Guaranteed Recovery of Planted Cliques and Dense Subgraphs by Convex Relaxation

Brendan P. W. Ames

Abstract We consider the problem of identifying the densest $k$-node subgraph in a given graph. We write this problem as an instance of rank-constrained cardinality minimization and then relax using the nuclear norm and one norm. Although the original combinatorial problem is NP-hard, we show that the densest $k$-subgraph can be recovered from the solution of our convex relaxation for certain program inputs. In particular, we establish exact recovery in the case that the input graph contains a single planted clique plus noise in the form of corrupted adjacency relationships. We also establish analogous recovery guarantees for identifying the densest subgraph of fixed size in a bipartite graph, and include results of numerical simulations for randomly generated graphs to demonstrate the efficacy of our algorithm.

Keywords Planted clique · Densest subgraph · Nuclear norm minimization · $l_1$ norm minimization

Mathematics Subject Classification 90C25 · 90C59 · 65K05 · 68Q25

1 Introduction

We consider the *densest $k$-subgraph problem*. Given input graph $G$ and integer $k$, the densest $k$-subgraph problem seeks the $k$-node subgraph of $G$ with maximum number of edges. The identification and analysis of dense subgraphs play significant roles...
in a wide range of applications, including information retrieval, pattern recognition, computational biology and image processing; for example, a group of densely connected nodes may correspond to a community of users in a social network, or cluster of similar items in a given data set. Unfortunately, the problem of finding a densest subgraph of given size is known to be both NP-hard (see [1]) and hard to approximate (see [2–4]).

Our results can be thought of as a generalization to the densest $k$-subgraph problem of those in [5] for the maximum clique problem. In [5], Ames and Vavasis show that the maximum clique in a graph consisting of a single large clique, called a planted clique, and a moderate amount of diversionary edges and nodes can be identified from the minimum nuclear norm solution of a particular system of linear inequalities. These linear constraints restrict all feasible solutions to be adjacency matrices of subgraphs with a desired number of nodes, say $k$, while the objective acts as a surrogate for the rank of the feasible solution; a rank-one solution would correspond to a $k$-clique in the input graph. We establish analogous recovery guarantees for a convex relaxation of the planted clique problem that is robust to noise in the form of both diversionary edge additions and deletions within the planted complete subgraph. In particular, we modify the relaxation of [5] by adding an $\ell_1$ norm penalty to measure the error between the rank-one approximation of the adjacency matrix of each $k$-subgraph and its true adjacency matrix.

This relaxation technique, and its accompanying recovery guarantee, mirrors that of several recent papers regarding convex optimization approaches for robust principal component analysis [6–8] and graph clustering [9–11]. Each of these papers establishes that a desired matrix or graph structure, represented as the sum of a low-rank and sparse matrix, can be recovered from the optimal solution of some convex program under certain assumptions on the input matrix or graph. In particular, our analysis and results are closely related to those of [11]. In [11], Chen et al. consider a convex optimization heuristic for identifying clusters in data, represented as collections of relatively dense subgraphs in a sparse graph, and provide bounds on the size and density of these subgraphs ensuring exact recovery using this method. We establish analogous guarantees for identifying a single dense subgraph when the cardinality of this subgraph is known a priori; for example, we will show that a planted clique of cardinality as small as $\Omega(N^{1/3})$ can be recovered in the presence of sparse random noise, where $N$ is the number of nodes in the input graph, significantly less than the bound $\Omega(N^{1/2})$ established in [5].

2 The Densest $k$-Subgraph Problem

The density of a graph $G = (V, E)$ is defined to be the average number of edges incident at a vertex or average degree of $G$: $d(G) = |E|/|V|$. The densest $k$-subgraph problem seeks a $k$-node subgraph of $G$ of maximum average degree or density: $\max\{d(H) : H \subseteq G, |V(H)| = k\}$. Although the problem of finding a subgraph with maximum average degree is polynomially solvable (see [12, Chapter 4]), the densest $k$-subgraph problem is NP-hard. Indeed, if a graph $G$ has a clique of size $k$, this clique would be the densest $k$-subgraph of $G$. Thus, any instance of the maximum
clique problem, known to be NP-hard [13], is equivalent to an instance of the densest
$k$-subgraph problem. Moreover, the densest $k$-subgraph problem is hard to approxi-
mate; specifically, it has been shown that the densest $k$-subgraph problem does not
admit a polynomial-time approximation scheme under various complexity theoretical
assumptions [2–4]. Due to, and in spite of, this intractability of the densest $k$-subgraph
problem, we consider relaxation to a convex program. Although we do not expect this
relaxation to provide a good approximation of the densest $k$-subgraph for every input
graph, we will establish that the densest $k$-subgraph can be recovered from the optimal
solution of this convex relaxation for certain classes of input graphs. In particular, we
will show that our relaxation is exact for graphs containing a single dense subgraph
obscured by noise in the form of diversionary nodes and edges.

Our relaxation is based on the observation that the adjacency matrix of a dense
subgraph is well approximated by the rank-one adjacency matrix of the complete graph
on the same node set. Let $V' \subseteq V$ be a subset of $k$ nodes of the graph $G = (V, E)$,
and let $v$ be its characteristic vector. That is, for all $i \in V$, $v_i = 1$ if $i \in V'$ and is
equal to 0 otherwise. The vector $v$ defines a rank-one matrix $X$ by the outer product of $v$
with itself: $X := vv^T$. Moreover, if $V'$ is a clique of $G$, then the nonzero entries of $X$
correspond to the $k \times k$ all-ones block of the perturbed adjacency matrix $\tilde{A}_G := A_G + I$
of $G$ with rows and columns indexed by $V'$; here, $A_G \in \mathbb{R}^{V \times V}$ denotes the adjacency
matrix of the graph $G$, defined by $[A_G]_{ij} = 1$ if $ij \in E$ and $[A_G]_{ij} = 0$ otherwise,
and $I \in \mathbb{R}^{V \times V}$ denotes the identity matrix with rows and columns indexed by $V$. If
$V'$ is not a clique of $G$, then the entries of $\tilde{A}_G(V', V')$ indexed by nonadjacent nodes
are equal to 0. Let $\tilde{Y} \in \mathbb{R}^{V \times V}$ be the matrix defined by

$$\tilde{Y}_{ij} := \begin{cases} -\tilde{X}_{ij}, & \text{if } ij \in \tilde{E}, \\ 0, & \text{otherwise}, \end{cases}$$

(1)

where $\tilde{E}$ is the complement of the edge set of $G$ given by $\tilde{E} := (V \times V) - E - \{uu : u \in V\}$. That is, $\tilde{Y} = -P_{\tilde{E}}(\tilde{X})$, where $P_{\tilde{E}}$ is the orthogonal projection onto the set of
matrices with support contained $\tilde{E}$ defined by

$$[P_{\tilde{E}}(M)]_{ij} := \begin{cases} M_{ij}, & \text{if } (i, j) \in \tilde{E}, \\ 0, & \text{otherwise}, \end{cases}$$

(2)

for all $M \in \mathbb{R}^{V \times V}$. The matrix $\tilde{Y}$ can be thought of as a correction for the entries of
$\tilde{X}$ indexed by nonedges of $G$. Indeed, $\tilde{X} + \tilde{Y}$ is exactly the adjacency matrix of the
subgraph of $G$ induced by $V'$, with ones in diagonal entries indicating loops at each
$v \in V'$; $\tilde{X}_{ij} + \tilde{Y}_{ij} = 1$ for all $(i, j) \in V' \times V'$ such that $ij \in E$ or $i = j$, and is 0
otherwise. Moreover, the density of $G(V')$ is equal to

$$d(G(V')) = \frac{1}{2k} (k(k - 1) - 0),$$

by the fact that the number of nonzero entries in $\tilde{Y}$ is exactly twice the number of
nonadjacent pairs of nodes in $G(V')$; here $\| \tilde{Y} \|_0$ denotes the so-called $\ell_0$ norm of $\tilde{Y}$,
defined as the cardinality of the support of \( \bar{Y} \). Maximizing the density of \( H \) over all \( k \)-node subgraphs of \( G \) is equivalent to minimizing \( \| Y \|_0 \) over all \( (X, Y) \) as constructed above. Consequently, the densest \( k \)-subgraph problem is equivalent to

\[
\min_{X, Y \in \Sigma V} \left\{ \| Y \|_0 : \text{rank}(X) = 1, \langle e, Xe \rangle = k^2, X_{ij} + Y_{ij} = 0 \forall ij \in \tilde{E}, X \in \{0, 1\}^{V \times V} \right\},
\]

where \( e \) is the all-ones vector in \( \mathbb{R}^V \), \( \Sigma V \) denotes the cone of \( |V| \times |V| \) symmetric matrices with rows and columns indexed by \( V \), and \( \langle \cdot, \cdot \rangle \) denotes the Euclidean inner product on \( \mathbb{R}^V \) defined by \( \langle x, y \rangle = \sum_{i \in V} x_i y_i \).

