Journal of Machine Learning Research 14 (2013) 3495-3536 Submitted 5/12; Revised 3/13; Published 12/13

Lovász ϑ function, SVMs and Finding Dense Subgraphs

Vinay Jethava JETHAVA@CHALMERS.SE
Anders Martinsson ANDEMAR@STUDENT.CHALMERS.SE
Computer Science & Engineering Department
Chalmers University of Technology
Rännvagen 6B
S 412 96, Göteborg, Sweden

Chiranjib Bhattacharyya CHIRU@CSA.IISC.ERNET.IN
Department of Computer Science and Automation
Indian Institute of Science
Bangalore, 560012
Karnataka, India

Devdatt Dubhashi DUBHASHI@CHALMERS.SE
Computer Science & Engineering Department
Chalmers University of Technology
Rännvagen 6B
S 412 96, Göteborg, Sweden

Editor: Francis Bach

Abstract

In this paper we establish that the Lovász ϑ function on a graph can be restated as a kernel learning problem. We introduce the notion of $\mathbf{SVM}-\vartheta$ graphs, on which Lovász ϑ function can be approximated well by a Support vector machine (SVM). We show that Erdös-Rényi random $G(n,p)$ graphs are $\mathbf{SVM}-\vartheta$ graphs for $\frac{\log^4 n}{n} \leq p < 1$. Even if we embed a large clique of size $\Theta\left(\sqrt{\frac{np}{1-p}}\right)$ in a $G(n,p)$ graph the resultant graph still remains a $\mathbf{SVM}-\vartheta$ graph. This immediately suggests an SVM based algorithm for recovering a large planted clique in random graphs. Associated with the ϑ function is the notion of orthogonal labellings. We introduce *common orthogonal labellings* which extends the idea of orthogonal labellings to multiple graphs. This allows us to propose a Multiple Kernel learning (MKL) based solution which is capable of identifying a large common dense subgraph in multiple graphs. Both in the planted clique case and common subgraph detection problem the proposed solutions beat the state of the art by an order of magnitude.

Keywords: orthogonal labellings of graphs, planted cliques, random graphs, common dense subgraph

1. Introduction

In a general graph many problems, such as computing the size of the largest clique or determining the chromatic number, are NP hard (Garey and Johnson, 1979). The ϑ function, introduced by Lovász (1979), is an extremely powerful tool for approximating such quantities in polynomial time. In some cases one can compute ϑ on large graphs efficiently, for example by exploiting symmetry (Bachoc et al., 2012), but in general efficient computation of ϑ function on large graphs remains a challenge. Evaluating the ϑ function on a graph requires solving a Semidefinite Program (SDP).

Using interior point methods one can solve SDPs, albeit with a high computational complexity of $O(n^6)$ (Boyd and Vandenberghe, 2004). Indeed numerical experiments show that computing ϑ on graphs consisting of more than 5000 vertices, using off the shelf SDP solvers, is impractical. Consider the problem of finding a large planted clique in a random graph. One could use an algorithm based on ϑ function computation (Feige and Krauthgamer, 2000) to recover such a clique. Unfortunately one cannot apply this algorithm to large graphs, say graphs of size more than $20,000$ vertices, due to high computational complexity of computing ϑ.

In this paper we establish that the ϑ function is equivalent to solving a kernel learning problem in the one class SVM setting. This surprising connection opens up lot of opportunities between graph theory and machine learning. Instead of trying to compute ϑ function exactly we show that by judicious choice of a kernel function, one can compute an upper-bound on the ϑ function by solving an SVM. We show that on random graphs this upper bound serves as a constant factor approximation to the ϑ function. We study how this bound can be exploited to identify large dense subgraphs in large graphs. In particular we study the problem of finding a *common dense subgraph* in multiple graphs (Pardalos and Rebennack, 2010), a computationally challenging problem for large graphs. We also study the problem of finding a hidden planted clique in a random graph. This is again an instance of computationally challenging problem (Jerrum, 1992).

1.1 The Importance of Studying Dense Subgraphs

Finding dense subgraphs in large graphs is an important problem, which has many applications in a variety of disciplines. In Computational Biology, mining for large dense subgraphs has important consequences for function discovery (Hu et al., 2005). Many other problems in Computational Biology can be posed as that of finding dense subgraphs (see, e.g., Spirin and Mirny, 2003; Jiang and Pei, 2009; Takahashi et al., 1987). In E-commerce one could find isolated submarkets, important for advertising, by finding dense subgraphs (Lang and Andersen, 2007). Many problems in social network analysis can also be posed as dense subgraph discovery problem (Newman et al., 2006). Dense subgraph discovery can also be useful in designing more secure systems (Applebaum et al., 2010). Recently a very interesting suggestion was made in Arora et al. (2010) where understanding the complexity of financial derivatives was linked to finding dense subgraphs. A comprehensive review of applications of dense subgraphs is beyond the scope of this paper and we refer the interested reader to the survey by Lee et al. (2010).

In this paper we target two difficult versions of dense subgraph recovery problem. The problem of planted clique in a random graph is an instance of dense subgraph discovery in a random graph. This problem is extensively studied by the Algorithms community. Though the focus of the study is mainly theoretical it also has practical implications in several disciplines including Machine Learning. To cite an example recently the problem of correlation detection was formulated as that of finding a planted clique in a large random graph (Devroye et al., 2011). Inspired by several applications in Computational Biology (see, e.g., Podolyan and Karypis, 2009), we study the problem of finding a large *common* dense subgraph in multiple graphs.

1.2 Contributions

In this paper we make several contributions. Lovász (1979) introduced the notion of orthogonal labellings and used it to define the ϑ function. We show that for any orthogonal labelling one can define a Kernel matrix, $\mathbf{K}$. Using this matrix $\mathbf{K}$, one can compute an upper-bound on the Lovász ϑ

function by solving a SVM. Furthermore we show that

$$\min_{\mathbf{K}\in\mathcal{K}(G)} \omega(\mathbf{K}) = \vartheta(G),$$

where $\omega(\mathbf{K})$ is the optimal SVM objective function. Finding a common dense region in multiple graphs is known to be computationally difficult problem. One of the main contribution of this paper is to show how the connection of ϑ to SVMs can be exploited to find a dense subgraph in multiple graphs. We extend the idea of orthogonal labelling to multiple graphs by introducing the notion of *common orthogonal labelling*. This allows us to use a formulation based on multiple kernel learning for this problem. The proposed method beats existing methods by an order of magnitude. Our results on the well-known benchmark DIMACS data set show that the proposed method can identify dense graphs in large variety of settings, while state of the art method fails. An important contribution of this paper is to introduce $\mathbf{SVM}-\vartheta$ graphs, on which Lovász ϑ function can be well approximated by SVM. It is interesting to note that $G(n,p)$ graphs are $\mathbf{SVM}-\vartheta$ graphs. In many approximation algorithms, the ϑ function needs to be computed on $G(n,p)$ graphs. An immediate consequence of our result is that one does not need to solve an SDP to compute the ϑ function but can potentially use an SVM to approximate it. An extremely difficult instance of dense subgraph recovery problem is to pose the question of finding a hidden clique in a random graph. State of the art approaches are not practical for large graphs as they use Lovász ϑ function (Feige and Krauthgamer, 2000). Another key contribution of this paper is show that one can find a planted clique by solving an SVM. In particular, we show that in a $G(n,1-p)$ graph even if we embed a clique of size $k=\Theta\big(\sqrt{n(1-p)/p}\big)$, the resultant graph is a $\mathbf{SVM}-\vartheta$ graph. Furthermore, even if embed a sparse *random* subgraph in a large random graph, the resultant graph turns out to be $\mathbf{SVM}-\vartheta$ graph. In both cases one can prove that the SVM solution can be used to recover the planted subgraph.

1.2.1 STRUCTURE OF THE PAPER

In Section 2 we review the definition of orthogonal labeling and establish a connection between Lovász ϑ function and a kernel learning problem on one class problem. We extend the notion of orthogonal labellings for single graphs to include multiple graphs. We introduce the notion of common orthogonal labeling to multiple graphs in Section 3. This leads to a MKL based formulation which is capable of finding a common dense region. Next we present one of the major contributions of this paper. In Section 4 we establish that there exists graphs on which the Lovász ϑ function can be well approximated by an SVM. Furthermore we show that the the graph associated with the planted clique problem also satisfies this property. In Section 5 we empirically evaluate the performance of the proposed algorithms on large variety of graphs.

1.2.2 NOTATION

We represent vectors using lower case bold letters $\mathbf{a},\mathbf{b},\ldots$, etc., and matrices using upper case bold letters $\mathbf{A},\mathbf{B},\ldots$ etc.; with a_i referring to i^{th} element of $\mathbf{a}$, and similarly A_{ij} referring to $(i,j)^{th}$ entry of matrix $\mathbf{A}$. We use notation $[n]$ to denote the set $\{1,2,\ldots,n\}$. For a vector in $\mathbb{R}^d$, we denote the Euclidean norm by $\|.\|$ and the infinity norm by $\|.\|_\infty$. The inequality $\mathbf{a}\geq 0$ is true if it the inequality holds element-wise. Let $\mathcal{S}^{d-1}=\{\mathbf{u}\in\mathbb{R}^{d-1}\,|\,\|\mathbf{u}\|_2=1\}$ denote a $(d-1)$ dimensional sphere. Let $\mathbf{S}_n$ denote the set of $n\times n$ square symmetric matrices and $\mathbf{S}_n^+$ denote $n\times n$ square symmetric positive semidefinite matrices. For any matrix $\mathbf{A}\in\mathbb{R}^{n\times n}$, we denote the eigenvalues $\lambda_1(\mathbf{A})\geq\ldots\geq\lambda_n(\mathbf{A})$,

and $\|\mathbf{A}\| = \sqrt{\lambda_1(\mathbf{A}^\top \mathbf{A})}$. $\texttt{diag}(\mathbf{r})$ will denote a diagonal matrix with diagonal entries defined by components of $\mathbf{r}$.

Support vector machines (SVMs) have emerged as a powerful tool for binary classification problems (Vapnik, 1995). SVMs are posed as a Convex Quadratic program (CQP) and can be solved in linear time (Hush et al., 2006). In this paper we will extensively use a variation of the SVM formulation, known as the one-class SVM (Schölkopf et al., 2001), and written as

$$\omega(\mathbf{K}) = \max_{\alpha_i \geq 0, i=1,\ldots,n} f(\alpha; \mathbf{K}) \left(= 2\sum_{i=1}^{n} \alpha_i - \sum_{i,j=1}^{n} \alpha_i \alpha_j K_{ij} \right), \tag{1}$$

where $\mathbf{K} \in \mathbf{S}_n^+$ is called the kernel matrix. This formulation can be solved in $O(n^2)$ time (Hush et al., 2006). In the sequel, for the sake of brevity we will denote (1) as SVM formulation.

Let $G = (V, E)$ be a graph of order n with vertex set $V = [n]$ and edge set $E \subseteq V \times V$. Let $\mathbf{A} \in \mathbf{S}_n$ denote the adjacency matrix of G where $A_{ij} = 1$ if edge $(i, j) \in E$, and 0 otherwise. Let $\bar{G}$ denote the complement graph of G. The adjacency matrix of $\bar{G}$ is $\bar{\mathbf{A}} = \mathbf{e}\mathbf{e}^\top - \mathbf{I} - \mathbf{A}$, where $\mathbf{e} = [1, 1, \ldots, 1]^\top$ is a vector of length n containing all 1's, and $\mathbf{I}$ denotes the identity matrix. We denote the indicator vector for some set $S \subseteq V$ as $\mathbf{e}_S$ which is one for all $i \in S$ and zero in other co-ordinates.

Let $N_i(G) = \{j : (i, j) \in E\}$ denote the neighbourhood of node $i \in V$; $d_i(G) = |N_i(G)|$ denote the degree of node i; and $\gamma(G) = |E| / \binom{|V|}{2}$ denote the density of graph G. Let $G_S = (S, E_S)$ denote the subgraph induced by $S \subseteq V$ in graph G. An independent set in G (a clique in $\bar{G}$) is a subset of vertices $S \subseteq V$ for which no (every) pair of vertices has an edge in G (in $\bar{G}$). Our notations are standard, see Bollobás (1998).

An event $\mathcal{A}_n$ holds with high probability if $P(\mathcal{A}_n)$ tends to 1 as n goes to ∞. The notations O, o, Ω, Θ will denote the standard measures defined in asymptotic analysis (see, e.g., Cormen et al., 2009, Chapter 3).

2. Lovász theta Function and Kernel Learning

Consider the problem of embedding a graph $G = (V, E)$ on an unit sphere S^{d-1}. The study of this problem was initiated by Lovász (1979) who introduced the idea of *orthogonal labelling*:

Definition 1 (Lovász, 1979) *An orthogonal labelling of graph $G = (V, E)$ with $|V| = n$, is a matrix $U = [\mathbf{u}_1, \ldots, \mathbf{u}_n] \in \mathbb{R}^{d \times n}$ such that $\mathbf{u}_i^\top \mathbf{u}_j = 0$ whenever $(i, j) \notin E$ and $\mathbf{u}_i \in S^{d-1}$ $\forall\, i \in [n]$.*

Let $Lab(G)$ denote the set of orthogonal labellings of graph G, given by:

$$Lab(G) := \{\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_n] \,|\, \mathbf{u}_i \in S^{d-1},\ \mathbf{u}_i^\top \mathbf{u}_j = 0 \,\forall\, (i, j) \notin E\}.$$

Using $Lab(G)$, Lovász (1979) defined $\vartheta(G)$ as follows:

$$\vartheta(G) = \min_{\mathbf{U} \in Lab(G)} \min_{\mathbf{c} \in S^{d-1}} \max_i \left(\mathbf{c}^\top \mathbf{u}_i\right)^{-2}.$$

In the sequel we will sometimes denote $\vartheta(G)$ by ϑ when the argument is clear from the context. There exist several other equivalent definitions of ϑ function, for a comprehensive discussion see monograph by Knuth (1994).

It can be shown that ϑ serves as an upper-bound on the size of maximum independent set, $\mathbf{ALPHA}(G)$,[1] of a graph G (Lovász, 1979). Indeed for any graph G,

$$\mathbf{ALPHA}(G) \leq \vartheta(G).$$

Computing $\mathbf{ALPHA}(G)$ is a classic NP-hard problem (Garey and Johnson, 1979), which is furthermore known to be very hard even to approximate (Håstad, 1999). However $\vartheta(G)$, which can be computed in polynomial time by solving an SDP, gives a polynomial time computable upper-bound on $\mathbf{ALPHA}(G)$. Since then Lovász ϑ function has been extensively used in solving a variety of algorithmic problems (see, e.g., Coja-Oghlan and Taraz, 2004; Krivelevich, 2002; Karger et al., 1998).

2.1 The Relationship between SVMs and $\vartheta(G)$

In this subsection we establish that the $\vartheta(G)$ function can be re-stated as a Kernel learning problem. An interesting characterization of ϑ function involving convex quadratic program (CQP) was given by Luz and Schrijver (2006), which we describe below:

Theorem 2 (Luz and Schrijver, 2006) *For a graph $G = (V,E)$ having n vertices, let $\mathbf{C} \in \mathbb{R}^{n\times n}$ be an $n \times n$ matrix with $C_{ij} = 0$ whenever $(i,j) \notin E$. Then,*

$$\begin{aligned} \vartheta(G) &= \min_{\mathbf{C}} \quad v(G,\mathbf{C}), \text{ where} \\ v(G,\mathbf{C}) &= \max_{\mathbf{x}\geq 0} \underbrace{2\mathbf{x}^\top \mathbf{e} - \mathbf{x}^\top \left(\frac{\mathbf{C}}{-\lambda_n(\mathbf{C})} + \mathbf{I}\right)\mathbf{x}}_{g_G(\mathbf{x})}. \end{aligned}$$

Proof See Luz and Schrijver (2006). ■

The above theorem can also be understood from a Kernel learning perspective in the SVM setting. Observe that for every feasible choice of $\mathbf{C}$, there exists an orthogonal labelling, $\mathbf{U}$ where $\mathbf{U}^\top\mathbf{U} = \mathbf{I} + \frac{\mathbf{C}}{-\lambda_n(\mathbf{C})}$. Taking a cue from this observation we state and prove the following theorem from first principles.

Theorem 3 *For an undirected graph $G = (V,E)$, with $|V| = n$, let*

$$\mathcal{K}(G) := \{\mathbf{K} \in \mathbf{S}_n^+ \mid K_{ii} = 1, i \in [n], K_{ij} = 0, (i,j) \notin E\}.$$

Then,

$$\vartheta(G) = \min_{\mathbf{K}\in\mathcal{K}(G)} \omega(\mathbf{K}).$$

Proof We begin by noting that any $\mathbf{K} \in \mathcal{K}(G)$ is positive semidefinite and hence there exists $\mathbf{U} \in \mathbb{R}^{d\times n}$ such that $\mathbf{K} = \mathbf{U}^\top\mathbf{U}$. Note that $K_{ij} = \mathbf{u}_i^\top \mathbf{u}_j$ where $\mathbf{u}_i$ is a column of $\mathbf{U}$. Hence by inspection it is clear that the columns of $\mathbf{U}$ define an orthogonal labelling on G, that is, $bU \in Lab(G)$. Using a similar argument we can show that for any $\mathbf{U} \in Lab(G)$, the matrix $\mathbf{K} = \mathbf{U}^\top\mathbf{U}$, is an element of $\mathcal{K}(G)$. The set of valid kernel matrices $\mathcal{K}(G)$ is thus equivalent to $Lab(G)$. Note that if $\mathbf{U}$

1. Usually in Algorithms literature, the size of the largest independent set in a graph G is denoted by $\alpha(G)$. We have chosen to denote it by $\mathbf{ALPHA}(G)$ to avoid conflict with the notation α_i for Support vectors.

is a labelling then $\mathbf{U} = \mathbf{U}\texttt{diag}(\varepsilon)$ is also an orthogonal labelling for any $\varepsilon^\top = [\varepsilon_1, \ldots, \varepsilon_n]$ where $\varepsilon_i \in \{1, -1\} \forall\, i \in [n]$. It thus suffices to consider only those labellings for which $\mathbf{c}^\top \mathbf{u}_i \geq 0 \forall\, i \in [n]$ holds. For a fixed $\mathbf{c}$, one can rewrite

$$\max_i \frac{1}{(\mathbf{c}^\top \mathbf{u}_i)^2} = \left(\min_t t^2 \text{ s.t. } \frac{1}{\mathbf{c}^\top \mathbf{u}_i} \leq t \; \forall\, i \in [n] \right).$$

Setting $\mathbf{w} = 2t\mathbf{c}$ yields the following relation

$$\min_{\mathbf{c} \in \mathcal{S}^{d-1}} \max_i \frac{1}{(\mathbf{c}^\top \mathbf{u}_i)^2} = \left(\min_{\mathbf{w} \in \mathbb{R}^d} \frac{\|\mathbf{w}\|^2}{4} \text{ s.t. } \mathbf{w}^\top \mathbf{u}_i \geq 2 \; \forall\, i \in [n] \right).$$

This establishes that given an orthogonal labelling $\mathbf{U}$, the minimization of $\mathbf{c}$ is obtained by solving a convex quadratic program (CQP) which is equivalent to solving a SVM. Application of strong duality immediately leads to the claim

$$\min_{\mathbf{c} \in \mathcal{S}^{d-1}} \max_i \frac{1}{(\mathbf{c}^\top \mathbf{u}_i)^2} = \omega(\mathbf{K})$$

where $\mathbf{K} = \mathbf{U}^\top \mathbf{U}$ and $\omega(\mathbf{K})$ is defined in (1). As there is a correspondence between each element of $Lab(G)$ and $\mathcal{K}$ minimization of $\omega(\mathbf{K})$ over $\mathcal{K}$ is equivalent to computing the $\vartheta(G)$ function. ■

Theorem 3 establishes a connection between $\vartheta(G)$ and $\omega(\mathbf{K})$, two well studied formulations in Graph theory and Machine Learning. As a consequence this immediately opens up the possibility of applying large-scale Kernel learning algorithms (see, e.g., Hu et al., 2011) for computing ϑ function. In this paper we do not explore this direction further, leaving it for future study, but instead focus on another important consequence of Theorem 3. The theorem establishes that if we choose not to optimize over all possible labellings but instead fix an orthogonal labelling and then, one can easily compute an upper-bound on $\mathbf{ALPHA}(G)$ by solving an SVM. See that for any graph G

$$\mathbf{ALPHA}(G) \leq \vartheta(G) \leq \omega(\mathbf{K}) \; \forall\, \mathbf{K} \in \mathcal{K}(G).$$

As SVMs have linear time complexity (Hush et al., 2006) this could be an efficient alternative for computing upper-bounds on size of the maximum independent set for any chosen labelling.

