf{Pr}(X > \xi) \leq \frac{2^{\ell+4}\ell^{\ell+1/2}e^r}{\xi^\ell r^{\ell/2}}.$$

Putting $\xi = C_6 n^{1/2}\log n$ and $\ell = (r+s+t)\log n$, we see that

$$\mathbf{Pr}(X(x,y,z) \geq C_6 n^{1/2}\log n) \leq n^{-20(r+s+t)}. \tag{11}$$

This completes the proof of Lemma 6.

4 Proof of the Corollaries

Corollary 2 follows from Theorem 1 and the following:

$$\left|\sum_{i,j,k\in P^{3*}} x_i y_j z_k\right| \leq \left|\left(\sum_{i\in P} x_i\right)\left(\sum_{j\in P} y_j\right)\left(\sum_{k\in P} z_k\right)\right| + |y\cdot z|\left|\sum_P x_i\right| + |x\cdot z|\left|\sum_P y_j\right| + |x\cdot y|\left|\sum_P z_k\right| + \left|\sum_{i\in P} x_i y_i z_i|\right| \leq 3p^{1/2}.$$

□

For Corollary 3 we write $x^* = (x^*\cdot u)u + x'$, where x' is orthogonal to u, similarly for y^*, z^*. This splits $A(x^*,y^*,z^*)$ into the sum of 8 parts. Using (3), we get

$$A(u,u,u) \leq A(x^*,y^*,z^*) \leq o(A(u,u,u)) + (x^*\cdot u)(y^*\cdot u)(z^*\cdot u)A(u,u,u),$$

and the corollary follows. □

5 Proof of Theorem 4

Now, we prove Theorem 4. Let v with $|v| = 1$ be the given vector. Define a vector w by: $w_i = \max(v_i, 0)$. Clearly, $\sum_{i\in P} w_i \geq \sum_P v_i$. For ease of notation, we re-number the indices of coordinates so that $w_1 \geq w_2 \geq \ldots w_n$. Since v is given, we can explicitly do this reordering. Also for convenience, we let $w_{n+1} = 0$. After this renumbering, we let

$$S_k = \{1,2,\ldots k\}, \qquad T_k = S_k \cap P, \qquad t_k = |T_k| \qquad k = 1,2,\ldots,n. \tag{12}$$

LEMMA 9. *If $\sum_{i\in P} v_i \geq C_8 \log n$, then for some integer k,*

$$t_k \geq C_8\sqrt{k\log n}/3.$$

Proof Assume for the sake of contradiction that $\sum_{i\in P} v_i \geq C_8 \log n$ and that for all k, $t_k < C_8\sqrt{k\log n}/3$.

$$\begin{aligned}\sum_{i\in P} w_i &= \sum_{k=1}^{n} t_k\,(w_k - w_{k+1}) \leq \frac{1}{3}C_8\sqrt{\log n}\sum_{k=1}^{n}\sqrt{k}(w_k - w_{k+1}) \\ &= \frac{1}{3}C_8\sqrt{\log n}\sum_{k=1}^{n} w_k(\sqrt{k}-\sqrt{k-1}) \leq \frac{2}{3}C_8\sqrt{\log n}\sum_{k=1}^{n}\frac{w_k}{\sqrt{k}} \\ &\leq \frac{2}{3}C_8\sqrt{\log n}|w|\left(\sum_{k=1}^{n}\frac{1}{k}\right)^{1/2} \leq \frac{3}{4}C_8\log n,\end{aligned}$$

using $\frac{2}{\sqrt{k}} \geq \sqrt{k}-\sqrt{k-1}$ and also the Cauchy-Scwartz inequality. This contradiction proves the Lemma. □

Let G be the graph we are given (the random graph plus the planted clique.) Let M be its adjacency matrix, where we put a +1 for an edge and -1 for a non-edge. For a subset S of V, let G^S denote the induced subgraph on S and M^S the $|S|\times|S|$ adjacency matrix of G^S. (In our definition of adjacency matrix, we have 1's on the diagonal). We may write

$$M = puu^T + \hat{M} - \tilde{M}, \tag{13}$$

where $\hat{M}$ is the adjacency matrix of the random graph and $\tilde{M}$ is the adjacency matrix of the sub-graph induced on P of the random graph. [$\tilde{M}$ has 0 entries outside $P\times P$.] We may similary write for any $S\subseteq V$,

$$M^S = tu^S u^{S^T} + \hat{M}^S - \tilde{M}^S, \tag{14}$$

where $|S\cap P| = t$ and u^S denotes the vector with $1/\sqrt{t}$ in the $S\cap P$ positions and 0 elsewhere.

