

Convex optimization for the planted k -disjoint-clique problem *

Brendan P.W. Ames [†] Stephen A. Vavasis[‡]

December 3, 2013

We consider the k -disjoint-clique problem. The input is an undirected graph G in which the nodes represent data items, and edges indicate a similarity between the corresponding items. The problem is to find within the graph k disjoint cliques that cover the maximum number of nodes of G . This problem may be understood as a general way to pose the classical ‘clustering’ problem. In clustering, one is given data items and a distance function, and one wishes to partition the data into disjoint clusters of data items, such that the items in each cluster are close to each other. Our formulation additionally allows ‘noise’ nodes to be present in the input data that are not part of any of the cliques.

The k -disjoint-clique problem is NP-hard, but we show that a convex relaxation can solve it in polynomial time for input instances constructed in a certain way. The input instances for which our algorithm finds the optimal solution consist of k disjoint large cliques (called ‘planted cliques’) that are then obscured by noise edges inserted either at random or by an adversary, as well as additional nodes not belonging to any of the k planted cliques.

1 Introduction

Given a set of data, *clustering* seeks to partition the data into sets of similar objects. These subsets are called ‘clusters’, and the goal is to find a few large clusters covering as much of the data as possible. Clustering plays a significant role in a wide range of applications; including, but not limited to, information retrieval, pattern recognition, computational biology, and image processing. For a recent survey of clustering techniques and algorithms with a particular focus on applications in data mining see [3].

In this paper, we consider the following graph-based representation of data. Given a set of data where each pair of objects is known to be similar or dissimilar, we consider the graph $G = (V, E)$ where the objects in the given data set are the set of nodes of G and any two

*Supported in part by a Discovery Grant and Postgraduate Scholarship (Doctoral) from NSERC (Natural Science and Engineering Research Council of Canada), MITACS (Mathematics of Information Technology and Complex Systems), and the US Air Force Office of Scientific Research.

[†]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@gmail.com

[‡]Department of Combinatorics and Optimization, University of Waterloo, 200 University Avenue W., Waterloo, Ontario N2L 3G1, Canada, vavasis@math.uwaterloo.ca

nodes are adjacent if and only if their corresponding objects are similar. Hence, for this representation of the data, clustering is equivalent to partitioning G into disjoint cliques. Therefore, for any integer k , the problem of identifying k clusters in the data containing the maximum number of objects is equivalent to the maximum node k -disjoint-clique problem of the corresponding graph G . Given an undirected graph $G = (V, E)$ and integer $k \in [1, |V|]$, the *maximum node k -disjoint-clique problem* refers to the problem of finding the subgraph K of G composed of a collection of k disjoint cliques, called a *k -disjoint-clique subgraph*, maximizing the number of nodes in K . Unfortunately, since the $k = 1$ case is exactly the maximum clique problem, well-known to be NP-hard [10], the maximum node k -disjoint-clique problem is NP-hard.

In Section 2, we relax the maximum node k -disjoint-clique problem to a semidefinite program. We show that this convex relaxation can recover the exact solution in two cases. In the first case, presented in Section 3, the input graph is constructed deterministically as follows. The input graph consists of k disjoint cliques C_1, \dots, C_k , each of size at least \hat{r} , plus a number of diversionary nodes and edges inserted by an adversary. We show that the algorithm can tolerate up to $O(\hat{r}^2)$ diversionary edges and nodes provided that, for each $i = 1, \dots, k$, each node in the clique C_i is adjacent to at most $O(\min\{|C_i|, |C_j|\})$ nodes in the clique C_j for each $j = 1, \dots, k$ such that $i \neq j$. In Section 4, we suppose that the graph contains a k -disjoint-clique subgraph K and some additional nodes, and the remaining nonclique edges are added to the graph independently at random with fixed probability p . We give a general formula for clique sizes that can be recovered by the algorithm; for example, if the graph contains N nodes total and $N^{1/4}$ planted cliques each of size $\Omega(N^{1/2})$, then the convex relaxation will find them. We develop the necessary optimality and uniqueness theorems in Section 2 and provide the necessary background on random matrices in Section 4.1.

The rationale for this line of analysis is that in real-world applications of clustering, it is often the case that the sought-after clusters are present in the input data but are hidden by the presence of noisy data. Therefore, it is of interest to find cases of clustering data in which the clusters are hidden by noise and yet can still be found in polynomial time.

Our analysis is related in an indirect manner to work on measuring ‘clusterability’ of data, e.g., Ostrovsky et al. [16]. In that work, the authors prove that a certain clustering algorithm works well if the data has k ‘good’ clusters. Our assumptions and analysis, however, differ substantially from [16] (for example, we do not require all the data items to be placed in clusters), so there is no direct relationship between our result and theirs.

Our results and techniques can be seen as an extension of those in [2] from the maximum clique problem to the maximum node k -disjoint-clique problem. Indeed, in the $k = 1$ case, our results agree with those presented in [2], as well as those found in earlier work by Alon et al. [1], and by Feige and Krauthgamer [8]. Recent papers by Oymak and Hassibi [17] and Jalali et al. [14] written subsequently to the initial prepublication release of this paper extend our results to a more general model for clustered data. Specifically, in [17] and [14] the authors independently propose new heuristics for partitioning an input graph G into densely connected subgraphs based on the heuristic for decomposition of a matrix into low-rank and sparse components considered in [7] and [5]. Under certain assumptions on the input graph, Oymak and Hassibi [17] and Jalali et al. [14] show that this approach successfully recovers the correct partition of the graph G into densely connected subgraphs. In particular, the

correct partition is recovered when the input graph is composed of several disjoint cliques obscured by noise in the form of random edge additions and *deletions*. It is important to note that our approach can only tolerate noise in the form of edge additions. The lower bound on the minimum clique size ensuring exact recovery provided by [17] and [14] is equal to that provided in Section 4, although the conditions ensuring exact recovery provided in [17] and [14] do not require any explicit upper bound on the number of planted cliques to be satisfied. On the other hand, exact recovery is guaranteed for our approach in the presence of up to $O(\hat{r}^2)$ diversionary nodes, while [17] and [14] do not consider noise of this form.

More generally, our results follow in the spirit of several recent papers, in particular Recht et al. [18] and Candès and Recht [6], which consider nuclear norm minimization, a special case of semidefinite programming, as a convex relaxation of matrix rank minimization. *Matrix rank minimization* refers to the problem of finding a minimum rank solution of a given linear system. These papers have results of the following general form. Suppose that it is known that the constraints of the given linear system are random in some sense and that it is known that a solution of very low rank exists. Then the nuclear norm relaxation recovers the (unique) solution of minimum rank. We will argue that, in the case that the graph G contains a planted k -disjoint-clique subgraph K and not too many diversionary edges, a rank k solution, corresponding to the adjacency matrix of K , of a system of linear equations defined by the input graph G can be recovered exactly by solving a semidefinite program.

Like many of the papers mentioned in the previous paragraph, the proof that the convex relaxation exactly recovers the combinatorial solution constructs multipliers to establish that the combinatorial solution satisfies the KKT optimality conditions of the convex problem. Herein we introduce a new technical method for the construction of multipliers. In [2], the multipliers are constructed according to simple formulas because of the fairly simple nature of the problem. On the other hand, in [6], the multipliers are constructed by projection (i.e., solving a linear least-squares problem), which entails a quite difficult analysis. This paper introduces a technique of intermediate complexity: we construct the multipliers as the solution to a system of invertible linear equations. We show that the equations are within a certain norm distance of much simpler (diagonal plus rank-one) linear equations. Finally, the result is obtained from standard bounds on the perturbation of the solution of a linear system due to perturbation in its coefficients.

2 The Maximum Node k -disjoint-clique Problem

Let $G = (V, E)$ be a simple graph. The **maximum node k -disjoint-clique** problem focuses on finding k disjoint cliques in G such that the total number of nodes in these cliques is maximized. We call a subgraph of G composed of k disjoint cliques a “ k -disjoint-clique subgraph”. This problem is clearly NP-hard since it is equivalent to the maximum clique problem in the case that $k = 1$.

The problem of maximizing the number of nodes in a k -disjoint-clique subgraph of G can

be formulated as the following combinatorial optimization problem

$$\max_{S=\{\mathbf{v}_1, \dots, \mathbf{v}_k\}} \sum_{i=1}^k \mathbf{v}_i^T \mathbf{e} \quad (1)$$

$$\text{s.t. } \mathbf{v}_i^T \mathbf{v}_j = 0, \quad \forall i, j = 1, \dots, k, i \neq j \quad (2)$$

$$[\mathbf{v}_i \mathbf{v}_i^T]_{uv} = 0, \quad \text{if } uv \notin E, u \neq v, \quad \forall i = 1, \dots, k \quad (3)$$

$$\mathbf{v}_i \in \{0, 1\}^V, \quad \forall i = 1, \dots, k. \quad (4)$$

Here \mathbf{e} denotes the all-ones vector in \mathbf{R}^N . A feasible solution $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ for (1) is the collection of characteristic vectors of a set of disjoint cliques of G . Indeed, the constraint (2) ensures that the sets of nodes indexed by S are disjoint and the constraint (3) ensures that the sets of nodes indexed by S induce complete subgraphs of G and, hence, are cliques of G . Note that a feasible solution $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ need not define a partition of V . That is, a feasible solution need not correspond to a k -disjoint-clique subgraph of G that contains every node in V . Unfortunately, finding the solution to a nonlinear program with binary constraints is NP-hard in general. The formulation (1) may be relaxed to the rank constrained semidefinite program

$$\begin{aligned} \max \quad & \sum_{i=1}^N \sum_{j=1}^N X_{ij} \\ \text{s.t.} \quad & X\mathbf{e} \leq \mathbf{e}, \\ & X_{ij} = 0, \quad \forall (i, j) \notin E \text{ s.t. } i \neq j \\ & \text{rank}(X) = k, \\ & X \succeq 0 \end{aligned} \quad (5)$$

where $N = |V|$, X is an $N \times N$ real symmetric matrix, and the notation “ $X \succeq 0$ ” means that X is positive semidefinite. To see that (5) is a relaxation of (1), suppose that $\{C_1, \dots, C_k\} \subseteq V$ defines a k -disjoint-clique subgraph of G . Let $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ be the set of characteristic vectors of $\{C_1, \dots, C_k\}$. The matrix

$$X = \sum_{i=1}^k \frac{\mathbf{v}_i \mathbf{v}_i^T}{|C_i|} \quad (6)$$

is positive semidefinite with rank equal to k . Note that $\sum_{i=1}^k \mathbf{v}_i \leq \mathbf{e}$ since $\mathbf{v}_1, \dots, \mathbf{v}_k$ are orthogonal binary vectors. It follows that

$$X\mathbf{e} = \sum_{i=1}^k \left(\frac{\mathbf{v}_i}{|C_i|} \right) \mathbf{v}_i^T \mathbf{e} = \sum_{i=1}^k \mathbf{v}_i \leq \mathbf{e} \quad (7)$$

since $\mathbf{v}_i^T \mathbf{e} = |C_i|$. Moreover,

$$\sum_{i=1}^N \sum_{j=1}^N X_{ij} = \mathbf{e}^T X \mathbf{e} = \sum_{i=1}^k \mathbf{v}_i^T \mathbf{e}. \quad (8)$$

