# Robust convex relaxation for the planted clique and densest $k$-subgraph problems

Brendan P.W. Ames [*]

June 22, 2013

### 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 and $\ell_1$ norms. 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 consider two constructions for this noise. In the first, noise is introduced by an adversary deterministically deleting edges within the planted clique and placing diversionary edges. In the second, these edge corruptions are performed at random. Analogous recovery guarantees for identifying the densest subgraph of fixed size in a bipartite graph are also established, and results of numerical simulations for randomly generated graphs are included to demonstrate the efficacy of our algorithm.

## 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 plays a significant role 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

---

[*]Institute for Mathematics and its Applications, College of Science and Engineering, University of Minnesota, 207 Church Street SE, 400 Lind Hall, Minneapolis, Minnesota, 55455, U.S.A., bpames@ima.umn.edu

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 [23] and hard to approximate [20, 34, 2].

Our results can be thought of as a generalization to the densest $k$-subgraph problem of those in [7] for the maximum clique problem. In [7], Ames and Vavasis establish that the maximum clique of a given graph can be recovered from the optimal solution of a particular convex program for certain classes of input graphs. Specifically, 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 [7] 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 [14, 13, 15] and graph clustering [40, 30, 16]. 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 [16]. In [16], 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 [7].

The remainder of the paper is structured as follows. We present our relaxation for the densest $k$-subgraph problem and state our theoretical recovery guarantees in Section 2. In particular, we will show that the densest $k$-subgraph can be recovered from the optimal solution of our convex relaxation in the case that the input graph $G = (V, E)$ consists of a planted $k$-clique $V^*$ that has been corrupted by the noise in the form of diversionary edge

2

additions and deletions, as well as diversionary nodes. We consider two cases. In the first, noise is introduced deterministically by an adversary adding diversionary edges and deleting edges between nodes within the planted clique. In the second, these edge deletions and additions are performed at random. We present an analogous relaxation for identifying the densest bipartite subgraph of given size in a bipartite graph in Section 3. A proof of the recovery guarantee for the densest $k$-subgraph problem in the random noise case comprises Section 4; the proofs of the remaining theoretical guarantees are similar and are omitted. We conclude with simulation results for synthetic data sets in Section 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:

$$\max\{d(H) : H \subseteq G, |V(H)| = k\}. \tag{2.1}$$

Although the problem of finding a subgraph with maximum average degree is polynomially solvable [36, 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 [33], is equivalent to an instance of the densest $k$-subgraph problem. Moreover, the densest $k$-subgraph problem is hard to approximate; specifically, it has been shown that the densest $k$-subgraph problem does not admit a polynomial-time approximation scheme under various complexity theoretic assumptions [20, 34, 2]. Due to, and in spite of, this intractability of the densest $k$-subgraph problem, we consider relaxation of (2.1) 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 $\bar{\mathbf{v}}$ be its characteristic vector. That is, for all $i \in V$, $\bar{\mathbf{v}}_i = 1$ if $i \in V'$ and is equal to 0 otherwise. The vector $\bar{\mathbf{v}}$ defines a rank-one matrix $\bar{X}$ by the outer product of $\bar{\mathbf{v}}$ with itself: $\bar{X} = \bar{\mathbf{v}}\bar{\mathbf{v}}^T$. Moreover, if $V'$ is a clique of $G$ then the nonzero entries of $\bar{X}$ correspond to the $k \times k$ all-ones

block of the perturbed adjacency matrix $\tilde{A}_G := A_G + I$ of $G$ indexed by $V' \times V'$. 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 $\bar{Y} \in \mathbf{R}^{V \times V}$ be the matrix defined by

$$\bar{Y}_{ij} = \begin{cases} -1, & \text{if } ij \in \tilde{E} \\ 0, & \text{otherwise,} \end{cases} \tag{2.2}$$

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, $\bar{Y} = -P_{\tilde{E}}(X)$, where $P_{\tilde{E}}$ is the orthogonal projection onto the set of matrices with support contained $\tilde{E}$ defined by

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

for all $M \in \mathbf{R}^{V \times V}$. The matrix $\bar{Y}$ can be thought of as a "correction" for the entries of $\bar{X}$ indexed by nonedges of $G$; $\bar{X} + \bar{Y}$ is exactly the adjacency matrix of the subgraph of $G$ induced by $V'$ (including loops). Moreover, the density of $G(V')$ is equal to

$$d(G(V')) = \frac{1}{2k}\big(k(k-1) - \|\bar{Y}\|_0\big),$$

by the fact that the number of nonzero entries in $\bar{Y}$ is exactly twice the number of nonadjacent pairs of nodes in $G(V')$, Here $\|\bar{Y}\|_0$ denotes the so-called $\ell_0$ *norm* of $\bar{Y}$, 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, (2.1) is equivalent to

$$\min_{X, Y \in \Sigma^V} \left\{ \|Y\|_0 : \text{rank}\,(X) = 1, \ \mathbf{e}^T X \mathbf{e} = k^2, \ X_{ij} + Y_{ij} = 0 \ \forall\, ij \in \tilde{E}, \ X \in \{0,1\}^{V \times V} \right\}, \tag{2.3}$$

where $\mathbf{e}$ is the all-ones vector in $\mathbf{R}^V$, $\Sigma^V$ denotes the cone of $|V| \times |V|$ symmetric matrices with rows and columns indexed by $V$. Indeed, the constraints $\text{rank}\,(X) = 1$, $\mathbf{e}^T X \mathbf{e} = k^2$, 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

$$\min_{X, Y \in \Sigma^V} \left\{ \text{rank}\,(X) + \gamma \|Y\|_0 : \ \mathbf{e}^T X \mathbf{e} = k^2, \ X_{ij} + Y_{ij} = 0 \ \forall\, ij \in \tilde{E}, \ X \in \{0,1\}^{V \times V} \right\}. \tag{2.4}$$

Here $\gamma > 0$ is a regularization parameter to be chosen later. We relax (2.4) to the convex

4

problem

$$\min \left\{ \|X\|_* + \gamma \|Y\|_1 : \ \mathbf{e}^T X \mathbf{e} = k^2, \ X_{ij} + Y_{ij} = 0 \ \forall \, ij \in \tilde{E}, \ X \in [0,1]^{V \times V} \right\} \qquad (2.5)$$

by replacing 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 ignoring the symmetry constraints on $X$ and $Y$. 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 (2.2), although this equality clearly does not hold in general. Our relaxation mirrors that proposed by Chandrasekaran et al. [14] for robust principal component analysis. Given matrix $M \in \mathbf{R}^{m \times n}$, the Robust PCA problem seeks a decomposition of the form $M = L + S$ where $L \in \mathbf{R}^{m \times n}$ has low rank and $S \in \mathbf{R}^{m \times n}$ is sparse. In [14], Chandrasekaran et al. establish 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 $M$. Several recent papers [13, 19, 15, 40, 30, 16] have extended this result to obtain conditions on the input matrix $M$ ensuring perfect decomposition under partial observation of $M$ and other linear constraints. Although these conditions do not translate immediately to our formulation for the densest $k$-subgraph problem and its relaxation, we will establish analogous exact recovery guarantees.

We consider a planted case analysis of (2.5). Suppose that the input graph $G$ contains a single dense subgraph $H$, plus diversionary edges and nodes. We are interested in the tradeoff 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 (2.5). 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 [7]. In [7], Ames and Vavasis provide conditions ensuring exact recovery of a planted clique from the optimal solution of the convex program

$$\min_X \left\{ \|X\|_* : \mathbf{e}^T X \mathbf{e} \geq k^2, \ X_{ij} = 0 \ \forall \, ij \in \tilde{E} \right\}. \qquad (2.6)$$

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 [7, Section 4.1].