Indeed, the constraints \( \text{rank}(X) = 1, \langle e, Xe \rangle = k^2, \) and \( X \in \Sigma V \cap \{0, 1\}^{V \times V} \) force any feasible \( X \) to be a rank-one symmetric binary matrix with exactly \( k^2 \) nonzero entries, while the requirement that \( X_{ij} + Y_{ij} = 0 \) if \( ij \in \tilde{E} \) ensures that every entry of \( X \) indexed by a nonadjacent pair of nodes is corrected by \( Y \). Moving the constraint \( \text{rank}(X) = 1 \) to the objective as a penalty term yields the nonconvex program

\[
\min_{X, Y \in \Sigma V} \left\{ \text{rank}(X) + \gamma \| Y \|_0 : \langle e, Xe \rangle = k^2, X_{ij} + Y_{ij} = 0 \forall ij \in \tilde{E}, X \in \{0, 1\}^{V \times V} \right\}.
\]

Here \( \gamma > 0 \) is a regularization parameter to be chosen later. We relax (4) to the convex problem

\[
\min \left\{ \| X \|_* + \gamma \| Y \|_1 : \langle e, Xe \rangle = k^2, X_{ij} + Y_{ij} = 0 \forall ij \in \tilde{E}, X \in \{0, 1\}^{V \times V} \right\}
\]

by replacing \( \text{rank} \) and \( \| \cdot \|_0 \) with their convex envelopes, the nuclear norm \( \| \cdot \|_* \) and the \( \ell_1 \) norm \( \| \cdot \|_1 \), relaxing the binary constraints on the entries of \( X \) to the corresponding box constraints, and omitting the symmetry constraints on \( X \) and \( Y \). Although ignoring the symmetry constraints is not necessary to obtain a tractable relaxation, our proofs of Theorems 2.1 and 2.2 suggest that the Lagrange multipliers corresponding to the symmetry constraints in (4) may be chosen to be equal to 0; we omit the symmetry constraints to eliminate these \( O(N^2) \) potentially redundant constraints and allow a simpler extension of (5) to the bipartite problem considered in Sect. 3. Here \( \| Y \|_1 \) denotes the \( \ell_1 \) norm of the vectorization of \( Y \): \( \| Y \|_1 := \sum_{i \in V} \sum_{j \in V} |Y_{ij}|. \) Note that \( \| Y \|_0 = \| Y \|_1 \) for the proposed choice of \( Y \) given by (1), although this equality clearly does not hold in general.

Our relaxation mirrors that proposed by Chandrasekaran et al. [6] for robust principal component analysis. Given matrix \( M \in \mathbb{R}^{m \times n} \), the robust PCA problem seeks a decomposition \( M = L + S \) where \( L \in \mathbb{R}^{m \times n} \) has low rank and \( S \in \mathbb{R}^{m \times n} \) is sparse. In [6], it is shown that such a decomposition can be obtained by solving the convex

\[
\min \left\{ \| X \|_* + \gamma \| Y \|_1 : \langle e, Xe \rangle = k^2, X_{ij} + Y_{ij} = 0 \forall ij \in \tilde{E}, X \in \{0, 1\}^{V \times V} \right\}
\]

by replacing \( \text{rank} \) and \( \| \cdot \|_0 \) with their convex envelopes, the nuclear norm \( \| \cdot \|_* \) and the \( \ell_1 \) norm \( \| \cdot \|_1 \), relaxing the binary constraints on the entries of \( X \) to the corresponding box constraints, and omitting the symmetry constraints on \( X \) and \( Y \). Although ignoring the symmetry constraints is not necessary to obtain a tractable relaxation, our proofs of Theorems 2.1 and 2.2 suggest that the Lagrange multipliers corresponding to the symmetry constraints in (4) may be chosen to be equal to 0; we omit the symmetry constraints to eliminate these \( O(N^2) \) potentially redundant constraints and allow a simpler extension of (5) to the bipartite problem considered in Sect. 3. Here \( \| Y \|_1 \) denotes the \( \ell_1 \) norm of the vectorization of \( Y \): \( \| Y \|_1 := \sum_{i \in V} \sum_{j \in V} |Y_{ij}|. \) Note that \( \| Y \|_0 = \| Y \|_1 \) for the proposed choice of \( Y \) given by (1), although this equality clearly does not hold in general.

Our relaxation mirrors that proposed by Chandrasekaran et al. [6] for robust principal component analysis. Given matrix \( M \in \mathbb{R}^{m \times n} \), the robust PCA problem seeks a decomposition \( M = L + S \) where \( L \in \mathbb{R}^{m \times n} \) has low rank and \( S \in \mathbb{R}^{m \times n} \) is sparse. In [6], it is shown that such a decomposition can be obtained by solving the convex
problem \( \min \{ \| L \|_* + \| S \|_1 : M = L + S \} \) under certain assumptions on the input matrix \( M \). This result was subsequently extended in [7–11, 14] to obtain conditions on the input matrix \( M \) ensuring perfect decomposition under partial observation of \( M \) and other linear constraints. These results can be thought of as generalizations of the results of [15], which established conditions under which a low-rank matrix can be recovered from linear samples of its entries and an appropriate nuclear norm minimization; this result, in turn, generalizes those of [16–18] establishing that the sparsest solution of some sets of linear equations can be recovered using relaxation of vector cardinality to the vector \( \ell_1 \) norm. Suppose that \( G \) contains at most \( r \) edges not in \( G \) indexed by this dense subgraph; therefore, we can always construct graphs where (5) fails to satisfy the restricted isometry property. Although these recovery guarantees do not translate immediately to our relaxation, we will establish analogous conditions ensuring exact recovery of the densest \( k \)-subgraph of \( G \) from (5).

We consider a planted case analysis of (5). Suppose that the input graph \( G \) contains a single dense subgraph \( H \), plus diversionary edges and nodes. We are interested in the trade-off between the density of \( H \), the size \( k \) of \( H \), and the level of noise required to guarantee recovery of \( H \) from the optimal solution of (5). In particular, we consider graphs \( G = (V, E) \) constructed as follows. We start by adding all edges between elements of some \( k \)-node subset \( V^* \subseteq V \) to \( E \). That is, we create a \( k \)-clique \( V^* \) by adding the edge set of the complete graph with vertex set \( V^* \) to \( G \). We then corrupt this \( k \)-clique with noise in the form of deletions of edges within \( V^* \times V^* \) and additions of potential edges in \((V \times V) - (V^* \times V^*)\).

We consider two cases. In the first, these additions and deletions are performed deterministically. In the second, the adjacency of each vertex pair is corrupted independently at random. In the absence of edge deletions, this is exactly the planted clique model considered in [5]. In [5], Ames and Vavasis provide conditions ensuring exact recovery of a planted clique from the optimal solution of the convex program

\[
\min_X \left\{ \| X \|_* : \langle e, X e \rangle \geq k^2, \ X_{ij} = 0 \ \forall \ i, j \in \tilde{E} \right\}. \tag{6}
\]

The following theorem provides a recovery guarantee for the densest \( k \)-subgraph in the case of adversarial edge additions and deletions, analogous to that of [5, Sect. 4.1].

**Theorem 2.1** Let \( V^* \) be a \( k \)-subset of nodes of the graph \( G = (V, E) \) and let \( v \) be its characteristic vector. Suppose that \( G \) contains at most \( r \) edges not in \( G(V^*) \) and \( G(V^*) \) contains at least \( \left( \frac{1}{2} \right) - s \) edges, such that each vertex in \( V^* \) is adjacent to at least \((1 - \delta_1)k \) nodes in \( V^* \) and each vertex in \( V - V^* \) is adjacent to at most \( \delta_2 k \) nodes in \( V^* \) for some \( \delta_1, \delta_2 \in (0, 1) \) satisfying \( 2\delta_1 + \delta_2 < 1 \). Let \((X^*, Y^*)\) be the feasible solution for (5) where \( X^* = vv^T \) and \( Y^* \) is constructed according to (1). Then, there exist scalars \( c_1, c_2 > 0 \), depending only on \( \delta_1 \) and \( \delta_2 \), such that if \( s \leq c_1 k^2 \) and \( r \leq c_2 k^2 \), then \( G(V^*) \) is the unique maximum density \( k \)-subgraph of \( G \) and \((X^*, Y^*)\) is the unique optimal solution of (5) for \( \gamma = 2((1 - 2\delta_1 - \delta_2)k)^{-1} \).
In Theorem 2.1, the constants $\delta_1$ and $c_1$ parametrize the density of the planted dense subgraph $G(V^*)$, while $\delta_2$ and $c_2$ control the number of edges in $G$ outside of $G(V^*)$. Specifically, $\delta_1$ denotes the minimum degree of a node in $G(V^*)$ and $\delta_2$ denotes the maximum number of neighbours each node in $V - V^*$ may have in $V^*$. Theorem 2.1 states that if $G(V^*)$ is sufficiently dense, then $G(V^*)$ is the densest $k$-subgraph of $G$ and can be recovered by solving the relaxation (5); here, “sufficiently dense” corresponds to $G(V^*)$ containing at least $\binom{k}{2} - c_1 k^2$ edges and $G$ containing at most $c_2 k^2$ edges total.

The bound on the number of adversarially added edges given by Theorem 2.1 matches that given in [5, Sect. 4.1] up to constants. Moreover, the noise bounds given by Theorem 2.1 are optimal in the following sense. Adding $k$ edges from any node $v'$ outside $V^*$ to $V^*$ would result in the creation of a $k$-subgraph (induced by $v'$ and $V^* - u$ for some $u \in V^*$) of greater density than $G(V^*)$. Similarly, if the adversary can add or delete $O(k^2)$ edges, then the adversary can create a $k$-subgraph with greater density than $G(V^*)$. In particular, a $k$-clique could be created by adding at most $\binom{k}{2}$ edges.

We also consider random graphs $G = (V, E)$ constructed in the following manner. Let $V^*$ be a subset of $V$ of size $k$ and add $ij$ to $E$ independently with fixed probability $1 - q$ for all $(i, j) \in V^* \times V^*$. Each of the remaining potential edges in $(V \times V) - (V^* \times V^*)$ is added independently to $E$ with fixed probability $p$. We say such a graph $G$ is sampled from the planted dense $k$-subgraph model. By construction, the subgraph $G(V^*)$ induced by $V^*$ will likely be substantially more dense than all other $k$-subgraphs of $G$ if $p + q < 1$. Note that $V^*$ is a $k$-clique of $G$ if $q = 0$. Theorem 7 of [5] states that a planted $k$-clique can be recovered from the optimal solution of (6) with high probability if $|V| = O(k^2)$ in this case. The following theorem generalizes this result for all $q \neq 0$.