Therefore, every feasible solution S for (1) defines a feasible solution of (5) with objective value equal to the number of nodes in the k -disjoint-clique subgraph defined by S . The

nonconvex program (5) may be relaxed further to a semidefinite program by replacing the nonconvex constraint $\text{rank}(X) = k$ with the linear constraint $\text{tr}(X) = k$:

$$\begin{aligned} \max \quad & \sum_{i=1}^N \sum_{j=1}^N X_{ij} \\ \text{s.t.} \quad & X\mathbf{e} \leq \mathbf{e}, \\ & X_{ij} = 0, \quad \forall (i, j) \notin E \text{ s.t. } i \neq j \\ & \text{tr}(X) = k, \\ & X \succeq 0. \end{aligned} \tag{9}$$

For every k -disjoint-clique subgraph of G composed of cliques of sizes r_1, \dots, r_k , (6) defines a feasible solution $X^* \in \mathbf{R}^{N \times N}$ for (9) such that

$$X_{ij}^* = \begin{cases} 1/r_\ell & \text{if both } i, j \text{ belong to clique } \ell \\ 0 & \text{otherwise.} \end{cases}$$

By construction, the objective value of (9) corresponding to X^* is equal to the number of nodes in the k -disjoint-clique subgraph and $\text{rank}(X^*) = k$. Using the Karush-Kuhn-Tucker conditions, we will derive conditions for which X^* corresponding to a k -disjoint-clique subgraph of G , as defined by (6), is optimal for the convex relaxation of maximum node k -disjoint-clique problem given by (9). In particular, these conditions are summarized by the following theorem.

Theorem 2.1 *Let X^* be feasible for (9). Suppose also that there exists $\lambda \in \mathbf{R}_+^N$, $\mu \in \mathbf{R}$, $\eta \in \mathbf{R}^{N \times N}$ and $S \in \Sigma_+^N$ such that*

$$-\mathbf{e}\mathbf{e}^T + \lambda\mathbf{e}^T + \mathbf{e}\lambda^T + \mu I + \sum_{\substack{(i,j) \notin E \\ i \neq j}} \eta_{ij} \mathbf{e}_i \mathbf{e}_j^T = S, \tag{10}$$

$$\lambda^T (X^* \mathbf{e} - \mathbf{e}) = 0, \tag{11}$$

$$\langle S, X^* \rangle = 0. \tag{12}$$

Here Σ_+^N is the cone of $N \times N$ positive semidefinite matrices, $\langle \cdot, \cdot \rangle$ is the trace inner product on $\mathbf{R}^{N \times N}$ defined by

$$\langle Y, Z \rangle = \text{tr}(YZ^T)$$

for all $Y, Z \in \mathbf{R}^{N \times N}$, \mathbf{e}_i denotes the i th column of the identity matrix in $\mathbf{R}^{N \times N}$ for all $i = 1, \dots, N$, and \mathbf{e} is the all-ones vector in \mathbf{R}^N . Then X^* is an optimal solution of (9).

We omit the proof of this theorem, as it is nothing more than the specialization of the KKT conditions [4] in convex programming to (9).

2.1 Construction of the auxiliary matrix \tilde{S}

Our proof technique to show that X^* is optimal for (6) is to construct multipliers to satisfy Theorem 2.1. The difficult multiplier to construct is S , the dual semidefinite matrix. The reason is that S must simultaneously satisfy homogeneous linear equations given by $\langle S, X^* \rangle = 0$, requirements on its entries given by the gradient equation (10), and positive semidefiniteness.

In this subsection, we will lay the groundwork for our definition of S ; in particular, we construct an auxiliary matrix \tilde{S} . The actual multipliers used to prove the optimality of X^* , as well as the proof itself, are in the next subsection.

Our strategy for satisfying the requirements on S is as follows. The matrix S will be constructed in blocks with sizes inherited from the blocks of X^* . In particular, let the nodes contained in the k planted cliques be denoted C_1, \dots, C_k , and let the remaining nodes be C_{k+1} . Then according to (6), X^* has diagonal blocks X_{C_q, C_q}^* for $q = 1, \dots, k$ consisting of multiples of the all 1's matrix. The remaining blocks of X^* are 0's. The diagonal blocks of S will be perturbations of the identity matrix, with the rank-one perturbation chosen so that each diagonal block of S , say S_{C_q, C_q} is orthogonal to X_{C_q, C_q}^* .

The entries of an off-diagonal block, say S_{C_q, C_s} must satisfy, first of all, (10). This constraint, however, is binding only on the entries corresponding to edges in G , since entries corresponding to absent edges are not constrained by (10) thanks to the presence of the unbounded multiplier η_{ij} on the left hand side. These entries that are free in (10) are chosen so that (12) is satisfied. It is a well known result in semidefinite programming that the requirements $\langle S, X \rangle = 0$, $X, S \in \Sigma_+^N$ together imply $SX = XS = 0$. Thus, the remaining entries of S must be chosen so that $X^*S = SX^* = 0$. Because of the special form of X^* , this is equivalent to requiring all row and column sums of S_{C_q, C_s} to equal zero.

We parametrize the entries of S_{C_q, C_s} that are not predetermined by (10) using the entries of two vectors $\mathbf{y}^{q,s}$ and $\mathbf{z}^{q,s}$. These vectors are chosen to be the solutions to systems of linear equations, namely, those imposed by the requirement that $X^*S = SX^* = 0$. We show that the system of linear equations may be written as a perturbation of a linear system with a known solution, and we can thus get bounds on $\mathbf{y}^{q,s}$ and $\mathbf{z}^{q,s}$. The bounds on $\mathbf{y}^{q,s}$ and $\mathbf{z}^{q,s}$ in turn translate to bounds on $\|S_{C_q, C_s}\|$, which are necessary to establish the positive semidefiniteness of S . This semidefiniteness is established by proving that the diagonal blocks, which are identity matrices plus rank-one perturbations, dominate the off-diagonal blocks.

Recalling our notation introduced earlier, $G = (V, E)$ has a k -disjoint-clique subgraph K composed of cliques C_1, C_2, \dots, C_k of sizes r_1, r_2, \dots, r_k respectively. Let $C_{k+1} := V \setminus (\cup_{i=1}^k C_i)$ be the set of nodes of G not in K and let $r_{k+1} := |C_{k+1}|$. Let $N := |V|$. Let $\hat{r} := \min\{r_1, r_2, \dots, r_k\}$. For each $v \in V$, let n_v^s denote the number of nodes adjacent to v in C_s for all $s \in \{1, \dots, k+1\}$, and let $cl(v)$ denote index $i \in \{1, \dots, k+1\}$ such that $v \in C_i$.

Let $A(\bar{G}) \in \mathbf{R}^{N \times N}$ be the adjacency matrix of the complement \bar{G} of G ; that is $[A(\bar{G})]_{i,j} = 1$ if $(i, j) \notin E$ and 0 otherwise. Next, fix $q, s \in \{1, \dots, k+1\}$ such that $q \neq s$. Let $H = H_{q,s} \in \mathbf{R}^{C_q \times C_s}$ be the block of $A(\bar{G})$ with entries indexed by the vertex sets C_q and C_s , and let $D = D_{q,s} \in \mathbf{R}^{C_q \times C_q}$ be the diagonal matrix such that, for each $i \in C_q$, the (i, i) th entry of D is equal to the number of nodes in C_s not adjacent to i . That is

$$D = r_s I - \text{Diag}(\mathbf{n}_{C_q}^s)$$

where $\mathbf{n}_{C_q}^s \in \mathbf{R}^{C_q}$ is the vector with i th entry equal to n_i^s for each $i \in C_q$. Let $F = F_{q,s} = D_{s,q}$.

Next, define the scalar

$$c = c_{q,s} := \begin{cases} \frac{\hat{r}}{2} \left(\frac{1}{r_q} + \frac{1}{r_s} \right), & \text{if } s \leq k \\ \frac{\hat{r}}{2} \left(\frac{1}{\hat{r}} + \frac{1}{r_q} \right), & \text{otherwise.} \end{cases}$$

Next, for each $q, s = 1, \dots, k+1$ such that $q \neq s$ let $\mathbf{b} = \mathbf{b}^{q,s} \in \mathbf{R}^{C_q \cup C_s}$ be defined by

$$[\mathbf{b}^{q,s}]_i = c \cdot \begin{cases} n_i^s, & \text{if } i \in C_q \\ n_i^q, & \text{if } i \in C_s. \end{cases}$$

Note that the matrix

$$\begin{pmatrix} D & H \\ H^T & F \end{pmatrix}$$

is weakly diagonally dominant since the i th row of H contains exactly $r_s - n_i^s$ 1's, and, hence, positive semidefinite. Further, let $\mathbf{y} = \mathbf{y}^{q,s}$ and $\mathbf{z} = \mathbf{z}^{q,s}$ be a solution of the perturbed system

$$\begin{pmatrix} D + \theta \mathbf{e} \mathbf{e}^T & H - \theta \mathbf{e} \mathbf{e}^T \\ H^T - \theta \mathbf{e} \mathbf{e}^T & F + \theta \mathbf{e} \mathbf{e}^T \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix} = \mathbf{b} \quad (13)$$

for some scalar $\theta > 0$ to be defined later.