**Theorem 2.1** *Let $V^*$ be a $k$-subset of nodes of the graph $G = (V, E)$ and let $\mathbf{v}$ be its characteristic vector. Suppose that $G$ contains at most $r$ edges not in $G(V^*)$ and $G(V^*)$ contains at least $\binom{k}{2} - s$ 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 (2.5) where $X^* = \mathbf{v}\mathbf{v}^T$ and $Y^*$ is constructed according to (2.2). 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 (2.5) for $\gamma = 2\big((1 - 2\delta_1 - \delta_2)k\big)^{-1}$.*

The bound on the number of adversarially added edges given by Theorem 2.1 matches that given in [7, Section 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 $V^*$ woud 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.

($\rho_1$) Fix subset $V^* \subseteq V$ of size $K$. Add $ij$ to $E$ independently with fixed probability $1 - q$ for all $(i, j) \in V^* \times V^*$.

($\rho_2$) 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 $p + q < 1$. We wish to determine which choices of $p, q$ and $k$ yield $G$ such that the planted dense $k$-subgraph $G(V^*)$ can be recovered from the optimal solution of (2.5). Note that $V^*$ is a $k$-clique of $G$ if $q = 0$. Theorem 7 of [7] states that a planted $k$-clique can be recovered from the optimal solution of (2.6) 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*

$$(1 - p)k \geq \max\{8p, 1\} \cdot 8 \log k, \quad \sqrt{pN} \geq (1 - p) \log N \tag{2.7}$$

$$(1 - p - q)k \geq 72 \max\left\{ (q(1 - q)k \log k)^{1/2}, \log k \right\} \tag{2.8}$$

6

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

$$(1 - p - q)(1 - p)k > c_1 \max\left\{ p^{1/2}, \left((1-p)k\right)^{-1/2} \right\} \cdot N^{1/2} \log N \qquad (2.9)$$

*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 (2.5) with high probability for*

$$\gamma \in \left( \frac{c_2}{(1 - p - q)k}, \frac{c_3}{(1 - p - q)k} \right). \qquad (2.10)$$

Here, and in the rest of the paper, an event is said to occur *with high probability (w.h.p.)* if it occurs with probability tending polynomially to 1 as $k$ (or $\min\{k_1, k_2\}$ in Section 3) approaches $+\infty$.

Suppose that the graph $G$ constructed according to $(\rho_1)$ and $(\rho_2)$ 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 (2.5) w.h.p. provided that $k$ is at least as large as $\Omega(\sqrt{N} \log n)$. This lower bound matches that of [7, Theorem 7], as well as [35, 3, 21, 38, 24, 17, 18], 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 [7, Theorem 7] shows that the planted dense $k$-subgraph can be recovered w.h.p. provided $k = \Omega(\sqrt{N})$ in the dense case. 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 difficult (and this presumed difficulty has been exploited in cryptographic applications [32] and complexity analysis [1, 28, 2, 8]). Moreover, a number of algorithmic approaches [31, 22, 39] 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 $N \to \infty$) this 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 Theorem 2.1 and Theorem 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 $p + q \le 1/2$, then we may take $\gamma = 6/k$.

# 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 $V$. We say that 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 [41], and hard to approximate [20, 25].

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 $X' = \mathbf{u}\mathbf{v}^T$, where $\mathbf{u}$ and $\mathbf{v}$ 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:

$$\min \left\{ \|Y\|_0 : \operatorname{rank} X = 1, \ \mathbf{e}^T X \mathbf{e} = k_1 k_2, \ X_{ij} + Y_{ij} = 0, \ \forall ij \in \tilde{E}, \ X \in \{0, 1\}^{U \times V} \right\}.$$

This problem is identical to (2.3) but for a slightly different definition of the set $\tilde{E}$, a different right-hand side in the sum constraint, and omission of symmetry constraints. As before, 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:

$$\min \left\{ \|X\|_* + \gamma \|Y\|_1 : \mathbf{e}^T X \mathbf{e} = k_1 k_2, \ X_{ij} + Y_{ij} = 0, \ \forall ij \in \tilde{E}, \ X \in [0, 1]^{U \times V} \right\}. \quad (3.1)$$

Again, except for superficial differences, this problem is identical to the convex relaxation of (2.3) given by (2.5). As can be expected, the recovery guarantees for the relaxation of the densest $k$-subgraph problem translate to similar guarantees for the convex relaxation (3.1) of the densest $(k_1, k_2)$-subgraph problem.

As in Section 2, we consider the performance of the relaxation (3.1) 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 (3.1).

**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 $\mathbf{u}$ and $\mathbf{v}$ denote the characteristic vectors of $U^*$ and $V^*$, and let $(X^*, Y^*) = (\mathbf{uv}^T, -P_{\tilde{E}}(\mathbf{uv}^T))$. Suppose that $G(U^*, V^*)$ contains at least $k_1 k_2 - s$ 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 $\beta_1 k_1$ nodes in $U^*$ and each node in $U - U^*$ is adjacent to at most $\beta_2 k_2$ 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 $r \leq c_1 k_1 k_2$ and $s \leq c_2 k_1 k_2$ 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 (3.1) for $\gamma = 2\left(\sqrt{k_1 k_2}\left(1 - \alpha_1 - \alpha_2 - \max\{\beta_1, \beta_2\}\right)\right)^{-1}$.*

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 [7, Section 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:

$(\psi_1)$ 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 $1 - q$.

$(\psi_2)$ Then each remaining possible edge is added independently to $E$ with probability $p$.

By construction $G(U^*, V^*)$ is dense in expectation, relative to its complement, if $p + q < 1$. Theorem 9 of [7] asserts that $G(U^*, V^*)$ is the densest $(k_1, k_2)$-subgraph of $G$ and can be recovered using a modification of (3.1), for sufficiently large $k_1$ and $k_2$, in the special case that $q = 0$. The following theorem 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 $(\Psi_1)$ and $(\Psi_2)$ such that $p + q < 1$ and*

$$(1 - p)k_i \geq \max\{8, 64p\} \log k_i$$

$$pN_i \geq (1 - p)^2 \log^2 N_i$$

$$(1 - p - q)k_i \geq 72 \max \left\{ \log k_i, \left(q(1 - q)k_i \log k_i\right)^{1/2} \right\}$$

*for $i = 1, 2$. Then there exist absolute constants $c_1, c_2, c_3 > 0$ such that if*

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

*where $\bar{N} = \max\{N_1, N_2\}$, then $G(U^*, V^*)$ is the densest $(k_1, k_2)$-subgraph of $G$ and $(X^*, Y^*)$ is the unique optimal solution of (3.1) for*

$$\gamma \in \left( \frac{c_2}{(1 - \alpha_1 - \alpha_2 - \max\{\beta_1, \beta_2\})\sqrt{k_1 k_2}}, \frac{c_3}{(1 - \alpha_1 - \alpha_2 - \max\{\beta_1, \beta_2\})\sqrt{k_1 k_2}} \right)$$

# 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 [5]. 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 $\mathbf{v}$ be its characteristic vector. Our goal is to show that the solution $(X^*, Y^*) := (\mathbf{v}\mathbf{v}^T, -P_{\tilde{E}}(\mathbf{v}\mathbf{v}^T))$, where $\tilde{E} := (V \times V) - (E \cup \{uu : u \in V\})$, is the unique optimal solution of (2.5) 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 (2.5) corresponding to a $k$-subgraph of $G$ and then establish that these sufficient conditions are satisfied at $(X^*, Y^*)$ with high probability if the assumptions of Theorem 2.2 are satisfied.

## 4.1 Optimality conditions

We will show that $(X^*, Y^*)$ is optimal for (2.5) and, consequently, $G(V^*)$ is the densest $k$-subgraph of $G$ by establishing that $(X^*, Y^*)$ satisfy the sufficient conditions for optimality given by the Karush-Kuhn-Tucker theorem ([11, Section 5.5.3]). The following theorem

provides the necessary specialization of these conditions to (2.5). A proof of Theorem 4.1 can be found in Appendix A.1.