**Theorem 2.2** Suppose that the $N$-node graph $G$ is sampled from the planted dense $k$-subgraph model with $p, q$ and $k$ satisfying $p + q < 1$ and

\[
(1 - p)k \geq \max\{8p, 1\} \cdot 8 \log k, \quad \sqrt{pN} \geq (1 - p) \log N \tag{7}
\]

\[
(1 - p - q)k \geq 72 \max \left\{ (q(1 - q)k \log k)^{1/2}, \log k \right\} \tag{8}
\]

Then, there exist absolute constants $c_1, c_2, c_3 > 0$ such that if

\[
(1 - p - q)(1 - p)k > c_1 \max \left\{ p^{1/2}, ((1 - p)k)^{-1/2} \right\} \cdot N^{1/2} \log N, \tag{9}
\]

then the $k$-subgraph induced by $V^*$ is the densest $k$-subgraph of $G$ and the proposed solution $(X^*, Y^*)$ is the unique optimal solution of (5) with high probability for

\[
\gamma \in \left( \frac{c_2}{(1 - p - q)k}, \frac{c_3}{(1 - p - q)k} \right). \tag{10}
\]
The constant $c_1$ places a lower bound on the size of planted clique recoverable, while $c_2$ and $c_3$ provided a range of acceptable regularization parameters. The assumption (9) places a lower bound on the density of the subgraph $G(V^*)$ induced by the planted clique in terms of its size $k$. On the other hand, while places bounds on the number of noise edges in $G$; in particular, (7) ensures that $p$ is not too close to 0 or 1, primarily for the sake of simplifying the proof of Theorem 2.2. If (7) and (9) are satisfied, then Theorem 2.2 states that if we use the regularization parameter $\gamma = \kappa / ((1 - p - q)k)$ for any $\kappa$ in the interval $(c_2, c_3)$, then we can recover a planted $k$-clique of $G$ w.h.p. from the optimal solution of (5) provided that the size of the planted clique $k$ is greater than a constant $c_1$ times the maximum of $\sqrt{pN \log N}$ and $\sqrt{N/((1 - p)k) \log N}$. That $\gamma$ is chosen from an interval and not a single value as in Theorem 2.1 is a consequence of the probabilistic analysis in Sect. 4. The fact that we deterministically construct the graph $G$ in Theorem 2.1 allows us to choose a single value of $\gamma$ in order to simplify our analysis substantially; the proof of Theorem 2.1 can be modified to ensure recovery for a range of $\gamma$ as in the probabilistic case (see the analysis in the supplemental material found in [19]).

To further clarify the implications of Theorem 2.2, we consider two cases. Suppose that the graph $G$ constructed according to the planted dense $k$-subgraph model is dense, that is, both $p$ and $q$ are fixed with respect to $k$ and $N$. In this case, Theorem 2.2 suggests that $G(V^*)$ is the densest $k$-subgraph and its matrix representation is the unique optimal solution of (5) w.h.p. provided that $k$ is at least as large as $\Omega(\sqrt{N \log N})$. This lower bound matches that of [5, Theorem 7], as well as [20–26], up to the constant and logarithmic factors, despite the presence of additional noise in the form of edge deletions. Moreover, modifying the proof of Theorem 2.2 to follow the proof of [5, Theorem 7] shows that the planted dense $k$-subgraph can be recovered w.h.p. provided $k = \Omega(\sqrt{N})$ in the dense case; see the remarks following Lemma 4.6. Whether planted cliques of size $o(\sqrt{N})$ can be recovered in polynomial time is still an open problem, although this task is widely believed to be intractable (and this presumed hardness has been exploited in cryptographic applications [27] and complexity analysis [4,28–30]). Moreover, several heuristics [31–33] have been shown to fail to recover planted cliques of size $o(\sqrt{N})$ in polynomial time.

When the noise obscuring the planted clique is sparse, i.e. both $p$ and $q$ are tending to 0 as $N \to \infty$, the lower bound on the size of a recoverable clique can be significantly improved; for example, if $p, q = O(1/k)$, then Theorem 2.2 states that the planted clique can be recovered w.h.p. if $k = \Omega(N^{1/3} \log N)$. On the other hand, if either $p$ or $q$ tends to 1 as $N \to \infty$, then the minimum size of $k$ required for exact recovery will necessarily increase.

It is important to note that the choice of $\gamma$ in both Theorems 2.1 and 2.2 ensuring exact recovery is not universal, but rather depends on the parameters governing edge addition and deletion. These quantities are typically not known in practice. However, under stronger assumptions on the edge corrupting noise, $\gamma$ independent of the unknown noise parameters may be identified; for example, if we impose the stronger assumption that $p + q \leq 1/2$, then we may take $\gamma = 6/k$. 

---

\cite{Springer}
3 The Densest \((k_1, k_2)\)-Subgraph Problem

Let \(G = (U, V, E)\) be a bipartite graph. That is, \(G\) is a graph whose vertex set can be partitioned into two independent sets \(U\) and \(V\). A bipartite subgraph \(H = (U', V', E')\) is a \((k_1, k_2)\)-subgraph of \(G\) if \(U' \subseteq U\) and \(V' \subseteq V\) such that \(|U'| \cdot |V'| = k_1 k_2\). Given bipartite graph \(G\) and integers \(k_1, k_2\), the densest \((k_1, k_2)\)-subgraph problem seeks the \((k_1, k_2)\)-subgraph of \(G\) containing maximum number of edges. This problem is NP-hard, by reduction from the maximum edge biclique problem [34], and hard to approximate [2,35].

As before, we consider a convex relaxation of the densest \((k_1, k_2)\)-subgraph problem motivated by the fact that the adjacency matrices of dense \((k_1, k_2)\)-subgraphs are closely approximated by rank-one matrices. If \((U', V')\) is a biclique of \(G\), i.e. \(ij \in E\) for all \(i \in U', j \in V'\), then the bipartite subgraph \(G(U', V')\) induced by \((U', V')\) is a \((k_1, k_2)\)-subgraph of \(G\) containing all \(k_1 k_2\) possible edges between \(U'\) and \(V'\). In this case, the \((U, V)\) block of the adjacency matrix of \(G(U', V')\) is equal to \(X' = uv^T\), where \(u\) and \(v\) are the characteristic vectors of \(U'\) and \(V'\), respectively. Note that \(X'\) has rank equal to one. If \((U', V')\) is not a biclique of \(G\), then there exists some \(i \in U', j \in V'\) such that \(ij \notin E\). In this case, the \((U, V)\) block of the adjacency matrix of \(G(U', V')\) has the form \(X' + Y'\), where \(Y' = -P_{\tilde{E}}(X')\). Here \(P_{\tilde{E}}\) denotes the orthogonal projection onto the set of matrices with support contained in the complement \(\tilde{E} := (U \times V) - E\) of the edge set \(E\). As such, the densest \((k_1, k_2)\)-subgraph problem may be formulated as the rank-constrained cardinality minimization problem:

\[
\min \left\{ \|Y\|_0 : \text{rank } X = 1, \ \langle e, X e \rangle = k_1 k_2, \ X_{ij} + Y_{ij} = 0, \ \forall ij \in \tilde{E}, \ X \in \{0, 1\}^{U \times V} \right\}.
\]

We obtain a tractable convex relaxation by moving the rank constraint to the objective as a regularization term, relaxing rank and the \(\ell_0\) norm with the nuclear norm and \(\ell_1\) norm, respectively, and replacing the binary constraints with appropriate box constraints:

\[
\min \left\{ \|X\|_* + \gamma \|Y\|_1 : \langle e, X e \rangle = k_1 k_2, \ X_{ij} + Y_{ij} = 0, \ \forall ij \in \tilde{E}, \ X \in \{0, 1\}^{U \times V} \right\}.
\]

(11)

Except for superficial differences, this problem is identical to the convex relaxation of (3) given by (5). As can be expected, the recovery guarantees for the relaxation of the densest \(k\)-subgraph problem translate to similar guarantees for the convex relaxation (11) of the densest \((k_1, k_2)\)-subgraph problem.

As in Sect. 2, we consider the performance of the relaxation (11) in the special case that the input graph contains an especially dense \((k_1, k_2)\)-subgraph. As before, we consider graphs constructed to contain such a subgraph as follows. First, all edges between \(U^*\) and \(V^*\) are added for a particular pair of subsets \(U^* \subseteq U\), \(V^* \subseteq V\) such that \(|U^*| = k_1\) and \(|V^*| = k_2\). Then, some of the remaining potential edges in \(U \times V\) are added, while some of the edges between \(U^*\) and \(V^*\) are deleted. As in the previous section, this introduction of noise is either performed deterministically by an adversary or at random with each edge added or deleted independently with fixed...
probability. The following theorem provides bounds on the amount of deterministic noise ensuring exact recovery of the planted dense \((k_1, k_2)\)-subgraph by \((11)\).

**Theorem 3.1** Let \(G = (U, V, E)\) be a bipartite graph and let \(U^* \subseteq U, V^* \subseteq V\) be subsets of cardinality \(k_1\) and \(k_2\), respectively. Let \(u\) and \(v\) denote the characteristic vectors of \(U^*\) and \(V^*\), and let \((X^*, Y^*) = (uv^T, -P_{\hat{E}}(uv^T))\). Suppose that \(G(U^*, V^*)\) contains at least \(k_1k_2 - s\) edges and that \(G\) contains at most \(r\) edges other than those in \(G(U^*, V^*)\). Suppose that every node in \(V^*\) is adjacent to at least \((1 - \alpha_1)k_1\) nodes in \(U^*\) and every node in \(U^*\) is adjacent to at least \((1 - \alpha_2)k_2\) nodes in \(V^*\) for some scalars \(\alpha_1, \alpha_2 > 0\). Further, suppose that each node in \(V - V^*\) is adjacent to at most \(\beta_1k_1\) nodes in \(U^*\) and each node in \(U - U^*\) is adjacent to at most \(\beta_2k_2\) nodes in \(V^*\) for some \(\beta_1, \beta_2 > 0\). Finally, suppose that the scalars \(\alpha_1, \alpha_2, \beta_1, \beta_2\) satisfy \(\alpha_1 + \alpha_2 + \max\{\beta_1, \beta_2\} < 1\). Then, there exist scalars \(c_1, c_2 > 0\), depending only on \(\alpha_1, \alpha_2, \beta_1\), and \(\beta_2\), such that if \(r \leq c_1k_1k_2\) and \(s \leq c_2k_1k_2\), then \(G(U^*, V^*)\) is the unique maximum density \((k_1, k_2)\)-subgraph of \(G\) and \((X^*, Y^*)\) is the unique optimal solution of \((11)\) for \(\gamma = 2(\sqrt{k_1k_2}(1 - \alpha_1 - \alpha_2 - \max\{\beta_1, \beta_2\}))^{-1}\).