The rationale for this system of equations (13) is as follows. Below in (18), we shall define the matrix \tilde{S}_{C_q, C_s} according to the formula that entries (i, j) corresponding to edges in E are set to $-c_{q,s}$, while entries (i, j) corresponding to absent edges are set to the sum $[\mathbf{y}^{q,s}]_i + [\mathbf{z}^{q,s}]_j$. Matrix S_{C_q, C_s} has the same formula; refer to (25) below.

As mentioned earlier, it is required that all row and column sums of S_{C_q, C_s} equal zero. Consider, e.g., the sum of the entries in a particular row $i \in C_q$. This sum consists of r_s terms; of these terms, n_i^s of them are $-c_{q,s}$ (corresponding to edges from i to C_s) while the other $r_s - n_i^s$ have the form $[\mathbf{y}^{q,s}]_i + [\mathbf{z}^{q,s}]_j$. Thus, the requirement that the row sum to zero is written

$$-n_i^s c_{q,s} + \sum_{j \in C_s; (i,j) \notin E} ([\mathbf{y}^{q,s}]_i + [\mathbf{z}^{q,s}]_j) = 0$$

which may be rewritten

$$(r_s - n_i^s)[\mathbf{y}^{q,s}]_i + \sum_{j \in C_s; (i,j) \notin E} [\mathbf{z}^{q,s}]_j = n_i^s c_{q,s}. \quad (14)$$

Equation (14) is exactly a row of (13) in the case $\theta = 0$ because of the formulas used to define D, F, H, \mathbf{b} .

In the case that θ is not zero, the equation for the i th row in (13) has an additional term of the form $\theta(\mathbf{e}^T \mathbf{y}^{q,s} - \mathbf{e}^T \mathbf{z}^{q,s})$. This additional term does not affect the result, as the following argument shows. The version of (13) with $\theta = 0$ is singular because the vector $(\mathbf{e}; -\mathbf{e})$ is in its null space. Now fix $\theta > 0$ such that (13) is nonsingular. By the fact that $(\mathbf{e}; -\mathbf{e})$ is in the null space of the coefficient matrix

$$\begin{pmatrix} D & H \\ H^T & F \end{pmatrix},$$

taking the inner product of each side of (13) with $(\mathbf{e}; -\mathbf{e})$ yields

$$\theta(r_q + r_s)(\mathbf{e}^T \mathbf{y} - \mathbf{e}^T \mathbf{z}) = \mathbf{b}_1^T \mathbf{e} - \mathbf{b}_2^T \mathbf{e},$$

where $\mathbf{b}_1 \in \mathbf{R}^{C_q}$, $\mathbf{b}_2 \in \mathbf{R}^{C_s}$ are the vectors of entries of \mathbf{b} corresponding to C_q and C_s respectively. Moreover,

$$\mathbf{b}_1^T \mathbf{e} - \mathbf{b}_2^T \mathbf{e} = \sum_{i \in C_q} n_i^s - \sum_{j \in C_s} n_j^q = 0$$

because the number of edges entering C_q from C_s is equal to the number of edges entering C_s from C_q . Therefore, if for some $\theta > 0$ we are able to show that (13) is nonsingular (which we shall establish in Section 3 and again in Section 4) then this particular $(\mathbf{y}^{q,s}, \mathbf{z}^{q,s})$ satisfying (13) will also be a solution to (14) since the additional term $\theta(\mathbf{e}^T \mathbf{y}^{q,s} - \mathbf{e}^T \mathbf{z}^{q,s})$ is zero.

For the remainder of this section, in order to formulate definitions for the remaining multipliers, assume that $\theta > 0$ and that (13) is nonsingular. Furthermore, assume that $D_{ii} > 0$ for all $i \in C_q$ and $F_{ii} > 0$ for all $i \in C_s$. Let

$$A = A(\theta) := \begin{pmatrix} D + \theta \mathbf{e} \mathbf{e}^T & 0 \\ 0 & F + \theta \mathbf{e} \mathbf{e}^T \end{pmatrix}, \quad P = P(\theta) := \begin{pmatrix} 0 & H - \theta \mathbf{e} \mathbf{e}^T \\ H^T - \theta \mathbf{e} \mathbf{e}^T & 0 \end{pmatrix};$$

then we have assumed that $A + P$ is nonsingular, and

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix} = (A + P)^{-1} \mathbf{b}.$$

The proof technique in Sections 3 and 4 is to show that $Q := (A + P)^{-1} - A^{-1}$ is small so that (\mathbf{y}, \mathbf{z}) is close to $A^{-1} \mathbf{b}$. Let $Q = (Q_1^T, Q_2^T)^T$ where $Q_1 \in \mathbf{R}^{C_q \times (C_q \cup C_s)}$ and $Q_2 \in \mathbf{R}^{C_s \times (C_q \cup C_s)}$. Then, under this notation,

$$\begin{aligned} \begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix} &= A^{-1} \mathbf{b} + \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} \mathbf{b} \\ &= \begin{pmatrix} (D + \theta \mathbf{e} \mathbf{e}^T)^{-1} & 0 \\ 0 & (F + \theta \mathbf{e} \mathbf{e}^T)^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} + \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} \mathbf{b}. \end{aligned}$$

Therefore, if D , F and $A + P$ are nonsingular,

$$\mathbf{y} = (D + \theta \mathbf{e} \mathbf{e}^T)^{-1} \mathbf{b}_1 + Q_1 \mathbf{b} = (I + \theta D^{-1} \mathbf{e} \mathbf{e}^T)^{-1} D^{-1} \mathbf{b}_1 + Q_1 \mathbf{b}$$

and

$$\mathbf{z} = (I + \theta F^{-1} \mathbf{e} \mathbf{e}^T)^{-1} F^{-1} \mathbf{b}_2 + Q_2 \mathbf{b}.$$

Let $\bar{\mathbf{y}} := \mathbf{y} - Q_1 \mathbf{b}$ and $\bar{\mathbf{z}} := \mathbf{z} - Q_2 \mathbf{b}$. In order to give explicit formulas for $\bar{\mathbf{y}}$ and $\bar{\mathbf{z}}$ we use the well-known Sherman-Morrison-Woodbury formula (see, for example, [12, Equation 2.1.4]), stated in the following lemma, to calculate $(I + \theta D^{-1} \mathbf{e} \mathbf{e}^T)^{-1}$ and $(I + \theta F^{-1} \mathbf{e} \mathbf{e}^T)^{-1}$.

Lemma 2.1 *If A is a nonsingular matrix in $\mathbf{R}^{n \times n}$ and $\mathbf{u}, \mathbf{v} \in \mathbf{R}^n$ satisfy $\mathbf{v}^T A^{-1} \mathbf{u} \neq -1$ then*

$$(A + \mathbf{u} \mathbf{v}^T)^{-1} = A^{-1} - \frac{A^{-1} \mathbf{u} \mathbf{v}^T A^{-1}}{1 + \mathbf{v}^T A^{-1} \mathbf{u}}. \quad (15)$$

As an immediate corollary of Lemma 2.1, notice that

$$\bar{\mathbf{y}} = \left(D^{-1} - \frac{\theta D^{-1} \mathbf{e} \mathbf{e}^T D^{-1}}{1 + \theta \mathbf{e}^T D^{-1} \mathbf{e}} \right) \mathbf{b}_1 = D^{-1} \left(I - \frac{\theta \mathbf{e} \mathbf{e}^T D^{-1}}{1 + \theta \mathbf{e}^T D^{-1} \mathbf{e}} \right) \mathbf{b}_1. \quad (16)$$

and

$$\bar{\mathbf{z}} = \left(F^{-1} - \frac{\theta F^{-1} \mathbf{e} \mathbf{e}^T F^{-1}}{1 + \theta \mathbf{e}^T F^{-1} \mathbf{e}} \right) \mathbf{b}_2 = F^{-1} \left(I - \frac{\theta \mathbf{e} \mathbf{e}^T F^{-1}}{1 + \theta \mathbf{e}^T F^{-1} \mathbf{e}} \right) \mathbf{b}_2. \quad (17)$$

Finally, we define the $(k+1) \times (k+1)$ block matrix $\tilde{S} \in \mathbf{R}^{N \times N}$ as follows:

($\tilde{\sigma}_1$) For all $q \in \{1, \dots, k\}$, let $\tilde{S}_{C_q, C_q} = 0$.

($\tilde{\sigma}_2$) For all $q, s \in \{1, \dots, k\}$ such that $q \neq s$, let

$$\tilde{S}_{C_q, C_s} = H_{q,s} \circ (\mathbf{y}^{q,s} \mathbf{e}^T + \mathbf{e}(\mathbf{z}^{q,s})^T) + c_{q,s}(H_{q,s} - \mathbf{e} \mathbf{e}^T). \quad (18)$$

($\tilde{\sigma}_3$) For all $q \in \{1, \dots, k\}$ and $i \in C_q, j \in C_{k+1}$, let

$$[\tilde{S}_{C_q, C_{k+1}}]_{ij} = [\tilde{S}_{C_{k+1}, C_q}]_{ji} = \begin{cases} -c_{q,k+1}, & \text{if } (i, j) \in E \\ c_{q,k+1} n_j^q / (r_q - n_j^q), & \text{otherwise.} \end{cases}$$

($\tilde{\sigma}_4$) Finally, for all $i, j \in C_{k+1}$, choose

$$[\tilde{S}_{C_{k+1}, C_{k+1}}]_{ij} = \begin{cases} -1, & \text{if } (i, j) \in E \text{ or } i = j \\ \gamma, & \text{if } (i, j) \notin E \end{cases} \quad (19)$$

for some scalar γ to be defined later.