**Theorem 4.1** *Let $G = (V, E)$ be a graph sampled from the planted dense $k$-subgraph model. Let $\bar{V}$ be a subset of $V$ of cardinality $k$ and let $\bar{\mathbf{v}}$ be the characteristic vector of $\bar{V}$. Let $\bar{X} = \bar{\mathbf{v}}\bar{\mathbf{v}}^T$ and let $\bar{Y}$ be defined as in (2.2). Suppose that there exist $F, W \in \mathbf{R}^{V \times V}$, $\lambda \in \mathbf{R}_+$, and $M \in \mathbf{R}_+^{V \times V}$ such that*

$$\bar{X}/k + W - \lambda \mathbf{e}\mathbf{e}^T - \gamma(\bar{Y} + F) + M = 0, \tag{4.1}$$

$$W\bar{\mathbf{v}} = W^T\bar{\mathbf{v}} = 0, \ \|W\| \le 1, \tag{4.2}$$

$$P_\Omega(F) = 0, \quad \|F\|_\infty \le 1, \tag{4.3}$$

$$F_{ij} = 0 \ \text{for all } (i, j) \in E \cup \{vv : v \in V\}, \tag{4.4}$$

$$M_{ij} = 0 \ \text{for all } (i, j) \in (V \times V) - (\bar{V} \times \bar{V}). \tag{4.5}$$

*Then $(\bar{X}, \bar{Y})$ is an optimal solution of (2.5) 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 (2.5) and $G(\bar{V})$ is the unique maximum density $k$-subgraph of $G$.*

It remains to show that multipliers $F, W \in \mathbf{R}^{V \times V}$, $\lambda \in \mathbf{R}_+$, and $M \in \mathbf{R}_+^{V \times V}$ corresponding to the proposed solution $(X^*, Y^*)$ and satisfying the hypothesis of Theorem 4.1 do indeed exist. In particular, 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 (4.1) is equal to 0.

($\omega_2$) If $(i, j) \in \Omega = V^* \times V^* \cap \tilde{E}$, then $F_{ij} = 0$ and choosing $W_{ij} = \tilde{\lambda} - \gamma - M_{ij}$ makes the left-hand side of (4.1) 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 (4.1) is equal to $W_{ij} - \lambda$. In this case, we choose $W_{ij} = \lambda$ to make both sides of (4.1) zero.

($\omega_4$) Suppose that $i, j \in V - V^*$ such that $(i, j) \in \tilde{E}$. 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 (4.1) is zero.

11

($\omega_5$) If $i \in V^*$, $j \in V - V^*$ such that $(i, j) \in \tilde{E}$ 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 $(i, j) \in \tilde{E}$ 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.

It remains to construct multipliers $M$, $\lambda$, and $\gamma$ so that $W$ and $F$ as chosen above satisfy (4.2) and (4.3) if $p, q$, and $k$ satisfy the hypothesis of Theorem 2.2. In this case, $(X^*, Y^*)$ is optimal for (2.5) and the corresponding densest $k$-subgraph of $G$ can be recovered from $X^*$. The remainder of the proof is structured as follows. In Section 4.2, we construct valid $\lambda \geq 0$ and $M \in \mathbf{R}_+^{V \times V}$ such that $W\mathbf{v} = W^T\mathbf{v} = 0$. 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 Section 4.3. We conclude in Section 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 $\lambda \in \mathbf{R}_+$ and $M \in \mathbf{R}_+^{V \times V}$ such that $W\mathbf{v} = W^T\mathbf{v} = 0$. Note that $[W\mathbf{v}]_i = \sum_{j \in V^*} W_{ij}$ for all $i \in V$. If $i \in V - V^*$, we have

$$[W\mathbf{v}]_i = n_i \lambda - (k - n_i) \left( \frac{n_i}{k - n_i} \right) \lambda = 0$$

by our choice of $W_{ij}$ in ($\omega_3$) and ($\omega_5$). By symmetry, $[W^T\mathbf{v}]_i = 0$ for all $i \in V - V^*$.

The conditions $W(V^*, V^*)\mathbf{e} = W(V^*, V^*)^T\mathbf{e} = 0$ define $2k$ equations for the $k^2$ unknown entries of $M$. To obtain a particular solution of this underdetermined system, we parametrize $M$ as $M = \mathbf{y}\mathbf{e}^T + \mathbf{e}\mathbf{y}^T$, for some $\mathbf{y} \in \mathbf{R}^V$. After this parametrization

$$\sum_{j \in V^*} W_{ij} = k\tilde{\lambda} - (k - 1 - n_i)\gamma - ky_i - \mathbf{e}^T\mathbf{y}.$$

Rearranging shows that $\mathbf{y}$ is the solution of the linear system

$$(kI + \mathbf{e}\mathbf{e}^T)\mathbf{y} = k\tilde{\lambda}\mathbf{e} - \gamma((k-1)\mathbf{e} - \mathbf{n}), \tag{4.6}$$

where $\mathbf{n} \in \mathbf{R}^{V^*}$ is the vector with $i$th entry $n_i$ equal to the degree of node $i$ in $G(V^*)$. By

12

the Sherman-Morrison-Woodbury formula [27, Equation (2.1.4)], we have

$$\mathbf{y} = \frac{1}{2k}\left(k\tilde{\lambda} - (k-1)\gamma\right)\mathbf{e} + \frac{\gamma}{k}\left(\mathbf{n} - \left(\frac{\mathbf{n}^T\mathbf{e}}{2k}\right)\mathbf{e}\right). \tag{4.7}$$

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

$$\|\mathbf{y} - \mathbb{E}[\mathbf{y}]\|_\infty = \frac{\gamma}{k}\left\|\mathbf{n} - \left(\frac{\mathbf{n}^T\mathbf{e}}{2k}\right)\mathbf{e} - \mathbb{E}\left[\mathbf{n} - \left(\frac{\mathbf{n}^T\mathbf{e}}{2k}\right)\mathbf{e}\right]\right\|_\infty \leq \frac{\gamma}{2k}\left(k\epsilon + q\right) \tag{4.8}$$

with high probability to establish that the entries of $M$ are nonnegative with high probability, by the fact that each component $y_i$ is bounded below by $\mathbb{E}[y_i] - \|\mathbf{y} - \mathbb{E}[\mathbf{y}]\|_\infty$. 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 $|s - pm| \leq 6\max\left\{\left(p(1-p)m\log m\right)^{1/2}, \log m\right\}$ with probability at least $1 - 2m^{-12}$.*

Lemma 4.1 is a specialization of the standard Bernstein inequality (see, for example, [37, Theorem 6]) to binomially distributed random variables; the proof is left to the reader.

We are now ready to state and prove the desired lower bound on the entries of $\mathbf{y}$.

**Lemma 4.2** *For each $i \in V^*$, we have*

$$y_i \geq \gamma\left(\frac{\epsilon}{2} - 12\max\left\{\left(\frac{q(1-q)\log k}{k}\right)^{1/2}, \frac{\log k}{k}\right\}\right). \tag{4.9}$$

*with high probability.*

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

$$|n_i - (1-q)(k-1)| \leq 6\max\left\{\left(q(1-q)k\log k\right)^{1/2}, \log k\right\} \tag{4.10}$$

for all $i \in V^*$ with high probability. On the other hand, $\mathbf{n}^T\mathbf{e} = 2|E(G(V^*))|$ because each entry of $\mathbf{n}$ is equal to the degree in the subgraph induced by $V^*$ of the corresponding node.

Therefore $\mathbf{n}^T \mathbf{e}$ 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

$$|\mathbf{n}^T\mathbf{e} - \mathbb{E}[\mathbf{n}^T\mathbf{e}]| \leq 12 \max \left\{ k\big(q(1-q)\log k\big)^{1/2},\, 2\log k \right\} \tag{4.11}$$

with high probability. Substituting (4.10) and (4.11) into the left-hand side of (4.8) and applying the triangle inequality shows that

$$\|\mathbf{y} - \mathbb{E}[\mathbf{y}]\|_\infty \leq \frac{12\gamma}{k} \max \left\{ (q(1-q)k\log k)^{1/2},\, \log k \right\} \tag{4.12}$$

for sufficiently large $k$ with high probability. Subtracting the right-hand side of (4.12) from $E[y_i] \geq \gamma\epsilon/2$ for each $i \in V^*$ completes the proof. ∎

In Section 4.3, we will choose $\epsilon = (1 - p - q)/3$ to ensure that $\|F\|_\infty \leq 1$ with high probability. Substituting this choice of $\epsilon$ in the right-hand side of (4.9) yields

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

with high probability. Therefore, the entries of the multiplier $M$ are nonnegative w.h.p. if $p, q$, and $k$ satisfy (2.8).