Here \((\alpha_1, \alpha_2)\) are analogous to \(\delta_1\) and \(\beta_1, \beta_2\) are analogous to \(\delta_2\) in Theorem 2.1. That is, Theorem 3.1 implies that we may recover the densest \((k_1, k_2)\)-subgraph of \(G\) provided this densest \((k_1, k_2)\)-subgraph is sufficiently dense. As in the earlier theorem, “sufficiently dense” is controlled by the minimum degree of nodes in \(G(U^*, V^*)\), parametrized by \((\alpha_1, \alpha_2)\) and \(c_2\), and the maximum number of neighbours outside of \(G(U^*, V^*)\) that each node in \(U^*\) and \(V^*\) may have, as parametrized by \((\beta_1, \beta_2)\) and \(c_1\).

As before, the bounds on the number of edge corruptions that guarantee exact recovery given by Theorem 3.1 are identical to those provided in [5, Sect. 5.1] (up to constants). Moreover, these bounds are optimal for reasons similar to those in the discussion immediately following Theorem 2.1.

A similar result holds for random bipartite graphs \(G = (U, V, E)\) constructed as follows. For some \(k_1\)-subset \(U^* \subseteq U\) and \(k_2\)-subset \(V^* \subseteq V\), we add each potential edge from \(U^*\) to \(V^*\) independently with probability \(1 - q\). Then, each remaining possible edge is added independently to \(E\) with probability \(p\). We say such a graph is sampled from the planted dense \((k_1, k_2)\)-subgraph model. By construction \(G(U^*, V^*)\) is dense in expectation, relative to its complement, if \(p + q < 1\). Theorem 9 of [5] asserts that \(G(U^*, V^*)\) is the densest \((k_1, k_2)\)-subgraph of \(G\) and can be recovered using a modification of \((11)\), for sufficiently large \(k_1\) and \(k_2\), in the special case that \(q = 0\). The following generalizes this result for all \(p\) and \(q\).

**Theorem 3.2** Suppose that the \((N_1, N_2)\)-node bipartite graph \(G = (U, V, E)\) is constructed according to the planted dense \((k_1, k_2)\)-subgraph model with \(p + q < 1\) and

\[
(1 - p)k_1 \geq \max\{8, 64p\} \log k_i \tag{12}
\]

\[
pN_i \geq (1 - p)^2 \log^2 N_i \tag{13}
\]

\[
(1 - p - q)k_i \geq 72 \max\left\{\log k_i, (q(1 - q)k_i \log k_i)^{1/2}\right\} \tag{14}
\]
for \( i = 1, 2 \). Then, there exist absolute constants \( c_1, c_2, c_3 > 0 \) such that

\[
c_1(1 - p - q)\sqrt{k_1k_2} \geq \tilde{N}^{1/2} \log \tilde{N} \cdot \max \left\{ p^{1/2}, ((1 - p) \min(k_1, k_2))^{-1/2} \right\},
\]

where \( \tilde{N} = \max\{N_1, N_2\} \); then, \( G(U^*, V^*) \) is the densest \((k_1, k_2)\)-subgraph of \( G \) and \((X^*, Y^*)\) is the unique optimal solution of (11) w.h.p. for

\[
\gamma \in \left( \frac{c_2}{(1 - p - q)\sqrt{k_1k_2}}, \frac{c_3}{(1 - p - q)\sqrt{k_1k_2}} \right).
\]

Theorem 3.2 is the bipartite analogue of Theorem 2.2. That is, Theorem 3.2 implies that we can recover the planted \((k_1, k_2)\)-subgraph of \( G \) from the optimal solution of (11) provided this subgraph is sufficiently large, as characterized by (15), and we choose the parameter \( \gamma \) from the interval of acceptable values given by (16).

### 4 Exact Recovery of the Densest \( k \)-Subgraph Under Random Noise

This section consists of a proof of Theorem 2.2. The proofs of Theorems 2.1, 3.1, and 3.2 follow a similar structure and are omitted; proofs of these theorems may be found in the supplemental material [19]. Let \( G = (V, E) \) be a graph sampled from the planted dense \( k \)-subgraph model. Let \( V^* \) be the node set of the planted dense \( k \)-subgraph \( G(V^*) \) and \( v \) be its characteristic vector. Our goal is to show that the solution \((X^*, Y^*) := (vv^T, -P_{\tilde{E}}(vv^T))\), where \( \tilde{E} := (V \times V) - (E \cup \{uu : u \in V\}) \), is the unique optimal solution of (5) and \( G(V^*) \) is the densest \( k \)-subgraph of \( G \) in the case that \( G \) satisfies the hypothesis of Theorem 2.2. To do so, we will apply the Karush–Kuhn–Tucker Theorem to derive sufficient conditions for optimality of a feasible solution of (5) corresponding to a \( k \)-subgraph of \( G \) and then establish that these sufficient conditions are satisfied at \((X^*, Y^*)\) w.h.p. if the assumptions of Theorem 2.2 are met.

#### 4.1 Optimality Conditions

The following theorem provides the necessary specialization of the Karush–Kuhn–Tucker conditions (see [36, Sect. 5.5.3]) to (5); a proof of Theorem 4.1 can be found in the supplemental material.

**Theorem 4.1** Let \( G = (V, E) \) be a graph sampled from the planted dense \( k \)-subgraph model. Let \( \tilde{V} \) be a subset of \( V \) of cardinality \( k \) and let \( \tilde{v} \) be the characteristic vector of \( \tilde{V} \). Let \( \tilde{X} = \tilde{v}\tilde{v}^T \) and let \( \tilde{Y} \) be defined as in (1). Suppose that there exist \( F, W \in \mathbb{R}^{V \times V}, \lambda \in \mathbb{R}, \) and \( M \in \mathbb{R}^{V \times V}_+ \) such that

\[
\tilde{X} + W - \lambda ee^T - \gamma(\tilde{Y} + F) + M = 0,
\]

\[
W\tilde{v} = W^T\tilde{v} = 0, \quad \|W\| \leq 1,
\]

\[
P_{\Omega}(F) = 0, \quad \|F\|_\infty \leq 1,
\]
\[ F_{ij} = 0 \quad \text{for all} \quad (i, j) \in E \cup \{vv : v \in V\}, \quad (20) \]
\[ M_{ij} = 0 \quad \text{for all} \quad (i, j) \in (V \times V) - (\bar{V} \times \bar{V}). \quad (21) \]

Then, \((\bar{X}, \bar{Y})\) is an optimal solution of (5) and the subgraph \(G(\bar{V})\) induced by \(\bar{V}\) is a maximum density \(k\)-subgraph of \(G\). Moreover, if \(\|W\| < 1\) and \(\|F\|_\infty < 1\), then \((\bar{X}, \bar{Y})\) is the unique optimal solution of (5) and \(G(\bar{V})\) is the unique maximum density \(k\)-subgraph of \(G\).

It remains to show that multipliers \(F, W \in \mathbb{R}^{V \times V}, \lambda \in \mathbb{R}\), and \(M \in \mathbb{R}^{V \times V}_+\) corresponding to the proposed solution \((X^*, Y^*)\) and satisfying the hypothesis of Theorem 4.1 do indeed exist. To do so, we consider \(W\) and \(F\) constructed according to the following cases:

- **(\(\omega_1\))** If \((i, j) \in V^* \times V^*\) such that \(ij \in E\) or \(i = j\), choosing \(W_{ij} = \tilde{\lambda} - M_{ij}\), where \(\tilde{\lambda} := \lambda - 1/k\), ensures that the left-hand side of (17) is equal to 0.

- **(\(\omega_2\))** If \((i, j) \in \Omega = V^* \times V^* \cap \bar{E}\), then \(F_{ij} = 0\) and choosing \(W_{ij} = \tilde{\lambda} - \gamma - M_{ij}\) makes the left-hand side of (17) equal to 0.

- **(\(\omega_3\))** Let \((i, j) \in (V \times V) - (V^* \times V^*)\) such that \(ij \in E\) or \(i = j\); then, the left-hand side of (17) is equal to \(W_{ij} - \lambda\). In this case, we choose \(W_{ij} = \lambda\) to make both sides of (17) zero.

- **(\(\omega_4\))** Suppose that \(i, j \in V - V^*\) such that \((i, j) \in \bar{E}\). We choose

\[ W_{ij} = -\lambda \left( \frac{p}{1 - p} \right), \quad F_{ij} = -\frac{\lambda}{\gamma} \left( \frac{1}{1 - p} \right). \]

Again, by our choice of \(W_{ij}\) and \(F_{ij}\), the left-hand side of (17) is zero.

- **(\(\omega_5\))** If \(i \in V^*, j \in V - V^*\) such that \((i, j) \in \bar{E}\), then we choose

\[ W_{ij} = -\lambda \left( \frac{n_j}{k - n_j} \right), \quad F_{ij} = -\frac{\lambda}{\gamma} \left( \frac{k}{k - n_j} \right) \]

where \(n_j\) is equal to the number of neighbours of \(j\) in \(V^*\).