We make a couple of remarks about ($\tilde{\sigma}_2$). As already noted earlier, this formula defines entries of \tilde{S}_{C_q, C_s} to be $-c_{q,s}$ in positions corresponding to edges, and $[\mathbf{y}^{q,s}]_i + [\mathbf{z}^{q,s}]_j$ in other positions. The vectors $\mathbf{y}^{q,s}$ and $\mathbf{z}^{q,s}$ are defined by (13) precisely so that, when used in this manner to define \tilde{S}_{C_q, C_s} , its row and column sums are all 0 (so that $X^* S = S X^* = 0$; the relationship $S_{C_q, C_s} \equiv \tilde{S}_{C_q, C_s}$ is given by (25) below). The system is square because the number of constraints on $S_{q,s}$ imposed by $X^* S = S X^* = 0$ after the predetermined entries are filled in is $|C_q| + |C_s|$ (one constraint for each row and column), which is the total number of entries in $\mathbf{y}^{q,s}$ and $\mathbf{z}^{q,s}$. As mentioned earlier, there is the slight additional complexity that these $|C_q| + |C_s|$ equations have a dependence of dimension 1, which explains why we needed to regularize (13) with the addition of the $\theta \mathbf{e} \mathbf{e}^T$ terms.

As a second remark about $\tilde{\sigma}_2$, we note that $\tilde{S}_{C_q, C_s} = \tilde{S}_{C_s, C_q}^T$. This is a consequence of our construction detailed above. In particular, $\mathbf{y}^{q,s} = \mathbf{z}^{s,q}$, $H_{q,s} = H_{s,q}^T$, and $D_{q,s} = F_{s,q}$ for all $q, s = 1, \dots, k$ such that $q \neq s$.

2.2 Definition of the multipliers, optimality and uniqueness

We finally come to the main theorem of this section, which provides a sufficient condition for when the k -disjoint-clique subgraph of G composed of the cliques C_1, \dots, C_k is the maximum node k -disjoint-clique subgraph of G .

Theorem 2.2 *Suppose that $G = (V, E)$ has a k -disjoint-clique subgraph G^* composed of the disjoint cliques C_1, \dots, C_k and let $C_{k+1} := V \setminus (\cup_{i=1}^k C_i)$. Let $r_i = |C_i|$ for all $i = 1, \dots, k+1$, and let $\hat{r} = \min_{i=1, \dots, k} \{r_i\}$. Let X^* be the matrix of the form (6) corresponding to the k -disjoint-clique subgraph generated by C_1, \dots, C_k . Moreover, suppose that the matrix \tilde{S} as defined by $(\tilde{\sigma}_1), \dots, (\tilde{\sigma}_4)$ satisfies*

$$\|\tilde{S}\| \leq \hat{r}. \quad (20)$$

Then X^ is optimal for (9), and G^* is the maximum node k -disjoint-clique subgraph of G . Moreover, if $\|\tilde{S}\| < \hat{r}$ and*

$$n_v^q < r_q \quad (21)$$

for all $v \in V$ and $q \in \{1, \dots, k\} - cl(v)$ then X^ is the unique optimal solution of (9) and G^* is the unique maximum node k -disjoint-clique subgraph of G .*

Proof: We will prove that (20) is a sufficient condition for optimality of X^* by defining multipliers μ , λ , η , and S and proving that if (20) holds then these multipliers satisfy the optimality conditions given by Theorem 2.1. Let us define the multipliers μ and λ by

$$\mu = \hat{r} = \min\{r_1, r_2, \dots, r_k\}, \quad (22)$$

$$\lambda_i = \frac{(1 - \hat{r}/r_q)}{2} \quad \text{for all } i \in C_q, \quad (23)$$

for all $q = 1, \dots, k$ and

$$\lambda_i = 0. \quad (24)$$

for all $i \in C_{k+1}$. Notice that by our choice of μ and λ we have

$$S_{C_q, C_q} = \hat{r}I - (\hat{r}/r_q)\mathbf{e}\mathbf{e}^T$$

for all $q = 1, \dots, k$ by (10). Moreover, we choose η such that

$$\eta_{ij} = \begin{cases} \tilde{S}_{ij} - \lambda_i - \lambda_j + 1, & \text{if } (i, j) \notin E, i \neq j \\ 0, & \text{otherwise} \end{cases}$$

for all $i, j \in V$. Note that, by our choice of η , we have

$$S_{C_q, C_s} = \begin{cases} \tilde{S}_{C_q, C_s}, & \text{if } q, s \in \{1, \dots, k+1\}, q \neq s \\ \tilde{S}_{C_{k+1}, C_{k+1}} + \hat{r}I, & \text{if } q = s = k+1. \end{cases} \quad (25)$$

by (10).

By construction, μ, λ, η , and S satisfy (10). Since the i th row sum of X^* is equal to 1 for all $i \in C_q$ for all $q = 1, \dots, k$ and is equal to 0 for all $i \in C_{k+1}$, X^* and λ satisfy the complementary slackness condition (11). Moreover,

$$\langle X^*, S \rangle = \sum_{q=1}^k \left(\sum_{i \in C_q} \sum_{j \in C_q} \frac{1}{r_q} [S_{C_q, C_q}]_{i,j} \right) = \sum_{q=1}^k \frac{1}{r_q} \left(r_q \hat{r} - r_q^2 \left(\frac{\hat{r}}{r_q} \right) \right) = 0,$$

and thus X^* and S satisfy (12). It remains to prove that (20) implies that S is positive semidefinite.

To prove that S is positive semidefinite we show that $\mathbf{x}^T S \mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbf{R}^N$ if \tilde{S} satisfies (20). Fix $\mathbf{x} \in \mathbf{R}^N$ and decompose \mathbf{x} as $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ where

$$\mathbf{x}_1(C_i) = \begin{cases} \phi_i \mathbf{e}, & i \in \{1, \dots, k\} \\ 0, & i = k+1 \end{cases}$$

for $\phi \in \mathbf{R}^k$ chosen so that $\mathbf{x}_2(C_i)^T \mathbf{e} = 0$ for $i = 1, \dots, k$, $\mathbf{x}_2(C_{k+1}) = \mathbf{x}(C_{k+1})$. Here, $\mathbf{x}(C_i)$, $\mathbf{x}_1(C_i)$, and $\mathbf{x}_2(C_i)$ denote the vectors in \mathbf{R}^{C_i} composed of the entries indexed by C_i of \mathbf{x} , \mathbf{x}_1 , and \mathbf{x}_2 respectively, for each $i = 1, \dots, k+1$. Note that \mathbf{x}_1 is in the column space of X^* . Then \mathbf{x}_1 is in the null space of S and we have

$$\mathbf{x}^T S \mathbf{x} = \mathbf{x}_2^T S \mathbf{x}_2 = \hat{r} \|\mathbf{x}_2\|^2 + \mathbf{x}_2^T \tilde{S} \mathbf{x}_2 \geq (\hat{r} - \|\tilde{S}\|) \|\mathbf{x}_2\|^2,$$

since $\mathbf{x}_2(C_i)$ is orthogonal to \mathbf{e} for all $i = 1, \dots, k$. Therefore, S is positive semidefinite, and, hence, X^* is optimal for (9) if $\|\tilde{S}\| \leq \hat{r}$.

Now suppose that $\|\tilde{S}\| < \hat{r}$ and, for all $i = 1, \dots, k$, no node in C_i is adjacent to every node in some other clique. Then X^* is optimal for (9). For all $i = 1, \dots, k$, let $\mathbf{v}_i \in \mathbf{R}^N$ be the characteristic vector of C_i . That is,

$$[\mathbf{v}_i]_j = \begin{cases} 1, & \text{if } j \in C_i \\ 0, & \text{otherwise.} \end{cases}$$

Notice that $X^* = \sum_{i=1}^k (1/r_i) \mathbf{v}_i (\mathbf{v}_i)^T$. Moreover, by complementary slackness, $\langle X^*, S \rangle = 0$ and, thus, \mathbf{v}_i is in the null space of S for all $i = 1, \dots, k$. On the other hand, consider nonzero $\mathbf{x} \in \mathbf{R}^N$ such that $\mathbf{x}^T \mathbf{v}_i = 0$ for all $i = 1, \dots, k$. That is, \mathbf{x} is orthogonal to the span of $\{\mathbf{v}_i : i = 1, \dots, k\}$. Then

$$\mathbf{x}^T S \mathbf{x} = \hat{r} \|\mathbf{x}\|^2 + \mathbf{x}^T \tilde{S} \mathbf{x} \geq (\hat{r} - \|\tilde{S}\|) \|\mathbf{x}\|^2 > 0.$$

Therefore, $\text{Null}(S) = \text{span}\{\mathbf{v}_i : i = 1, \dots, k\}$ and $\text{rank}(S) = N - k$.