## 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 $(i, j) \in (V \times V) - (V^* \times V^*)$.

We first consider the case when $i, j \in V - V^*$ and $ij \notin E$. In this case, 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

$$\frac{1}{\gamma k} + \epsilon + p + q \leq 1. \tag{4.13}$$

Taking $\epsilon = (1 - p - q)/3$ and $\gamma \geq 1/(\epsilon k)$ ensures that (4.13) is satisfied in this case.

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

$$\frac{1}{\gamma k} + \epsilon + q + \frac{n_j}{k} \leq 1. \tag{4.14}$$

Applying Lemma 4.1 and the union bound over all $j \in V - V^*$ shows that

$$|n_j - pk| \leq 6 \max\left\{ (p(1-p)k \log k)^{1/2}, \log k \right\}$$

for all $j \in V - V^*$ with high probability. Thus, the left-hand side of (4.14) is bounded above by

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

with high probability, which is bounded above by 1 for sufficiently large $k$ for our choice of $\epsilon$ and $\gamma$. Therefore, our choice of $F$ satisfies $\|F\|_\infty < 1$ with high probability.

## 4.4 A bound on $\|W\|$

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

**Lemma 4.3** *Suppose that $p, q$, and $k$ satisfy (2.7). Then*

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

*with high probability.*

Taking $\gamma = O\left( ((1-p-q)k)^{-1} \right)$ 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

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

15

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 $a_{ij}$ having $\sigma^2$ 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 [43, Theorem 1.4] and is included as Appendix A.2.

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

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

**Proof:** We have $\|Q\| = \|Q(V^*, V^*)\|$ by the block structure of $Q$. Let

$$Q_1 = H(V^*, V^*) - \left(\frac{k-1}{k}\right) q\mathbf{e}\mathbf{e}^T, \quad Q_2 = \frac{1}{k}\left(\mathbf{n}\mathbf{e}^T - (1-q)(k-1)\mathbf{e}\mathbf{e}^T\right), \quad Q_3 = Q_2^T$$

$$Q_4 = \frac{1}{k}\left(\mathbf{n}^T\mathbf{e} - (1-q)(k-1)k\right),$$

where $H$ is the adjacency matrix of the complement of $G(V^*, V^*)$. 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 obtain the desired bound on $\|Q\|$.

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 $\mathbb{E}[\tilde{H}] = q\mathbf{e}\mathbf{e}^T$ and $\tilde{H} - q\mathbf{e}\mathbf{e}^T$ 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 $\tilde{H} - q\mathbf{e}\mathbf{e}^T$ has magnitude bounded above by $B = \max\{q, 1-q\} \leq 1$. Therefore, applying Lemma 4.4 shows that $\|\tilde{H} - q\mathbf{e}\mathbf{e}^T\| \leq 6 \max\{\sqrt{q(1-q)k \log k}, \log^2 k\}$ with high probability. It follows immediately that

$$\|Q_1\| \leq \|\tilde{H} - q\mathbf{e}\mathbf{e}^T\| + \|(q/k)\mathbf{e}\mathbf{e}^T\| + \|\mathrm{Diag}\,(\mathrm{diag}\,\tilde{H})\|$$
$$\leq 6 \max\left\{\left(q(1-q)k \log k\right)^{1/2}, \log^2 k\right\} + q + 1 \tag{4.15}$$

with high probability by the triangle inequality.

16

We next bound $\|Q_2\|$ and $\|Q_3\|$. By (4.10), we have

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

with high probability. It follows that

$$\|Q_2\| = \|Q_3\| \leq \frac{1}{k}\|\mathbf{n} - \mathbb{E}[\mathbf{n}]\|\|\mathbf{e}\| \leq 6 \max\left\{(q(1-q)k \log k)^{1/2}, \log k\right\} \qquad (4.16)$$

with high probability. Finally,

$$\|Q_4\| \leq \frac{1}{k^2}\left|\mathbf{n}^T\mathbf{e} - \mathbb{E}[\mathbf{n}^T\mathbf{e}]\right| \|\mathbf{e}\mathbf{e}^T\| \leq 12 \max\left\{(q(1-q)\log k)^{1/2}, 2\log k/k\right\} \qquad (4.17)$$

with high probability, where the last inequality follows from (4.11). Combining (4.15), (4.16), and (4.17) and applying the union bound we have $\|Q\| \leq 24\gamma \max\{(q(1-q)k \log k)^{1/2}, \log^2 k\}$ with high probability. $\blacksquare$

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

**Lemma 4.6** *Suppose that $p$ and $k$ satisfy (2.7) Then*

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

*with high probability.*

**Proof:** We decompose $R$ as in the proof of Theorem 7 in [7]. Specifically, we let $R = \lambda(R_1 + R_2 + R_3 + R_4 + R_5)$ as follows.

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 $(i,j) \in V^* \times V^*$ we take $[R_1]_{ij}$ 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

$$\|R_1\| \leq 6 \max\left\{B \log^2 N, \left(\left(\frac{p}{1-p}\right) N \log N\right)^{1/2}\right\} \qquad (4.18)$$

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

17

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

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

with high probability by Lemma 4.4. We define $R_3$ to be the correction matrix for diagonal entries of $R_1$: $\lambda[R_3]_{ii} = R_{ii} - \lambda[R_1]_{ii}$ 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 $\|R_3\| \le 1/1 - p$.

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

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

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

$$
\begin{aligned}
\|R_4\|^2 &\le \frac{36(N - k) \max\{ p(1-p)k \log k, \log^2 k \}}{(1-p)^2 \left( (1-p)k - \max\{ \sqrt{p(1-p)k \log k}, \log k \} \right)} \\
&\le \left( \frac{144N}{(1-p)^3 k} \right) \max\{ p(1-p)k \log k, \log^2 k \}
\end{aligned}
$$

with high probability, where the last inequality follows from Assumption (2.7). Combining the upper bounds on each $\|R_i\|$ shows that

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

with high probability, provided $p, k$, and $N$ satisfy (2.7). This completes the proof. ∎

# 5   Experimental results

In this section, we empirically evaluate the performance of our relaxation for the planted densest $k$-subgraph problem. Specifically, we apply our relaxation (2.5) to $N$-node random graphs sampled from the planted dense $k$-subgraph model for a variety of planted clique sizes $k$.

For each randomly generated program input, we apply the Alternating Directions Method of Multipliers (ADMM) to solve (2.5). 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 well beyond the scope of this paper; we direct the reader to the recent survey [10] and the references within for more details.