- **(\(\omega_6\))** If \(i \in V - V^*, j \in V^*\) such that \((i, j) \in \bar{E}\), we choose \(W_{ij}, F_{ij}\) symmetrically according to (\(\omega_5\)), that is, we choose \(W_{ij} = W_{ji}, F_{ij} = F_{ji}\) in this case.

In Sect. 4.2, we construct valid \(\lambda \in \mathbb{R}\) and \(M \in \mathbb{R}^{V \times V}_+\) such that \(Wv = W^Tv = 0\). We next establish that \(\|F\|_\infty < 1\) w.h.p. for this choice of \(\lambda\) and \(M\) and a particular choice of regularization parameter \(\gamma\) in Sect. 4.3. We conclude in Sect. 4.4 by showing that \(\|W\| < 1\) w.h.p. provided the assumptions of Theorem 2.2 are satisfied.

### 4.2 Choice of the Multipliers \(\lambda\) and \(M\)

In this section, we construct multipliers \(\lambda \in \mathbb{R}\) and \(M \in \mathbb{R}^{V \times V}_+\) such that \(Wv = W^Tv = 0\). Note that \([Wv]_i = \sum_{j \in V^*} W_{ij}\) for all \(i \in V\). If \(i \in V - V^*\), we have

\[ [Wv]_i = n_i \lambda - (k - n_i) \left( \frac{n_i}{k - n_i} \right) \lambda = 0 \]
by our choice of $W_{ij}$ in (ω3) and (ω5). By symmetry, $[W^T v]_i = 0$ for all $i \in V - V^*$.

The conditions $W(V^*, V^*) e = W(V^*, V^*)^T e = 0$ define $2k$ equations for the $k^2$ unknown entries of $M$. To obtain a particular solution of this underdetermined system, we parametrize $M$ as $M = ye^T + ey^T$, for some $y \in \mathbb{R}^V$. After this parametrization $\sum_{j \in V^*} W_{ij} = k\tilde{\lambda} - (k - 1 - n_i)\gamma - ky_i - \langle e, y \rangle$. Rearranging shows that $y$ is the solution of the linear system $(kI + e e^T)y = k\tilde{\lambda}e - \gamma((k - 1)e - n)$, where $n \in \mathbb{R}^{V^*}$ is the vector with $i$th entry $n_i$ equal to the degree of node $i$ in $G(V^*)$. By the Sherman-Morrison-Woodbury formula [37, Equation(2.1.4)], we have

$$y = \frac{1}{2k} \left( k\tilde{\lambda} - (k - 1)\gamma \right) e + \frac{\gamma}{k} \left( n - \left( \frac{\langle n, e \rangle}{2k} \right) e \right).$$

and $\mathbb{E}[y] = (k\tilde{\lambda} - (k - 1)\gamma q)e/(2k)$ by the fact that $\mathbb{E}[n] = (k - 1)(1 - q)e$. Taking $\tilde{\lambda} = \gamma(e + q) + 1/k$ yields $\mathbb{E}[y] = (k\epsilon + q)e/(2k)$. Therefore, each entry of $y$ and, consequently, each entry of $M$ is positive in expectation for all $\epsilon > 0$. Therefore, it suffices to show that

$$\|y - \mathbb{E}[y]\|_\infty = \frac{\gamma}{2k} \left\| n - \left( \frac{\langle n, e \rangle}{2k} \right) e - \mathbb{E} \left[ n - \left( \frac{\langle n, e \rangle}{2k} \right) e \right] \right\|_\infty \leq \frac{\gamma}{2k} (k\epsilon + q) \tag{22}$$

w.h.p. to establish that the entries of $M$ are nonnegative w.h.p. since each component $y_i$ is bounded below by $\mathbb{E}[y_i] - \|y - \mathbb{E}[y]\|_\infty$. To do so, we will use the following concentration bound on the sum of independent Bernoulli variables.

**Lemma 4.1** Let $x_1, \ldots, x_m$ be a sequence of $m$ independent Bernoulli trials, each succeeding with probability $p$ and let $s = \sum_{i=1}^m x_i$ be the binomially distributed variable describing the total number of successes. Then, $|s - pm| \leq 6 \max \left\{ \sqrt{p(1 - p)m \log m}, \log m \right\}$ with probability at least $1 - 2m^{-12}$.

Lemma 4.1 is a specialization of the standard Bernstein inequality (see [38, Theorem 6]) to binomially distributed random variables; the proof is left to the reader. We are now ready to state the desired lower bound on the entries of $y$.

**Lemma 4.2** For each $i \in V^*$, we have w.h.p.

$$y_i \geq \gamma \left( \frac{\epsilon}{2} - 12 \max \left\{ \left( q(1 - q) \log k \right)^{1/2}, \frac{\log k}{k} \right\} \right). \tag{23}$$

**Proof** Each entry of $n$ corresponds to $k - 1$ independent Bernoulli trials, each with probability of success $1 - q$. Applying Lemma 4.1 and the union bound shows that

$$|n_i - (1 - q)(k - 1)| \leq 6 \max \left\{ q(1 - q)k \log k \right\} \tag{24}$$
for all $i \in V^*$ w.h.p. On the other hand, $\langle \mathbf{n}, \mathbf{e} \rangle = 2|E(G(V^*))|$ because each $n_i$ is equal to the degree of the node $i$ in $G(V^*)$. Therefore, $\langle \mathbf{n}, \mathbf{e} \rangle$ is a binomially distributed random variable corresponding to $\binom{k}{2}$ independent Bernoulli trials, each with probability of success $1 - q$. As before, Lemma 4.1 implies that

$$|\langle \mathbf{n}, \mathbf{e} \rangle - \mathbb{E}[\langle \mathbf{n}, \mathbf{e} \rangle]| \leq 12 \max \left\{ k(q(1 - q) \log k)^{1/2}, \ 2 \log k \right\}$$

(25)

with high probability. Substituting (24) and (25) into the left-hand side of (22) and applying the triangle inequality shows that

$$\|y - \mathbb{E}[y]\|_\infty \leq \frac{12\gamma}{k} \max \left\{ (q(1 - q)k \log k)^{1/2}, \ \log k \right\}$$

(26)

for sufficiently large $k$ with high probability. Subtracting the right-hand side of (26) from $E[y_i] \geq \gamma \epsilon / 2$ for each $i \in V^*$ completes the proof. \[\square\]

In Sect. 4.3, we will choose $\epsilon = (1 - p - q)/3$ to ensure that $\|F\|_\infty \leq 1$ w.h.p. Substituting this choice of $\epsilon$ in the right-hand side of (23) yields

$$\min_{i \in V^*} y_i \geq \frac{\gamma}{6} \left( (1 - p - q) - \frac{72}{k} \max \left\{ k(q(1 - q) \log k)^{1/2}, \ 2 \log k \right\} \right)$$

w.h.p. Therefore, the entries of the multiplier $M$ are nonnegative w.h.p. if $p, q$ and $k$ satisfy (8).

### 4.3 A Bound on $\|F\|_\infty$

We next establish that $\|F\|_\infty < 1$ w.h.p. under the assumptions of Theorem 2.2. Recall that all diagonal entries, entries corresponding to edges in $G$ and entries indexed by $V^* \times V^*$ of $F$ are chosen to be equal to $0$. It remains to bound $|F_{ij}|$ when $ij \notin E$, and $(i, j) \in (V \times V) - (V^* \times V^*)$.

When $i, j \in V - V^*$ and $ij \notin E$, we choose $F_{ij}$ according to ($\omega_4$): $F_{ij} = -\lambda/(\gamma(1 - p))$. Substituting $\lambda = \gamma(\epsilon + q) + 1/k$, we have $|F_{ij}| \leq 1$ if and only if $1/(\gamma k) + \epsilon + p + q \leq 1$. Taking $\epsilon = (1 - p - q)/3$ and $\gamma \geq 1/(\epsilon k)$ ensures that this is satisfied.

We next consider $i \in V^*$, $j \in V - V^*$ such that $ij \notin E$. The final case, $i \in V - V^*$, $j \in V^*$, $ij \notin E$, follows immediately by symmetry. In this case, we take $F_{ij} = -\lambda k/(\gamma(k - n_j))$ by ($\omega_5$). Clearly, $|F_{ij}| \leq 1$ if and only if $1/(\gamma k) + \epsilon + q + n_j/k \leq 1$. Applying Lemma 4.1 and the union bound over all $j \in V - V^*$ shows that $|n_j - pk| \leq 6 \max \{ (p(1 - p)k \log k)^{1/2}, \ \log k \}$ for all $j \in V - V^*$ w.h.p. Thus,

$$\frac{1}{\gamma k} + \epsilon + q + \frac{n_j}{k} \leq \frac{1}{\gamma k} + \epsilon + q + p + 6 \max \left\{ (p(1 - p)k \log k)^{1/2}, \log k \right\} \leq 1$$

w.h.p., for sufficiently large $k$ and our choice of $\epsilon$ and $\gamma$. Therefore, we have $\|F\|_\infty < 1$ w.h.p.
4.4 A Bound on $\|W\|$  

We complete the proof by establishing that $\|W\|$ is bounded above by a multiple of $\sqrt{N} \log N / k$ w.h.p. for $\gamma$, $\lambda$, and $M$ chosen as in Sects. 4.2 and 4.3. Specifically, we have the following bound on $\|W\|$.

**Lemma 4.3** Suppose that $p$, $q$ and $k$ satisfy (7). Then, with high probability,

$$
\|W\| \leq 24 \gamma \max \left\{ \left( q(1-q)k \log k \right)^{1/2}, \log^2 k \right\} 
+ 36 \lambda \max \left\{ 1, \left( p(1-p)k \right)^{1/2} \right\} \left( \frac{N}{(1-p)^3 k^3} \right)^{1/2} \log N.
$$

Taking $\gamma = O\left( (1-p-q)k^{-1/2} \right)$ and $\lambda = 1/k + \gamma((1-p-q)/3 + q)$ shows that

$$
\|W\| = O\left( \max \left\{ 1, (p(1-p)k)^{1/2} \right\} \left( \frac{N}{(1-p)^3 k^3} \right)^{1/2} \log N \right)
$$

with high probability. Therefore, $\|W\| < 1$ w.h.p. if $p$, $q$ and $k$ satisfy the assumptions of Theorem 2.2 for appropriate choice of constants $c_1$ and $c_3$.