In Theorem 2, if we fix $\mathbf{C} = \mathbf{A}$, the adjacency matrix, we obtain a very interesting orthogonal labelling, which we will refer to as **LS** labelling as it was first introduced by Luz and Schrijver (2006). In Q graphs, introduced by Luz (1995), the **LS** labelling recovers $\mathbf{ALPHA}(G)$ by solving a CQP. The CQP is obtained by setting $\mathbf{C} = \mathbf{A}$ in Theorem 2.

Definition 4 (Luz, 1995) *A graph G, with adjacency matrix $\mathbf{A}$, is called a Q graph whenever $\mathbf{ALPHA}(G) = v(G, \mathbf{A})$ where $v(G, \mathbf{A})$ is defined in Theorem 2.*

Indeed, on a Q graph, computation of the ϑ function reduces to computing the minimum eigenvalue and solving a CQP, which in general has a complexity of $O(n^3)$. By Theorem 3, one can now state that on Q graphs evaluating the ϑ function is equal to solving an SVM, a special case of CQP, which can be solved in $O(n^2)$ time (Hush et al., 2006). Keeping computational advantages in mind it is interesting to characterize Q graphs.

Theorem 5 (Luz, 1995) *For a simple unweighted graph $G = (V,E)$, let $\mathbf{A}$ denote the adjacency matrix , S be the largest independent set, $v(G,\mathbf{A})$ and $g_G(x)$ be defined as in Theorem 2. The graph G is a Q graph iff*

$$-\lambda_n(A) \leq \min_{i \in V \setminus S} |N_i(G) \cap S|.$$

Furthermore $v(G,A) = g_G(\mathbf{e}_S)$, where $\mathbf{e}_S$ is the indicator vector on S.

Proof See Theorem 4 in Luz (1995). ■

In Section 4 we will use this characterization to show that random graphs with planted cliques are not Q graphs.

Inspired by the computational simplicity of the **LS** labelling, we study it more closely. As a labelling is completely defined by the associated kernel matrix, we refer to the following matrix

$$\mathbf{K} = \frac{\mathbf{A}}{\rho} + \mathbf{I} \text{ where } \rho \geq -\lambda_n(\mathbf{A}) \tag{2}$$

as the **LS** labelling. The KKT conditions for $\omega(\mathbf{K})$ in (1) are given by

$$\alpha_i + \frac{1}{\rho} \sum_{(i,j) \in E} A_{ij}\alpha_j = 1 + \mu_i,\ \mu_i\alpha_i = 0,\ \mu_i \geq 0. \tag{3}$$

Direct algebra yields

$$\omega(\mathbf{K}) = \sum_{i=1}^{n} \alpha_i^* \tag{4}$$

when α^* satisfies the KKT conditions.

2.2 On Regular Graphs

Before ending this section we would like to discuss the case of regular graphs. A graph is said to be $d-$regular if all nodes haves the same degree d. For such graphs one can compute an upper-bound, popularly known as Hoffman bound, on $\mathbf{ALPHA}(G)$ via the minimum eigenvalue of G. More specifically,

Theorem 6 (Hoffman and Howes, 1970) *For a d-regular graph, G, on n vertices and adjacency matrix $\mathbf{A}$,*

$$\mathbf{ALPHA}(G) \leq \frac{n}{1 - \frac{d}{\lambda_n}}$$

where λ_n is the smallest eigenvalue of $\mathbf{A}$.

Computation of this bound involves solving an eigenvalue problem, a significantly cheaper option than computing the ϑ function. However ϑ yields a tighter bound than the Hoffman bound. It was indeed proved (Lovász, 1979) that

$$\vartheta(G) \leq \frac{n}{1 - \frac{d}{\lambda_n}}.$$

It is interesting to note that $\omega(\mathbf{K})$ equals the Hoffman bound on d-regular graphs. To this end we have,

Lemma 7 *Let $G = (V,E)$, be a $d-$regular graph with adjacency matrix* $\mathbf{A}$*, and* $|V| = n$*. Then,*

$$\omega(\mathbf{K}) = \frac{n}{1+\frac{d}{\rho}}$$

where $\mathbf{K}$ *and* ρ *are defined by* (2)*.*

This was also derived in Luz (1995) as Corollary 1. To make the paper self-contained we give a short proof.

Proof For any d-regular graph the largest eigenvalue is d and the corresponding eigenvector is $\mathbf{e}$. It is easy to verify that $\alpha_i = \frac{1}{1+\frac{d}{\rho}}$ satisfies KKT conditions (3) and hence by (4) the lemma is proved. ■

The lemma holds for any $\rho \geq -\lambda_n$ and consequently, for regular graphs, $\omega(\mathbf{K})$ equals the Hoffman bound whenever $\rho = -\lambda_n$.

Before we discuss the applicability of the results obtained here to random graphs we study the interesting problem of finding a dense common subgraph in multiple graphs.

3. Finding Large Dense Regions in Multiple Graphs

The problem of finding a dense subgraph in a single graph is a computationally challenging problem (Pardalos and Rebennack, 2010). In this section we attempt an even more difficult version of the problem, namely that of finding a *common* large dense region in multiple graphs. Most methods which apply to single graphs (Lee et al., 2010) do not extend to the case of multiple graphs. An interesting method was proposed by Jiang and Pei (2009), which uses an enumerative strategy for finding a common dense subgraph in multiple graphs. This section presents one of the main contributions of this paper. It introduces the notion of *common orthogonal labelling* and proposes a Multiple Kernel Learning (MKL) (Lanckriet et al., 2004) inspired formulation. Later on we will see experimentally how this formulation achieves an order of magnitude scalability when compared with Jiang and Pei (2009).

3.1 Dense Common Subgraph Detection

Jiang and Pei (2009) studied the problem of finding *all* possible common subgraphs in graphs $G^{(1)},\ldots,G^{(m)}$ for a given choice of parameters $\gamma^{(1)},\ldots,\gamma^{(m)}$; such that the subgraphs have density at least $\gamma^{(l)}$ in graph $G^{(l)}$ for all $l \in \{1,\ldots,m\}$ respectively. The maximal quasiclique size depends on the choice of parameters $\gamma^{(1)},\ldots,\gamma^{(m)}$. In fact, if the parameters are not chosen properly the algorithm may fail to return any solution at all. A different choice of parameters requires solving the problem again. As a consequence one might have to run several iterations of the algorithm with different parameter choices before obtaining a desired subgraph.

The approach is essentially enumerative in nature, and consequently, the space and time complexity of the algorithm is exponential in the size of the largest possible clique, rendering it impractical for finding large dense regions. For example, finding subgraphs of size 60 requires 11.5 hours (see Figure 17 in Jiang and Pei, 2009). Clearly this algorithm is not suitable for finding large subgraphs.

In practice one might wish to quickly find a single subset of vertices which is large (some fraction of the overall graph) and which induces a dense subgraph in each of the original graphs

$G^{(l)} \in \mathbb{G}$ without fine tuning of parameters; or multiple runs of the algorithm. In other words, one would like to have a subgraph finding algorithm which is *parameter-less*. To this extent, we define the problem formally as follows:

Definition 8 (Problem definition) *Let* $\mathbb{G} := \{G^{(1)}, \ldots, G^{(m)}\}$ *be a set of simple, undirected graphs* $G^{(l)} = (V, E^{(l)})$ *defined on vertex set* $V = \{1, \ldots, n\}$. *Find a* common *subgraph which is* dense *in all the graphs.*

The remainder of this section is organized as follows: we develop the notion of *common orthogonal labelling* and establish the connection to MKL formulation in Section 3.1.1. Section 3.1.2 presents our algorithm for recovery of large common dense subgraph from multiple graphs.

3.1.1 COMMON ORTHOGONAL LABELLING AND MKL FORMULATION

We begin by defining common orthogonal labelling below:

Definition 9 *Given set of simple undirected graphs* $\mathbb{G}$ *on a common vertex set* $V = \{1, \ldots, n\}$, *the common orthogonal labelling* $\mathbf{u}_1, \ldots, \mathbf{u}_n$ *is given by* $\mathbf{u}_i \in \mathcal{S}^{d-1}$ *such that* $\mathbf{u}_i^\top \mathbf{u}_j = 0$ *if* $(i, j) \notin E^{(l)}\ \forall l = \{1, \ldots, m\}$.

Let $\mathcal{K}(\mathbb{G})$ denote the set of common orthogonal labellings of $\mathbb{G}$ so that

$$\mathcal{K}(\mathbb{G}) = \left\{\mathbf{K} : \mathbf{K} \in \mathbf{S}_+^n,\ \mathbf{K}_{ij} = 0\ if\ (i, j) \notin E^{(l)}\ \forall 1 \leq l \leq m\right\}.$$

The common orthogonal labellings of $\mathbb{G}$ is related to the *union graph* $G^\cup = (V, E^\cup)$ constructed from $\mathbb{G}$ as follows: an edge (i, j) is present in the union graph $G^\cup$ if it is present in at least one of the original graphs $G^{(l)} \in \mathbb{G}$, and absent otherwise, that is, $(i, j) \notin E^\cup$ iff $(i, j) \notin E^{(l)} \forall_l$. By construction, we see

$$\mathcal{K}(\mathbb{G}) = \mathcal{K}(G^\cup).$$

Let $\Upsilon(\mathbb{G})$ denote the size of the *maximum common independent set*, that is, subset of vertices $CS \subseteq V$ of maximum possible cardinality for which the subgraph $G_{CS}^{(l)}$ induced by CS in graph $G^{(l)}$ is an independent set for all $G^{(l)} \in \mathbb{G}$. It is immediate that $\Upsilon(\mathbb{G})$ is equal to the size of the maximum independent set in the union graph $\mathbf{ALPHA}(G^\cup)$. Following the arguments in Theorem 3, one can show

$$\mathbf{ALPHA}(G^\cup) = \Upsilon(\mathbb{G}) \leq \vartheta(G^\cup) \leq \omega(\mathbf{K})\ \forall\, \mathbf{K} \in \mathcal{K}(\mathbb{G}). \tag{5}$$

As noted in the previous section, the optimization problem in (5) is a SDP, which cannot be solved for large graphs. Let $\mathbf{K}^{(l)} = \frac{\mathbf{A}^{(l)}}{\rho^{(l)}} + \mathbf{I}$ and $\rho^{(l)} \geq -\lambda_n(\mathbf{A}^{(l)})$ be the **LS** labelling for graph $G^{(l)}$. In the remainder of this section, we use the notation $\mathbb{K} := \{\mathbf{K}^{(1)}, \ldots, \mathbf{K}^{(m)}\}$ to denote the set of orthogonal labellings corresponding to the graphs $\mathbb{G}$.

We consider the set of convex combinations of the labellings of the original graphs $\mathbb{K}$ such that

$$conv(\mathbb{K}) := \left\{\mathbf{K} : \mathbf{K} = \sum_{l=1}^{m} \delta^{(l)} \mathbf{K}^{(l)} \text{ with } \delta^{(l)} \geq 0,\ \sum_{l=1}^{m} \delta^{(l)} = 1\right\}.$$

Note that $K_{ij}^{(l)}$ is zero whenever edge (i, j) is absent in $G^\cup$ for all graphs $G^{(l)}$ and consequently $\mathbf{K}_{ij} = 0$. Thus $\mathbf{K} \in conv(\mathbb{K})$ is a common orthogonal labelling and the following is immediate

$$conv(\mathbb{K}) \subseteq \mathcal{K}(\mathbb{G}).$$

Instead of solving the original problem, we consider the following problem

$$\psi(\mathbb{K}) = \min_{\mathbf{K} \in conv(\mathbb{K})} \omega(\mathbf{K}). \tag{6}$$

An immediate advantage of this formulation over the SDP formulation is that it could be solved efficiently by standard MKL solvers (Rakotomamonjy et al., 2008; Aflalo et al., 2011). The quantity, $\psi(\mathbb{K})$ also defines an upper bound on the size of the maximum common independent set. More precisely

$$\Upsilon(\mathbb{G}) \leq \min_{\mathbf{K} \in \mathcal{K}(\mathbb{G})} \omega(\mathbf{K}) \leq \psi(\mathbb{K}).$$

Rewriting (6), we get

$$\begin{aligned}
\psi(\mathbb{K}) &= \min_{\mathbf{K} \in conv(\mathbb{K})} \omega(\mathbf{K}) = \min_{\delta^{(l)} \geq 0, \sum \delta^{(l)} = 1} \omega\big(\sum_l \delta^{(l)} \mathbf{K}^{(l)}\big) \\
&= \min_{\delta^{(l)} \geq 0, \sum \delta^{(l)} = 1} \max_{\alpha \geq 0} \sum_l \delta^{(l)} f(\alpha; \mathbf{K}^{(l)}) \\
&= \max_{\alpha \geq 0} \min_{\delta^{(l)} \geq 0, \sum \delta^{(l)} = 1} \sum_l \delta^{(l)} f(\alpha; \mathbf{K}^{(l)}).
\end{aligned}$$

The optimization problem is linear in δ and strictly feasible in both δ and α. One can interchange $\min_\delta \max_\alpha$ to $\max_\alpha \min_\delta$, Sion (1958) yielding the last equality. For any vector, $\mathbf{x} = [x_1, \ldots, x_d]^\top$,

$$\min_{0 \leq \delta_i \leq 1, \sum_{i=1}^d \delta_i = 1} \delta^\top x = \min(x_1, \ldots, x_d) = \max_{t, x_i \geq t} t.$$

An alternative re-statement of (6) is

$$\psi(\mathbb{K}) = \max_{t \in \mathbb{R}, \alpha_i \geq 0} t \quad \text{s.t.} \;\; f(\alpha; \mathbf{K}^{(l)}) \geq t \; \forall \; 1 \leq l \leq m. \tag{7}$$

The Lagrange dual of (7) is given by

$$\mathcal{L}(t, \alpha, \lambda, \delta) = t + \sum_{l=1}^m \lambda_l \big(f(\alpha; \mathbf{K}^{(l)}) - t\big) + \sum_{i=1}^n \delta_i \alpha_i,$$

where $\lambda \in \mathbb{R}_+^m$ and $\delta \in \mathbb{R}_+^n$ denote the dual variables. The KKT conditions yield

$$2 \sum_{l=1}^m \lambda_l^* \big(1 - \alpha_i^* - \frac{1}{\rho^{(l)}} \sum_j A_{ij}^{(l)} \alpha_j^*\big) + \delta_i^* = 0, \quad \delta_i^* \alpha_i^* = 0, \quad \delta_i^*, \alpha_i^* \geq 0 \forall i \in [n], \quad \text{and} \tag{8}$$

$$\lambda_l^* (f(\alpha^*; \mathbf{K}^{(l)}) - t^*) = 0, \quad \sum_l \lambda_l^* = 1, \quad \lambda_l^* \geq 0 \forall l \in [m]. \tag{9}$$

The above optimization can be readily solved by state of the art MKL solvers. The obvious question that arises is when, or precisely, for which sets of graphs does solving (6) yield good approximation to the original problem of computing $\Upsilon(\mathbb{G})$.

In order to address this, we begin by defining the family of *Common Quadratically-stable* ($\mathcal{CQ}$) sets of graphs.

Definition 10 *A set of graphs $\mathbb{G} = \{G^{(1)}, \ldots, G^{(m)} : G^{(i)} = (V, E^{(i)})\}$ having a common vertex set V is Common Quadratically-stable if the optimal value $\psi(\mathbb{K})$ in* (6) *is equal to the size of the maximum common independent set $\Upsilon(\mathbb{G})$, that is,*

$$\Upsilon(\mathbb{G}) = \psi(\mathbb{K}), \tag{10}$$

where $\mathbb{K}$ is the set of **LS** *labellings of graphs $\mathbb{G}$.*

Remark 11 *We use the notation $\mathbb{G} \in \mathcal{CQ}$ to denote that the set of graphs $\mathbb{G}$ satisfies* (10)*, that is, it is Common Quadratically-stable.*

Let $Y \subseteq V$ denote the maximum common independent set in $\mathbb{G}$; with indicator vector $\mathbf{e}_Y$. We now characterize $\mathcal{CQ}$ family (of sets of graphs having a common vertex set) in the following result.