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 [26, Example 3]. We iteratively solve the linearly constrained optimization problem

$$\min \quad \|X\|_* + \gamma\|Y\|_1 + \mathbf{1}_{\Omega_Q}(Q) + \mathbf{1}_{\Omega_W}(W) + \mathbf{1}_{\Omega_Z}(Z)$$
$$\text{s.t.} \quad X + Y = Q, \ \ X = W, \ \ X = Z,$$

where $\Omega_Q := \{Q \in \mathbf{R}^{V \times V} : P_{\tilde{E}}(Q) = 0\}$, $\Omega_W := \{W \in \mathbf{R}^{V \times V} : \mathbf{e}^T W \mathbf{e} = k^2\}$, and $\Omega_Z := \{Z \in \mathbf{R}^{V \times V} : Z_{ij} \leq 1 \ \forall (i,j) \in V \times V\}$. Here $\mathbf{1}_S : \mathbf{R}^{V \times V} \to \{0, +\infty\}$ is the indicator function of the set $S \subseteq \mathbf{R}^{V \times V}$, defined by $\mathbf{1}_S(X) = 0$ if $X \in S$ and $+\infty$ otherwise. During each iteration we sequentially update each primal decision variable by minimizing the augmented Lagrangian

$$\begin{aligned} L_\tau = &\|X\|_* + \gamma\|Y\|_1 + + \mathbf{1}_{\Omega_Q}(Q) + \mathbf{1}_{\Omega_W}(W) + \mathbf{1}_{\Omega_Z}(Z) \\ &+ \operatorname{Tr}(\lambda_Q(X + Y - Q)) + \operatorname{Tr}(\Lambda_W(X - W)) + \operatorname{Tr}(\Lambda_Z(X - Z)) \\ &+ \frac{\tau}{2}\left(\|X + Y - Q\|^2 + \|X - W\|^2 + \|X - Z\|^2\right) \end{aligned}$$

in Gauss-Seidel fashion with respect to each primal variable and then updating the dual variables $\lambda_Q, \lambda_W, \Lambda_Z$ using the updated primal variables. Here $\tau$ is a regularization parameter chosen so that $L_\tau$ is strongly convex in each primal variable. Equivalently, we update each of $X, Y, Q, W$, and $Z$ by evaluation of an appropriate proximity operator during each iteration. Minimizing the augmented Lagrangian with respect to each of the artificial primal variables $Q, W$ and $Z$ is equivalent to projecting onto each of the sets $\Omega_Q, \Omega_W$, and $\Omega_Z$, 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 $S_\phi : \mathbf{R}^n \to \mathbf{R}^n$ defined by

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

It has recently been shown that ADMM converges linearly when applied to the minimization

---

**Algorithm 1** ADMM for solving (2.5)

---

**Input:** $G = (V, E)$, $k \in \{1, \ldots, N\}$, where $N = |V|$, and error tolerance $\epsilon$.

**Initialize:** $X^{(0)} = W^{(0)} = (k/n)^2 \mathbf{e}\mathbf{e}^T$, $Y^{(0)} = -X$, $Q^{(0)} = \Lambda_Q^{(0)} = \Lambda_W^{(0)} = \Lambda_Z^{(0)} = 0$.

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

    **Step 1:** Update $Q^{(\ell+1)}$

$$Q^{(\ell+1)} = P_{\tilde{E}}\left(X^{(\ell)} + Y^{(\ell)} - \Lambda_Q^{(\ell)}\right).$$

    **Step 2:** Update $X^{(\ell+1)}$

        Let $\tilde{X}^{(\ell)} = Q^{(\ell+1)} + 2X^{(\ell)} - Z^{(\ell)} - W^{(\ell)} - \Lambda_W^{(\ell)}$.

        Take singular value decomposition $\tilde{X}^{(\ell)} = U\left(\text{Diag}\,\mathbf{x}\right)V^T$.

        Apply soft thresholding: $X^{(\ell+1)} = U\left(\text{Diag}\,S_\tau(\mathbf{x})\right)V^T$

    **Step 3:** Update $Y^{(\ell+1)}$

        $Y^{(\ell+1)} = S_{\tau\gamma}\left(Y^{(\ell)} - \tau Q^{(\ell+1)}\right)$.

    **Step 4:** Update $W^{(\ell+1)}$

        Let $\tilde{W}^{(\ell)} = X^{(\ell+1)} - \Lambda_W^{(\ell)}$.

        Let $\beta_k = \left(k^2 - \mathbf{e}^T\tilde{W}^{(\ell)}\mathbf{e}\right)/n^2$.

        Update $W^{(\ell+1)} = \tilde{W}^{(\ell)} + \beta_k\mathbf{e}\mathbf{e}^T$.

    **Step 5: Update** $Z^{(\ell+1)}$

        Let $\tilde{Z}^{(\ell)} = X^{(\ell+1)} - \Lambda_Z^{(\ell)}$.

        For each $i, j \in V$: $Z_{ij}^{(\ell+1)} = \min\{\max\{\tilde{Z}_{ij}^{(\ell)}, 0\}, 1\}$

    **Step 6:** Update dual variables

        $\Lambda_Z^{(\ell+1)} = \Lambda_Z^{(\ell)} - \left(X^{(\ell+1)} - Z^{(\ell+1)}\right)$.

        $\Lambda_W^{(\ell+1)} = \Lambda_W^{(\ell)} - \left(X^{(\ell+1)} - W^{(\ell+1)}\right)$.

        $\Lambda_P^{(\ell+1)} = P_{V\times V - \tilde{E}}\left(\Lambda_P^{(\ell)} - \left(X^{(\ell+1)} + Y^{(\ell+1)}\right)\right)$

    **Step 7:** Check convergence

        $r_p\max\{\|X^{(\ell)} - W^{(\ell)}\|_F, \|X^k - Z^{(\ell)}\|_F\}$

        $r_d = \max\{\|W^{(\ell+1)} - W^{(\ell)}\|_F, \|Z^{(\ell+1)} - Z^{(\ell)}\|_F, \|\Lambda_Q^{(\ell+1)} - \Lambda_Q^{(\ell)}\|\}$.

        **if** $\max\{r_p, r_d\} < \epsilon$ **then**

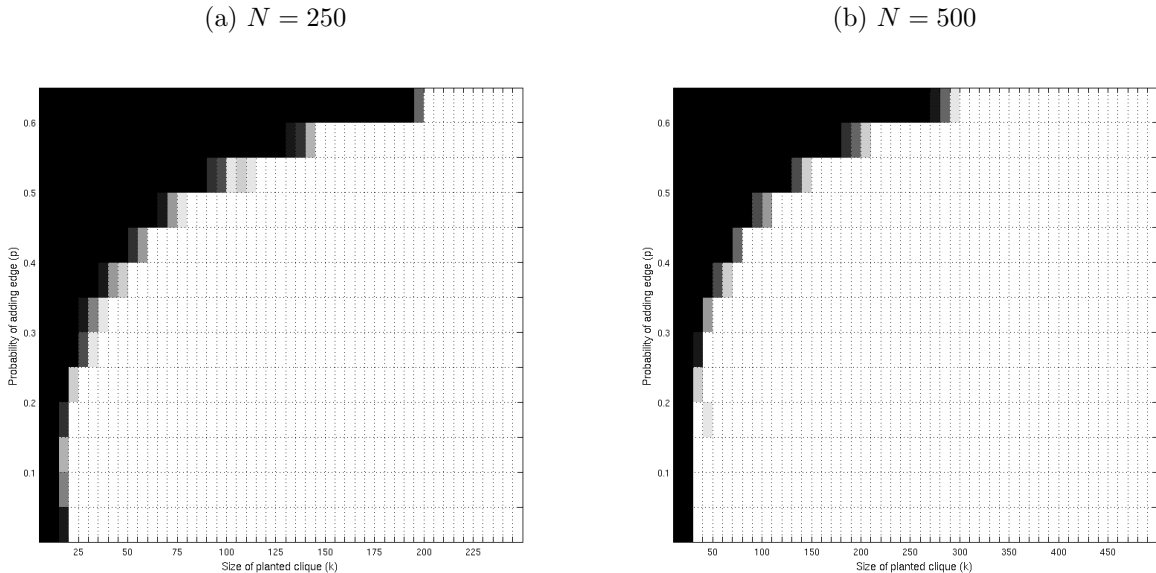            **Stop:** algorithm converged.

---

of convex separable functions, under mild assumptions on the program input (see [29]), and, as such, Algorithm 1 can be expected to converge to the optimal solution of (2.5); We stop Algorithm 1 when the primal and dual residuals

$$\|X^{(\ell)} - W^{(\ell)}\|_F, \ \|X^k - Z^{(\ell)}\|_F, \ \|W^{(\ell+1)} - W^{(\ell)}\|_F, \ \|Z^{(\ell+1)} - Z^{(\ell)}\|_F, \ \|\Lambda_Q^{(\ell+1)} - \Lambda_Q^{(\ell)}\|_F$$

are smaller than a desired error tolerance.