The remainder of this section comprises a proof of Lemma 4.3. We decompose $W$ as $W = Q + R$, where

$$
Q_{ij} = \begin{cases} 
W_{ij}, & \text{if } i, j \in V^*, \\
0, & \text{otherwise},
\end{cases}
\quad R_{ij} = \begin{cases} 
0, & \text{if } i, j \in V^*, \\
W_{ij}, & \text{otherwise}.
\end{cases}
$$

We will bound $\|Q\|$ and $\|R\|$ separately and then apply the triangle inequality to obtain the desired bound on $\|W\|$. To do so, we will make repeated use of the following bound on the norm of a random symmetric matrix with i.i.d. mean zero entries.

**Lemma 4.4** Let $A = [a_{ij}] \in \Sigma_n$ be a random symmetric matrix with i.i.d. mean zero entries $a_{ij}$ with variance $\sigma^2$ and satisfying $|a_{ij}| \leq B$. Then, $\|A\| \leq 6 \max \left\{ \sigma \sqrt{n \log n}, B \log^2 n \right\}$ with probability at least $1 - n^{-8}$.

The proof of Lemma 4.4 follows from an application of the noncommutative Bernstein inequality [39, Theorem 1.4] and is included in the supplemental material.

The following lemma gives the necessary bound on $\|Q\|$.

**Lemma 4.5** The matrix $Q$ satisfies $\|Q\| \leq 24 \gamma \max \left\{ \left( q(1-q)k \log k \right)^{1/2}, \log^2 k \right\}$ with high probability.

**Proof** Note that $\|Q\| = \|Q(V^*, V^*)\|$ by the block structure of $Q$. Let

$$
Q_1 = H(V^*, V^*) - \left( \frac{k-1}{k} \right) qee^T, \quad Q_2 = \frac{1}{k} (ne^T - (1-q)(k-1)ee^T),
$$

$$
Q_3 = Q_2^T, \quad Q_4 = \frac{1}{k} (\langle n, e \rangle - (1-q)(k-1)ee^T).
$$
where $H$ is the adjacency matrix of the complement of $G(V^*)$. Note that $Q(V^*, V^*) = \sum_{i=1}^4 \gamma Q_i$. We will bound each $Q_i$ separately and then apply the triangle inequality to bound $\|Q\|$.

We begin with $\|Q_1\|$. Let $\tilde{H} \in \Sigma V^*$ be the random matrix with off-diagonal entries equal to the corresponding entries of $H$ and whose diagonal entries are independent Bernoulli variables, each with probability of success equal to $q$. Then, $E[\tilde{H}] = qee^T$ and $\tilde{H} - qee^T$ is a random symmetric matrix with i.i.d. mean zero entries with variance equal to $\sigma^2 = q(1 - q)$. Moreover, each entry of $\tilde{H} - qee^T$ has magnitude bounded above by $B = \max\{ q, 1 - q \} \leq 1$. Therefore, applying Lemma 4.4 shows that $\|\tilde{H} - qee^T\| \leq 6 \max\{ \sqrt{q(1-q)k \log k}, \log^2 k \}$ w.h.p. It follows immediately that

$$\|Q_1\| \leq \|\tilde{H} - qee^T\| + \|(q/k)ee^T\| + \|\text{Diag} (\text{diag} \tilde{H})\| \leq 6 \max\{ (q(1-q)k \log k)^{1/2}, \log^2 k \} + q + 1 \quad (27)$$

w.h.p. by the triangle inequality.

We next bound $\|Q_2\|$ and $\|Q_3\|$. By (24), we have

$$\|n - E[n]\|^2 \leq k\|n - E[n]\|_\infty^2 \leq 36 \max\{ q(1-q)k^2 \log k, k \log^2 k \}$$

w.h.p. It follows that

$$\|Q_2\| = \|Q_3\| \leq \frac{1}{k} \|n - E[n]\|\|e\| \leq 6 \max\{ (q(1-q)k \log k)^{1/2}, \log k \} \quad (28)$$

w.h.p. Finally,

$$\|Q_4\| \leq \frac{1}{k^2} |\langle n, e \rangle - E[\langle n, e \rangle]| \|ee^T\| \leq 12 \max\{ (q(1-q) \log k)^{1/2}, 2 \log k/k \} \quad (29)$$

w.h.p., where the last inequality follows from (25). Combining (27), (28) and (29) and applying the union bound, we have $\|Q\| \leq 24\gamma \max\{ (q(1-q)k \log k)^{1/2}, \log^2 k \}$ w.h.p. \(\square\)

The following lemma provides the necessary bound on $\|R\|$.

**Lemma 4.6** Suppose that $p$ and $k$ satisfy (7). Then, with high probability, we have

$$\|R\| \leq 36\lambda \max\left\{ 1, (p(1-k)k)^{1/2} \right\} \left( \frac{N}{(1-p)^3k} \right)^{1/2} \log N.$$ 

*Proof* We decompose $R$ as $R = \lambda(R_1 + R_2 + R_3 + R_4 + R_5)$ as in the proof of Theorem 7 in [5].

We first define $R_1$ by considering the following cases. In case $(\omega_3)$ we take $[R_1]_{ij} = W_{ij}$. In cases $(\omega_4)$, $(\omega_5)$ and $(\omega_6)$ we take $[R_1]_{ij} = -p/(1 - p)$. Finally, for all
\[(i, j) \in V^* \times V^* \text{ we take } [R_1]_{ij} \text{ to be a random variable sampled independently from the distribution}

\[[R_1]_{ij} = \begin{cases} 
1, & \text{with probability } p, \\
-p/(1-p), & \text{with probability } 1-p.
\end{cases}\]

By construction, the entries of \(R_1\) are i.i.d. random variables taking value 1 with probability \(p\) and value \(-p/(1-p)\) otherwise. Applying Lemma 4.4 shows that

\[
\|R_1\| \leq 6 \max \left\{ B \log^2 N, \left( \frac{p}{1-p} \right) N \log N \right\}^{1/2}
\]

(30)

with high probability, where \(B := \max\{1, p/(1-p)\}\).

We next define \(R_2\) to be the correction matrix for the \((V^*, V^*)\) block of \(R\). That is, \([R_2]_{ij} = -[R_1]_{ij}\) for all \(i, j \in V^*\) and \([R_2]_{ij} = 0\) otherwise. Then,

\[
\|R_2\| = \|R_1(V^*, V^*)\| \leq 6 \max \left\{ B \log^2 k, \left( \frac{p}{1-p} \right) k \log k \right\}^{1/2}
\]

with high probability by Lemma 4.4. We define \(R_3\) to be the correction matrix for diagonal entries of \(R_1\), i.e. \(\lambda[R_3]_{ii} = R_{ii} - \lambda[R_1]_{ii}\) for all \(i \in V^*\). By construction, \(R_3\) is a diagonal matrix with diagonal entries taking value either 0 or \(1/(1-p)\). Therefore, \(\|R_3\| \leq 1/(1-p)\).

Finally, we define \(R_4\) and \(R_5\) to be the correction matrices for cases \((\omega_5)\) and \((\omega_6)\), respectively. That is, we take \([R_4]_{ij} = p/(1-p) - n_j/(k-n_j)\) for all \(i \in V^*, j \in V - V^*\) such that \(ij \notin E\) and is equal to 0 otherwise, and take \(R_5 = R_4^T\) by symmetry. Note that

\[
\|R_4\|^2 \leq \|R_4\|_F^2 = \sum_{j \in V - V^*} (k-n_j) \left( \frac{pk-n_j}{(1-p)(k-n_j)} \right)^2
\]

\[
= \sum_{j \in V - V^*} \frac{(n_j - pk)^2}{(1-p)^2(k-n_j)}.\]

By Lemma 4.1, we have \(|n_j - pk| \leq 6 \max\{\sqrt{p(1-p)k \log k}, \log k\}\) with high probability. Therefore,

\[
\|R_4\|^2 \leq \frac{36(N-k) \max\{p(1-p)k \log k, \log^2 k\}}{(1-p)^2 \left( (1-p)k - 6 \max\{\sqrt{p(1-p)k \log k}, \log k\} \right)} \leq \left( \frac{144N}{(1-p)^3k} \right) \max\{p(1-p)k \log k, \log^2 k\}
\]
with high probability, where the last inequality follows from (7), which implies that

\[(1 - p)k - 6 \max\{p(1 - p)k \log k, \log^2 k\} \geq \frac{1}{4}(1 - p)k.\]

Combining the upper bounds on each \(\|R_i\|\) shows that

\[\|R\| \leq 36\lambda \max\left\{1, (p(1 - k)k)^{1/2}\right\} \left(\frac{N}{(1 - p)^3k}\right)^{1/2} \log N\]

with high probability, provided that \(p, k\) and \(N\) satisfy (7). This completes the proof.

The construction of the matrix \(R\) is essentially identical to that of \(W\) in [5, Theorem 7]. In the dense case, i.e. when \(p\) is independent of \(k\) and \(N\), we may apply the proof of [5, Theorem 7] to show that \(\|R\| = O(\sqrt{N}/k)\). This, in turn, suggests that we have exact recovery w.h.p. if \(k = \Omega(\sqrt{N})\) in this case.