Theorem 12 *Given set of graphs $\mathbb{G} = \{G^{(1)}, \ldots, G^{(m)}\}$ having common vertex set V and* **LS** *labellings $\mathbf{K}^{(l)} = \frac{\mathbf{A}^{(l)}}{\rho^{(l)}} + \mathbf{I}$ and $\rho^{(l)} \geq -\lambda_n(\mathbf{A}^{(l)})\ \forall l \in [m]$; the optimal value $\psi(\mathbb{K})$ in* (6) *is equal to the size of the maximum common independent set Y, that is, the set $\mathbb{G}$ is Common Quadratically-stable ($\mathbb{G} \in \mathcal{CQ}$) if there exists non-empty $\mathbb{G}_Q \subseteq \mathbb{G}$ such that*

$$\rho^{(l)} \leq \min_{i \notin CS} |N_i(G^{(l)}) \cap Y| \qquad \forall G^{(l)} \in \mathbb{G}_Q. \tag{11}$$

Proof Let $L := |\mathbb{G}_Q|$ denote the number of graphs which satisfy the property

$$\rho^{(l)} \leq \min_{i \notin Y} |N_i(G^{(l)}) \cap Y| \quad \forall G^{(l)} \in \mathbb{G}_Q.$$

We consider primal solution $\alpha^* = \mathbf{e}_Y$ and dual solution given by

$$\lambda_l^* = \begin{cases} \frac{1}{L} & \text{if } l : G^{(l)} \in \mathbb{G}_Q \\ 0 & \text{otherwise} \end{cases}, \text{ and } \quad \delta_i^* = \begin{cases} 0 & \text{if } i \in Y \\ \sum_l 2\lambda_l^* (1 - \frac{|N_i(G^{(l)}) \cap Y|}{\rho^{(l)}}) & \text{otherwise} \end{cases}.$$

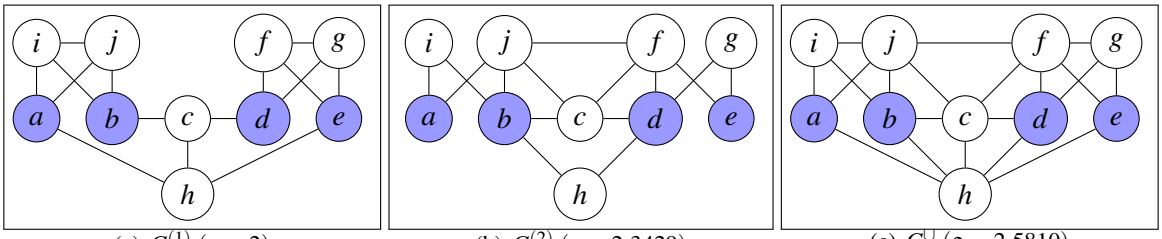
This solution satisfies the KKT conditions in (8) and (9). Further, the optimization in (7) is convex; and the KKT conditions are sufficient for optimality. The optimal value is given by $t^* = \sum_{i=1}^n \alpha_i^*$ where $\alpha^* = \mathbf{e}_Y$ yielding $\psi(\mathbb{K}) = \Upsilon(\mathbb{G})$. ■

Recall as in (5) that $\min_{\mathbf{K} \in \mathcal{K}(\mathbb{G})} \omega(\mathbf{K})$ is equivalent to computing Lovász ϑ function of $G^\cup$. Therefore, one can easily solve it whenever $G^\cup$ is a $\mathcal{Q}$ graph by solving $\omega(\mathbf{K}^\cup)$ where $\mathbf{K}^\cup = \frac{A^\cup}{\rho} + \mathbf{I}$ and $\rho \geq -\lambda_n(\mathbf{A}^\cup)$ whenever $\rho \leq \min_{i \notin Y} |N_i(G^\cup) \cap Y|$. It is of interest to find when $G^\cup$ is *not* a $\mathcal{Q}$ graph yet $\mathbb{G}$ is Complex Quadratically-stable ($\mathbb{G} \in \mathcal{CQ}$). One can easily see by construction of $G^\cup$ that

$$\min_{i \notin Y} |N_i(G^{(l)}) \cap Y| \leq \min_{i \notin Y} |N_i(G^\cup) \cap Y| \quad \forall G^{(l)} \in \mathbb{G}.$$

However, the relationship between minimum eigenvalue $\rho^{(l)} = -\lambda_n(\mathbf{A}^{(l)})$ of original graphs $G^{(l)} \in \mathbb{G}$, and minimum eigenvalue $\rho^\cup = -\lambda_n(A^\cup)$ of union graph $G^\cup$ is not clear, that is, $\rho^\cup$ can be greater or less than $\rho^{(l)}$ (see, e.g., Brouwer and Haemers, 2012, 3.1-3.2). We illustrate this in the following example.

Example 1 *Consider the graphs $G^{(1)}$ and $G^{(2)}$ shown in Figure 1. The maximum common independent set is $Y = \{a, b, d, e\}$ and $\min_{i \notin Y} |N_i \cap Y| = 2$. One can show that $\mathbb{G} \in \mathcal{CQ}$ since $\rho^{(1)} \leq \min_{i \notin Y} |N_i \cap Y|$, while the union graph $G^\cup$ is not a $\mathcal{Q}$ graph. Therefore, one can compute $\Upsilon(\mathbb{G})$ by solving the MKL optimization in* (7) *even though the maximum independent set problem in $G^\cup$ cannot be solved using QP. Thus, the MKL formulation is advantageous whenever that there exists some graph $G^{(l)} \in \mathbb{G}$ which satisfies* (11)*, even though $G^\cup$ is not a $\mathcal{Q}$ graph.*



(a) $G^{(1)}$ ($\rho = 2$) (b) $G^{(2)}$ ($\rho = 2.3429$) (c) $G^{\cup}$ ($\rho = 2.5810$)

Graph(s)	ρ	$\min_{i \notin Y} \lvert N_i \cap Y \rvert$	$\omega(\mathbf{K})$	$\psi(\mathbb{K})$
$G^{\cup}$	2.5810	2	4.3189	-
$\mathbb{G}$	-	-	-	**4.0**

Figure 1: Example of $\mathcal{CQ}$ set of graphs. Here $\mathbb{G} = \{G^{(1)}, G^{(2)}\}$ and maximum common independent set is $Y = \{a,b,d,e\}$ (highlighted in blue). The graph $G^{(1)}$ satisfies $\rho \leq \min_{i \notin Y} |N_i \cap Y|$ in (11); while $G^{(2)}$ and $G^{\cup}$ do not. Hence $\mathbb{G} \in \mathcal{CQ}$, while union graph $G^{\cup}$ is not a $\mathcal{Q}$ graph. The MKL optimization yields $\psi(\mathbb{K}) = 4$ which is equal to size of maximum common independent set $Y = \{a,b,d,e\}$.

3.1.2 SUBGRAPH DETECTION BASED ON MULTIPLE GRAPHS

In the remainder of this section, we relate the optimal solution (support vectors) of $\omega(\mathbf{K})$ and the density of related induced subgraph for the single graph case; which we later extend to multiple graphs. We first recall an interesting property of optimal solution α^* which maximizes $f(\alpha;\mathbf{K})$ in (1) when G is a $\mathcal{Q}$ graph.

Remark 13 (Luz, 1995) *Let $G = (V,E)$ be a $\mathcal{Q}$ graph having unique maximum independent set $S \subseteq V$. Then, the optimal solution α^* which maximizes objective $f(\alpha;\mathbf{K})$ in (1) is given by $\alpha^* = \mathbf{e}_S$, that is,*

$$\alpha_i^* = \begin{cases} 1 & \text{if } i \in S \\ 0 & \text{otherwise} \end{cases}.$$

The above claim follows from observing that α^* satisfies the Karush-Kuhn-Tucker (KKT) conditions, which is sufficient for optimality. The key idea here is that choosing the vertices having support vectors with high numerical values yields a subgraph with low density (or more precisely, independent set with zero density for $\mathcal{Q}$ graphs). We extend this notion to general graphs by relating the density of the induced subgraph obtained by choosing vertices having "high" support through the KKT conditions.

We now consider a general graph $G = (V,E)$ with adjacency matrix $\mathbf{A}$, and let α^* be the optimal solution of (1) when $\mathbf{K} = \frac{\mathbf{A}}{\rho} + \mathbf{I}$ with $\rho \geq -\lambda_n(\mathbf{A})$. We wish to relate the density of the subgraph G_{S_c} induced by the "high" support vectors $S_c := \{i : \alpha_i^* \geq c\}$ for some threshold $c \in (0,1)$.

Let $\bar{\alpha}_i^*(S)$ denote the average of the support vectors α_j^* over the neighbourhood $N_i(G_S)$ of node i in subgraph G_S induced by $S \subseteq V$ in graph G; and $\bar{\alpha}_S^*$ be the minimum $\bar{\alpha}_i^*(S)$ over all $i \in S$, that is,

$$\bar{\alpha}_i^*(S) = \frac{\sum_{j \in S} A_{ij} \alpha_i^*}{d_i(G_S)}, \quad \text{and} \quad \bar{\alpha}_S^* = \min_{i \in S} \bar{\alpha}_i^*(S). \tag{12}$$

With the above notation, one can show the following:

Lemma 14 *Let $G = (V,E)$ be a simple graph with at least one edge. Let α^* be the optimal solution of* (1) *where $\mathbf{K} = (\frac{\mathbf{A}}{\rho} + \mathbf{I})$ denotes the* **LS** *labelling of G and $\rho \geq -\lambda_n(\mathbf{A})$. The set $S_c = \{i : \alpha_i^* > c\}$ with cardinality n_c induces a subgraph G_{S_c} with density $\gamma(G_{S_c})$ where*

$$\gamma(G_{S_c}) \leq \frac{\rho(1-c)}{\bar{\alpha}^*_{S_c}(n_c-1)} \leq \frac{\rho(1-c)}{c(n_c-1)}.$$

Proof The Karush-Kuhn Tucker (KKT) conditions for (1) are given in (3). Using definition in (12), one can obtain

$$\bar{\alpha}^*_{S_c} \sum_{i \in S_c} d_i(G_{S_c}) \leq \sum_{i \in S_c} \bar{\alpha}^*_i(S_c) d_i(G_{S_c}) = \sum_{i,j \in S_c} \mathbf{A}_{ij} \alpha^*_j \leq \rho(n_c - \sum_{i \in S_c} \alpha^*_i) \leq \rho n_c(1-c).$$

Observing that $\sum_{i \in S_c} d_i(G_{S_c}) = 2|E_{S_c}|$, and dividing by $\binom{n_c}{2}$ yields the desired result, that is,

$$\gamma(G_{S_c}) \leq \frac{\rho(1-c)}{\bar{\alpha}^*_{S_c}(n_c-1)} \leq \frac{\rho(1-c)}{c(n_c-1)} \quad (\because \bar{\alpha}^*_{S_c} \geq c \text{ by defn. in (12)}).$$

■

This result provides an upper bound on the density $\gamma(G_{S_c})$ of the subgraph induced by set S_c in graph G for general c. Two special cases of interest are the set of non-zero support vectors $SV = \{i : \alpha_i^* > 0\}$ and the set of support vectors with support one $S_1 := \{i : \alpha_i^* = 1\}$ respectively.

Setting $c = 0$, one can show the set of (non-zero) support vectors $SV = \{i : \alpha_i^* > 0\}$ with cardinality $n_{SV} = |SV|$ induces a subgraph G_{SV} having density $\gamma(G_{SV})$ at most $\rho/\bar{\alpha}^*_{SV}(n_{SV}-1)$, that is,

$$\gamma(G_{SV}) \leq \frac{\rho}{\bar{\alpha}^*_{SV}(n_{SV}-1)}.$$

This provides a simple procedure for finding a sparse subgraph by selection the subgraph G_{SV} induced by the set of non-zero support vectors SV. It also gives an upper bound on density.

Setting c arbitrarily close to 1, one can show the set $S_1 = \{i : \alpha_i^* = 1\}$ of support vectors having support 1 is an independent set in G. An extreme case is when G is a Q graph and all support vectors are 1 over the maximum independent set and 0 otherwise.

We now consider the problem of common dense subgraph recovery from multiple graphs based on the MKL formulation in (7). Let (α^*, t^*) be the optimal solution of (7). One can partition the set $\mathbb{G}$ into two sets: *active* graphs $\mathbb{G}_A \subseteq \mathbb{G}$ for which the constraint in (7) is tight, that is,

$$f(\alpha^*, \mathbf{K}^{(l)}) = t^* \quad \forall\, G^{(l)} \in \mathbb{G}_a,$$

and *inactive* graphs $\mathbb{G}_i := \mathbb{G} \backslash \mathbb{G}_a$, where the constraint is not tight. Consequently, the analysis in the single graph case based on selection of support vector above certain threshold cannot be directly extended to multiple graph case. In the remainder, we address this using a two step procedure.

Let $SV := \{i : \alpha_i^* > 0\}$ and $S_c := \{i : \alpha_i^* > c\}$ denote the set of support vectors, and the set of support vectors having "high" support respectively for some appropriate choice of $c \in [0,1)$. Let

$\bar{\alpha}_i^{(l)}(S)$ denote the average of the support vectors α_j^* over the neighbourhood $N_i(G_S^{(l)})$ of node i in subgraph $G_S^{(l)}$ induced by $S \subseteq V$ in graph $G^{(l)}$; and $\bar{\alpha}_S^*$ be the minimum $\bar{\alpha}_i^*(S)$ over all $i \in S$, that is,

$$\bar{\alpha}_i^{(l)}(S) = \frac{\sum_{j\in S} A_{ij}\alpha_i^*}{d_i(G_S^{(l)})}, \quad \text{and,} \quad \bar{\alpha}_S^{(l)} = \min_{i\in S} \bar{\alpha}_i^{(l)}(S).$$

Notice that since $\bar{\alpha}_S^{(l)}$ is minimum (over all vertices in the set S) of average value of support vectors in the neighbourhood $N_i(S)$ for a vertex $i \in S$, it is equal or greater than the minimum support vector $\min_{i\in S}\alpha_i^*$ in S. Let $\alpha_{\min} = \min_{i\in SV}\alpha_i^*$ denote the minimum non-zero support vector. We define the sets $T^{(l)} \subseteq V$ and $T \subseteq V$ as

$$T^{(l)} := T^{(l)}(SV) = \left\{ i \in SV : 1 - \alpha_l - d_i(G_{SV}^{(l)})\bar{\alpha}_{SV}^{(l)}/\rho^{(l)} > 0 \right\}, \quad T = \cap_{l=1}^m T^{(l)}. \tag{13}$$

Then, one can show the following

Lemma 15 *Given set of graphs $\mathbb{G} = \{G^{(l)} : G^{(l)} = (V, E^{(l)}) \forall\, l \in [m]\}$ defined on common vertex set V with* **LS** *labelling $\mathbb{K}$ where $\mathbf{K}^{(l)} = \frac{\mathbf{A}^{(l)}}{\rho^{(l)}} + \mathbf{I}$ and $\rho^{(l)} \geq -\lambda_n(\mathbf{A}^{(l)})$. Let (α^*, t^*) be the optimal solution of* (7)*, and T be defined as in* (13)*. The set T with cardinality $n_T = |T|$ induces a subgraph $G_T^{(l)}$ in graph $G^{(l)} \in \mathbb{G}$ having density at most $\gamma(G_T^{(l)})$ given by*

$$\gamma(G_T^{(l)}) \leq \frac{(1-\alpha_{\min})\rho^{(l)}}{\bar{\alpha}_{SV}^{(l)}(n_T - 1)} \quad \forall\, G^{(l)} \in \mathbb{G}.$$

Proof The KKT conditions in (8) and (9) yield $\lambda_l^* = 0$ for $G^{(l)} \in \mathbb{G}_i$. Further,

$$\begin{aligned}
0 &= \sum_{i=1}^n \alpha_i^* \left(2\sum_{l=1}^m \lambda_l^* \big(1 - \alpha_i^* - \frac{1}{\rho^{(l)}} \sum_j A_{ij}^{(l)} \alpha_j^*\big) + \delta_i^* \right) \\
&= \sum_{l: G^{(l)} \in \mathbb{G}_a} \lambda_l^* \sum_i \alpha_i^* \big(1 - \alpha_i^* - \frac{1}{\rho^{(l)}} \sum_j A_{ij}^{(l)} \alpha_j^*\big) && (\because \alpha_i^* \delta_i^* = 0 \forall_i) \\
&= \sum_{l: G^{(l)} \in \mathbb{G}_a} \lambda_l^* (t^* - \sum_i \alpha_i^*) = t^* - \sum_i \alpha_i^* && (\because \sum_l \lambda_l^* = 1).
\end{aligned}$$

This yields $t^* = \sum_i \alpha_i^*$, that is, the objective is equal to sum of the support vectors. We can rewrite feasibility condition $f(\alpha^*; \mathbf{K}^{(l)}) \geq t^*$ for all graphs $G^{(l)} \in \mathbb{G}$:

$$\begin{aligned}
0 &\leq \sum_{i\in SV} \alpha_i^* (2 - \alpha_i^* - \sum_{j\neq i} A_{ij}^{(l)} \alpha_j^*) - t^* \quad = \sum_{i\in SV} \alpha_i^* (1 - \alpha_i^* - \sum_{j\neq i} A_{ij}^{(l)} \alpha_j^*) \\
&= \sum_{i\in T} \alpha_i^* \underbrace{\big(1 - \alpha_i^* - \frac{d_i(G_{SV}^{(l)})}{\rho^{(l)}} \bar{\alpha}_{SV}^{(l)}\big)}_{>0} + \sum_{i\notin T} \alpha_i^* \underbrace{\big(1 - \alpha_i^* - \frac{d_i(G_{SV}^{(l)})}{\rho^{(l)}} \bar{\alpha}_{SV}^{(l)}\big)}_{<0} \\
&\leq \sum_{i\in T} \alpha_i^* \big(1 - \alpha_i^* - \frac{d_i(G_{SV}^{(l)})}{\rho^{(l)}} \bar{\alpha}_{SV}^{(l)}\big) \quad \leq \sum_{i\in T} \alpha_i^* \big(1 - \alpha_i^* - \frac{d_i(G_T^{(l)})}{\rho^{(l)}} \bar{\alpha}_{SV}^{(l)}\big) \\
&\leq \sum_{i\in T} \alpha_i^* \big(1 - \alpha_{\min} - \frac{d_i(G_T^{(l)})}{\rho^{(l)}} \bar{\alpha}_{SV}^{(l)}\big) \quad \leq \sum_{i\in T} \big(1 - \alpha_{\min} - \frac{d_i(G_T^{(l)})}{\rho^{(l)}} \bar{\alpha}_{SV}^{(l)}\big) \\
&\leq n_T(1 - \alpha_{\min}) - \frac{\bar{\alpha}_{SV}^{(l)}}{\rho^{(l)}} \sum_{i\in T} d_i(G_T^{(l)}) \quad = n_T(1 - \alpha_{\min}) - \frac{\bar{\alpha}_{SV}^{(l)}}{\rho^{(l)}} \gamma(G_T^{(l)}) n_T(n_T - 1).
\end{aligned}$$

Algorithm 1 $T = CSS(G^{(1)}, \ldots, G^{(M)})$

Get α^* using MKL solver to solve eqn. (7)
$T = \cap_{l=1}^{m} T^{(l)}$ {eqn. (13)}
Return T

Dividing by $\binom{n_T}{2}$ and rewriting, we get

$$\gamma(G_T^{(l)}) \leq \frac{(1-\alpha_{\min})\rho^{(l)}}{\bar{\alpha}_{SV}^{(l)}(n_T-1)},$$

which completes the proof. ■

The above result allows us to build a *parameter-less* common sparse subgraph (CSS) algorithm shown in Algorithm 1 having following advantages: it provides a theoretical bound on subgraph density; and, it requires no parameters from the user beyond the set of graphs $\mathbb{G}$.

The size of the induced subgraph n_T is important in the overall quality of the solution. Ideally, one would like n_T to be some large fraction of the overall number of nodes N, typically $n_T/N \leq 1/2$. However, if n_T is very large, that is, $n_T/N \simeq 1$, the density of the induced subgraph is close to the average graph density. More generally, one might be interested in a trade-off between the subgraph size n_T and subgraph density $\gamma(G_T^{(l)})$. Analogous to the simple graph case, we can improve the subgraph density is obtained by choosing smaller region nodes $T_c := \{i \in T : \alpha_i^* > c\} \subseteq T$. We discuss this further in Section 5.2.