We evaluate the performance of our algorithm for a variety of random program inputs. We generate random $N$-node graph $G$ constructed according to ($\rho_1$) and ($\rho_2$) for $q = 0.25$ and various clique sizes $k \in (0, N)$ and edge addition probabilities $p \in [0, q]$. Each graph $G$ is represented by a random symmetric binary matrix $A$ with entries in the $(1 : k) \times (1 : k)$

Figure 5.1: Simulation results for $N$-node graphs with planted dense $k$-subgraph Each entry gives the average number of recoveries of the planted subgraph per set of 10 trials for the corresponding choice of $k$ and probability of adding noise edges $p$. Fixed probability of deleting clique edge $q = 0.25$ was used in each trial. A higher rate of recovery is indicated by lighter colours.

(a) $N = 250$  (b) $N = 500$



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 $\tau = 0.35$ and stopping tolerance $\epsilon = 10^{-4}$ is 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 $X_0 = \mathbf{v}\mathbf{v}^T$ and $\mathbf{v}$ 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 Figure 5.1. The 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 $p$.

# 6    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

tradeoffs 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 (2.5). 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.

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. [42] 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 if 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 [6] and clustering [4] problems

# 7    Acknowledgments

# A   Appendices

## A.1   Proof of Theorem 4.1

The convex program (2.5) admits a strictly feasible solution and, hence, (2.5) satisfies Slater's constraint qualification (see [9, Equation (3.2.7)]). Therefore, the Karush-Kuhn-Tucker conditions applied to (2.5) state that a feasible solution $(X, Y)$ of (2.5) is optimal if and only if there exist multipliers $\lambda \in \mathbf{R}_+$, $H \in \mathbf{R}^{V \times V}$, $M_1, M_2 \in \mathbf{R}_+^{V \times V}$ and subgradients $\phi \in \partial \|X\|_*$, $\psi \in \partial \|Y\|_1$ such that

$$\phi - \lambda \mathbf{e}\mathbf{e}^T + \sum_{(i,j) \in \tilde{E}} H_{ij} \mathbf{e}_i \mathbf{e}_j^T + M_1 - M_2 = 0 \tag{A.1}$$

$$\gamma \psi + \sum_{(i,j) \in \tilde{E}} H_{ij} \mathbf{e}_i \mathbf{e}_j^T = 0 \tag{A.2}$$

$$[M_1]_{ij}(X_{ij} - 1) = 0 \quad \forall\, i, j \in V \tag{A.3}$$

$$[M_2]_{ij} X_{ij} = 0 \quad \forall\, i, j \in V \tag{A.4}$$

Taking $M_2 = 0$ ensures that (A.4) is satisfied for all $X$. Since $\bar{X}_{ij} = 1$ if $(i,j) \in \bar{V} \times \bar{V}$ and is 0 otherwise, (A.3) is equivalent to (4.5) when $X = \bar{X}$. It is known (see, for example, [12, Section 3.4]) that $\partial \|\bar{Y}\|_1 = \{\bar{Y} + F : P_\Omega F = 0, \ \|F\|_\infty \le 1\}$. We can substitute $\psi = \mathrm{sign}\,(\bar{Y}) + F$ in (A.2) for some matrix $F$ such that $P_\Omega F = 0$ and $\|F\|_\infty \le 1$. Moreover, since $\bar{Y} = 0$ for all $(i,j) \notin \tilde{E}$, (A.2) implies that $F_{ij} = 0$ for all $(i,j) \notin \tilde{E}$. Since the complement of $\tilde{E}$ is exactly $E \cup \{vv : v \in V\}$, this yields (4.4). Similarly, the subdifferential of the nuclear norm at $\bar{X}$ is equal to the set $\partial \|\bar{X}\|_* = \{\bar{\mathbf{v}}\bar{\mathbf{v}}^T / k + W : W\bar{\mathbf{v}} = W^T\bar{\mathbf{v}} = 0, \ \|W\| \le 1\}$; see [44, Example 2]. Combining (A.1) and (A.2) and substituting this formula for the subgradients of $\|\bar{X}\|_*$ into the resulting equation yields (4.1) and (4.2). Thus, the conditions (4.1), (4.2), (4.3), (4.4), and (4.5) are exactly the Karush-Kuhn-Tucker conditions for (2.5) applied at $(\bar{X}, \bar{Y})$, with the Lagrange multiplier $M_2$ taken to be 0.

We next show that $G(\bar{V})$ has maximum density among all $k$-subgraphs of $G$ if $(\bar{X}, \bar{Y})$ is optimal for (2.5)). Fix some subset of nodes $\hat{V} \subseteq V$ of cardinality $k$. Let $\hat{X} = \hat{\mathbf{v}}\hat{\mathbf{v}}^T$ where $\hat{\mathbf{v}}$ is the characteristic vector of $\hat{V}$ and let $\hat{Y}$ be the matrix constructed according to (2.2) for $\hat{V}$. Note that both $\bar{X}$ and $\hat{X}$ are rank-one matrices with nonzero singular value equal to $k$. By the optimality of $(\bar{X}, \bar{Y})$, we have $\|\hat{X}\|_* + \gamma \|\hat{Y}\|_1 = k + \gamma \|\hat{Y}\|_1 \ge k + \gamma \|\bar{Y}\|_1$. Consequently, $\|\hat{Y}\|_1 \ge \|\bar{Y}\|_1$ and $d(G(\hat{V})) \le d(G(\bar{V}))$ as required.

It remains to show that the conditions $\|W\| < 1$ and $\|F\|_\infty < 1$ imply that $(\bar{X}, \bar{Y})$ is the unique optimal solution of (2.5). The relaxation (2.5) can be written as the semidefinite

program

$$\min \quad \frac{1}{2}\big(\operatorname{Tr}(R_1) + \operatorname{Tr}(R_2)\big) + \gamma\langle \mathbf{e}\mathbf{e}^T, Z\rangle$$

$$\text{s.t.} \quad R = \begin{pmatrix} R_1 & X \\ X^T & R_2 \end{pmatrix} \succeq 0$$

$$-Z_{ij} \le Y_{ij} \le Z_{ij}, \quad \forall\, i,j \in V$$

$$\mathbf{e}^T X \mathbf{e} = k^2$$  (A.5)

$$X_{ij} + Y_{ij} = 0 \quad \forall\, (i,j) \in \tilde{E}$$

$$X_{ij} \le 1 \quad \forall\, i,j \in V.$$

This problem is strictly feasible and, hence, strong duality holds. The dual of (A.5) is

$$\max \quad k^2 \lambda + \operatorname{Tr}(\mathbf{e}\mathbf{e}^T M)$$

$$\text{s.t.} \quad Q = \begin{pmatrix} I & -\lambda \mathbf{e}\mathbf{e}^T - \sum_{(i,j)\in\tilde{E}} H_{ij} + M \\ -\lambda \mathbf{e}\mathbf{e}^T - \sum_{(i,j)\in\tilde{E}} H_{ij} + M & I \end{pmatrix} \succeq 0$$

$$H_{ij} - S^1_{ij} + S^2_{ij} = 0 \quad \forall\, (i,j) \in \tilde{E}$$  (A.6)

$$S^1_{ij} - S^2_{ij} = 0 \quad \forall\, (i,j) \in (V \times V) - \tilde{E}$$

$$S^1_{ij} + S^2_{ij} = 0 \quad \forall\, i,j \in V.$$

$$M, S^1, S^2 \in \mathbf{R}^{N\times N}_+, \ H \in \mathbf{R}^{N\times N}, \ \lambda \in \mathbf{R},.$$