LEMMA 10. *With probability at least $1-n^{-3}$, we have that for all $S\subseteq V$,*

$$\max\{\lambda_1(\hat{M}^S), \lambda_1(\tilde{M}^S)\} \leq 100\sqrt{|S|\log n}$$

where λ_1 denotes the largest absolute value of an eigenvalue.

Proof For each fixed S, the matrix $\hat{M}^S$ is a random symmetric matrix. It is known [2] that with probability at least $1-4e^{-10|S|\log n}$, we have that $|\lambda_1(\hat{M}^S)| \leq 100\sqrt{|S|\log n}$. For each $s\in\{1,2,\ldots n\}$, there are at most n^s subsets S of V with $|S|=s$. So the probability that the assertion of the Lemma does not hold is at most $\sum_{s=1}^{n} n^s e^{-10s\log n} \leq 1/(2n^3)$. $\tilde{M}^S$ is dealt with similarly.

□

For notational convenience, we let M^k denote M^{S_k} (see (12)) and similarly for $\hat{M}^k, \tilde{M}^k$. The first step of our algorithm is to run through $k=1,2,\ldots n$, find $\lambda_1(M^k)$ and stop when for the first time, we find a k such that

$$\lambda_1(M^k) \geq 1000\sqrt{k\log n}. \tag{15}$$

LEMMA 11.
(i) *If $C_8 \geq 3000$ then the algorithm will find a k satisfying (15).*
(ii) *For any k satisfying (15), we have:*
(a) *if a is the top eigenvector of M^k, then $\left|\sum_{i\in T_k} a_i\right| \geq 0.8\sqrt{t_k}$ and*
(b) $t_k \geq 800\sqrt{k\log n}$.

Proof Let u^k be a vector defined by $u^k_i = 1/\sqrt{t_k}$ for $i \in T_k$ and 0 elsewhere. Then, $u^{k^T} M^k u^k = t_k$; this implies that $\lambda_1(M^k) \geq t_k$. Now (i) follows from Lemma 9.

(ii) Suppose now k satisfies (15) and a is the top eigenvector of M^k. Then, we have (recalling (14) and using Lemma 10),

$$1000\sqrt{k\log n} \leq a^T M^k a = t_k(u^k\cdot a)^2 + a^T\hat{M}^k a - a^T\tilde{M}^k a \leq t_k + 200\sqrt{k\log n}.$$

Thus,

$$t_k \geq 800\sqrt{k\log n}.$$

Also,

$$t_k \leq \lambda_1(M^k) \leq t_k(u^k\cdot a)^2 + 200\sqrt{k\log n} \leq t_k\left((u^k\cdot a)^2 + \frac{1}{4}\right)$$

which implies $(u^k\cdot a)^2 \geq 3/4$. This proves (ii).

□

LEMMA 12. *There is a polynomial time algorithm which given $S \subseteq V$ and a unit length vector a with support S, finds a $P' \subseteq V$ with the following property:*
If $|S\cap P| \geq 800\sqrt{|S|\log n}$ and $\sum_{i\in S\cap P} a_i \geq 0.8\sqrt{|S\cap P|}$, then $P' = P$.