4. SVM $-$ ϑ Graphs: Graphs Where $\vartheta(G)$ Can Be Approximated by SVM

Computing the Lovász function and related relaxations involves solving a semidefinite program. Off the shelf SDP solvers are computationally very demanding and do not scale to graphs of more than 5000 vertices. In this section we study Erdös-Rényi graphs, parametrized as $G(n,p)$ where p is probability of an edge and n is the number of vertices. As noted in Section 2, one can evaluate the ϑ function in $O(n^3)$ on $\mathcal{Q}$ graphs. Further, we have equality $\mathbf{ALPHA}(G) = \vartheta(G) = \omega(\mathbf{K})$ for $\mathcal{Q}$ graphs, where $\mathbf{K}$ is a **LS** labelling. It is well known that in $G(n,1/2)$ graph, with high probability, $\mathbf{ALPHA}(G) = \Theta(\log n)$ whereas $\vartheta(G) = \Theta(\sqrt{n})$ (Juhász, 1982; Coja-Oghlan, 2005). This immediately establishes a negative result that $G(n,1/2)$ graphs are not $\mathcal{Q}$ graphs. In the following we will show that despite the above negative result one can still obtain constant factor approximation to ϑ function on random graphs by solving a SVM problem. We begin by introducing a class of graphs, called $\mathbf{SVM}-\vartheta$ graphs, where the gap between ϑ and $\omega(\mathbf{K})$ for $\mathbf{K}$ defined by (2) is not too large. Subsequently we show that $G(n,p)$ graphs and $G(n,p)$ graphs with planted cliques are $\mathbf{SVM}-\vartheta$ graphs. This results immediately show that one can identify planted cliques or planted subgraphs. Furthermore we prove a concentration result on $\omega(\mathbf{K})$ for $G(n,p)$ graphs.

Definition 16 *A family of graphs $\mathcal{G} = \{G = (V,E)\}$ is said to be $\mathbf{SVM}-\vartheta$ graph family if there exist a constant γ, such that for any graph $G \in \mathcal{G}$ with $|V| \geq n_0$, the following holds:*

$$\omega(\mathbf{K}) \leq \gamma\vartheta(G),$$

where $\omega(\mathbf{K})$ is defined in (1) *and $\mathbf{K}$ is defined on G by* (2)*.*

Such classes of graphs are interesting because of two reasons. Firstly on these class of graphs one can approximate the Lovász function well without resorting to solving a SDP, and secondly the ϑ function in turn can be used in the design and analysis of approximation algorithms. We will demonstrate examples of such families of random graphs: the Erdös–Rényi random graph $G(n,p)$ and a planted variation. Here the relaxation $\omega(\mathbf{K})$ could be used in place of $\vartheta(G)$, resulting in algorithms with the same quality guarantees but with faster running time—in particular, this will allow the algorithms to be scaled to large graphs.

4.1 $G(n,p)$ Graphs Are SVM $-\vartheta$ Graphs

In this section we show that $G(n,p)$ graphs are indeed **SVM** $-\vartheta$. We begin with some preliminaries.

4.1.1 PRELIMINARIES

The following lemma is well known (see Boyd and Vandenberghe, 2004, Section 9.1.2):

Lemma 17 *A function $g : C \subset \mathbb{R}^d \to \mathbb{R}$ is said to be strongly concave over C if there exists $t > 0$ such that $\nabla^2 g(\mathbf{x}) \preceq -t\mathbf{I}\ \forall\ \mathbf{x} \in C$. For such functions one can show that if $p^* = g(\mathbf{x}^*) = \max_{\mathbf{x}\in C} g(\mathbf{x}) < \infty$ then*

$$\forall \mathbf{x} \in C: \quad \frac{t}{2}\|\mathbf{x}-\mathbf{x}^*\|^2 \ \le\ p^* - g(\mathbf{x}) \ \le\ \frac{1}{2t}\|\nabla g(\mathbf{x})\|^2.$$

The classical Erdös-Rényi random graph $G(n,p)$ has n vertices and each edge (i,j) is present independently with probability p. (In the closely related the $G(n,M)$ model, a graph is chosen uniformly at random from the collection of all graphs which have n nodes and M edges, or, equivalently, a set of M edges is chosen uniformly at random without replacement from the set of all possible $\binom{n}{2}$ edges. With $M = \binom{n}{2}p$, the two models are essentially equivalent.) For many types of random distributions, $G(n,p)$ is considered a paradigm choice for input instances and hence both the combinatorial structure and the algorithmic theory of $G(n,p)$ are of fundamental interest (Bollobás, 2001; Janson et al., 2000; Frieze and McDiarmid, 1997). We list a few well known facts about $G(n,p)$ that will be used repeatedly.

Remark 18 *For $G(n,p)$ for any $0 \le p < 1$,*

- *With probability $1 - O(1/n)$, the degree of each vertex is in the range $np \pm \sqrt{np\log n}$.*
- *With probability $1 - e^{-n^c}$ for some $c > 0$, the maximum eigenvalue is $np(1+o(1))$. The the minimum eigenvalue is in the range $[-2\sqrt{np(1-p)}, 2\sqrt{np(1-p)}]$ (Füredi and Komlós, 1981),*
- *With high probability, $\vartheta(G(n,p)) = \Theta(\sqrt{\frac{n(1-p)}{p}})$ [Coja-Oghlan, 2005; Juhász, 1982].*

4.1.2 $G(n,p)$ GRAPHS ARE SVM $-\vartheta$

We are now ready to state our main result.

Theorem 19 *Let $G = G(n,p)$, with $p(1-p) = \Omega(n^{-1}\log^4 n)$. For every constant $\delta > 0$,*

$$\omega(\mathbf{K}) \le (1+O(1))\vartheta(G)$$

holds with probability $1 - O(1/n)$, whenever $\mathbf{K}$ is defined in (2) with $\rho = (1+\delta)2\sqrt{np(1-p)}$.

Proof By Remark 18 for all choices of $\delta > 0$, the minimum eigenvalue of $\frac{1}{\rho}\mathbf{A} + \mathbf{I}$ is, almost surely, greater than 0 which implies that $f(\alpha, \mathbf{K})$ (see (1)) is strongly concave. As $\mathbf{A}$ is random we begin by analyzing the KKT conditions (3) for $\mathbb{E}(\mathbf{A})$, the expectation of $\mathbf{A}$. For $G(n,p)$ graph $\mathbb{E}(\mathbf{A}) = p(\mathbf{e}\mathbf{e}^\top - \mathbf{I})$. For the given choice of ρ, the matrix $\tilde{\mathbf{K}} = \frac{\mathbb{E}(\mathbf{A})}{\rho} + \mathbf{I}$ is positive definite. More importantly $f(\alpha, \tilde{\mathbf{K}})$ is again strongly concave and attains maximum at a KKT point. By direct verification $\hat{\alpha} = \hat{\beta}\mathbf{e}$ where $\hat{\beta} = \frac{\rho}{(n-1)p+\rho}$ satisfies the KKT conditions. More precisely

$$\hat{\alpha} + \frac{1}{\rho}\mathbb{E}(\mathbf{A})\hat{\alpha} = \mathbf{e}. \tag{14}$$

Thus $\hat{\alpha}$ is the optimal for the expected case with the optimal value, $\bar{f}$, given by

$$\bar{f} = \max_{\alpha \geq 0} f(\alpha, \tilde{\mathbf{K}}) = 2\sum_{i=1}^{n} \hat{\alpha}_i - \hat{\alpha}^\top \left(\frac{\mathbb{E}(\mathbf{A})}{\rho} + \mathbf{I} \right) \hat{\alpha} = n\hat{\beta}. \tag{15}$$

By choice of ρ, for any p in the regime $np \geq 1$ one notes that

$$\hat{\beta} = \frac{\rho}{(n-1)p+\rho} = \frac{\rho}{np}(1 + o(1)) = 2(1+\delta)\sqrt{\frac{1-p}{np}}. \tag{16}$$

The last equality holds by neglecting the $o(1)$ term. Using the fact about degrees of vertices in $G(n,p)$, in the regime of interest,

$$\mathbf{a}_i^\top \mathbf{e} = (n-1)p + \Delta_i \text{ with } |\Delta_i| \leq \sqrt{np\log n}, \tag{17}$$

where $\mathbf{a}_i^\top$ is the i^{th} row of the adjacency matrix $\mathbf{A}$. It is interesting to note that $\hat{\alpha}$ is an approximate KKT point for (3). Indeed, for all $i \in V$, application of (17) and (16) alongwith the choice of ρ as given in the statement of theorem we obtain

$$\left| \hat{\alpha}_i + \frac{1}{\rho}\sum_j A_{ij}\hat{\alpha}_j - 1 \right| = \left| \frac{\hat{\beta}}{\rho}\Delta_i \right| \leq \sqrt{\frac{\log n}{np}}. \tag{18}$$

We would like to exploit this property to approximate the $\omega(\mathbf{K})$ for a random graph. To this end note that

$$f(\hat{\alpha}; \mathbf{K}) = \hat{\alpha}^\top \mathbf{e} + \sum_{i=1}^{n} \hat{\alpha}_i \left(1 - \hat{\alpha}_i - \frac{\mathbf{a}_i^\top \mathbf{e}}{\rho} \right) = \bar{f} - \frac{\hat{\beta}^2}{\rho}\sum_{i=1}^{n} \Delta_i,$$

which on application of (18) and (16) yield

$$|f(\hat{\alpha}; \mathbf{K}) - \bar{f}| \leq \hat{\beta}\sum_{i=1}^{n} \left| \frac{\hat{\beta}}{\rho}\Delta_i \right| \leq n\hat{\beta}\sqrt{\frac{\log n}{np}} = \left(2(1+\delta)\sqrt{\frac{(1-p)}{p}} \right) \sqrt{\frac{\log n}{p}}. \tag{19}$$

As noted before the function $f(\alpha; \mathbf{K})$ is strongly concave with $\nabla^2_\alpha f(\alpha; \mathbf{K}) \preceq -\frac{\delta}{1+\delta}\mathbf{I}$ for all feasible α. Recalling a useful result from convex optimization, see Lemma 17, we obtain

$$\omega(\mathbf{K}) - f(\hat{\alpha}; \mathbf{K}) \leq \frac{1}{2}\left(1 + \frac{1}{\delta} \right) \|\nabla f(\hat{\alpha}; \mathbf{K})\|^2. \tag{20}$$

Observing that $\nabla f(\alpha;\mathbf{K}) = 2(\mathbf{e}-\alpha-\frac{\mathbf{A}}{\rho}\alpha)$ and using the relation between $\|\cdot\|_\infty$ and $\|.\|$ along with (18) and (17) gives $\|\nabla f(\hat{\alpha};\mathbf{K})\| \leq \sqrt{n}\|\nabla f(\hat{\alpha};\mathbf{K})\|_\infty \leq 2\sqrt{\frac{\log n}{p}}$. Plugging this estimate in (20) and using equation (15) along with (19) we obtain

$$\omega(\mathbf{K}) \leq n\hat{\beta} + \left(1+\frac{1}{\delta}\right)\left(\frac{\log n}{p}\right) + 2(1+\delta)\sqrt{\frac{1-p}{p}}\sqrt{\frac{\log n}{p}}$$

$$= 2(1+\delta)\sqrt{\frac{n(1-p)}{p}} + O\left(\frac{\log n}{p}\right).$$

The second inequality is true because δ is constant. One notes that

$$\text{for any } p(1-p) \geq \frac{(\log n)^2}{n},\ \left(\frac{\log n}{p}\right) \leq \sqrt{\frac{n(1-p)}{p}} \tag{21}$$

holds. The result follows by dividing the first inequality by p^2 and taking square roots. By choice of p as stated in the theorem and for large enough n, one obtains

$$\omega(\mathbf{K}) \leq (1+O(1))\sqrt{\frac{n(1-p)}{p}}$$

and the theorem follows from Remark 18. ■

In the next section we show that when a large independent set is hidden in a random graph one can still view it as a $\mathbf{SVM}-\vartheta$ graph. This property can be very useful in detecting planted cliques in dense graphs.

4.2 Finding Planted Cliques in Random Graphs

Finding large cliques or independent sets is a computationally difficult problem even in random graphs. While it is known that the size of the largest clique or independent set in $G(n,1/2)$ is $2\log n$ with high probability, there is no known efficient algorithm to find a clique of size significantly larger than $\log n$ - even a cryptographic application was suggested based on this (see the discussion and references in the introduction of Feige and Krauthgamer, 2000). Motivated by this, the following problem was introduced (Jerrum, 1992; Kucera, 1995).

Definition 20 (Hidden Planted Clique) *A random $G(n,q)$ graph is chosen first and then a clique of size k is introduced in the first $1,\ldots,k$ vertices. The problem is to identify the clique.*

The case of $q=\frac{1}{2}$ is extensively studied in the literature. Kucera (1995) observed that if $k = \Omega(\sqrt{n\log n})$, then the hidden clique can be easily discovered by examining the high degree vertices. Alon et al. (1998) and Feige and Krauthgamer (2000) showed that if $k=\Omega(\sqrt{n})$, then the hidden clique can be discovered in polynomial time. No efficient algorithm is known to discover the hidden clique if $k=o(\sqrt{n})$.

We consider the (equivalent) complement model $\bar{G}(n,1-p,k)$ where an independent set is planted on the first k vertices and apply the SVM based approach. We show that

Theorem 21 *For $p(1-p) = \Omega(n^{-1}\log^4 n)$, the graph $G = \bar{G}(n,\ 1-p,\ k)$ is a Q graph almost surely if $k = \Omega\left(n^{2/3}p^{-1/3}\ln^{1/3} n\right)$.*

Proof See Appendix B.1 ■

The result shows that the SVM based formulation yields an integral solution for $k \geq c \cdot n^{2/3}\ln^{1/3} n$. More precisely the optimal α for a given G has the property that $\alpha_i = 1$ whenever $i \in \{1,\ldots,k\}$ otherwise $\alpha_i = 0$ whenever k is in the stated regime. Unfortunately this is interesting but not very competitive with state of the art. For $p = \frac{1}{2}$ in the regime, $k = \Omega(\sqrt{n\log n})$, already the highest degree vertices form a clique in the complement $G(n,1/2,k)$ (Kucera, 1995). However, we show next that in the regime $k = \Theta\left(\sqrt{\frac{n(1-p)}{p}}\right)$, the graph $\bar{G}(n,1-p,k)$ is a $\mathbf{SVM}-\vartheta$ graph. Moreover, we can identify the hidden independent set with high probability. In contrast to Feige and Krauthgamer (2000), our algorithm does not involve a SDP and hence will scale to large problems. In Feige and Krauthgamer (2000) the case of $p = 1/2$ was studied and it was conjectured that the results could possibly be extended for a general $p \leq \frac{1}{2}$. In this paper we establish that for large planted clique $\bar{G}(n,1-p,k)$ is indeed a $\mathbf{SVM}-\vartheta$ graph. Moreover the proof motivates an algorithm capable of recovering the clique.

To begin the investigation we will need to compute the ϑ function for $\bar{G}(n,1-p,k)$. For $p = 1/2$, it was shown that the ϑ function is k (Feige and Krauthgamer, 2000). For a general p the ϑ function is also the same as k. To this end we present the following theorem.

Theorem 22 *Let $G = \bar{G}(n,1-p,k)$ where p satisfies $p(1-p) = \Omega(n^{-1}\log^4 n)$. If*

$$k \geq 2\sqrt{\frac{(1-p)}{p}n}(1+o(1)),$$

then, with high probability, $\vartheta(G) = k$.

Proof See Appendix B.2. ■

We study the planted clique problem where k is in the above regime. Indeed for such a choice of k, the graph $\bar{G}(n,1-p,k)$ is a $\mathbf{SVM}-\vartheta$ graph. One of the main contribution of this paper is the following theorem.

Theorem 23 *Let $p(1-p) = \Omega(n^{-1}\log^4 n)$ and $\delta > 0$ be a given constant. For any $G = \bar{G}(n,1-p,k)$ and $k = 2t\sqrt{\frac{n(1-p)}{p}}$, for large enough constant $t \geq 1$ with $\mathbf{K}$ as in (2) and*

$$\rho = \left(2\sqrt{np(1-p)} + kp\right)(1+\delta),$$

the following holds:

$$\omega(\mathbf{K}) \leq \left(1 + \frac{1}{t}(1+\delta) + \delta\right)\vartheta(G) + o(1), \tag{22}$$

with probability at least $1 - O(1/n)$.

Note that we need a stronger regime for p in the above theorem when compared to Theorem 19. This is necessary in view of the conditions in Theorem 21 and Theorem 22.