Suppose that there exists multipliers $F, W, \lambda$, and $M$ satisfying the hypothesis of Theorem 4.1 such that $\|W\| < 1$ and $\|F\|_\infty$. Note that $\bar{X} = \bar{R}_1 = \bar{R}_2 = \bar{\mathbf{v}}\bar{\mathbf{v}}^T$, $\bar{Y}$ as constructed according to (2.2), and $\bar{Z} = \operatorname{sign}(\bar{Y})$ defines a primal feasible solution for (A.5). We define a dual feasible solution as follows. If $(i,j) \in \Omega$ then $F_{ij} = 0$ and we take $\bar{H}_{ij} = -\gamma \bar{Y}_{ij} = \gamma$. In this case, we choose $\bar{S}^1_{ij} = \gamma$ and $\bar{S}^2_{ij} = 0$. If $(i,j) \in \tilde{E} - \Omega$, we choose $\bar{H}_{ij} = -\gamma F_{ij}$ and take $\bar{S}^1_{ij} = \gamma(1 - F_{ij})/2$, $\bar{S}^2_{ij} = \gamma(1 + F_{ij})/2$. Finally, if $(i,j) \notin \tilde{E}$, we take $\bar{S}^1_{ij} = \bar{S}^2_{ij} = \gamma/2$. Note that, since $|F_{ij}| < 1$ for all $i,j \in V$ and $\gamma > 0$, the entries of $\bar{S}^1$ are strictly positive and those of $\bar{S}^2$ are nonnegative with $\bar{S}^2_{ij} = 0$ if and only if $Y_{ij} - Z_{ij} < 0$. Therefore, the dual solution $(\bar{Q}, \bar{H}, \bar{S}^1, \bar{S}^2)$ defined by the multipliers $F, W, \lambda, M$ is feasible and satisfies complementary slackness by construction. Thus, $(\bar{R}, \bar{Y}, \bar{Z})$ is optimal for (A.6) and $(\bar{Q}, \bar{H}, \bar{S}^1, \bar{S}^2)$ is optimal for the dual problem (A.6).

We next establish that $(\bar{X}, \bar{Y})$ is the unique solution of (2.5). By (4.1) and our choice of $\bar{H}$,

$$\bar{Q} = \begin{pmatrix} I & -W - \bar{X}/k \\ -W^T - \bar{X}/k & I \end{pmatrix}.$$

Note that $\bar{R}\bar{Q} = 0$ since $W\bar{X} = W^T\bar{X} = 0$ and $\bar{X}^2/k = \bar{X}$. This implies that the column space of $\bar{R}$ is contained in the null space of $\bar{Q}$. Since $\bar{R}$ has rank equal to 1, $\bar{Q}$ has rank at most $2N - 1$. Moreover, $W + \bar{X}/k$ has maximum singular value equal to 1 with multiplicity

1 since $\|W\| < 1$. Therefore, $\bar{Q}$ has exactly one zero singular value, since $\omega$ is an eigenvalue of $\bar{Q} - I$ if and only if $\omega$ or $-\omega$ is an eigenvalue of $W + \bar{X}/k$. Thus $\bar{Q}$ has rank equal to $2N - 1$.

To see that $(\bar{X}, \bar{Y})$ is the unique optimal solution of (2.5), suppose on the contrary that $(\hat{R}, \hat{Y}, \hat{Z})$ is also optimal for (A.5). In this case, $(\hat{X}, \hat{Y})$ is optimal for (2.5). Since $(\bar{Q}, \bar{H}, \bar{S}^1, \bar{S}^2)$ is optimal for (A.6), we have $\hat{R}\bar{Q} = 0$ by complementary slackness. This implies that $\hat{R} = t\bar{R}$ and $\hat{X} = t\bar{X}$ for some scalar $t \geq 0$ by the fact that the column and row spaces of $\hat{R}$ lie in the null space of $\bar{Q}$, which is spanned by $[\mathbf{v}; \mathbf{v}]$. Moreover, $\hat{Y}, \hat{Z}, \bar{H}, \bar{S}^1, \bar{S}^2$ also satisfy complementary slackness. In particular, $\hat{Y}_{ij} = -\hat{Z}_{ij}$ for all $ij \in \Omega$ since $\bar{S}_{ij}^1 \neq 0$. On the other hand, $\bar{S}_{ij}^1 \neq 0$, $\bar{S}_{ij}^2 \neq 0$ for all $(i, j) \notin \Omega$ and $\hat{Y}_{ij} = \hat{Z}_{ij} = 0$ in this case. It follows that $\mathrm{supp}\,(\hat{Y}) \subseteq \mathrm{supp}\,(\bar{Y}) = \Omega$ and $\hat{Y} = -P_\Omega \hat{X} = -tP_\Omega \bar{X}$ by the fact that $P_\Omega(\hat{X} + \hat{Y}) = 0$. Finally, since $(\bar{X}, \bar{Y})$ and $(\hat{X}, \hat{Y})$ are both optimal for (2.5),

$$\|\bar{X}\|_* + \gamma\|\bar{Y}\|_1 = \|\hat{X}\|_* + \gamma\|\hat{Y}\|_1 = t(\|\bar{X}\|_* + \gamma\|\bar{Y}\|_1).$$

Therefore $t = 1$ and $(\bar{X}, \bar{Y})$ is the unique optimal solution for (2.5). ∎

## A.2 Proof of Lemma 4.4

In this section, we establish the concentration bound on the norm of a mean zero matrix given by Lemma 4.4. To do so, we will show that Lemma 4.4 is a special case of the following bound on the largest eigenvalue of a sum of random matrices.

**Theorem A.1 ([43, Theorem 1.4])** *Let $\{X_k\} \in \Sigma^d$ be a sequence of independent, random, symmetric matrices of dimension $d$ satisfying $\mathbb{E}[X_k] = 0$ and $\|X\| \leq R$, and let $S = \sum X_k$. Then, for all $t \geq 0$,*

$$P\left(\|S\| \geq t\right) \leq d \cdot \exp\left(-t^2/2\right) \tilde{\sigma}^2 + Rt/3 \quad \text{where } \tilde{\sigma}^2 := \left\|\sum_k \mathbb{E}(X_k^2)\right\|. \quad (A.7)$$

To see that Lemma 4.4 follows as a corollary of Theorem A.1, let $A \in \Sigma^n$ be a random symmetric matrix with i.i.d. mean zero entries having variance $\sigma^2$ such that $|a_{ij}| \leq B$ for all $i, j$. Let $\{X_{ij}\}_{1 \leq i \leq j \leq n} \in \Sigma^n$ be the sequence defined by

$$X_{ij} = \begin{cases} a_{ij}(\mathbf{e}_i\mathbf{e}_j^T + \mathbf{e}_j\mathbf{e}_i^T), & \text{if } i \neq j \\ a_{ii}\mathbf{e}_i\mathbf{e}_i^T, & \text{if } i = j, \end{cases}$$

where $\mathbf{e}_k$ is the $k$-th standard basis vector in $\mathbf{R}^n$. Note that $A = \sum X_{ij}$. It is easy to see that $\|X_{ij}\| \leq |a_{ij}| \leq B$ for all $1 \leq i \leq j \leq n$. On the other hand,

$$M = \sum_{1 \leq i \leq j \leq n} \mathbb{E}(X_{ij}^2) = \sum_{i=1}^{n} \left( \mathbb{E}(a_{ii}^2)\mathbf{e}_i\mathbf{e}_i^T + \sum_{j=i+1}^{n} \mathbb{E}(a_{ij}^2)(\mathbf{e}_i\mathbf{e}_i^T + \mathbf{e}_j\mathbf{e}_j^T) \right) = \sigma^2 n \cdot I,$$

by the independence of the entries of $A$. Therefore, $\tilde{\sigma}^2 = \|M\| = \sigma^2 n$. Substituting into (A.7) shows that

$$P(\|A\| \geq t) \leq n \exp\left( -\frac{\sigma^2 n \log n/2}{\sigma^2 n + Bt/3} \right) \tag{A.8}$$

for all $t \geq 0$. To complete the proof, we take $t = 6 \max\{\sigma\sqrt{n \log n}, B \log^2 n\}$ and consider the following cases.