Now suppose that \hat{X} is also optimal for (9). Then, by complementary slackness, $\langle \hat{X}, S \rangle = 0$ which holds if and only if $\hat{X} S = 0$. Therefore, the row and column spaces of \hat{X} lie in the null space of S . It follows immediately, since $\hat{X} \succeq 0$, that \hat{X} can be written in the form

$$\hat{X} = \sum_{i=1}^k \sigma_i \mathbf{v}_i \mathbf{v}_i^T + \sum_{i=1}^k \sum_{\substack{j=1 \\ j \neq i}}^k \omega_{i,j} \mathbf{v}_i \mathbf{v}_j^T$$

for some $\sigma \in \mathbf{R}_+^k$ and $\omega \in \Sigma^k$, where Σ^k denotes the set of $k \times k$ symmetric matrices. Now, if $\omega_{i,j} \neq 0$ for some $i \neq j$ then every entry in the block $\hat{X}(C_i, C_j) = \hat{X}(C_j, C_i)^T$ must be equal to $\omega_{i,j}$. Since each of these entries is nonzero, this implies that each node in C_i is adjacent to every node in C_j , contradicting Assumption (21). Therefore, \hat{X} has singular value decomposition $\hat{X} = \sigma_1 \mathbf{v}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{v}_k \mathbf{v}_k^T$. Moreover, since \hat{X} is optimal for (9) it must have objective value equal to that of X^* and thus

$$\sum_{i=1}^k r_i = \sum_{i=1}^N \sum_{j=1}^N X_{i,j}^* = \sum_{i=1}^N \sum_{j=1}^N \hat{X}_{i,j} = \sum_{i=1}^k \sigma_i r_i^2. \quad (26)$$

Further, since \hat{X} is feasible for (9),

$$\sigma_i r_i \leq 1 \quad (27)$$

for all $i = 1, \dots, k$. Combining (26) and (27) shows that $\sigma_i = 1/r_i$ for all $i = 1, \dots, k$ and, hence, $\hat{X} = X^*$ as required. \blacksquare

3 The Adversarial Case

Suppose that the graph $G = (V, E)$ is generated as follows. We first add k disjoint cliques with vertex sets C_1, \dots, C_k of size r_1, r_2, \dots, r_k respectively. Then, an adversary is allowed to add a set C_{k+1} of additional vertices and a number of the remaining potential edges to the graph. We will show that our adversary can add up to $O(\hat{r}^2)$ noise edges where $\hat{r} = \min\{r_1, \dots, r_k\}$ and the k -disjoint-clique subgraph composed of C_1, \dots, C_k will still be the unique maximum k -disjoint-clique subgraph of G .

The main theorem concerning the adversarial case is as follows.

Theorem 3.1 *Consider an instance of the k -disjoint-clique problem constructed according to the preceding description, namely, G contains a k -disjoint-clique graph G^* whose nodes are partitioned as $C_1 \cup \dots \cup C_k$ where $|C_q| = r_q$, $q = 1, \dots, k$, plus additional nodes denoted C_{k+1} and additional edges (which may have endpoints chosen from any of C_1, \dots, C_{k+1}). Let $\hat{r} = \min(r_1, \dots, r_k)$. Assume the following conditions are satisfied:*

1. *For all $q = 1, \dots, k$, $i \in C_q$, for all $s \in \{1, \dots, k+1\} - q$,*

$$n_i^s \leq \delta \min(r_q, r_s). \quad (28)$$

Here, δ is a scalar satisfying (29) below.

2. *$|E(G) \setminus E(G^*)| \leq \rho \hat{r}^2$, where ρ is a positive scalar depending on δ .*

Then X^ given by (6) is the unique optimal solution to (9), and G^* is the unique optimal solution of the k -disjoint clique problem.*

We remark that two of the conditions imposed in this theorem are, up to the constant factors, the best possible according to the following information-theoretic arguments. First, if $n_i^s = r_s$, then node i could be inserted into clique s , so the partitioning between C_s and C_q would no longer be uniquely determined. This shows the necessity of the condition $n_i^s \leq O(r_s)$. The condition that $|E(G) \setminus E(G^*)| \leq \rho \hat{r}^2$ is necessary, up to the constant factor, because if $|E(G) \setminus E(G^*)| \geq \hat{r}(\hat{r} - 1)/2$, then we could interconnect an arbitrary set of \hat{r} nodes chosen from among the existing cliques with edges to make a new clique out of them, again spoiling the uniqueness of the decomposition.

An argument for the necessity of the condition that $n_i^s \leq \delta r_q$ is not apparent, so possibly there is a strengthening of this theorem that drops that condition.

The remainder of this section is the proof of this theorem. As might be expected, the proof hinges on establishing (20); once this inequality is established, then Theorem 2.2 completes the argument.

As before, let r_{k+1} denote $|C_{k+1}|$. For the remainder of the proof, to simplify the notation, we assume that $r_{k+1} \leq 2\rho \hat{r}^2$. The reason is that since $|E(G) \setminus E(G^*)| \leq \rho \hat{r}^2$ by assumption, if $r_{k+1} > 2\rho \hat{r}^2$ then C_{k+1} would include one or more isolated nodes (i.e, nodes of degree 0), and these nodes can simply be deleted in a preliminary phase of the algorithm. (The algorithm still works with an arbitrary number of isolated nodes in $G \setminus G^*$, but the notation in the proof requires some needless additional complexity.)

Recall that the construction of the multipliers presented in Section 2 depended on two scalars θ in (13) and γ in (19): choose $\theta = 1$ and $\gamma = 0$.

We impose the assumption that $\delta \in (0, 0.382)$. The constant 0.382 is chosen so that

$$0 < \delta < (1 - \delta)^2. \quad (29)$$

We will show that, under this assumption, there exists some $\beta > 0$ depending only on δ such that

$$\|\tilde{S}_{C_q, C_s}\|^2 \leq \beta \|\mathbf{b}^{q,s}\|_1$$

for all $q, s \in \{1, \dots, k\}$ such that $q \neq s$.

Choose $q, s \in \{1, \dots, k\}$ such that $q \neq s$ and let D , F , H , \mathbf{b} , and c be defined as in Section 2.1. Without loss of generality we may assume that $r_q \leq r_s$. Moreover, let \mathbf{y} and \mathbf{z} be the solution of the system (13) and define A, Q, P as in Section 2.1.

We begin by showing that, under this assumption, \mathbf{y} and \mathbf{z} are uniquely determined. Note that, since $n_i^s = r_s - D_{ii} \leq \delta r_s$ for all $i \in C_q$ and $n_i^q = r_q - F_{ii} \leq \delta r_q$ for all $i \in C_s$ by Assumption (28), D and F are nonsingular and hence A is nonsingular. Moreover,

$$A + P = A(I + A^{-1}P)$$

and, hence, $A + P$ is nonsingular if $\|A^{-1}P\| < 1$. Note that, for all $t > 0$, we have

$$\lambda_{\min}(D + tee^T) \geq \lambda_{\min}(D) = \min_{i \in C_q} D_{ii} \quad (30)$$

since $\mathbf{e}\mathbf{e}^T \succeq 0$ where $\lambda_{\min}(D + tee^T)$ is the smallest eigenvalue of the symmetric matrix $D + tee^T$. Taking $t = 1$ in (30) shows that

$$\|(D + \mathbf{e}\mathbf{e}^T)^{-1}\| \leq \|D^{-1}\| = \frac{1}{\min_{i \in C_q} D_{ii}} \leq \frac{1}{(1 - \delta)r_s} \quad (31)$$

since, for each $i \in C_q$, we have

$$(1 - \delta)r_s \leq D_{ii} \leq r_s$$

by Assumption (28). Similarly,

$$\|(F + \mathbf{e}\mathbf{e}^T)^{-1}\| \leq \|F^{-1}\| = \frac{1}{\min_{j \in C_s} F_{jj}} \leq \frac{1}{(1 - \delta)r_q}. \quad (32)$$

Combining (31) and (32) we have

$$\begin{aligned} \|A^{-1}\| &= \max\{\|(D + \mathbf{e}\mathbf{e}^T)^{-1}\|, \|(F + \mathbf{e}\mathbf{e}^T)^{-1}\|\} \\ &\leq \frac{1}{(1 - \delta) \min\{r_q, r_s\}} \\ &= \frac{1}{(1 - \delta)r_q}. \end{aligned} \quad (33)$$

On the other hand,

$$\|P\| = \|H - \mathbf{e}\mathbf{e}^T\| \leq \|H - \mathbf{e}\mathbf{e}^T\|_F = \left(\sum_{i \in C_q} \sum_{j \in C_s} (H_{ij} - 1)^2 \right)^{1/2} \leq \sqrt{\delta} r_q \quad (34)$$

since $H_{ij} - 1$ is equal to -1 in the case that $(i, j) \in E$ and 0 otherwise and there at most δr_q^2 edges between C_q and C_s by Assumption (28). Therefore, since $\delta < (1 - \delta)^2$ by Assumption (29), we have

$$\|A^{-1}P\| \leq \|A^{-1}\| \|P\| \leq \frac{\sqrt{\delta}}{1 - \delta} < 1$$

and, thus, $A + P$ is nonsingular and \mathbf{y} and \mathbf{z} are uniquely determined.

Now, recall that

$$\tilde{S}_{C_q, C_s} = H \circ (\mathbf{y}\mathbf{e}^T + \mathbf{e}\mathbf{z}^T) - c(\mathbf{e}\mathbf{e}^T - H).$$

In order to calculate an upper bound on $\|\tilde{S}_{C_q, C_s}\|$ we write \tilde{S}_{C_q, C_s} as

$$\tilde{S}_{C_q, C_s} = m_1 + m_2 + m_3 + m_4 + m_5 \quad (35)$$

where

$$\begin{aligned} m_1 &:= H \circ (\bar{\mathbf{y}}\mathbf{e}^T), & m_2 &:= H \circ (\mathbf{e}\bar{\mathbf{z}}^T), & m_3 &:= H \circ (Q_1\mathbf{b}\mathbf{e}^T), \\ m_4 &:= H \circ (\mathbf{e}(Q_2\mathbf{b})^T), & m_5 &:= -c(\mathbf{e}\mathbf{e}^T - H) \end{aligned} \quad (36)$$

and apply the triangle inequality to obtain

$$\|\tilde{S}_{C_q, C_s}\| \leq \sum_{i=1}^5 \|m_i\|. \quad (37)$$

Throughout our analysis of $\|\tilde{S}_{C_q, C_s}\|$ we will use the following series of inequalities. For any $W \in \mathbf{R}^{m \times n}$, $\mathbf{u} \in \mathbf{R}^m$ and $\mathbf{v} \in \mathbf{R}^n$, we have

$$\begin{aligned} \|W \circ \mathbf{u}\mathbf{v}^T\| &= \|\text{Diag}(\mathbf{u}) \cdot W \cdot \text{Diag}(\mathbf{v})\| \leq \|\text{Diag}(\mathbf{u})\| \|\text{Diag}(\mathbf{v})\| \|W\| \\ &= \|\mathbf{u}\|_\infty \|\mathbf{v}\|_\infty \|W\|. \end{aligned} \quad (38)$$

On the other hand,

$$\begin{aligned}\|W \circ \mathbf{u}\mathbf{v}^T\| &= \|\text{Diag}(\mathbf{u}) \cdot W \cdot \text{Diag}(\mathbf{v})\| \\ &\leq \|\mathbf{v}\|_\infty \|\text{Diag}(\mathbf{u}) \cdot W\| \leq \|\mathbf{v}\|_\infty \|\text{Diag}(\mathbf{u}) \cdot W\|_F \\ &= \|\mathbf{v}\|_\infty \left(\sum_{i=1}^m \mathbf{u}_i^2 \|W(i, :)\|^2 \right)^{1/2} \end{aligned} \quad (39)$$

$$\leq \|\mathbf{u}\| \|\mathbf{v}\|_\infty \max_{i=1, \dots, m} \|W(i, :)\| \quad (40)$$

and

$$\|W \circ \mathbf{u}\mathbf{v}^T\| \leq \|\mathbf{u}\|_\infty \|\mathbf{v}\| \max_{j=1, \dots, n} \|W(:, j)\| \quad (41)$$

where $W(i, :)$ and $W(:, j)$ denote the i th and j th row and column of W .