As a preliminary, we need a bound on the minimum eigenvalue of $\bar{G}(n,1-p,k)$:

Lemma 24 *With high probability, the minimum eigenvalue of $\bar{G}(n,1-p,k)$ is bounded in absolute value by $2\sqrt{np(1-p)}+kp$.*

Proof We write the adjacency matrix $\mathbf{A}$ of $\bar{G}(n,p,k)$ as $\mathbf{A}' = \mathbf{A}+\mathbf{E}$ where $\mathbf{A}'$ is the adjacency matrix of $G(n,1-p)$ and observe that E is zero except for a $k\times k$ block where it is the adjacency matrix of $G(k,1-p)$. Using the Weyl perturbation inequality (Horn and Johnson, 1990), we deduce that $\lambda_n(\mathbf{A}) \geq \lambda_n(\mathbf{A}') - \lambda_1(\mathbf{E})$. By Remark 18 we see that with high probability $\lambda_n(\mathbf{A}') \geq -2\sqrt{np(1-p)}$ and $\lambda_1(\mathbf{E}) = kp$ ■

Proof (of Theorem 23) The proof is analogous to that of Theorem 19. Let $\mathbf{A}$ be the adjacency matrix of G. By definition of ρ and by Lemma 24 we see that $\rho = (1+\delta)|\lambda_n(\mathbf{A})|$. For any $\delta > 0$, we see that the minimum eigenvalue of $\frac{1}{\rho}\mathbf{A}+\mathbf{I}$ is almost surely *strictly* greater than 0. As a consequence the function $f(\alpha,\mathbf{K})$ (see (1)) is strongly concave for any instance of G. We begin by analysing the KKT conditions, which are necessary and sufficient for optimising strongly concave problems. The KKT conditions in (3) specialises to the graph G as follows

$$\begin{array}{rcl} \alpha_i + \frac{1}{\rho}\sum_{j=k+1}^{n} A_{ij}\alpha_j & = & 1+\mu_i, \quad \mu_i\alpha_i = 0, \quad \mu_i \geq 0 \quad \forall 1 \leq i \leq k; \quad \text{and} \\ \alpha_i + \frac{1}{\rho}\sum_{j=1}^{n} A_{ij}\alpha_j & = & 1+\mu_i, \quad \mu_i\alpha_i = 0, \quad \mu_i \geq 0 \quad \forall k+1 \leq i \leq n. \end{array} \tag{23}$$

Let us analyse the average case. By the conditions of the theorem, the expectation of $\mathbf{A}$ is defined as follows

$$\mathbb{E}(\mathbf{A})_{ij} = \begin{cases} 0 & \text{whenever } 1 \leq i,j \leq k \text{ and } i = j, \\ p & \text{otherwise.} \end{cases}$$

By direct verification the matrix $\tilde{\mathbf{K}} = \frac{\mathbb{E}(\mathbf{A})}{\rho} + \mathbf{I}$ is seen to be positive definite. More importantly $f(\alpha,\tilde{\mathbf{K}})$ is again strongly concave and attains maximum at a KKT point. A KKT point for the problem

$$\max_{\alpha \geq 0} f(\alpha,\tilde{\mathbf{K}})$$

is given by $\hat{\alpha} = [\beta_1 \mathbf{e}_k^\top \ \beta_2 \mathbf{e}_{(n-k)}^\top]$ where β_1 and β_2 satisfies

$$\beta_1 + \frac{(n-k)p}{\rho}\beta_2 = 1, \quad \text{and,} \quad \beta_2 + \frac{k}{\rho}p\beta_1 + \frac{(n-k-1)}{\rho}p\beta_2 = 1;$$

which leads to

$$\beta_1 = \frac{1}{1+\frac{np}{\rho}(1-r)}, \quad \text{and,} \quad \beta_2 = \frac{1}{\frac{1}{1-r}+\frac{np}{\rho}}.$$

For k and ρ as given, the value of r is defined as

$$r = \frac{k}{\rho}p = \frac{t}{(1+t)(1+\delta)}, \qquad 1-r \approx \frac{1}{t+1}.$$

For small δ, the second approximation holds. Further noting that $k = o(n)$ allows us to write

$$\beta_1 = \frac{(t+1)^2}{(t+1)^2 + \frac{1}{2}\sqrt{\frac{np}{1-p}}\frac{1}{1+\delta}} \qquad \text{and} \qquad \beta_2 = \frac{(t+1)}{(t+1)+\frac{1}{2}\sqrt{\frac{np}{1-p}}\frac{1}{1+\delta}}. \tag{24}$$

By construction $\hat{\alpha}$ is the optimal solution for the expected case and the optimal value, $\bar{f}$ is given by

$$\bar{f} = k\beta_1 + (n-k)\beta_2 \;\leq\; o(k) + 2\frac{n(t+1)(1+\delta)}{\sqrt{\frac{np}{1-p}}} \;=\; \left(1+\delta+\frac{1}{t}(1+\delta)+o(1)\right)k. \qquad (25)$$

As $\beta_1 < 1$, the first term is $o(k)$. The second inequality follows by noting that second term can be upper-bounded by neglecting $(t+1)^2$ in the denominator of β_2. The last equality follows from definition of k. Now as in the case of $G(n,p)$ graphs, we can indeed show that $\hat{\alpha}$ is an approximate KKT solution for a random graph G.

Let $deg(i,T)$ = number of edges from i to T where $T \subset V$. Since $deg(j,S) \sim Bin(k,p)$ for any $j \in \bar{S}(= V \backslash S)$ random variable, we can deduce that

$$|deg(j,S) - kp| \leq \sqrt{kp\log k} \text{ with probability at least } 1 - O(1/k)$$

by application of Chernoff bound. Similarly the following is true for any $j \in V$

$$|deg(j,\bar{S}) - (n-k)p| \leq \sqrt{(n-k)p\log(n-k)}$$

with high probability. Using in $\hat{\alpha}$ in (23) we obtain

$$\begin{aligned}
\forall i \in S \qquad & |\beta_1 + \tfrac{deg(i,\bar{S})}{\rho}\beta_2 - 1| \leq && \frac{\sqrt{(n-k)p\log(n-k)}}{\rho}\beta_2, \quad \text{and} \\
\forall i \in \bar{S} \quad & |\beta_2 + \tfrac{deg(i,S)}{\rho}\beta_1 + \tfrac{deg(i,\bar{S})}{\rho} - 1| \leq && \frac{\sqrt{kp\log k}}{\rho}\beta_1 + \frac{\sqrt{(n-k)p\log(n-k)}}{\rho}\beta_2.
\end{aligned}$$

Plugging in the values of β_1 and β_2, see (24), gives us the following claim.

Lemma 25 *Let β_1 and β_2 be defined in* (24)*. The vector $\bar{\alpha}^\top = [\beta_1 \mathbf{e}_k^\top \; \beta_2 \mathbf{e}_{n-k}^\top]$ satisfies the following relationship with $\varepsilon = O\left(\sqrt{\frac{\log n}{np}}\right)$.*

$$\begin{aligned}
1 \leq i \leq k \qquad & |\hat{\alpha}_i + \frac{1}{\rho}\sum_{j=k+1}^{n} A_{ij}\hat{\alpha}_j - 1| \leq \varepsilon, && \textit{and} \\
k+1 \leq i \leq n \qquad & |\hat{\alpha}_i + \frac{1}{\rho}\sum_{j=1}^{n} A_{ij}\hat{\alpha}_j - 1| \leq \varepsilon.
\end{aligned}$$

Proof By direct verification. ■

The above lemma establishes that $\hat{\alpha}$ is an ε-approximate KKT point. Indeed

$$|f(\hat{\alpha};\mathbf{K}) - \bar{f}| \leq (k\beta_1 + (n-k)\beta_2)\varepsilon = O(k\varepsilon) = O\left(\frac{1}{\sqrt{p}}\sqrt{\frac{(1-p)}{p}}\sqrt{\log n}\right). \qquad (26)$$

The first equality follows by using Lemma 25, and the second equality follows from (25). The last equality follows from definition of k and ε. Again using the strong concavity condition $\nabla^2_\alpha f(\alpha;\mathbf{K}) \preceq -\frac{\delta}{1+\delta}\mathbf{I}$ for all feasible α and using Lemma 17, we obtain

$$\omega(\mathbf{K}) - f(\hat{\alpha};\mathbf{K}) \leq \frac{1}{2}\left(1+\frac{1}{\delta}\right)\|\nabla f(\hat{\alpha};\mathbf{K})\|^2. \qquad (27)$$

Algorithm 2 Input: $\mathbf{A}(\bar{G}(n,1-p,k)),k$ Output: $S \subset \{1,\ldots,n\}, |S| = k$

$\alpha^* =$ Use SVM solver to solve (1) with $\mathbf{K}$ as defined in Theorem 23

Return $S = \{i_j | j = 1,\ldots,k\ \alpha^*_{i_1} \geq \alpha^*_{i_2} \ldots \geq \alpha^*_{i_n}\}$

Recalling that $\nabla f(\alpha;\mathbf{K}) = 2(\mathbf{e} - \alpha - \frac{\mathbf{A}}{\rho}\alpha)$ and using the relation between $\|\cdot\|_\infty$ and 2 norm along with Lemma 25 gives $\|\nabla f(\hat{\alpha};\mathbf{K})\| \leq \sqrt{n}\|\nabla f(\hat{\alpha};\mathbf{K})\|_\infty \leq O(\sqrt{\frac{\log n}{p}})$. Plugging this estimate in (27) and using equation (26) we obtain

$$\omega(\mathbf{K}) \leq \hat{f} + O\left(\frac{\log n}{p}\right) + O\left(\frac{1}{\sqrt{p}}\sqrt{\frac{(1-p)}{p}}\sqrt{\log n}\right).$$

For the given choice of p one can use (21) to write

$$\omega(\mathbf{K}) \leq \left(1 + \frac{1}{t}(1+\delta) + \delta + o(1)\right)k.$$

By Theorem 22, $\vartheta(\bar{G}(n,1-p,k) = k$ holds with high probability, proving (22). ■

The above theorem establishes that $\hat{\alpha}$ is a very good approximation for the optimal solution with high probability.

We note that $\hat{\alpha}_i$, for any vertex i in the independent set $S = \{1,\ldots,k\}$, is $(t+1)$ times larger than any other $\hat{\alpha}_j \notin S$. Since α^* is very close to $\hat{\alpha}$ one might expect to see the same property in α^*. This motivates Algorithm 2 for finding planted clique in a random graph.

It is expected that with high probability the largest k values in the optimal solution would correspond to nodes in the independent set. In a later section we will present empirical results which show that the algorithm is extremely successful in recovering the clique. The runtime of this algorithm is dependent on the availability of an efficient SVM solver. Current SVM solvers have complexity less than $O(n^2)$, hence the proposed procedure is indeed scalable to large graphs. In contrast the best known algorithm for solving Lovász ϑ has a runtime complexity of $O(n^5 \log n)$ (Chan et al., 2009). The algorithm in Feige and Krauthgamer (2000) is based on computation of the ϑ function and hence will not scale well to large graphs.

4.3 Finding Planted Subgraphs in Random Graphs

The above results show that the SVM procedure can recover a planted independent set in a sparse random graph, which is later exploited to solve the planted clique problem in a dense graph. Moreover we have also established that on random graphs the SVM objective function is tightly concentrated. This points to a more general question that of identifying a planted subgraph, which is much sparser than the original graph.

Let $G(n,p,p',k)$ be a graph which has a planted $G(k,p')$ graph on the first k vertices of a random $G(n,p)$ graph. We consider the problem of recovering the planted subgraph.

An interesting property of ϑ function is that deletion of edges always results in an increases of the ϑ function. As a consequence for any k

$$\vartheta(\bar{G}(n,1-p,k)) \geq \vartheta(G(n,p,p',k)) \geq \vartheta(G(n,p))$$

whenever $p' < p$. By Remark 18 and by Theorem 22 we have $\vartheta(\bar{G}(n,1-p,k))$ and $\vartheta(G(n,p))$ are both $\Theta\left(\sqrt{\frac{n(1-p)}{p}}\right)$ whenever $k = \Theta\left(\sqrt{\frac{n(1-p)}{p}}\right)$. Indeed one can show that the graph with a planted subgraph is also a $\mathbf{SVM}-\vartheta$ graph. More formally

Theorem 26 *Let $G = G(n,p,p',k)$ with $p(1-p) = \Omega(n^{-1}\log^4 n)$ and $p' < p$. If*

$$k = 2t\left(\sqrt{\frac{n(1-p)}{p}}\frac{1-2p'}{1-\frac{p'}{p}}\right),$$

then the graph G satisfies

$$\omega(\mathbf{K}) \leq \left(1+\frac{1}{t}(1+\delta)\right)\vartheta(G),$$

with probability at least $1-O(1/n)$ where $\mathbf{K}$ as in (2) and $\rho = \left(2\sqrt{np(1-p)}+k(p-p')\right)(1+\delta)$ for any constant $\delta > 0$.

The proof can be constructed as in Theorem 23 so we do not repeat it here. Instead we make some observations which would help to construct a formal proof by repeating the arguments in Theorem 23. One would need to compute an upper-bound on $\lambda_n(\mathbf{A})$ where $\mathbf{A}$ is the adjacency matrix of $G(n,p,p',k)$. We write $\mathbf{A}' = \mathbf{A}+\mathbf{E}$ where $\mathbf{A}'$ is the adjacency matrix of $G(n,p)$ graph. The matrix $\mathbf{E}$ is defined as follows. Set all entries of $\mathbf{E}$ to be 0 except the first $k \times k$ submatrix. On that block set

$$E_{ij} = 1 \text{ with probability } s = \tfrac{p-p'}{1-2p'} \text{ whenever } A_{ij} = 0 \text{ for } 1 \leq i,j \leq k,$$

otherwise $E_{ij} = 0$. By such a choice we see that $\mathbf{E}$ is the adjacency matrix of $G(k,s)$ on the first $k \times k$ block and 0 everywhere else. To check if $P(A'_{ij} = 1) = p$ we note that all entries, except in the leading $k \times k$ submatrix, are $Ber(p)$ distributed. In the leading $k \times k$ submatrix we note that $A'_{ij} = 1$ is the union of two mutually exclusive events, namely $A_{ij} = 1, E_{ij} = 0$ or $A_{ij} = 0, E_{ij} = 1$. Substituting the value of s confirms

$$P(A'_{ij} = 1) = P(A_{ij} = 1)P(E_{ij} = 0) + P(A_{ij} = 0)P(E_{ij} = 1) = p'(1-s) + (1-p')s = p.$$

Now by Weyl Perturbation (Horn and Johnson, 1990) we can write

$$\lambda_n(\mathbf{A}) \geq \lambda_n(\mathbf{A}') - \lambda_1(\mathbf{E}),$$

which leads to $|\lambda_n(\mathbf{A})| \leq 2\sqrt{np(1-p)} + k\frac{p-p'}{1-2p'}$. This explains the choice of ρ. As before the expected case analysis leads to the following equation

$$\beta_1 + \frac{kp'}{\rho}\beta_1 + \frac{(n-k)p}{\rho}\beta_2 = 1 \quad \text{and} \quad \beta_2 + \frac{kp}{\rho}\beta_1 + \frac{(n-k-1)p}{\rho}\beta_2 = 1.$$

The theorem follows by retracing the steps in the proof of Theorem 23. As in the planted clique case the Algorithm 2 recovers the planted subgraph. The proofs in this section suggest that $\omega(\mathbf{K})$ maybe concentrated. Before we close the section we would like to show that indeed this is the case.

4.4 Concentration of $\omega(\mathbf{K})$

The SVM objective function, computed on the random graphs in the previous sections, is also highly concentrated. Using the celebrated Talagrand's inequality (Talagrand, 1995), one can prove the following theorem

Theorem 27 *Let $\omega(\mathbf{K})$ be defined as in* (1)*. Let G and ρ satisfy either of*

- $G = G(n, p)$, $\rho = (1+\delta)2\sqrt{np(1-p)}$
- $G = \bar{G}(n, 1-p, k)$, $\rho = (1+\delta)(2\sqrt{np(1-p)}+kp)$ *where* $k = 2t\sqrt{\frac{n(1-p)}{p}}$,

where $p(1-p) = \Omega(n^{-1}\log^2 n)$. Then there exists a constant M such that

$$|\omega(\mathbf{K})) - M| \leq O\sqrt{\frac{\ln^3 n}{np^3(1-p)}}$$

with probability $1 - O(\frac{1}{n})$ whenever $\mathbf{K}$ is defined in (2)*.*

Before we begin the proof we collect some results. Let $\omega(\mathbf{K}) = f(\alpha^*, \mathbf{K})$, see (1). For any graph G satisfying the conditions of the theorem we can establish that

Lemma 28 *For α^* be defined as above we have $\|\alpha^*\| = O\left(\sqrt{\frac{\log n}{p}}\right)$ with probability $1 - \frac{1}{O(n)}$*

Proof For $G(n,p)$ case, application of Lemma 17 allows us to rewrite (20)

$$\frac{1}{2}\frac{\delta}{1+\delta}\|\hat{\alpha} - \alpha^*\|^2 \leq \omega(\mathbf{K}) - f(\hat{\alpha}; \mathbf{K}) \leq \frac{1}{2}\left(1 + \frac{1}{\delta}\right)\|\nabla f(\hat{\alpha}; \mathbf{K})\|^2 \leq O\left(\sqrt{\frac{\log n}{p}}\right)$$

where $\hat{\alpha}$ satisfies (14). Note that this holds with probability at least $1 - \frac{1}{O(n)}$ This yields the result $\|\hat{\alpha} - \alpha^*\| \leq O\left(\sqrt{\frac{\log n}{p}}\right)$ and the lemma follows by noting that $\|\hat{\alpha}\| = O(1)$. Similarly for $G = G(n, 1-p, k)$, application of Lemma 17 to (27) yields the lemma. ■

We employ the following version of Talagrand's inequality

Theorem 29 (Dubhashi and Panconesi, 2009) *Let $x = (x_1, \ldots, x_d)$ where $x_1, \ldots, x_d$ are independent random variables with $x_i \in \Omega_i$. Let $f : \Omega = (\Omega_1 \times \ldots \times \Omega_d) \to \mathbb{R}$ be a function with the property that if for each x in the domain of f, there exists a vector $c = c(x)$ such that $\|c\| \leq B$ and*

$$f(x) \leq f(y) + \sum_{x_i \neq y_i} c_i, \ \forall\, y \in \Omega,$$

then for any $s > 0$:

$$Pr[|f(x) - M| > Bs] \leq 4e^{-s^2/4},$$

where M denotes the median of f.

Proof See Theorem 11.2 in Dubhashi and Panconesi (2009). ■

The application of the above theorem establishes the result

Lemma 30 *Let* $\mathbf{A}$ *be the adjacency matrix of G, as defined in Theorem 27. Let* α^* *denote the solution to the optimization problem*

$$v(\mathbf{A}) = \max_{\alpha \geq 0} 2e^T\alpha - \alpha^T\left(I + \frac{\mathbf{A}}{\rho}\right)\alpha \tag{28}$$

for some constant ρ*. Then for any* $b > 0$*, there exists a constant* M *such that*

$$Pr\left[|v(\mathbf{A}) - M| > \frac{\sqrt{2}b^2 s}{\rho}\right] \leq 4e^{-s^2/4} + Pr[\|\alpha^*\| > b] \tag{29}$$

for any $s > 0$*.*

Proof Let β^* be the optimal solution of

$$f(\mathbf{A}) := \max_{\beta \geq 0, \|\beta\| \leq b} 2e^T\beta - \beta^T\left(I + \frac{\mathbf{A}}{\rho}\right)\beta.$$

Since (28) can be seen as a relaxation of this, it follows that if α^* is feasible, that is, $\|\alpha^*\| \leq b$, then $v(\mathbf{A}) = f(\mathbf{A})$. For any $D > 0$, this means that if for some M, the inequality $|v(\mathbf{A}) - M| > D$ holds then either $|f(\mathbf{A}) - M| > D$ or $\|\alpha^*\| > b$, so

$$Pr[|v(\mathbf{A}) - M| > D] \leq Pr[|f(\mathbf{A}) - M| > D] + Pr[\|\alpha^*\| > b]. \tag{30}$$

By the definition of f we know that for any adjacency matrix $\mathbf{A}'$

$$\begin{aligned} f(\mathbf{A}') &\geq 2e^T\beta^* - (\beta^*)^T\left(\mathbf{I} + \frac{\mathbf{A}'}{\rho}\right)\beta^* \\ &= f(\mathbf{A}) + \frac{1}{\rho}\sum_{i,j}(A_{ij} - A'_{ij})\beta_i^*\beta_j^* \\ &\geq f(\mathbf{A}) - \frac{1}{\rho}\sum_{A_{ij} \neq A'_{ij}} 2\beta_i^*\beta_j^*, \end{aligned}$$

where the last sum goes over all $i > j$. Let $c_{ij} = \frac{2\beta_i^*\beta_j^*}{\rho}$ for $i > j$. We note that

$$\sum_{i>j} c_{ij}^2 \leq 2\frac{1}{\rho^2}\sum_{i,j}(\beta_i^*)^2(\beta_j^*)^2 = \frac{2\|\beta^*\|^4}{\rho^2} \leq \frac{2b^4}{\rho^2}.$$

The function f is defined on $A_{ij}, i < j$ a total of $d = \binom{n}{2}$ Bernoulli random variables for the $G(n,p)$ case. For the planted clique case one needs to consider $d = \binom{n}{2} - \binom{k}{2}$ random variables. Applying Theorem 29 on f with $c_{ij} = \frac{2\beta_i^*\beta_j^*}{\rho}$ for $i > j$ and $B = \sqrt{2}b^2/\rho$ we get that there exists an M such that for any $s > 0$

$$Pr\left[|f(\mathbf{A}) - M| > \frac{\sqrt{2}b^2 s}{\rho}\right] \leq 4e^{-s^2/4}.$$

Using this bound together with (30) where $D = \sqrt{2}b^2 s/\rho$ implies (29). ■

Proof [Proof of Theorem 27.] By Lemma 28 With probability $1 - \frac{1}{O(n)}$, for either one of $G = G(n, p)$ or $G = \bar{G}(n, 1-p, k)$, there exists a constant $C > 0$ such that $\|\alpha^*\| \leq C\sqrt{\frac{\ln n}{p}}$. Applying Lemma 30 with $b = C\sqrt{\frac{\ln n}{p}}$ and $s = 2\sqrt{\ln n}$ yields

$$Pr\left[|v(\mathbf{A}) - M| > C'\frac{\sqrt{\ln^3 n}}{\rho p}\right] = O\left(\frac{1}{n}\right),$$

for some constant $C' > 0$. By assumption $\rho = \Theta\left(\sqrt{np(1-p)}\right)$, and hence the above equation establishes that $|v(\mathbf{A}) - M| \leq O\sqrt{\frac{\ln^3 n}{np^3(1-p)}}$ with probability $1 - O(\frac{1}{n})$. ■

5. Experimental Evaluation

In Section 3 an algorithm was proposed which was capable of discovering a large common dense subgraph in a collection of graphs. In Section 4.2 an algorithm for discovering a large planted clique in a single graph was discussed. In this section we examine the performance of the proposed algorithms. The code and data for the experiments reported here is available online at `http://www.cse.chalmers.se/~jethava/svm-theta.html`.