First, suppose that $\sigma\sqrt{n \log n} \geq B \log^2 n$. In this case, we take $t = 6\sigma\sqrt{n \log n}$. Let $f(t) = (t^2/2)/(\sigma^2 n + Bt/3)$. Then, we have

$$f(t) = \frac{18\sigma^2 n \log n}{\sigma^2 n + 2B\sigma\sqrt{n \log n}} \geq \frac{18\sigma^2 n \log n}{\sigma^2 n + 2\sigma^2 n \log n/\log^2 n} = \frac{18 \log n}{1 + 2/logn} \geq 9 \log n$$

by the assumption that $B \leq \sigma\sqrt{n \log n}/\log^2 n$. On the other hand, if $B \log^2 n > \sigma\sqrt{n \log n}$ we take $t = B \log^2 n$ and

$$f(t) = \frac{18B^2 \log^4 n}{\sigma^2 n + 2B^2 \log^2 n} > \frac{18B^2 \log^4 n}{B^2 \log^2(\log n + 2)} > 9 \log n.$$

In either case, $P(\|A\| \geq t) \leq \exp(-f(t)) \leq n \exp(-9 \log n) = n^{-8}$. ∎

# References

[1] N. Alon, A. Andoni, T. Kaufman, K. Matulef, R. Rubinfeld, and N. Xie. Testing k-wise and almost k-wise independence. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 496–505. ACM, 2007. 7

[2] N. Alon, S. Arora, R. Manokaran, D. Moshkovitz, and O. Weinstein. Inapproximability of densest $\kappa$-subgraph from average-case hardness, 2011. 2, 3, 7

[3] N. Alon, M. Krivelevich, and B. Sudakov. Finding a large hidden clique in a random graph. In *Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, pages 594–598. Society for Industrial and Applied Mathematics, 1998. 7

[4] B. Ames. Guaranteed clustering and biclustering via semidefinite programming. *Arxiv preprint arXiv:1202.3663*, 2012. 22

[5] B. Ames. Robust convex relaxation for the planted clique and densest *k*-subgraph problems: additional proofs. 2013. Available from `http://ima.umn.edu/~bpames/papers.html`. 10

[6] B. Ames and S. Vavasis. Convex optimization for the planted k-disjoint-clique problem. *arXiv preprint arXiv:1008.2814*, 2010. 22

[7] B. Ames and S. Vavasis. Nuclear norm minimization for the planted clique and biclique problems. *Mathematical Programming*, 129(1):1–21, 2011. 2, 5, 6, 7, 9, 17

[8] Q. Berthet and P. Rigollet. Computational lower bounds for sparse pca. *arXiv preprint arXiv:1304.0828*, 2013. 7

[9] J.M. Borwein and A.S. Lewis. *Convex analysis and nonlinear optimization: theory and examples*, volume 3. Springer Verlag, 2006. 23

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011. 19

[11] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, UK, 2004. 10

[12] S. Boyd and L. Vandenberghe. Subgradients. *Lecture Notes for EE364b, Stanford University, Winter 2006-07*, 2008. Available from `http://see.stanford.edu/materials/lsocoee364b/01-subgradients_notes.pdf`. 23

[13] E.J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Arxiv preprint arXiv:0912.3599*, 2009. 2, 5

[14] V. Chandrasekaran, S. Sanghavi, P.A. Parrilo, and A.S. Willsky. Rank-sparsity incoherence for matrix decomposition. *Arxiv preprint ArXiv:0906.2220*, 2009. 2, 5

[15] Y. Chen, A. Jalali, S. Sanghavi, and C. Caramanis. Low-rank matrix recovery from errors and erasures. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pages 2313–2317. IEEE, 2011. 2, 5

[16] Y. Chen, S. Sanghavi, and H. Xu. Clustering sparse graphs. *arXiv preprint arXiv:1210.3335*, 2012. 2, 5

[17] Y. Dekel, O. Gurel-Gurevich, and Y. Peres. Finding hidden cliques in linear time with high probability. *arXiv preprint arXiv:1010.2997*, 2010. 7

[18] Y. Deshpande and A. Montanari. Finding hidden cliques of size\ sqrt {N/e} in nearly linear time. *arXiv preprint arXiv:1304.7047*, 2013. 7

[19] X.V. Doan and S.A. Vavasis. Finding approximately rank-one submatrices with the nuclear norm and l1 norm. *Arxiv preprint arXiv:1011.1839*, 2010. 5

[20] U. Feige. Relations between average case complexity and approximation complexity. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 534–543. ACM, 2002. 2, 3, 8

[21] U. Feige and R. Krauthgamer. Finding and certifying a large hidden clique in a semi-random graph. *Random Structures and Algorithms*, 16(2):195–208, 2000. 7

[22] U. Feige and R. Krauthgamer. The probable value of the lovász–schrijver relaxations for maximum independent set. *SIAM Journal on Computing*, 32(2):345–370, 2003. 7

[23] U. Feige, D. Peleg, and G. Kortsarz. The dense k-subgraph problem. *Algorithmica*, 29(3):410–421, 2001. 2

[24] U. Feige and D. Ron. Finding hidden cliques in linear time. *DMTCS Proceedings*, (01):189–204, 2010. 7

[25] A. Goerdt and A. Lanka. An approximation hardness result for bipartite clique. In *Electronic Colloquium on Computational Complexity, Report*, number 48, 2004. 8

[26] D. Goldfarb, S. Ma, and K. Scheinberg. Fast alternating linearization methods for minimizing the sum of two convex functions. *Mathematical Programming*, pages 1–34, 2010. 19

[27] G.H. Golub and C.F. Van Loan. *Matrix computations*. Johns Hopkins University Press, 1996. 13

[28] E. Hazan and R. Krauthgamer. How hard is it to approximate the best nash equilibrium? *SIAM Journal on Computing*, 40(1):79–91, 2011. 7

[29] M. Hong and Z.Q. Luo. On the linear convergence of the alternating direction method of multipliers. *arXiv preprint arXiv:1208.3922*, 2012. 20

[30] A. Jalali, Y. Chen, S. Sanghavi, and H. Xu. Clustering partially observed graphs via convex optimization. *Arxiv preprint arXiv:1104.4803*, 2011. 2, 5

[31] M. Jerrum. Large cliques elude the metropolis process. *Random Structures & Algorithms*, 3(4):347–359, 1992. 7

[32] A. Juels and M. Peinado. Hiding cliques for cryptographic security. *Designs, Codes and Cryptography*, 20(3):269–280, 2000. 7

[33] R.M. Karp. Reducibility among combinatorial problems. *Complexity of Computer Computations*, 40(4):85–103, 1972. 3

[34] S. Khot. Ruling out ptas for graph min-bisection, densest subgraph and bipartite clique. In *Foundations of Computer Science, 2004. Proceedings. 45th Annual IEEE Symposium on*, pages 136–145. IEEE, 2004. 2, 3

[35] L. Kučera. Expected complexity of graph partitioning problems. *Discrete Applied Mathematics*, 57(2):193–212, 1995. 7

[36] E. Lawler. Combinatorial optimization: Networks and matroids. 1976. *Holt, Rinehart and Winston, New York.* 3

[37] G. Lugosi. Concentration-measure inequalities, 2009. Available from `http://www.econ.upf.edu/~lugosi/anu.pdf`. 13

[38] F. McSherry. Spectral partitioning of random graphs. In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 529–537. IEEE, 2001. 7

[39] R.R. Nadakuditi. On hard limits of eigen-analysis based planted clique detection. In *Statistical Signal Processing Workshop (SSP), 2012 IEEE*, pages 129–132. IEEE, 2012. 7

[40] S. Oymak and B. Hassibi. Finding dense clusters via "low rank + sparse" decomposition. *Arxiv preprint arXiv:1104.5186*, 2011. 2, 5

[41] R. Peeters. The maximum edge biclique problem is np-complete. *Discrete Applied Mathematics*, 131(3):651–654, 2003. 8

[42] K. Rohe, T. Qin, and H. Fan. The highest dimensional stochastic blockmodel with a regularized estimator. *arXiv preprint arXiv:1206.2380*, 2012. 22

[43] J.A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, pages 1–46, 2011. 16, 25

[44] G.A. Watson. Characterization of the subdifferential of some matrix norms. *Linear Algebra and its Applications*, 170:33–45, 1992. 23