Proof Re-number the coordinates, so that $a_1 \geq a_2 \geq \ldots \geq a_n$. In particular this implies that if $\ell \leq |S|$ then $[\ell] \subseteq S$. We wish to prove that there is an integer ℓ such that

$$|[\ell]\cap P| \geq \max\{\ell/100, 10\log n\} \tag{16}$$

First, if $|S\cap P| \geq |S|/10$, then we can take $\ell = |S|$. So assume that $t = |S\cap P| < |S|/10$ and let $\ell = 4t$. Now

$$\sum_{i\leq\ell; i\in P} a_i \leq \sqrt{|[\ell]\cap P|}$$

and so

$$\sum_{i\geq\ell+1; i\in P} a_i \geq 0.8\sqrt{|S\cap P|} - \sqrt{|[\ell]\cap P|} \text{ and } \sum_{i\leq\ell} a_i \geq \frac{\ell}{t}\left(0.8\sqrt{|S\cap P|} - \sqrt{|[\ell]\cap P|}\right).$$

But,

$$\sum_{i\leq\ell} a_i \leq \sqrt{\ell}.$$

This implies

$$\sqrt{|[\ell]\cap P|} \geq 0.8\sqrt{|S\cap P|} - 0.25\sqrt{\ell} = .15\sqrt{\ell}. \tag{17}$$

Also, we have $|S\cap P|^2 \geq 640000|S|\log n$ and so $|S\cap P| \geq 640000\log n$ and then (16) follows from (17) and $|[\ell]\cap P| \geq 4(.15)^2|S\cap P|$.

Now to construct P we try all values of ℓ. For each value of ℓ, we pick a random set Q_1 of $10\log n$ from $[\ell]$. For ℓ satisfying (16) there is at least a $10^{-20\log n}$ chance that $Q_1 \subseteq P$. Now **whp** no set of $10\log n$ vertices in P have more than $2\log n$ common neighbours outside P. Indeed the probability of the contrary event is at most

$$\binom{p}{10\log n}\binom{n}{2\log n}2^{-20(\log n)^2} = o(1).$$

So let Q_2 be the set of common neighbours of Q_1. By assumption we have $P \subseteq Q_2$ and $|Q_2 \setminus P| \leq 2\log n$. Also, **whp** for every $10\log n$-subset Q of P, no common neighbour outside P has $3p/4$ neighbours in P. Indeed the probability of the contrary event is at most

$$n\binom{p}{10\log n}\binom{n}{2\log n}2^{-p/12} = o(1).$$

Thus P is the set of vertices of degree at least $7p/8$ in the subgraph of G induced by Q_2. □

Acknowledgement We thank Santosh Vempala for interesting discussions on this problem.

References

[1] N. Alon, M. Krivelevich and B. Sudakov, Finding a large hidden clique in a random graph, *Random Structures and Algorithms* 13 (1998) 457-466.

[2] N. Alon, M. Krivelevich and V. H. Vu, On the concentration of eigenvalues of random symmetric matrices", Israel Journal of Mathematics 131 (2002) 259-267.

[3] N. Alon and J.H. Spencer, *The Probabilistic Method,* Wiley, (second edition) 2000.

[4] U. Feige and R.Krauthgamer, Finding and certifying a large hidden clique in a semi-random graph, *Random Structures and Algorithms*, 13 (1998) 457-466.

[5] J. Hastad, Clique is hard to approximate within $n^{1-\epsilon}$, *Proceedings of the 37th Annual IEEE Symposium on Foundations of Computing*, (1997) 627-636.

[6] M. Jerrum, Large cliques elude the Metropolis process, *Random Structures and Algorithms* 3 (1992) 347-359.

[7] A. Juels and M. Peinado, Hiding Cliques for Cryptographic Security, Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms, (1998) 678-684.

[8] R. Karp, Reducibility among combinatorial problems, in *The complexity of computer computations*, R. Miller and J. Thatcher (eds.) Plenum Press, New York (1972) 85-103.

[9] R. Karp, The probabilistic analysis of some combinatorial search algorithms, in *Algorithms and Complexity: New Directions and Recent Results*, J.F. Traub, ed., Academic Press (1976) 1-19.

[10] L. Kucera, Expected complexity of graph partitioning problems, *Discrete Applied Mathematics* 57 (1995) 193-212.

[11] F. McSherry, Spectral Partitioning of random graphs, FOCS 2001, 529-537.