5.1 Common Dense Subgraph Detection

This subsection presents an experimental evaluation of the *CSS* algorithm for finding large dense subgraphs across multiple graphs. We study an abstraction of the problem using the DIMACS'94 data set which consists of graphs motivated from a number of different practical problems (Johnson and Trick, 1996). The key property of the algorithm is that it is parameter-less, and that it finds large common dense regions - which was not possible using enumerative approach (Jiang and Pei, 2009). We also investigate a thresholding heuristic which improves induced subgraph density at the cost of subgraph size.

5.1.1 DATA SET

We evaluate our algorithm for finding large dense regions on the DIMACS Challenge graphs,[2] which is a comprehensive benchmark for testing of clique finding and related algorithms. Each of the graph families in DIMACS (*brock*, *c-fat*, *p_hat*, *san*, *sanr*) is motivated by carefully selected real world problems, for example, fault diagnosis (*c-fat*), etc.; thus covering a wide range of practical scenarios (Johnson and Trick, 1996).

We evaluate the algorithm on following class of graphs *c-fat* graphs which are based on fault diagnosis problems and are relatively sparse; and, *p_hat* graphs which are generalizations of Erdös-Rényi random graphs; and are characerized by wider degree spread compared to classical $G(n, p)$

2. The DIMACS benchmark data set is available online at `ftp://dimacs.rutgers.edu/pub/challenge/graph/benchmarks/clique/`.

model. For the families of dense graphs (*brock*, *san*, *sanr*), we focus on finding large dense region in the complement of the original graphs.

5.1.2 EVALUATION

We run Algorithm 1 using off-the-shelf MKL solver SimpleMKL,[3] to find large common dense subgraph. In order to evaluate the performance of our algorithm, we compute $\bar{a} = \max_l a^{(l)}$ and $\underline{a} = \min_l a^{(l)}$ where $a^{(l)} = \gamma(G_T^{(l)})/\gamma(G^{(l)})$ is relative density of induced subgraph (compared to original graph density); and n_T/N is relative size of induced subgraph compared to original graph size. We want a high value of n_T/N; while $\underline{a}$ should not be lower than 1.

Graph family	N	M	n_{SV}	n_T	$\bar{a}$	$\underline{a}$
c-fat200	200	3	100	90	**2.3613**	0.99165
c-fat500	500	4	152	140	**3.8846**	1.0182
brock200‡	200	4	165	164	1.0704	0.9954
brock400‡	400	4	397	349	1.0214	1.0084
brock800‡	800	4	795	686	1.0129	1.0062
p_hat300	300	3	158	157	**1.5296**	1.1456
p_hat500	500	3	239	238	**1.5551**	1.1722
p_hat700	700	3	313	312	**1.5782**	1.1818
p_hat1000	1000	3	429	428	**1.5976**	1.1879
p_hat1500	1500	3	574	573	**1.6313**	1.2011
san200‡	200	5	200	185	1.0458	1.0029
san400‡	400	3	359	358	1.0411	0.9947
sanr200‡	200	2	157	156	1.1402	0.9949
sanr400‡	400	2	343	342	1.0355	1.0017

Table 1: Common dense subgraph recovery on multiple graphs in DIMACS data set. Here $\bar{a}$ and $\underline{a}$ denote the maximum and minimum relative density of the induced subgraph (relative to density of the original graph); n_{SV} and n_T denotes the number of support vectors and size of subset $T \subseteq SV$ returned by Algorithm 1. *The enumerative algorithm does not run for large dense regions* (see Jiang and Pei, 2009, Figure 17).

Table 1 shows evaluation of Algorithm 1 on DIMACS data set. We note that our algorithm finds a large subgraph (large n_T/N) with higher density compared to original graph in all of DIMACS graph classes making it suitable for finding large dense regions in multiple graphs. We note that traditional set enumerative methods fail to handle dense subgraph recovery for the case when n_T/N is large. For example, finding quasicliques of size $n_T \simeq 60$ requires 11.5 hours (see Jiang and Pei, 2009, Figure 17); in contrast to MKL-based approach which takes less than 1 minute.

The results in Table 1 show that in case of *c-fat* and *p_hat* graph families, the induced subgraph density is significantly improved (evidenced by high $\bar{a}$ and $\underline{a}$); and, the number of support vectors n_{SV} is a large fraction of N ($n_{SV}/N \simeq 1/2$). Thus, the algorithm recovers a large common dense subgraph.

3. The SimpleMKL solver is available online at `http://asi.insa-rouen.fr/enseignants/~arakotom/code/mklindex.html`.

Graph family	N	M	n_T	n_{S_c}	$\bar{a}(T)$	$\underline{a}(T)$	$\bar{a}(S_c)$	$\underline{a}(S_c)$
brock200‡	200	4	164	83	1.0704	0.9954	1.36	0.99
brock400‡	400	4	349	199	1.0214	1.0084	1.15	1.05
brock800‡	800	4	686	398	1.0129	1.0062	1.08	1.01
san200‡	200	5	185	100	1.0458	1.0029	1.51	1.08
san400‡	400	3	358	180	1.0411	0.9947	1.19	1.02
sanr200‡	200	2	156	79	1.1402	0.9949	1.86	1.04
sanr400‡	400	2	342	172	1.0355	1.0017	1.20	1.02

Table 2: Heuristic rule: If $n_{SV} \geq 0.8n$ then choose c such that $|S_c| = n_{SV}/2$. Here $\bar{a}$ and $\underline{a}$ denote the maximum and minimum relative density of the induced subgraph (relative to density of the original graph). Induced subgraph density improves by choosing S_c instead of T, that is, $\bar{a}(S_c) \geq \bar{a}(T)$ and $\underline{a}(S_c) \geq \underline{a}(T)$ for all graphs.

On the other hand, for *brock* and *san* graph families, the number of support vectors is equal to the overall graph size $n_{SV} \simeq N$; and consequently the relative density is 1, that is, $\gamma(G^{(l)}_{SV}) = \gamma(G^{(l)})$ which is not interesting. In the following subsection, we discuss a heuristic for improving subgraph detection performance when all nodes are support vectors.

5.2 Thresholding Heuristic

In this section, we discuss the impact of choosing support vectors with *high* support by choosing some $c > 0$ and selecting the set $S_c = \{i : \alpha^*_i > c\}$. Figure 2 shows the densities of the induced subgraph $\gamma(G^{(l)}_{S_c})$ relative to original graph density $\gamma(G^{(l)})$ for all graphs $G^{(l)} \in \mathbb{G}$, that is,

$$a^{(l)} = \gamma(G^{(l)}_{S_c})/\gamma(G^{(l)}) \forall\, G^{(l)} \in \mathbb{G}$$

at different c thresholds varying between $c = 0$ ($|S_c| = |SV|$) and $c = 1$ ($|S_c| = 0$) (and correspondingly different subgraph sizes $|S_c|$ which is shown on x-axis).

Figure 2 shows the variation in density of the induced subgraph $\gamma(G^{(l)}_{S_c})$ relative to original graph density $\gamma(G^{(l)})$ for all graphs $G^{(l)} \in \mathbb{G}$ at increasing subgraph sizes for the largest graph (resp. *c-fat500, p_hat1500, brock800, san400* and *sanr400*) in each graph family (resp. *c-fat, p_hat, brock, san* and *sanr*). Figure 5 in Appendix A presents the variation for other DIMACS graphs.

Notice that in case of *c-fat* and *p_hat* graph families (Figures 2(a) and 2(c)), one can further improve graph densities across all graphs $G^{(l)} \in \mathbb{G}$ by choosing a higher value of c (and correspondingly, a smaller induced subgraph $|S_c|$). In the remaining examples, choosing a higher value of c (and correspondingly lower $|S_c|$) improves the density in at least one $G^{(l)} \in \mathbb{G}$.

Based on performance on the DIMACS data set, we suggest a simple rule for using the heuristic whenever $n_{SV} \geq 0.8n$, then we choose c such that $|S_c| = \lceil n_{SV}/2 \rceil$. Table 2 shows the improvement by using the heuristic rule whenever $n_{SV} \geq 0.8n$. We note that minimum and maximum induced subgraph density improves by choosing S_c instead of T, that is, $\bar{a}(S_c) \geq \bar{a}(T)$ and $\underline{a}(S_c) \geq \underline{a}(T)$ for all graph families.

It can be seen from Figure 2 that the induced subgraph density is not strictly monotonic with induced subgraph size $|S_c|$. However, for medium to large $|S_c|$, the subgraph density shows a general

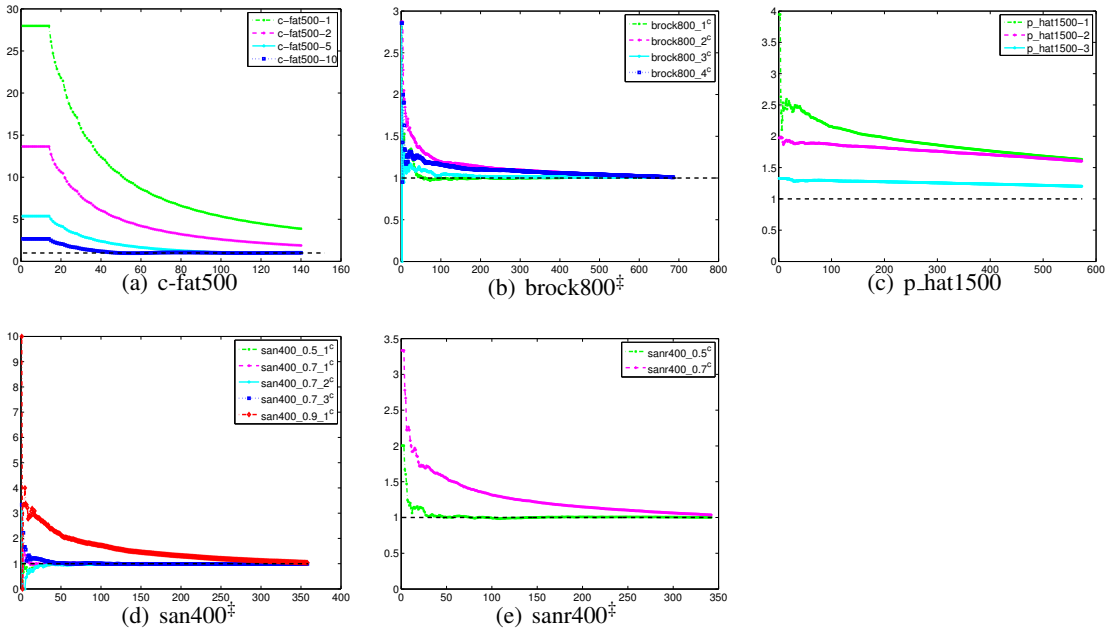


Figure 2: Common dense subgraph recovery. Figures $(a)-(e)$ show the densities of the induced subgraph $\gamma(G^{(l)}_{S_c})$ relative to original graph density $\gamma(G^{(l)})$ for all graphs $G^{(l)} \in \mathbb{G}$ at different values of $c \in [0,1]$ (i.e., different subgraph sizes $|S_c|$) for different DIMACS graph families.

decreasing trend, that is, if $|S_c| > |S_{c'}|$, then the induced subgraph density $\gamma(G^{(l)}_{S_c}) < \gamma(G^{(l)}_{S_{c'}})$ for some regime $c, c' \in [c_l, c_h]$. Thus, one can choose a smaller induced subgraph S_c having higher induced subgraph density by selecting higher value of threshold c.

It is instructive to note that in all graph families, the graph with maximum relative density, for example, *c-fat500* in Figure 2(a) is the graph with minimum average density among all graph $\mathbb{G}$. In other words, *MKL-based approach tries to find a dense region in the sparsest subgraph* $G^{(l)} \in \mathbb{G}$ *while making sure it is compatible with remaining graphs in* $\mathbb{G}$.

5.3 Planted Clique Recovery on Random Graphs

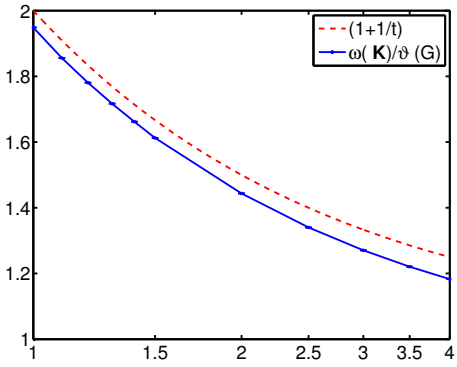
We consider the case for Erdos-Renyi graph with general $p \neq 1/2$ and planted clique of size k, that is, $G(n,p,k)$.

5.3.1 DATA SET

We generate $n_s = 100$ random graphs based on $\bar{G}(n, 1-p, k)$ with planted independent set of size $k = 2t\sqrt{n(1-p)/p}$ and $p = \frac{1}{2}n^{-\alpha}$ where $n = 20000$ and $\alpha \geq 0$. We choose $\alpha = c/30$ where c lies in the set $\{0, 1, 2, \ldots, 10\}$; and perform n_s experiments for each c. Note that the case of $\alpha = 0$ yields the familiar planted clique model $G(n, 1/2, k)$ with $k = 2t\sqrt{n}$.

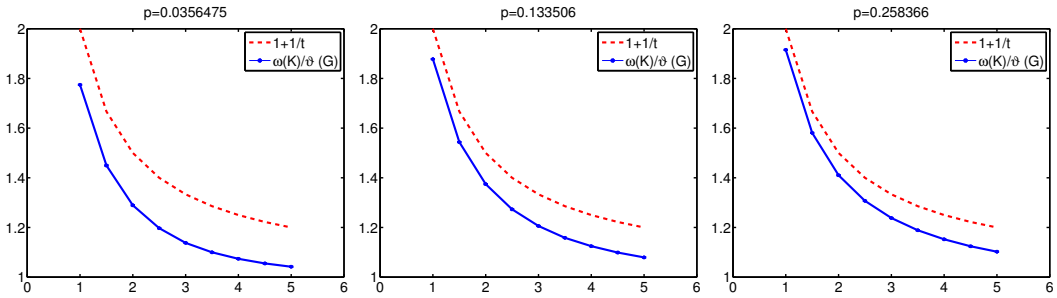
t	f_r	F_1
1	0	0.007
2	1	1
3	1	1
4	1	1

(a)



(b)

Figure 3: (a) shows f_r the fraction of graphs for which the hidden clique is recovered exactly; and, the average F_1-score measuring quality of recovered subset over n_s trials at each t ($k = 2t\sqrt{n}$). (b) shows $\omega(\mathbf{K})/k$ (blue) is bounded by $(1+1/t)$ (red) (Theorem 23). This allows approximation of Lovász ϑ function for large $\mathbf{SVM}-\vartheta$ graphs without SDP.



(a) (b) (c)

Figure 4: Approximation of Lovász ϑ function for general p. This figure shows $\omega(\mathbf{K})/k$ (blue) is bounded by $(1+1/t)$ (red) (Theorem 23). This allows approximation of Lovász ϑ function for large $\mathbf{SVM}-\vartheta$ graphs without SDP for general p.

5.3.2 EVALUATION

We consider the **LS** labelling $\mathbf{K} = \frac{A}{\rho} + I$ of $G(n, 1-p, k)$ for $\rho = 2\sqrt{np(1-p)} + kp$ as discussed in Section 4.2. We solve $\omega(\mathbf{K})$ using libsvm solver,[4] and return the top-k support vectors $\tilde{S}_k$ as independent set as discussed in Algorithm 2.

In order to evaluate algorithm accuracy, we compute the fraction of graphs for which the independent set is recovered exactly using Algorithm 2, that is, $f_r = n_r/n_s$ where n_r is the number of graphs for which independent set is recovered exactly for each $\bar{G}(n, 1-p, k)$. We also compute the average (over n_s trials) *F1* score which measures how different is our solution $\tilde{S}_k$ to maximum independent set S as: $F_1 = \frac{2pr}{p+r}$, $p = \frac{|\tilde{S}_k \cap S|}{|\tilde{S}_k|}$, $r = \frac{|\tilde{S}_k \cap S|}{|S|}$.

4. LibSVM solver is available online at `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`.

5.3.3 DISCUSSION

As predicted by Theorem 23, there exists some $t_0 > 0$ for which Lovász ϑ function is bounded by $\omega(\mathbf{K})$; and the planted clique can be recovered perfectly by selecting the top k support vectors in sorted descending order of α_i^*. We find experimentally that this approach recovers the planted clique *exactly* for $t \geq 2$ for all $c \in \{0,\ldots,10\}$, that is, random graph $\bar{G}(n,1-p,k)$ with $p = \frac{1}{2}n^{-\alpha}$ and planted independent set of size $k = 2t\sqrt{n(1-p)/p}$.