We begin with $\|m_1\|$. Applying the bound (40) with $W = H$, $\mathbf{u} = \bar{\mathbf{y}}$, and $\mathbf{v} = \mathbf{e}$ we have

$$\|m_1\|^2 \leq \max_{i \in C_q} D_{ii} \|\bar{\mathbf{y}}\|^2. \quad (42)$$

Here, we used the fact that $\max_{i \in C_q} \|H(i, :)\| = \max_{i \in C_q} D_{ii}^{1/2}$ since the i th row of H contains exactly $r_s - n_i^s$ 1's. Thus, since

$$\|\bar{\mathbf{y}}\| \leq \|(D + \mathbf{e}\mathbf{e}^T)^{-1}\| \|\mathbf{b}_1\| \leq \frac{\|\mathbf{b}_1\|}{\min_{i \in C_q} D_{ii}} \leq \frac{\|\mathbf{b}_1\|}{(1 - \delta)r_s},$$

it follows immediately that

$$\|m_1\|^2 \leq \frac{1}{(1 - \delta)^2 r_s} \|\mathbf{b}_1\|^2 \quad (43)$$

since $D_{ii} \leq r_s$ for all $i \in C_q$. By an identical calculation, we have

$$\|m_2\|^2 \leq \frac{1}{(1 - \delta)^2 r_q} \|\mathbf{b}_2\|^2. \quad (44)$$

Next, applying (40) with $W = H$, $\mathbf{u} = Q_1 \mathbf{b}$, $\mathbf{v} = \mathbf{e}$ yields

$$\|m_3\|^2 \leq \max_{i \in C_q} D_{ii} \|Q_1 \mathbf{b}\|^2 \leq r_s \|Q_1 \mathbf{b}\|^2 \leq r_s \|Q_1\|^2 \|\mathbf{b}\|^2$$

since $\max_i D_{ii} \leq r_s$. To derive an upper bound on $\|m_3\|^2$ we first derive an upper bound on $\|Q_1\|$.

Note that

$$Q = (A + P)^{-1} - A^{-1} = ((I + A^{-1}P)^{-1} - I)A^{-1} = \sum_{\ell=1}^{\infty} (-A^{-1}P)^\ell A^{-1} \quad (45)$$

since $(I + X)^{-1} = \sum_{\ell=0}^{\infty} (-X)^\ell$ for all X such that $\|X\| < 1$ by Taylor's Theorem. Notice that

$$A^{-1}P = \begin{pmatrix} 0 & P_1 \\ P_2 & 0 \end{pmatrix}$$

where

$$P_1 = (D + \theta \mathbf{e} \mathbf{e}^T)^{-1} (H - \theta \mathbf{e} \mathbf{e}^T), \quad P_2 = (F + \theta \mathbf{e} \mathbf{e}^T)^{-1} (H^T - \theta \mathbf{e} \mathbf{e}^T).$$

It follows immediately that

$$Q = \sum_{\ell=0}^{\infty} \left(\begin{pmatrix} (P_1 P_2)^{\ell+1} & 0 \\ 0 & (P_2 P_1)^{\ell+1} \end{pmatrix} - \begin{pmatrix} 0 & (P_1 P_2)^{\ell} P_1 \\ (P_2 P_1)^{\ell} P_2 & 0 \end{pmatrix} \right) A^{-1} \quad (46)$$

since, for any integer $\ell \geq 1$

$$\begin{pmatrix} 0 & P_1 \\ P_2 & 0 \end{pmatrix}^{\ell} = \begin{cases} \begin{pmatrix} (P_1 P_2)^{\ell/2} & 0 \\ 0 & (P_2 P_1)^{\ell/2} \end{pmatrix}, & \text{if } \ell \text{ even} \\ \begin{pmatrix} 0 & (P_1 P_2)^{(\ell-1)/2} P_1 \\ (P_2 P_1)^{(\ell-1)/2} P_2 & 0 \end{pmatrix}, & \text{if } \ell \text{ odd.} \end{cases}$$

Therefore,

$$\|Q_1\| \leq \|(D + \theta \mathbf{e} \mathbf{e}^T)^{-1}\| \sum_{\ell=1}^{\infty} \|P_1 P_2\|^{\ell} + \|P_1\| \|(F + \theta \mathbf{e} \mathbf{e}^T)^{-1}\| \sum_{\ell=0}^{\infty} \|P_1 P_2\|^{\ell} \quad (47)$$

and

$$\|Q_2\| \leq \|(F + \theta \mathbf{e} \mathbf{e}^T)^{-1}\| \sum_{\ell=1}^{\infty} \|P_1 P_2\|^{\ell} + \|P_2\| \|(D + \theta \mathbf{e} \mathbf{e}^T)^{-1}\| \sum_{\ell=0}^{\infty} \|P_1 P_2\|^{\ell}. \quad (48)$$

Substituting (31), (32) and (34) into (47) yields

$$\|Q_1\| \leq \frac{1}{(1-\delta)r_s} \sum_{\ell=1}^{\infty} \left(\frac{\delta}{(1-\delta)^2} \right)^{\ell} + \frac{\delta^{1/2}}{(1-\delta)^2 r_s} \sum_{\ell=0}^{\infty} \left(\frac{\delta}{(1-\delta)^2} \right)^{\ell} \leq \tilde{c}/r_s \quad (49)$$

where

$$\tilde{c} = \frac{2 \max\{\delta/(1-\delta), \sqrt{\delta}\}}{(1-\delta)^2 - \delta}$$

since

$$\|P_1 P_2\| \leq \|H - \mathbf{e} \mathbf{e}^T\|^2 \|D^{-1}\| \|F^{-1}\| \leq \frac{\delta}{(1-\delta)^2}.$$

Note that Assumption (29) ensures that the infinite series in (49) converge. It follows that

$$\|m_3\|^2 \leq \frac{\tilde{c}^2}{r_q} \|\mathbf{b}\|^2. \quad (50)$$

On the other hand,

$$\begin{aligned} \|Q_2\| &\leq \frac{1}{(1-\delta)r_q} \sum_{\ell=1}^{\infty} \left(\frac{\delta}{(1-\delta)^2} \right)^{\ell} + \frac{\sqrt{\delta}}{(1-\delta)^2 r_s} \sum_{\ell=0}^{\infty} \left(\frac{\delta}{(1-\delta)^2} \right)^{\ell} \\ &\leq \tilde{c}/r_q \end{aligned} \quad (51)$$

since $r_s \geq r_q$. Thus, applying (41) with $W = H$, $\mathbf{u} = \mathbf{e}$, $\mathbf{v} = Q_2 \mathbf{b}$ we have

$$\|m_4\|^2 \leq r_q \|Q_2\|^2 \|\mathbf{b}\|^2 \leq \frac{\tilde{c}^2}{r_q} \|\mathbf{b}\|^2. \quad (52)$$

Finally,

$$\begin{aligned} \|m_5\|^2 &= \|c(H - \mathbf{e}\mathbf{e}^T)\|^2 \leq \|c(H - \mathbf{e}\mathbf{e}^T)\|_F^2 \\ &= c \sum_{i \in C_q} \sum_{j \in C_s} (H_{ij} - 1)^2 \\ &= c \sum_{i \in C_q} n_i^s = \|\mathbf{b}_1\|_1. \end{aligned} \quad (53)$$

Therefore, there exists $\beta \in \mathbf{R}$ such that

$$\|\tilde{S}_{C_q, C_s}\|^2 \leq \beta \frac{\|\mathbf{b}\|^2}{r_q} + \|\mathbf{b}\|_1 \quad (54)$$

where β depends only on δ . Moreover, since $\|\mathbf{b}\|^2 \leq \|\mathbf{b}\|_1 \|\mathbf{b}\|_\infty$ and

$$\|\mathbf{b}\|_\infty = c \cdot \max \left\{ \max_{i \in C_q} n_i^s, \max_{i \in C_s} n_i^q \right\} \leq \delta c \min\{r_q, r_s\} = \delta c r_q \quad (55)$$

by Assumption (28), there exists $\tilde{\beta}$ depending only on δ such that

$$\|\tilde{S}_{C_q, C_s}\|^2 \leq \tilde{\beta} \|\mathbf{b}\|_1 \quad (56)$$

as required.