5 Experimental Results

In this section, we empirically evaluate the performance of our relaxation for the densest \(k\)-subgraph problem. Specifically, we apply our relaxation (5) to \(N\)-node random graphs sampled from the planted dense \(k\)-subgraph model. For each randomly generated program input, we apply the alternating direction method of multipliers (ADMM) to solve (5). ADMM has recently gained popularity as an algorithmic framework for distributed convex optimization, in part, due to its being well suited to large-scale problems arising in machine learning and statistics. A full overview of ADMM and related methods is beyond the scope of this paper; we direct the reader to the recent survey [40] and the references within for more details. We may also solve (5) by reformulating as a semidefinite program and solve using interior point methods when the graph \(G\) is small. However, the memory requirements needed to formulate and solve the Newton system are prohibitive for graphs containing more than a few hundred nodes.

A specialization of ADMM to our problem is given as Algorithm 1; specifically, Algorithm 1 is a modification of the ADMM algorithm for robust PCA given by [41, Example 3]. We iteratively solve the linearly constrained optimization problem

\[
\min \|X\|_* + \gamma\|Y\|_1 + 1_{\Omega_Q}(Q) + 1_{\Omega_W}(W) + 1_{\Omega_Z}(Z)
\]

s.t. \(X + Y = Q\), \(X = W\), \(X = Z\),

where \(\Omega_Q\), \(\Omega_W\) and \(\Omega_Z\) are the sets \(\Omega_Q := \{Q \in \mathbb{R}^{V \times V} : P_{\tilde{E}}(Q) = 0\}\), \(\Omega_W := \{W \in \mathbb{R}^{V \times V} : e^T We = k^2\}\) and \(\Omega_Z := \{Z \in \mathbb{R}^{V \times V} : Z_{ij} \leq 1 \ \forall (i, j) \in V \times V\}\). Here \(1_S : \mathbb{R}^{V \times V} \to \{0, +\infty\}\) is the indicator function of the set \(S \subseteq \mathbb{R}^{V \times V}\), defined by \(1_S(X) = 0\) if \(X \in S\), and \(+\infty\) otherwise. During each iteration, we sequentially update each primal decision variable by minimizing the augmented Lagrangian
\[ L_\tau = \|X\|_* + \gamma \|Y\|_1 + + 1_{\Omega_Q} (Q) + 1_{\Omega_W} (W) + 1_{\Omega_Z} (Z) \]
\[ + \text{Tr} (\lambda_Q (X + Y - Q)) + \text{Tr} (A_W (X - W)) + \text{Tr} (A_Z (X - Z)) \]
\[ + \frac{\tau}{2} (\|X + Y - Q\|^2 + \|X - W\|^2 + \|X - Z\|^2) \]

in Gauss–Seidel fashion with respect to each primal variable and then updating the dual variables \(\lambda_Q, \lambda_W, A_Z\) using the updated primal variables. Here \(\tau\) is a regularization parameter chosen so that \(L_\tau\) is strongly convex in each primal variable. Minimizing the augmented Lagrangian with respect to each of the artificial primal variables \(Q, W, Z\) is equivalent to projecting onto each of the sets \(\Omega_Q, \Omega_W, \Omega_Z\), respectively; each of these projections can be performed analytically. On the other hand, the subproblems for updating \(X\) and \(Y\) in each iteration allow closed-form solutions via the elementwise soft thresholding operator \(S_\phi : \mathbb{R}^n \to \mathbb{R}^n\) defined by

\[
[S_\phi(x)]_i = \begin{cases} 
  x_i - \phi, & \text{if } x_i > \phi \\
  0, & \text{if } -\phi \leq x_i \leq \phi \\
  x_i + \phi, & \text{if } x_i < -\phi.
\end{cases}
\]

It has recently been shown that ADMM converges linearly when applied to the minimization of convex separable functions, under mild assumptions on the program input (see [42]), and, as such, Algorithm 1 can be expected to converge to the optimal solution of (5); we stop Algorithm 1 when the primal and dual residuals \(\|X^{(\ell)} - \tilde{W}^{(\ell)}\|_F, \|X^k - Z^{(\ell)}\|_F, \|W^{(\ell+1)} - \tilde{W}^{(\ell)}\|_F, \|Z^{(\ell+1)} - \tilde{Z}^{(\ell)}\|_F, \text{ and } \|A_Q^{(\ell+1)} - \tilde{A}_Q^{(\ell)}\|_F\) are smaller than a desired error tolerance.

We evaluate the performance of our algorithm by sampling random \(N\)-node graphs from the planted dense \(k\)-subgraph model for \(q = 0.25\) and various clique sizes \(k \in (0, N)\) and edge addition probabilities \(p \in [0, 1 - q]\). Each graph \(G\) is represented by a random symmetric binary matrix \(A\) with entries in the \((1 : k) \times (1 : k)\) block set equal to 1 with probability \(1 - q = 0.75\) independently and remaining entries set independently equal to 1 with probability \(p\). For each graph \(G\), we call Algorithm 1 to obtain solution \((X^*, Y^*)\); regularization parameter \(\gamma = 4/(1 - p - q)k\), augmented Lagrangian parameter \(\tau = 0.35\), and stopping tolerance \(\varepsilon = 10^{-4}\) are used in each call to Algorithm 1. We declare the planted dense \(k\)-subgraph to be recovered if \(\|X^* - X_0\|_F/\|X_0\|_F < 10^{-3}\), where \(X_0 = vv^T\) and \(v\) is the characteristic vector of the planted \(k\)-subgraph. The experiment was repeated 10 times for each value of \(p\) and \(k\) for \(N = 250\) and \(N = 500\). The empirical probability of recovery of the planted \(k\)-clique is plotted in Fig. 1, where a higher rate of recovery is represented by brighter colours. The graphs of the functions

\[
f(p, q, N) = \frac{\sqrt{pN} \log N}{4(1 - p - q)(1 - p)}, \quad g(p, q, N) = \left( \frac{\sqrt{N} \log N}{4(1 - p - q)(1 - p)^{3/2}} \right)^{2/3},
\]

are plotted as the solid and dashed lines, respectively, and approximate the theoretical thresholds for exact recovery given by (9) (with the estimate of the scaling constant \(c_1 \approx 1/4\)); we should expect perfect recovery for all \(k\) to the right of both curves.

\(\odot\) Springer
Algorithm 1 ADMM for solving (5)

Input: $G = (V, E)$, $k \in \{1, \ldots, N\}$, where $N = |V|$, and error tolerance $\epsilon$.
Initialize: $X^{(0)} = W^{(0)} = (k/N)^2 ee^T$, $Y^{(0)} = -X$, $Q^{(0)} = A^{(0)} = A_W^{(0)} = A_Z^{(0)} = 0$.

for $i = 0, 1, \ldots$, until converged do

1. Update $Q^{(i+1)}$: $Q^{(i+1)} = P_{\tilde{E}} \left( X^{(i)} + Y^{(i)} - A_Q^{(i)} \right)$.

2. Update $X^{(i+1)}$: Take SVD $\tilde{X}^{(i)} = Q^{(i+1)} + 2X^{(i)} - W^{(i)} - A_W^{(i)} = U (\text{Diag} \ x) V^T$.

   Apply soft thresholding: $X^{(i+1)} = U (\text{Diag} S_{\tau} (x)) V^T$.

3. Update $Y^{(i+1)}$: $Y^{(i+1)} = S_{\tau\gamma} \left( Y^{(i)} - \tau Q^{(i+1)} \right)$.

4. Update $W^{(i+1)}$: Let $\tilde{W}^{(i)} = X^{(i+1)} - A_W^{(i)}$ and $\beta^{(i)} = \left( k^2 - e^T \tilde{W}^{(i)} e \right) / N^2$.

   Update $W^{(i+1)} = \tilde{W}^{(i)} + \beta^{(i)} ee^T$.

5. Update $Z^{(i+1)}$: Let $\tilde{Z}^{(i)} = X^{(i+1)} - A_Z^{(i)}$. For each $i, j \in V$: $Z_{ij}^{(i+1)} = \min[\max[\tilde{Z}_{ij}^{(i)}], 0, 1]$.

6. Update dual variables:

   $A_Z^{(i+1)} = \frac{1}{2} \left( X^{(i+1)} - Z^{(i+1)} \right)$, $A_W^{(i+1)} = \frac{1}{2} \left( X^{(i+1)} - W^{(i+1)} \right)$, and

   $A_P^{(i+1)} = P_{V \times V - \tilde{E}} \left( A_P^{(i)} - \frac{1}{2} \left( X^{(i+1)} + Y^{(i+1)} \right) \right)$.

7. Check convergence:

   $r_p = \max[\|X^{(i)} - W^{(i)}\|_F, \|X^k - Z^{(i)}\|_F]$,

   $r_d = \max[\|W^{(i+1)} - W^{(i)}\|_F, \|Z^{(i+1)} - Z^{(i)}\|_F, \|A_Q^{(i+1)} - A_Q^{(i)}\|]$.

if $\max[r_p, r_d] < \epsilon$ then stop: algorithm converged.

---

Fig. 1 Empirical recovery for $N$-node graphs with planted dense $k$-subgraph, a $N = 250$, b $N = 500$

observed performance of our heuristic closely matches that predicted by Theorem 2.2, with sharp transition to perfect recovery as $k$ increases past a threshold depending on $p$ and $N$. However, our simulation results suggest that the constants governing exact recovery in Theorem 2.2 may be overly conservative; we have perfect recovery for smaller choices of $k$ than those predicted by Theorem 2.2 for almost all choices of $p$. 
We repeated these experiments for bipartite graphs. Specifically, we sampled random \((M, N)\)-node bipartite graphs from the planted dense \((k_1, k_2)\)-subgraph model with \(q = 0.25\) for a variety of biclique sizes \((k_1, k_2)\) and \(p\). We call Algorithm 1 (with minor modifications to address the lack of symmetry in \(G\)) to solve (11) for each graph \(G\); in particular, we initialize \(X^{(0)} = W^{(0)} = \frac{k_1 k_2}{MN}\), update the auxiliary variable \(\beta\) by \(\beta^{(\ell)} = \frac{(k_1 k_2 - e^T \tilde{W}^{(\ell)} e)}{MN}\), and leave the rest of the algorithm unaltered. We use \(\gamma = \frac{4}{((1 - p - q)k_1 k_2)}, \tau = 0.35,\) and \(\epsilon = 10^{-4}\) in each trial. The obtained solutions were compared to the planted solutions as before to obtain the empirical probability of recovery of the planted \((k_1, k_2)\)-biclique over 10 trials for each choice of \(p\) and \((k_1, k_2)\). Figure 2 plots the empirical recovery rates over 10 repetitions for each choice of \(p\) and \(k_1\), as well as the theoretical recovery thresholds

\[
 f_1(p, q, N) = \frac{\sqrt{pN} \log N}{10(1-p-q)(1-p)} \quad \text{and} \quad g_1(p, q, N) = \left( \frac{\sqrt{N} \log N}{10(1-p-q)(1-p)^{3/2}} \right)^{2/3},
\]

with estimate of scaling factor \(c_1 \approx 10\). As before, we observe a sharp transition to perfect recovery as \(k_1\) increases past some threshold depending on \(p, M\) and \(N\). Again, it seems that the predicted threshold may be overly conservative when compared to that observed empirically.