In particular we discuss the case $c = 0$ which yields the Erdös-Rényi graph with $p = 1/2$ and planted clique of size $k = 2t\sqrt{n}$. Figure 3(a) shows the fraction of graphs for which the hidden clique is recovered exactly using above procedure. Figure 3(b) shows $\omega(\mathbf{K})/k$ (shown in blue), and consequently $\omega(\mathbf{K})/\vartheta(G)$ is bounded by $(1+1/t)$ (shown in red) for $G(n,1/2,k)$ case. Figure 4 shows that $\omega(\mathbf{K})/\vartheta(G)$ is bounded by $(1+1/t)$ for case when $p \neq 1/2$. Thus, *one can use* $\omega(\mathbf{K})$ *for approximating Lovász* ϑ *function for large* $\mathbf{SVM}-\vartheta$ *graphs without solving a SDP problem*.

6. Conclusion

The results in this paper establish that there is a close connection between the Lovász ϑ function, well studied in graph theory and SVM formulation. This link allows us to design scalable algorithms for computationally difficult problems. In particular we show that on random graphs, the Lovász ϑ function can be well approximated by solving an SVM. Furthermore this property is not destroyed even when one plants a large clique in such graphs, allowing extremely scalable algorithms for identifying it. Using tools from MKL we further describe a algorithm for finding a common large dense region in large graphs. This algorithm achieves an order of magnitude scalability compared to state-of-the-art method based on exhaustive search of frequent quasi-cliques.

Acknowledgments

We would like to thank the anonymous reviewers for their useful comments. The authors also thank Fredrik Johansson and Christos Dimitrikakis for critically reading the manuscript.

Appendix A. Extended Results on DIMACS Graph Families

Figure 5 shows the densities of the induced subgraph $\gamma(G_{S_c}^{(l)})$ relative to original graph density $\gamma(G^{(l)})$ for all graphs $G^{(l)} \in \mathbb{G}$ at different values of $c \in [0,1]$ (i.e., different subgraph sizes $|S_c|$) for remaining graphs (other than those presented in Figure 2) in different DIMACS graph families.

Appendix B. Some Results Related to $\bar{G}(n,1-p,k)$

In this section we derive two results related to the planted clique problem. In particular we derive conditions when $\bar{G}(n,1-p,k)$ is a $\mathcal{Q}$ graph and we compute the ϑ function for the same graph.

B.1 When Is $\bar{G}(n,1-p,k)$ a $\mathcal{Q}$ Graph?

To prove Theorem 21 we will need a lower bound on the minimum eigenvalue of $\bar{G}$. One can prove the following

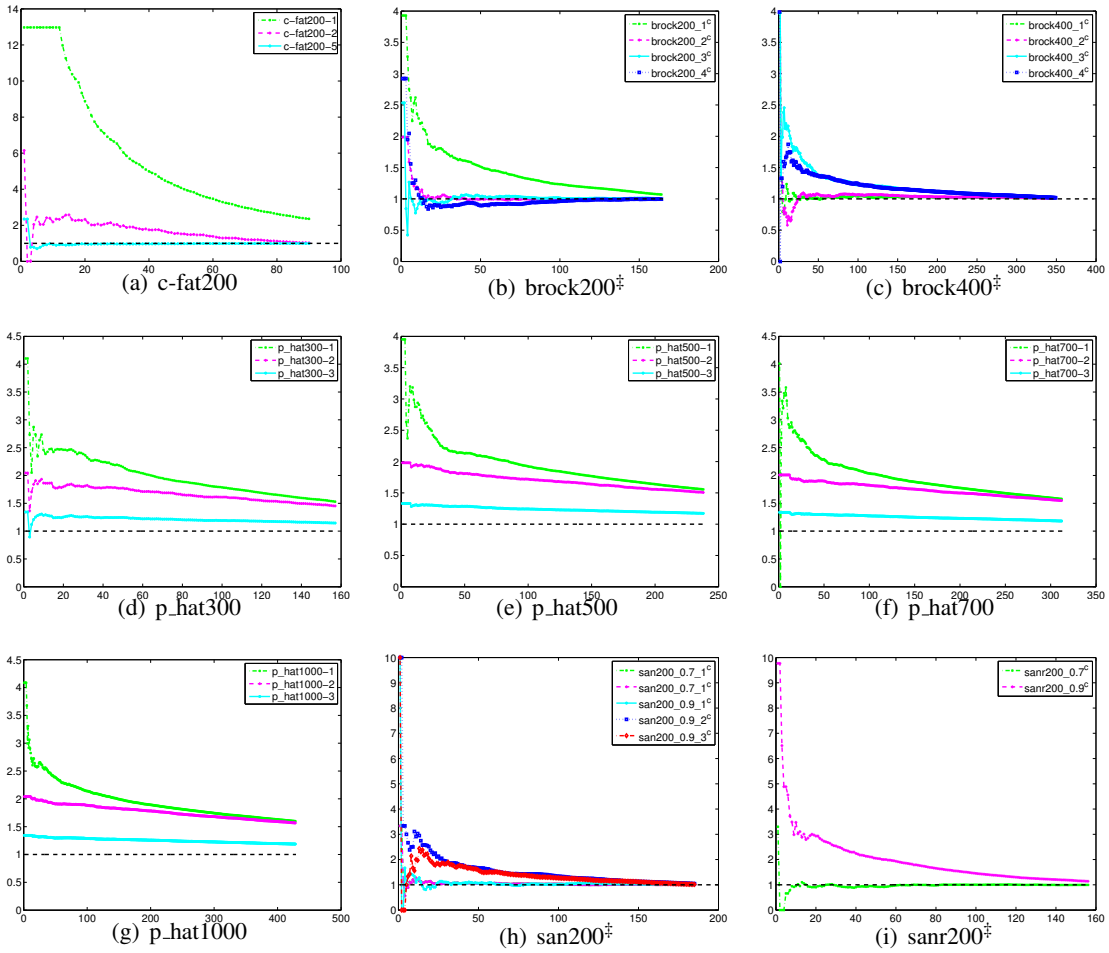


Figure 5: Common dense subgraph recovery. Figures $(a)-(i)$ show the densities of the induced subgraph $\gamma(G^{(l)}_{S_c})$ relative to original graph density $\gamma(G^{(l)})$ for all graphs $G^{(l)} \in \mathbb{G}$ at different values of $c \in [0,1]$ (i.e., different subgraph sizes $|S_c|$) for remaining graphs in different DIMACS graph families.

Lemma 31 *Let* $\mathbf{A}$ *be the adjacency matrix of* $\bar{G}(n, 1-p, k)$. *Then with probability at least* $1-\frac{1}{n}$, *the following holds:*

$$-\lambda_{\min}(\mathbf{A}) \leq kp - \frac{9k^2p}{16n} + \frac{\|\mathbf{A}-\mathbb{E}\mathbf{A}\|^2}{kp} + \sqrt{\ln n} + p.$$

The proof is crucially dependent on the following result.

Lemma 32 *Let* $\mathbf{A}$ *be the adjacency matrix of* $\bar{G}(n, 1-p, k)$. *For any fix* $\mathbf{x}$ *and any* $\delta > 0$ *we have*

$$Pr\left[\mathbf{x}^\top(\mathbf{A}-\mathbb{E}\mathbf{A})\mathbf{x} \leq -\delta\|\mathbf{x}\|^2\right] \leq e^{-\delta^2}.$$

Proof Let $\hat{\mathbf{A}} = \mathbf{A} - \mathbb{E}\mathbf{A}$ and consider the random variable

$$X = -\mathbf{x}^\top \hat{\mathbf{A}}\mathbf{x} = \sum_{i>j} -2x_i x_j \hat{A}_{ij},$$

where we note that the sum goes over independent random variables. Since $\hat{A}_{ij}$ varies by at most by 1, the term $2x_i x_j \hat{A}_{ij}$ varies at most $|2x_i x_j|$ where

$$\sum_{i>j} (2x_i x_j)^2 = 2\sum_{i\neq j} x_i^2 x_j^2 \leq 2\|\mathbf{x}\|^4.$$

Thus Hoeffding's inequality (McDiarmid, 1998) applied on X states that for any $t > 0$

$$Pr[X \geq t] \leq \exp\left(-\frac{t^2}{\|\mathbf{x}\|^4}\right).$$

The statement is obtained by letting $t = \delta\|\mathbf{x}\|^2$. ■

Proof [Proof of Lemma 31] Consider the matrix $\mathbf{A}' = \mathbf{A} + \mathbf{D}$ where $D_{ij} = p$ if $i = j \notin S$ and 0 otherwise, that is, adding p to all diagonal elements not in S. We note that this perturbation gives $\mathbb{E}\mathbf{A}'$ a 2-dimensional column space as all vectors in the column space are constant on S and $\bar{S}$ respectively. Let $\mathbf{x}$ be any unit vector in the column space where we write $\frac{\sin t}{\sqrt{k}}$ on S and $\frac{\cos t}{\sqrt{n-k}}$ on $\bar{S}$. Since all eigenvectors corresponding to non-zero eigenvalues must be in this vector space, the minimum eigenvalue of $\mathbb{E}\mathbf{A}'$ is either 0 or the minimum of

$$\begin{aligned}
\mathbf{x}^\top \mathbb{E}[\mathbf{A}']\mathbf{x} &= 2\sqrt{k(n-k)}p\sin t\cos t + (n-k)p\cos^2 t \\
&= \frac{(n-k)p}{2} + \sqrt{k(n-k)}p\sin 2t - \frac{(n-k)p}{2}\cos 2t \\
&\geq \frac{(n-k)p}{2} - \frac{p}{2}\sqrt{4k(n-k)+(n-k)^2} \\
&\geq \frac{(n-k)p}{2} - \frac{p}{2}\left(n + \frac{2nk-3k^2}{2n}\right) = -kp + \frac{3k^2 p}{4n}
\end{aligned}$$

where the last inequality follows by Taylor expansion of the square root at n^2. Since the last expression is non-positive we conclude that $\bar{\lambda} := -kp + \frac{3k^2 p}{4n}$ is a lower bound on the eigenvalues of $\mathbb{E}\mathbf{A}'$.

Now, let $\mathbf{v}$ be a normalized eigenvector corresponding to the minimum eigenvalue of $\mathbb{E}\mathbf{A}'$ and let $\mathbf{u} = \mathbf{x} + \mathbf{y}$ be any unit vector, where $\mathbf{x} \parallel \mathbf{v}$ and $\mathbf{y} \perp \mathbf{v}$. Since $\mathbb{E}\mathbf{A}'$ is a rank 2 matrix and clearly $\mathrm{Tr}(\mathbb{E}\mathbf{A}') \geq 0$, $\mathbb{E}\mathbf{A}'$ can at most have one negative eigenvalue and thus $\mathbf{y}^\top \mathbb{E}[\mathbf{A}']\mathbf{y} \geq 0$. Furthermore, Lemma 32 states that $\mathbf{v}^\top \hat{\mathbf{A}}\mathbf{v} \geq -\sqrt{\ln n}$ with probability at least $\left(1 - \frac{1}{n}\right)$. Thus,

$$\begin{aligned}
\lambda_{\min}(\mathbf{A}) &= \min_{\|\mathbf{u}\|=1} \mathbf{u}^\top \left(\mathbb{E}\mathbf{A}' + \hat{\mathbf{A}} - \mathbf{D}\right)\mathbf{u} \\
&\geq \min_{\|\mathbf{u}\|=1} \bar{\lambda}\|\mathbf{x}\|^2 + \mathbf{x}^\top \hat{\mathbf{A}}\mathbf{x} + (2\mathbf{x}+\mathbf{y})^\top \hat{\mathbf{A}}\mathbf{y} - \mathbf{u}^\top \mathbf{D}\mathbf{u} \\
&\geq -\sqrt{\ln n} - p + \min_{x_1^2 + x_2^2 = 1} \left(\bar{\lambda}x_1^2 - \|\hat{\mathbf{A}}\|\sqrt{4x_1^2 + x_2^2}\, x_2\right),
\end{aligned}$$

where we again use $\hat{\mathbf{A}} = \mathbf{A} - \mathbb{E}\mathbf{A}$. By substituting $x_2 = \frac{2}{\sqrt{3}}\cos t$ in the last term we get

$$\begin{aligned}\bar{\lambda}(1-x_2^2) - \|\hat{\mathbf{A}}\|\sqrt{4-3x_2^2}\,x_2 &= \frac{\bar{\lambda}}{3} - \frac{2}{3}\left(\bar{\lambda}\cos 2t + \sqrt{3}\|\hat{\mathbf{A}}\|\sin 2t\right) \\ &\geq \frac{\bar{\lambda}}{3} - \frac{2}{3}\sqrt{\bar{\lambda}^2 + 3\|\hat{\mathbf{A}}\|^2}\end{aligned}$$

so by Taylor expanding the square root at k^2p^2 we conclude that

$$\begin{aligned}\min_{x_1^2+x_2^2=1}\left(\bar{\lambda}x_1^2 - \|\hat{\mathbf{A}}\|\sqrt{4x_1^2+x_2^2}\,x_2\right) &\geq \frac{\bar{\lambda}}{3} - \frac{2}{3}\left(kp + \frac{\bar{\lambda}^2 + 3\|\hat{\mathbf{A}}\|^2 - k^2p^2}{2kp}\right) \\ &= -kp + \frac{k^2p}{n}\left(\frac{3}{4} - \frac{3k}{16n}\right) - \frac{\|\hat{\mathbf{A}}\|}{kp}.\end{aligned}$$

Thus with probability at least $\left(1-\frac{1}{n}\right)$, we get

$$\lambda_{\min}(\mathbf{A}) \geq -\sqrt{\ln n} - p - kp + \frac{k^2p}{n}\left(\frac{3}{4} - \frac{3k}{16n}\right) - \frac{\|\hat{\mathbf{A}}\|}{kp}.$$

■

Before we prove the main theorem we will need one more ingredient, a bound on the norm of the difference between the adjacency matrix and its expectation.

Lemma 33 *If* $\mathbf{A}$ *is the adjacency matrix of* $\bar{G}(n, 1-p, k)$ *and* $p(1-p) = \Omega(n^{-1}\log^4 n)$*, then almost surely* $\|\mathbf{A} - \mathbb{E}\mathbf{A}\| = O(\sqrt{np(1-p)})$.

Proof The Lemma is easily derived from Theorem 34, due to Vu (2007). In particular in the given regime of p, one can choose $\mathbf{R} = \mathbf{A} - \mathbb{E}\mathbf{A}$, $\sigma^2 = p(1-p)$ and $K = 1$. For such a choice, one can show $C(K\sigma)^{1/2}n^{1/4}\ln n = O(\sigma\sqrt{n})$ and the lemma follows by direct application of following result in Theorem 34. ■

We next state a Theorem, without proof, due to Vu (2007).

Theorem 34 *(Vu, 2007) There are constants C and C′ such that the following holds. Let* R_{ij}, $1 \leq i \leq j \leq n$ *be independent random variables, with* $\mathbb{E}(R_{ij}) = 0, \mathbb{E}(R_{ij}^2) \leq \sigma^2, |R_{ij}| \leq K$ *and* $\sigma \geq C'n^{-1/2}K\ln^2 n$*. Then with probability tending to 1 as* $n \to \infty$

$$\|\mathbf{R}\| \leq 2\sigma\sqrt{n} + C(K\sigma)^{1/2}n^{1/4}\ln n.$$

We are now ready to begin the proof of the main result.

Proof [Proof of Theorem 21] According to Theorem 5, G is a Q graph if

$$-\lambda_{\min}(\mathbf{A}) \leq |N(i) \cap S| \text{ for all } i \notin S, \tag{31}$$

where S denotes a maximum independent set. We will prove that this equation is satisfied almost surely for S denoting the planted independent set. In the specified parameter regime, one can argue that the planted set is the maximum independent set almost surely. It should however be noted

that this is strictly speaking not needed. As (31) is essentially reformulated optimality criteria, any independent set satisfying the equation must be a maximum independent set.

For each $i \notin S$, $|N(i) \cap S|$ has $Bin(k, p)$ distribution, so using Chernoff bound (Frieze and Reed, 1998) together with a union bound one obtains $|N(i) \cap S| \geq kp - \sqrt{6kp \ln n}$ for all $i \notin S$ with probability at least $1 - \frac{1}{n}$. Using this together with Lemma 31, we get that G is a Q graph with probability $1 - O(\frac{1}{n})$ if

$$kp - \frac{9k^2}{16n} + \frac{\|\mathbf{A} - \mathbb{E}\mathbf{A}\|^2}{kp} + \sqrt{\ln n} + p \leq kp - \sqrt{6kp \ln n}.$$

Application of Lemma 33 yields $\|\mathbf{A} - \mathbb{E}\mathbf{A}\|^2 = O(np(1-p))$. Clearly then almost surely G is a Q graph if the following holds,

$$\frac{9k^2 p}{16n} \geq \sqrt{6kp \ln n} + \sqrt{\ln n} + O\left(\frac{n}{k}\right). \tag{32}$$

Next we note that for $k \geq \frac{8}{3} n^{2/3} p^{-1/3} (\ln n)^{1/3}$ the following holds

$$\begin{aligned}
\sqrt{6kp \ln n} + \sqrt{\ln n} + O\left(\frac{n}{k}\right) &= \sqrt{6kp \ln n}\left(1 + O\left(\frac{1}{k^{1/2} p^{1/2}} + \frac{n}{k^{3/2} p^{1/2}}\right)\right) \\
&\leq \sqrt{6kp \ln n}\left(1 + O(\frac{1}{n^{1/3} p^{1/3} (\ln n)^{1/6}} + \frac{1}{(\ln n)^{1/2}})\right) \\
&= \sqrt{6kp \ln n}\,(1 + o(1))
\end{aligned}$$

where $np = \Omega(\log^4 n)$ by assumption in the theorem. This means that equation (32) is satisfied for

$$k \geq (1 + o(1)) \frac{8}{3} n^{2/3} p^{-1/3} (\ln n)^{1/3}$$

and the theorem follows. ■

B.2 Computation of ϑ Function for $\bar{G}(n, 1-p, k)$

Proof [of Theorem 22] We will base the proof on the following definition of ϑ function (Knuth, 1994),

$$\vartheta(G) = \min_{\mathbf{M}} \lambda_1(\mathbf{M})$$

where $\mathbf{M}$ goes over all symmetric $n \times n$ matrices such that $M_{ij} = 1$ whenever $A_{ij} = 0$, where $\mathbf{A}$ denotes the adjacency matrix of G. For the given graph $\bar{G}$ we will construct a qualified guess of $\mathbf{M}$ at optimality and use this to show a tight bound on $\vartheta(\bar{G})$. Specifically, let $\mathbf{A}$ be the adjacency matrix of $\bar{G}$ and consider the matrix $\mathbf{M}$ defined by

$$M_{ij} = \begin{cases} 1 & \text{if } 1 \leq i, j \leq k \\ 1 - r_j A_{ij} & \text{if } 1 \leq i \leq k < j \leq n \\ 1 - r_i A_{ij} & \text{if } 1 \leq j \leq k < i \leq n \\ \frac{1}{p}(p - A_{ij}) & \text{if } k < i, j \leq n \end{cases} \tag{33}$$

where r_i is chosen such that $\sum_{j=1}^{k} M_{ij} = 0$ for all $i > k$. Equivalently, r_i satisfies

$$k - S_i r_i = 0 \tag{34}$$

where S_i is the number of neighbours of i which shares an edge with the planted independent set. The case $S_i = 0$ for some $i > k$ may be resolved arbitrarily. We note that $S_i \sim Bin(k, p)$.