Next, consider $\tilde{S}_{C_q, C_{k+1}}$ for some $q \in \{1, \dots, k\}$. Recall that

$$[\tilde{S}_{C_q, C_{k+1}}]_{ij} = \begin{cases} -c, & \text{if } (i, j) \in E \\ cn_j/(r_q - n_j), & \text{otherwise} \end{cases}$$

where $n_j = n_j^q$ is the number of edges from $j \in C_{k+1}$ to C_q for each $j \in C_{k+1}$. Hence,

$$\begin{aligned} \|\tilde{S}_{C_q, C_{k+1}}\|^2 &\leq \|\tilde{S}_{C_q, C_{k+1}}\|_F^2 \\ &= \sum_{j \in C_{k+1}} \left(n_j c^2 + (r_q - n_j) \left(\frac{n_j c}{r_q - n_j} \right)^2 \right) \\ &\leq c^2 \sum_{j \in C_{k+1}} \left(n_j + \frac{\delta n_j}{(1 - \delta)} \right) \\ &= \frac{c^2}{1 - \delta} |E(C_q, C_{k+1})| \end{aligned} \quad (57)$$

where $E(C_q, C_{k+1})$ is the set of edges from C_q to C_{k+1} . Similarly, by our choice of $\gamma = 0$ in $(\tilde{\sigma}_4)$, we have

$$\|\tilde{S}_{C_{k+1}, C_{k+1}}\|^2 \leq \|\tilde{S}_{C_{k+1}, C_{k+1}}\|_F^2 = r_{k+1} + 2|E(C_{k+1}, C_{k+1})|. \quad (58)$$

Let B be the vector obtained by concatenating $\mathbf{b}^{q,s}$ for all $q, s \in \{1, \dots, k\}$. Then, there exist scalars $\hat{c}_1, \hat{c}_2 \in \mathbf{R}$ depending only on δ such that

$$\begin{aligned} \sum_{q=1}^{k+1} \sum_{s=1}^{k+1} \|\tilde{S}_{C_q, C_s}\|^2 &= \sum_{\substack{q, s \in \{1, \dots, k\} \\ q \neq s}} \|\tilde{S}_{C_q, C_s}\|^2 + 2 \sum_{q=1}^k \|\tilde{S}_{C_q, C_{k+1}}\|^2 + \|\tilde{S}_{C_{k+1}, C_{k+1}}\|^2 \\ &\leq \hat{c}_1 \|B\|_1 + \hat{c}_2 \sum_{q=1}^{k+1} |E(C_q, C_{k+1})| + r_{k+1} \end{aligned}$$

by (56), (57) and (58). It follows that, since $\|\mathbf{b}^{q,s}\|_1 \leq 2|E(C_q, C_s)|$ for all $q, s \in \{1, \dots, k\}$ such that $q \neq s$, there exists $\hat{c}_3 \geq 0$ depending only on δ such that

$$\sum_{q=1}^{k+1} \sum_{s=1}^{k+1} \|\tilde{S}_{C_q, C_s}\|^2 \leq \hat{c}_3 R + r_{k+1}$$

where $R := |E(G) \setminus E(G^*)|$ is the number of edges of G not contained in the k -disjoint-clique subgraph G^* composed of C_1, \dots, C_k . The hypothesis of the theorem is that $R \leq \rho \hat{r}^2$. We have also assumed earlier that $r_{k+1} \leq 2\rho \hat{r}^2$. Hence the sum of the squares of the 2-norms of the blocks of \tilde{S} is at most $(\hat{c}_3 + 2)\rho \hat{r}^2$. Therefore, there exists some $\rho > 0$ depending only on δ such that the preceding inequality implies $\|\tilde{S}\| \leq \hat{r}$. This proves the theorem.

4 The Randomized Case

Let C_1, C_2, \dots, C_{k+1} be disjoint vertex sets of sizes r_1, \dots, r_{k+1} respectively, and let $V = \cup_{i=1}^{k+1} C_i$. We construct the edge set of the graph $G = (V, E)$ as follows:

- (Ω_1) For each $q = 1, \dots, k$, and each $i \in C_q, j \in C_q$ such that $i \neq j$ we add (i, j) to E .
- (Ω_2) Each of the remaining possible edges is added to E independently at random with probability $p \in (0, 1)$.

Notice that, by our construction of E , the graph $G = (V, E)$ has a k -disjoint-clique subgraph G^* with cliques indexed by the vertex sets C_1, \dots, C_k . We wish to determine which random graphs G generated according to (Ω_1) and (Ω_2) have maximum node k -disjoint-clique subgraph equal to G^* and can be found with high probability via solving (9). We begin by providing a few results concerning random matrices with independently identically distributed (i.i.d.) entries of mean 0.

4.1 Results on norms of random matrices

Consider the probability distribution \mathbf{P} for a random variable x defined as follows:

$$x = \begin{cases} 1 & \text{with probability } p, \\ -p/(1-p) & \text{with probability } 1-p. \end{cases}$$

It is easy to check that the mean of x is 0 and the variance of x is $\sigma^2 = p/(1-p)$. In this section we provide a few results concerning random matrices with entries independently identically distributed (i.i.d.) according to \mathbf{P} . We first recall a theorem of Geman [11] which provides a bound on the largest singular value of a random matrix with independently identically distributed (i.i.d.) entries of mean 0.

Theorem 4.1 *Let A be a $\lceil yn \rceil \times n$ matrix whose entries are chosen according to \mathbf{P} for fixed $y \in \mathbf{R}_+$. Then, with probability at least $1 - c_1 \exp(-c_2 n^{c_3})$ where $c_1 > 0$, $c_2 > 0$, and $c_3 > 0$ are constants depending on p and y ,*

$$\|A\| \leq c_4 \sqrt{n}$$

for some constant c_4 .

Note that this theorem is not stated exactly in this form in [11], but can be deduced by taking $k = n^q$ for a q satisfying $(2\alpha + 4)q < 1$ in the equations on pp. 255-256. A similar theorem due to Füredi and Komlós [9] exists for symmetric matrices A with entries distributed according to \mathbf{P} .

Theorem 4.2 *For all integers i, j , $1 \leq j \leq i \leq n$, let A_{ij} be distributed according to \mathbf{P} . Define symmetrically $A_{ij} = A_{ji}$ for all $i < j$.*

Then the random symmetric matrix $A = [A_{ij}]$ satisfies

$$\|A\| \leq 3\sigma\sqrt{n}$$

with probability at least to $1 - \exp(-cn^{1/6})$ for some $c > 0$ that depends on σ .

As in Theorem 4.1, the theorem is not stated exactly in this manner in [9]; the stated form of the theorem can be deduced by taking $k = (\sigma/K)^{1/3} n^{1/6}$ and $v = \sigma\sqrt{n}$ in the inequality

$$P(\max |\lambda| > 2\sigma\sqrt{n} + v) < \sqrt{n} \exp(-kv/(2\sqrt{n} + v))$$

on p. 237.

Next, we state a version of the well-known Chernoff bounds which provides a bound on the tail distribution of a sum of independent Bernoulli trials (see [15, Theorems 4.4 and 4.5]).

Theorem 4.3 (Chernoff Bounds) *Let X_1, \dots, X_k be a sequence of k independent Bernoulli trials, each succeeding with probability p so that $E(X_i) = p$. Let $S = \sum_{i=1}^k X_i$ be the binomially distributed variable describing the total number of successes. Then for $\delta > 0$*

$$P(S > (1 + \delta)pk) \leq \left(\frac{e^\delta}{(1 + \delta)^{(1 + \delta)}} \right)^{pk}. \quad (59)$$

It follows that for all $a \in (0, p\sqrt{k})$,

$$P(|S - pk| > a\sqrt{k}) \leq 2 \exp(-a^2/p). \quad (60)$$

The final theorem of this section is as follows (see [2, Theorem 4]).

Theorem 4.4 *Let A be an $n \times N$ matrix whose entries are chosen according to \mathbf{P} . Suppose also that $\log N \leq \sqrt{n}$. Let \tilde{A} be defined as follows. For (i, j) such that $A_{ij} = 1$, we define $\tilde{A}_{ij} = 1$. For entries (i, j) such that $A_{ij} = -p/(1-p)$, we take $\tilde{A}_{ij} = -n_j/(n - n_j)$, where n_j is the number of 1's in column j of A . Then there exist $c_1 > 0$ and $c_2 \in (0, 1)$ depending on p such that*

$$P(\|A - \tilde{A}\|_F^2 \leq c_1 N) \geq 1 - (2/3)^N - Nc_2^n. \quad (61)$$

4.2 A bound on $\|\tilde{S}\|$ in the randomized case

Suppose that the random graph $G = (V, E)$ containing k -disjoint-clique subgraph G^* composed of cliques C_1, \dots, C_k is constructed according to (Ω_1) and (Ω_2) with probability p . Let C_{k+1} , r_1, \dots, r_{k+1} , \hat{r} , and N be defined as in Section 2.1. Further, let $\theta = 1 - p$ in (13) and $\gamma = p/(1 - p)$ in (19). We begin by stating the main theorem of the section.

Theorem 4.5 *Suppose that $G = (V, E)$ has a k -disjoint-clique subgraph G^* composed of the cliques C_1, \dots, C_k and let $C_{k+1} := V \setminus (\cup_{i=1}^k C_i)$. Let $r_i = |C_i|$ for all $i = 1, \dots, k+1$ and suppose that $r_q \leq \hat{r}^{3/2}$ for all $q = 1, 2, \dots, k$ where $\hat{r} = \min_{i=1, \dots, k} \{r_i\}$. Then there exists some $\beta_1, \beta_2 > 0$ depending only on p such that*

$$\|\tilde{S}\| \leq \beta_1 \left(\sum_{s=1}^k r_s^2 \right)^{1/2} \left(\sum_{q=1}^k \frac{1}{r_q} \right)^{1/2} + \beta_2 \sqrt{N} \quad (62)$$

with probability tending exponentially to 1 as \hat{r} approaches ∞ .

This theorem is meant to be used in conjunction with Theorem 2.2. In particular, if the right-hand side of (62) is less than \hat{r} , then the planted graph G^* may be recovered.

It is clear from (20) and the second term on the right-hand side of (62) that the sufficient conditions for uniqueness and optimality given by Theorem 2.2 cannot be satisfied unless $N = O(\hat{r}^2)$. We now give a few examples of values for r_1, \dots, r_{k+1} that fulfill (62).