6 Perspectives

These results suggest several possible avenues for future research. First, although our recovery guarantees match those previously identified in the literature, these bounds may not be the best possible. Rohe et al. [43] recently established that an \(N\)-node
random graph sampled from the stochastic blockmodel can be partitioned into dense subgraphs of size $\Omega(\log^4 N)$ using a regularized maximum likelihood estimator. It is unclear whether such a bound can be attained for our relaxation. It would also be interesting to see if similar recovery guarantees exist for more general graph models; for example, can we find the largest planted clique in a graph with several planted cliques of varying sizes? Other potential areas of future research may also involve post-processing schemes for identifying the densest $k$-subgraph in the case that the optimal solution of our relaxation does not exactly correspond to the sum of a low-rank and sparse matrix, and if a similar relaxation approach and analysis may lead to stronger recovery results for other intractable combinatorial problems, such as the planted $k$-disjoint-clique [44] and clustering [45] problems.

7 Conclusions

We have considered a convex optimization heuristic for identifying the densest $k$-node subgraph of a given graph, with novel recovery properties. In particular, we have identified trade-offs between the size and density of a planted subgraph ensuring that this subgraph can be recovered from the unique optimal solution of the convex program (5). Moreover, we establish analogous results for the identification of the densest bipartite $(k_1, k_2)$-subgraph in a bipartite graph. In each case, the relaxation relies on the decomposition of the adjacency matrices of candidate subgraphs as the sum of a dense and sparse matrix, and is closely related to recent results regarding robust principal component analysis.

Acknowledgments This research was supported in part by the Institute for Mathematics and its Applications with funds provided by the National Science Foundation. We are also grateful to Inderjit Dhillon, Stephen Vavasis and Teng Zhang for their helpful comments and suggestions, and to Shiqian Ma for his insight and help implementing the ADMM algorithm used in Sect. 5.

References

1. Feige, U., Peleg, D., Kortsarz, G.: The dense $k$-subgraph problem. Algorithmica 29(3), 410–421 (2001)
2. Feige, U.: Relations between average case complexity and approximation complexity. In: Proceedings of the thirty-fourth annual ACM symposium on Theory of computing, pp. 534–543. ACM (2002)
3. Khot, S.: Ruling out PTAS for graph min-bisection, densest subgraph and bipartite clique. SIAM J. Comput. 36(4), 1025–1071 (2006)
4. Alon, N., Arora, S., Manokaran, R., Moshkovitz, D., Weinstein, O.: Inapproximability of densest $\kappa$-subgraph from average-case hardness (2011)
5. Ames, B., Vavasis, S.: Nuclear norm minimization for the planted clique and biclique problems. Math. Program. 129(1), 1–21 (2011)
6. Chandrasekaran, V., Sanghavi, S., Parrilo, P.A., Willsky, A.S.: Rank-sparsity incoherence for matrix decomposition. SIAM J. Optim. 21(2), 572–596 (2011)
7. Candès, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? J. ACM (JACM) 58(3), 11 (2011)
8. Chen, Y., Jalali, A., Sanghavi, S., Caramanis, C.: Low-rank matrix recovery from errors and erasures. IEEE Trans. Inf. Theory 59(7), 4324–4337 (2013)
9. Oymak, S., Hassibi, B.: Finding dense clusters via “low rank + sparse” decomposition. Arxiv preprint arXiv:1104.5186 (2011)
10. Chen, Y., Jalali, A., Sanghavi, S., Xu, H.: Clustering partially observed graphs via convex optimization. J. Mach. Learn. Res. 15(1), 2213–2238 (2014)
11. Chen, Y., Sanghavi, S., Xu, H.: Clustering sparse graphs. In: Advances in neural information processing systems, pp. 2204–2212 (2012)
12. Lawler, E.L.: Combinatorial optimization: networks and matroids. Courier Corporation (1976)
13. Karp, R.: Reducibility among combinatorial problems. Complex. Comput. Comput. 40(4), 85–103 (1972)
14. Doan, X.V., Vavasis, S.: Finding approximately rank-one submatrices with the nuclear norm and $\ell_1$-norm. SIAM J. Optim. 23(4), 2502–2540 (2013)
15. Recht, B., Fazel, M., Parrilo, P.A.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev. 52(3), 471–501 (2010)
16. Gilbert, A.C., Guha, S., Indyk, P., Muthukrishnan, S., Strauss, M.: Near-optimal sparse fourier representations via sampling. In: STOC ’02: Proceedings of the thirty-fourth annual ACM symposium on Theory of computing, pp. 152–161. ACM, New York, NY, USA (2002). doi:10.1145/509907.509933
17. Donoho, D.: Compressed sensing. IEEE Trans. Inf. Theory 52(4), 1289–1306 (2006)
18. Candès, E., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inf. Theory 52(2) (2006)
19. Ames, B.: Robust convex relaxation for the planted clique and densest $k$-subgraph problems: additional proofs (2013). Available from http://bpames.people.ua.edu/uploads/3/9/0/0/39000767/dks_appendices.pdf
20. Kučera, L.: Expected complexity of graph partitioning problems. Discret. Appl. Math. 57(2), 193–212 (1995)
21. Alon, N., Krivelevich, M., Sudakov, B.: Finding a large hidden clique in a random graph. In: Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms, pp. 594–598. Society for Industrial and Applied Mathematics (1998)
22. Feige, U., Krauthgamer, R.: Finding and certifying a large hidden clique in a semirandom graph. Random Struct. Algorithms 16(2), 195–208 (2000)
23. McSherry, F.: Spectral partitioning of random graphs. In: Proceedings of the 42nd IEEE symposium on Foundations of Computer Science, pp. 529–537. IEEE Computer Society (2001)
24. Feige, U., Ron, D.: Finding hidden cliques in linear time. DMTCS Proc. 01, 189–204 (2010)
25. Dekel, Y., Gurel-Gurevich, O., Peres, Y.: Finding hidden cliques in linear time with high probability. Comb. Probab. Comput. 23(01), 29–49 (2014)
26. Deshpande, Y., Montanari, A.: Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. Found. Comput. Math. pp. 1–60 (2013)
27. Juels, A., Peinado, M.: Hiding cliques for cryptographic security. Des. Codes Crypt. 20(3), 269–280 (2000)
28. Alon, N., Andoni, A., Kaufman, T., Matulef, K., Rubinfeld, R., Xie, N.: Testing k-wise and almost k-wise independence. In: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing, pp. 496–505. ACM (2007)
29. Hazan, E., Krauthgamer, R.: How hard is it to approximate the best nash equilibrium? SIAM J. Comput. 40(1), 79–91 (2011)
30. Berthet, Q., Rigollet, P.: Computational lower bounds for sparse PCA. arXiv preprint arXiv:1304.0828 (2013)
31. Jerrum, M.: Large cliques elude the metropolis process. Random Struct. Algorithms 3(4), 347–359 (1992)
32. Feige, U., Krauthgamer, R.: The probable value of the lovász-schrijver relaxations for maximum independent set. SIAM J. Comput. 32(2), 345–370 (2003)
33. Nadakuditi, R.: On hard limits of eigen-analysis based planted clique detection. In: Statistical Signal Processing Workshop (SSP), 2012 IEEE, pp. 129–132. IEEE (2012)
34. Peeters, R.: The maximum edge biclique problem is NP-complete. Discret. Appl. Math. 131(3), 651–654 (2003)
35. Goerdt, A., Lanka, A.: An approximation hardness result for bipartite clique. In: Electronic Colloquium on Computational Complexity, Report, 48 (2004)
36. Boyd, S., Vandenberghe, L.: Convex optimization. Cambridge University Press, Cambridge (2004)
37. Golub, G., Van Loan, C.: Matrix computations. Johns Hopkins University Press, Baltimore (1996)
38. Lugosi, G.: Concentration-measure inequalities (2009). Available from http://www.econ.upf.edu/~lugosi/anu.pdf
39. Tropp, J.: User-friendly tail bounds for sums of random matrices. Foundations of Computational Mathematics pp. 1–46 (2011)
40. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. 3(1), 1–122 (2011)
41. Goldfarb, D., Ma, S., Scheinberg, K.: Fast alternating linearization methods for minimizing the sum of two convex functions. Math. Program. pp. 1–34 (2010)
42. Hong, M., Luo, Z.: On the linear convergence of the alternating direction method of multipliers. arXiv preprint arXiv:1208.3922 (2012)
43. Rohe, K., Qin, T., Fan, H.: The highest dimensional stochastic blockmodel with a regularized estimator. arXiv preprint arXiv:1206.2380 (2012)
44. Ames, B., Vavasis, S.: Convex optimization for the planted k-disjoint-clique problem. Math. Program. 143(1–2), 299–337 (2014)
45. Ames, B.: Guaranteed clustering and biclustering via semidefinite programming. Math. Program. 147(1–2), 429–465 (2014)