Let $\mathbf{e}_k$ denote the n-dimensional vector where the first k elements are 1 and the rest 0. Note that we constructed $\mathbf{M}$ such that it has $\mathbf{e}_k$ as an eigenvector with corresponding eigenvalue k. Using Lemma 35 we see that this is the maximum eigenvalue almost surely. To conclude we note that since there is an independent set of size k in $\bar{G}$, $k \leq \alpha(\bar{G}) \leq \vartheta(\bar{G})$. As noted above, $\vartheta(\bar{G}) \leq \lambda_1(\mathbf{M}) = k$ almost surely, so $k \leq \vartheta(\bar{G}) \leq k$, and thus $\vartheta(\bar{G}) = k$ almost surely. ■

Lemma 35 *Let* $\mathbf{M}$ *be defined as in above. If* $k > 2\sqrt{\frac{n(1-p)}{p}}\,(1+o(1))$ *and* $p(1-p) = \Omega(n^{-1}\log^4 n)$, *then* $\lambda_2(\mathbf{M}) < k$ *almost surely.*

Let $\mathbf{A}$, as before, denote the adjacency matrix of $\bar{G}$. Consider the two matrices,

$$U_{ij} = \begin{cases} 0 & \text{if } 1 \leq i, j \leq k \\ \frac{1}{p}(p - A_{ij}) & \text{if } 1 \leq i \leq k < j \leq n \\ \frac{1}{p}(p - A_{ij}) & \text{if } 1 \leq j \leq k < i \leq n \\ \frac{1}{p}(p - A_{ij}) & \text{if } k < i, j \leq n, \end{cases} \tag{35}$$

$$V_{ij} = \begin{cases} 0 & \text{if } 1 \leq i, j \leq k \\ \left(r_j - \frac{1}{p}\right)(p - A_{ij}) & \text{if } 1 \leq i \leq k < j \leq n \\ \left(r_i - \frac{1}{p}\right)(p - A_{ij}) & \text{if } 1 \leq j \leq k < i \leq n \\ 0 & \text{if } k < i, j \leq n \end{cases} \tag{36}$$

where r_i are the same as in (33).

The following corollary is a direct consequence of Theorem 34

Corollary 36 *For any symmetric* $n \times n$ *matrix* $\mathbf{R}$ *with entries* R_{ij}, *independent with mean* 0 *and variance bounded above by* $\frac{1-p}{p}$ *and* $|R_{ij}| \leq \frac{1}{p}$ *the following holds*

$$\|\mathbf{R}\| \leq 2\sqrt{\frac{n(1-p)}{p}}\,(1+o(1))$$

whenever $p(1-p) = \Omega(n^{-1}\log^4 n)$.

Proof The corollary follows by noticing that the matrix $\mathbf{R}$ satisfies the conditions of Theorem 34 with parameters $\sigma^2 = \frac{1-p}{p}$ and $K = \frac{1}{p}$. ■

We will use the corollary to obtain bounds on the largest eigenvalue of $\mathbf{U}$ and $\mathbf{V}$.

Lemma 37 *Let* $\mathbf{U}$ *be defined in* (35), *then if* $p(1-p) = \Omega(n^{-1}\log^4 n)$

$$\lambda_1(\mathbf{U}) \leq 2\sqrt{\frac{n(1-p)}{p}}\,(1+o(1))\,.$$

Proof Let $\mathbf{D}$ be diagonal matrix with $D_{ii} = 0$ whenever $1 \leq i \leq k$ and $D_{ii} = 1$ whenever $i = k+1, \ldots, n$. The matrix $\mathbf{U} - \mathbf{D}$ satisfies the condition of Corollary 36 and hence

$$\lambda_1(\mathbf{U}) \leq \|\mathbf{U} - \mathbf{D}\| + \|\mathbf{D}\| \leq 2\sqrt{\frac{n(1-p)}{p}}(1 + o(1)) + 1$$

where we have used the fact that $\|\mathbf{D}\| = 1$. ■

Lemma 38 *Let* $\mathbf{V}$ *be defined in* (36)*, then if* $p(1-p) = \Omega(n^{-1}\log^4 n)$ *and* $k > 2\sqrt{\frac{n(1-p)}{p}}$

$$\lambda_1(\mathbf{V}) \leq o\left(\sqrt{\frac{n(1-p)}{p}}\right).$$

Proof Consider the $n \times n$ matrix $\mathbf{V}'$ defined by

$$V'_{ij} = \begin{cases} 0 & \text{if } 1 \leq i, j \leq k \\ \frac{1}{p}(p - A_{ij}) & \text{if } 1 \leq i \leq k < j \leq n \\ \frac{1}{p}(p - A_{ij}) & \text{if } 1 \leq j \leq k < i \leq n \\ 0 & \text{if } k < i, j \leq n. \end{cases}$$

We can apply Corollary 36 to obtain $\|\mathbf{V}'\| = O\sqrt{\frac{n(1-p)}{p}}$.

Let $\mathbf{x}$ be a unit eigenvector corresponding to the maximum eigenvalue of $\mathbf{V}$. This means that

$$\begin{aligned} \lambda_1(\mathbf{V}) &= \mathbf{x}^\top \mathbf{V}\mathbf{x} \\ &= 2\sum_{i=k+1}^{n} x_i \left(r_i - \frac{1}{p}\right) \sum_{j=1}^{k} (p - A_{ij}) x_j \end{aligned}$$

and by Cauchy-Schwartz inequality

$$\begin{aligned} &\leq 2\left(\sum_{i=k+1}^{n} x_i^2 \cdot \sum_{i=k+1}^{n} \left(r_i - \frac{1}{p}\right)^2 \left(\sum_{j=1}^{k} (p - A_{ij}) x_j\right)^2\right)^{1/2} \\ &\leq 2\left(\sum_{i=k+1}^{n} (pr_i - 1)^2 \left(\sum_{j=1}^{k} \frac{1}{p}(p - A_{ij}) x_j\right)^2\right)^{1/2} \\ &= 2\left(\sum_{i=k+1}^{n} (pr_i - 1)^2 \left(\sum_{j=1}^{k} V'_{ij} x_j\right)^2\right)^{1/2} \leq 2\max_i |pr_i - 1| \cdot \left(\sum_{i=k+1}^{n} \left(\sum_{j=1}^{k} V'_{ij} x_j\right)^2\right)^{1/2} \end{aligned}$$

since $V'_{ij} = 0$ for $i, j > k$

$$
\begin{aligned}
&= 2\max_i |pr_i - 1| \cdot \left(\sum_{i=k+1}^{n} \left(\sum_{j=1}^{n} V'_{ij} x_j \right)^2 \right)^{1/2} \\
&\leq 2\max_i |pr_i - 1| \cdot \left(\sum_{i=1}^{n} \left(\sum_{j=1}^{n} V'_{ij} x_j \right)^2 \right)^{1/2} \\
&= 2\max_i |pr_i - 1| \cdot \|\mathbf{V}'\mathbf{x}\| = \max_i |pr_i - 1| \cdot O\left(\sqrt{\frac{n(1-p)}{p}} \right),
\end{aligned}
$$

where the last step follows from that $\|\mathbf{V}'\mathbf{x}\| \leq \|\mathbf{V}'\| \cdot \|\mathbf{x}\| = \|\mathbf{V}'\|$.

By the definition of $\mathbf{M}$, see (34), $r_i = \frac{k}{S_i}$, where $S_i \sim Bin(k, p)$. For such random variables, we have the following Chernoff bound

Lemma 39 (McDiarmid, 1998) *For every $0 < a \leq kp$ we have*

$$Pr\left[|S_i - kp| \geq a\right] \leq 2e^{-a^2/3kp}.$$

Let $a = \sqrt{6kp \ln n}$. Using the assumptions in the lemma it is easy to verify that $kp > 2\sqrt{np(1-p)} \gg \ln^2 n \gg \ln n$. Thus for large enough n we have $a \leq kp$, so in that case we know that $|S_i - kp| \geq a$ for some $i > k$ happens with at most probability $n \cdot e^{-2\ln n} = o(1)$. This means that, almost surely

$$\max_i |pr_i - 1| = |\frac{kp}{kp + O\sqrt{kp \ln n}} - 1| = \frac{O\sqrt{kp \ln n}}{kp + O\sqrt{kp \ln n}} = O\sqrt{\frac{\ln n}{kp}} = o(1),$$

where we use that $\ln n \ll kp$ as noted above. Thus $\lambda_1(V) = o\sqrt{\frac{n(1-p)}{p}}$ almost surely. ■

We are now in a position to begin the proof of Lemma 35.

Proof [Proof of Lemma 35] Let $\mathbf{e}_k$ denote the n-dimensional vector where the first k elements are 1 and the rest 0. By the variational inequality

$$
\begin{aligned}
\lambda_2(\mathbf{M}) &= \min_{\mathbf{v}} \max_{\|\mathbf{x}\|=1 \, \mathbf{x} \perp \mathbf{v}} \mathbf{x}^\top \mathbf{M} \mathbf{x} \\
&\leq \max_{\|\mathbf{x}\|=1 \, \mathbf{x} \perp \mathbf{e}_k} \mathbf{x}^\top \mathbf{M} \mathbf{x} \\
&\leq \lambda_1(\mathbf{U}) + \lambda_1(\mathbf{V}) + \max_{|\mathbf{x}|=1 \, \mathbf{x} \perp \mathbf{e}_k} \mathbf{x}^\top (\mathbf{M} - \mathbf{U} - \mathbf{V}) \mathbf{x}.
\end{aligned}
$$

Recalling the definitions of $\mathbf{M}$, $\mathbf{U}$ and $\mathbf{V}$, see (33), (35), (36), we note that for all $\mathbf{x} \perp \mathbf{e}_k$

$$\begin{aligned}
&\mathbf{x}^\top(\mathbf{M}-\mathbf{U}-\mathbf{V})\mathbf{x} \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k} x_i(M_{ij}-U_{ij}-V_{ij})x_j + 2\sum_{i=k+1}^{n}\sum_{j=1}^{k} x_i(M_{ij}-U_{ij}-V_{ij})x_j \\
&\quad + \sum_{i=k+1}^{n}\sum_{j=k+1}^{n} x_i(M_{ij}-U_{ij}-V_{ij})x_j \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k} x_i x_j + 2\sum_{i=k+1}^{n}\sum_{j=1}^{k} x_i\left(1 - r_i A_{ij} - \frac{1}{p}(p-A_{ij}) - \left(r_i - \frac{1}{p}\right)(p-A_{ij})\right)x_j \\
&= \sum_{i=1}^{k}\sum_{j=1}^{k} x_i x_j + 2\sum_{i=k+1}^{n}\sum_{j=1}^{k} x_i(1-pr_i)x_j \\
&= \sum_{i=1}^{k} x_i \underbrace{\left(\sum_{j=1}^{k} x_j\right)}_{=0} + 2\sum_{i=k+1}^{n} x_i(1-pr_i)\underbrace{\left(\sum_{j=1}^{k} x_j\right)}_{=0} = 0
\end{aligned}$$

so $\max_{\|\mathbf{x}\|=1,\,\mathbf{x}\perp\mathbf{e}_k} x^\top(\mathbf{M}-\mathbf{U}-\mathbf{V})\mathbf{x} = 0$. Thus if we assume that $k > 2\sqrt{\frac{n(1-p)}{p}}$, Lemmas 37 and 38 imply that

$$\lambda_2(\mathbf{M}) \le \lambda_1(\mathbf{U}) + \lambda_1(\mathbf{V}) \le 2\sqrt{\frac{n(1-p)}{p}}\,(1+o(1))\,.$$

The Lemma follows by letting k be strictly greater than the maximum of this value and $2\sqrt{\frac{n(1-p)}{p}}$. ■

References

J. Aflalo, A. Ben-Tal, C. Bhattacharyya, J.S. Nath, and S. Raman. Variable sparsity kernel learning. *Journal of Machine Learning Research*, 12:565–592, 2011.

N. Alon, M. Krivelevich, and B. Sudakov. Finding a large hidden clique in a random graph. *Random Structures and Algorithms*, pages 457–466, 1998.

B. Applebaum, B. Barak, and A. Wigderson. Public-key cryptography from different assumptions. In *Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC '10)*, pages 171–180, 2010.

S. Arora, B. Barak, M. Brunnermeier, and R. Ge. Computational complexity and information asymmetry in financial products (extended abstract). In *First Symposium on Innovations in Computer Science*, ICS, pages 49–65, 2010.

C. Bachoc, D. Gijswijt, A. Schrijver, and F. Vallentin. Invariant semidefinite programs. In M. F. Anjos and J. B. Lasserre, editors, *Handbook on Semidefinite, Conic and Polynomial Optimization*, volume 166 of *International Series in Operations Research & Management Science*, pages 219–269. Springer US, 2012.

B. Bollobás. *Modern Graph Theory*. Springer Verlag, 1998.

B. Bollobás. *Random Graphs*. Cambridge University Press, 2001.

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

A. E. Brouwer and W. H. Haemers. *Spectra of Graphs*. Springer Verlag, 2012.

T.-H. H. Chan, K. L. Chang, and R. Raman. An sdp primal-dual algorithm for approximating the lovász-theta function. In *IEEE International Symposium on Information Theory*, ISIT '09, pages 2808–2812, 2009.

A. Coja-Oghlan. The lovász number of random graphs. *Combinatorics, Probability & Computing*, 14(4):439–465, 2005.

A. Coja-Oghlan and A. Taraz. Exact and approximative algorithms for coloring g(n, p). *Random Structures and Algorithms*, 24(3):259–278, 2004.

T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. MIT Press, 3rd edition, 2009.

L. Devroye, A. György, G. Lugosi, and F. Udina. High-dimensional random geometric graphs and their clique number. *Electronic Journal of Probability*, 16, 2011.

D. P. Dubhashi and A. Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, 2009.

U. Feige and R. Krauthgamer. Finding and certifying a large hidden clique in a semirandom graph. *Random Structures and Algorithms*, 16:195–208, 2000.

A. M. Frieze and C. McDiarmid. Algorithmic theory of random graphs. *Random Structures and Algorithms*, 10(1-2):5–42, 1997.

A. M. Frieze and B. Reed. Probabilistic analysis of algorithms. In M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, volume 16 of *Algorithms and Combinatorics*, pages 36–92. Springer-Verlag, Berlin, 1998.

Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1:233–241, 1981.

M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., 1979.

J. Håstad. Clique is hard to approximate within $n^{1-\varepsilon}$. *Acta Mathematica*, 182(1):105–142, 1999.

A. J. Hoffman and L. Howes. On eigenvalues and colorings of graphs, ii. *Annals of the New York Academy of Sciences*, 175(1):238–242, 1970.

R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

E. Hu, B. Wang, and S. Chen. Bcdnpkl: Scalable non-parametric kernel learning using block coordinate descent. In *Proceedings of the 28th International Conference on Machine Learning*, ICML '11, pages 209–216, 2011.

H. Hu, X. Yan, Y. Huang, J. Han, and X. J. Zhou. Mining coherent dense subgraphs across massive biological networks for functional discovery. *Bioinformatics*, 21(suppl 1):i213–i221, 2005.

D. R. Hush, P. Kelly, C. Scovel, and I. Steinwart. Qp algorithms with guaranteed accuracy and run time for support vector machines. *Journal of Machine Learning Research*, 7:733–769, 2006.

S. Janson, T. Luczak, and A. Rucinski. *Random Graphs*. John Wiley & Sons, Inc., 2000.

M. Jerrum. Large cliques elude the metropolis process. *Random Structures and Algorithms*, 3(4): 347–360, 1992.

D. Jiang and J. Pei. Mining frequent cross-graph quasi-cliques. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2(4):16, 2009.

D. S. Johnson and M. A. Trick. *Cliques, Coloring, and Satisfiability: Second DIMACS Implementation Challenge, October 11-13, 1993*, volume 26. Amer Mathematical Society, 1996.

F. Juhász. The asymptotic behaviour of lovász' theta-function for random graphs. *Combinatorica*, 2(2):153–155, 1982.

D. Karger, R. Motwani, and M. Sudan. Approximate graph coloring by semidefinite programming. *J. ACM*, 45:246–265, March 1998. ISSN 0004-5411.

D. Knuth. The sandwich theorem. *Electronic Journal of Combinatorics*, 1(A1), 1994.

M. Krivelevich. Deciding k-colorability in expected polynomial time. *Information Processing Letters*, 81, pages 1–6, 2002.

L. Kucera. Expected complexity of graph partitioning problems. *Discrete Applied Mathematics*, 57 (2-3):193–212, 1995.

G. R. G. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan. Learning the kernel matrix with semidefinite programming. *Journal of Machine Learning Research*, 5:27–72, 2004.

K. J. Lang and R. Andersen. Finding dense and isolated submarkets in a sponsored search spending graph. In *Proceedings of the ACM Sixteenth Conference on Information and Knowledge Management (CIKM '07)*, pages 613–622, 2007.

V. E. Lee, N. Ruan, R. Jin, and C. Aggarwal. A survey of algorithms for dense subgraph discovery. *Managing and Mining Graph Data*, pages 303–336, 2010.

L. Lovász. On the Shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25(1): 1–7, 1979.

C. J. Luz. An upper bound on the independence number of a graph computable in polynomial-time. *Operations Research Letters*, 18(3):139 – 145, 1995.

C. J. Luz and A. Schrijver. A convex quadratic characterization of the lovász theta number. *SIAM Journal on Discrete Mathematics*, 19(2):382–387, 2006.

C. McDiarmid. Concentration. In M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, volume 16 of *Algorithms and Combinatorics*, pages 1–46. Springer-Verlag, Berlin, 1998.

M. E. J. Newman, A. L. Barabasi, and D. J. Watts. *The Structure and Dynamics of Networks*. Princeton Univ Pr, 2006.

P. Pardalos and S. Rebennack. Computational challenges with cliques, quasi-cliques and clique partitions in graphs. *Experimental Algorithms*, pages 13–22, 2010.

Y. Podolyan and G. Karypis. Common pharmacophore identification using frequent clique detection algorithm. *Journal of Chemical Information and Modeling*, 49(1):13–21, 2009.

A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet. SimpleMKL. *Journal of Machine Learning Research*, 9:2491–2521, 2008.

B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001.

M. Sion. On general minimax theorems. *Pac. J. Math.*, 8:171–176, 1958.

V. Spirin and L. A. Mirny. Protein complexes and functional modules in molecular networks. *Proceedings of the National Academy of Sciences*, 100(21):12123, 2003.

Y. Takahashi, Y. Satoh, H. Suzuki, and S. Sasaki. Recognition of largest common structural fragment among a variety of chemical structures. *Analytical Sciences*, 3(1):23–28, 1987.

M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathematiques de L'IHES*, 81:73–205, 1995.

V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.

V. H. Vu. Spectral norm of random matrices. *Combinatorica*, 27(6):721–736, 2007.