1. Consider the case $k = 1$, i.e., a single large clique. In this case, taking $r_1 = \text{const} \cdot N^{1/2}$ satisfies (20) since the first term on the right is $O(N^{1/4})$.
2. Suppose $k > 1$ and $r_1 = \dots = r_k = \text{const} \cdot N^\alpha$. In this case, the first parenthesized factor on the right in (62) is $O(k^{1/2} N^\alpha)$ while the second is $O(k^{1/2} N^{-\alpha/2})$, and therefore the first term is $O(k N^{\alpha/2})$. For (20) to hold, we need this term to be $O(\hat{r}) = O(N^\alpha)$, which is valid as long as $k = O(N^{\alpha/2})$. We also need $\alpha \geq 1/2$ as noted above to handle the second term on the right. For example, for $\alpha = 1/2$ the algorithm can find as many as $O(N^{1/4})$ cliques of this size. For $\alpha = 2/3$, the algorithm can find as many as $O(N^{1/3})$ cliques of this size, which is the maximum possible since the cliques are disjoint and N is the number of nodes.
3. The cliques may also be of different sizes. For example, if there is one large clique of size $O(N^{2/3})$ and $N^{1/6}$ smaller cliques of size $O(N^{1/2})$, then $\hat{r} = O(N^{1/2})$, the first parenthesized factor in (62) is $N^{2/3}$ while the second is $N^{-1/6}$, so the entire first factor is $O(N^{1/2}) = O(\hat{r})$.

We note that the results for random noise in the k -disjoint-clique problem are much better than the results for adversary-chosen noise. In the case of adversary-chosen noise, the number of allowable noise edges is bounded above by a constant times the number of edges in the smallest clique. In the case of random noise, the number of allowable noise edges is bounded above by a scalar multiple of the number of potential edges in the complement of the planted k -disjoint-clique subgraph. Thus, the number of allowable random noise edges can be as much as the square of the number of edges in the smallest clique (e.g., if there are $N^{1/4}$ cliques each of size $N^{1/2}$, then the smallest clique has $O(N)$ edges versus $O(N^2)$ random noise edges).

We do not know whether the bound in (20) is the best possible. For instance, there is no obvious barrier preventing the algorithm from recovering as many as $N^{1/2}$ planted cliques of size $N^{1/2}$ in a random graph, but our analysis does not carry through to this case.

The remainder of this section is devoted to the proof of Theorem 4.5. We write \tilde{S} as

$$\tilde{S} = \tilde{S}_1 + \tilde{S}_2 + \tilde{S}_3 + \tilde{S}_4 + \tilde{S}_4^T$$

where $\tilde{S}_i \in \mathbf{R}^{N \times N}$, $i = 1, \dots, 4$ are $(k+1)$ by $(k+1)$ block matrices such that

$$\begin{aligned} [\tilde{S}_1]_{C_q, C_s} &= \begin{cases} \tilde{S}_{C_q, C_s}, & \text{if } q, s \in \{1, \dots, k\}, q \neq s \\ 0, & \text{otherwise} \end{cases} \\ [\tilde{S}_2]_{C_q, C_s} &= \begin{cases} R_{C_q, C_s}, & \text{if } q, s \in \{1, \dots, k\} \\ \hat{S}_{C_q, C_{k+1}}, & \text{if } s = k+1 \\ \hat{S}_{C_{k+1}, C_s}, & \text{if } q = k+1 \\ \tilde{S}_{C_{k+1}, C_{k+1}}, & \text{if } q = s = k+1 \end{cases} \\ [\tilde{S}_3]_{C_q, C_s} &= \begin{cases} -R_{C_q, C_s}, & \text{if } q, s \in \{1, \dots, k\} \\ 0, & \text{otherwise} \end{cases} \\ [\tilde{S}_4]_{C_q, C_s} &= \begin{cases} \tilde{S}_{C_q, C_{k+1}} - \hat{S}_{C_q, C_{k+1}}, & \text{if } s = k+1, q \in \{1, \dots, k\} \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

where $R \in \mathbf{R}^{N \times N}$ is a symmetric random matrix with independently identically distributed entries such that

$$R_{ij} = \begin{cases} -1, & \text{with probability } p \\ p/(1-p), & \text{with probability } 1-p \end{cases}$$

and $\hat{S} \in \mathbf{R}^{N \times N}$ such that

$$\hat{S}_{ij} = \begin{cases} -1, & \text{if } (i, j) \in E \\ p/(1-p), & \text{otherwise.} \end{cases}$$

Notice that, by Theorem 4.2, there exists some $\kappa_1, \kappa_2, \kappa_3 > 0$ such that

$$P\left(\|\tilde{S}_2\| + \|\tilde{S}_3\| \geq \kappa_1 \sqrt{N}\right) \leq \kappa_2 \exp(-\kappa_3 N^{1/6}). \quad (63)$$

Moreover, by Theorem 4.4, there exists $\kappa_4 > 0$ and $\kappa_5, \kappa_6 \in (0, 1)$ such that

$$P\left(\|\tilde{S}_4\| \geq \kappa_4 \sqrt{N}\right) \leq \kappa_5^N + N\kappa_6^N. \quad (64)$$

Hence, there exists some scalar β_2 depending only on p such that

$$\|\tilde{S}\| \leq \|\tilde{S}_1\| + \beta_2 \sqrt{N}$$

with probability tending exponentially to 1 as $\hat{r} \rightarrow \infty$. It remains to prove that

$$\|\tilde{S}_1\| = O \left(\left(\sum_{s=1}^k r_s^2 \right)^{1/2} \left(\sum_{q=1}^k \frac{1}{r_q} \right)^{1/2} \right)$$

with probability tending exponentially to 1 as \hat{r} approaches ∞ .

To do so, consider two vertex sets C_q and C_s such that $q \neq s$. Without loss of generality we may assume that $r_q \leq r_s$. Define $H = H_{q,s}$, $D = D_{q,s}$, $F = D_{s,q}$, $\mathbf{b} = \mathbf{b}^{q,s}$, $c = c_{q,s}$, \mathbf{y} , \mathbf{z} , A , and P as in Section 2.1. The following theorem provides an upper bound on the spectral norm of \tilde{S}_{C_q, C_s} for $q \neq s$, that holds with probability tending exponentially to 1 as \hat{r} approaches ∞ .

Theorem 4.6 *Suppose that r_q and r_s satisfy*

$$r_q \leq r_s \leq r_q^{3/2}. \quad (65)$$

Then there exists $\tilde{B}_1 > 0$ depending only on p such that

$$\|[\tilde{S}_1]_{C_q, C_s}\| = \|\tilde{S}_{C_q, C_s}\| \leq \tilde{B}_1 \frac{r_s}{\sqrt{r_q}} \quad (66)$$

with probability tending exponentially to 1 as \hat{r} approaches ∞ .

The remainder of this section is devoted to the proof of Theorem 4.6; Theorem 4.5 will then be established as an immediate consequence.

Recall that $\tilde{S}_{C_q, C_s} = H \circ (\mathbf{y}\mathbf{e}^T + \mathbf{e}\mathbf{z}^T) - c(\mathbf{e}\mathbf{e}^T - H)$. We begin by showing that $A + P$ is nonsingular and, hence, \mathbf{y} and \mathbf{z} are uniquely determined. Let $\delta := (1-p)/(2p)$. Recall that $n_i^s = r_s - D_{ii}$ corresponds to r_s independent Bernoulli trials each succeeding with probability equal to p and, hence,

$$P(n_i^s \geq (1+\delta)pr_s) = P(r_s - D_{ii} \geq (1+\delta)pr_s) \leq \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}} \right)^{pr_s} \quad (67)$$

for each $i \in C_q$ by Theorem 4.3. Rearranging, we have that $D_{ii} \geq (\theta - \delta p)r_s$ with probability at least

$$1 - \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}} \right)^{pr_s}$$

for each $i \in C_q$. Similarly,

$$P(n_i^q \leq (1+\delta)pr_q) = P(F_{ii} \geq (\theta - \delta p)r_q) \geq 1 - \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}} \right)^{pr_q} \quad (68)$$

for all $i \in C_s$. Therefore, by the union bound, $r_s - D_{ii} \leq (1 + \delta)pr_s$ for all $i \in C_q$ and $r_q - F_{ii} \leq (1 + \delta)pr_q$ for all $i \in C_s$, and, hence, D, F are nonsingular with probability at least

$$\begin{aligned} & 1 - r_s \left(\frac{e^\delta}{(1 + \delta)^{(1 + \delta)}} \right)^{pr_q} - r_q \left(\frac{e^\delta}{(1 + \delta)^{(1 + \delta)}} \right)^{pr_s} \\ & \geq 1 - (r_q + r_s) \left(\frac{e^\delta}{(1 + \delta)^{(1 + \delta)}} \right)^{p\hat{r}}. \end{aligned} \quad (69)$$

Moreover, applying (30) shows that $D + \theta \mathbf{e}\mathbf{e}^T$ and $F + \theta \mathbf{e}\mathbf{e}^T$ are nonsingular and

$$\|(D + \theta \mathbf{e}\mathbf{e}^T)^{-1}\| \leq \|D^{-1}\| \leq \frac{1}{(\theta - \delta p)r_s}, \quad (70)$$

$$\|(F + \theta \mathbf{e}\mathbf{e}^T)^{-1}\| \leq \|F^{-1}\| \leq \frac{1}{(\theta - \delta p)r_q}, \quad (71)$$

with probability at least (69). It follows immediately that A is nonsingular and

$$\begin{aligned} \|A^{-1}\| &= \max\{\|(D + \theta \mathbf{e}\mathbf{e}^T)^{-1}\|, \|(F + \theta \mathbf{e}\mathbf{e}^T)^{-1}\|\} \\ &\leq \frac{1}{(\theta - \delta p) \min\{r_q, r_s\}} = \frac{1}{(\theta - \delta p)r_q} \end{aligned} \quad (72)$$

with probability at least (69).

Recall that, in the case that A is nonsingular, it suffices to prove that $\|A^{-1}\|\|P\| < 1$ to show that $A + P$ is nonsingular. Moreover, recall that $\theta = 1 - p$ is chosen to ensure that the entries of $H - \theta \mathbf{e}\mathbf{e}^T$ have expected value equal to 0. We can extend $H - \theta \mathbf{e}\mathbf{e}^T$ to an $r_s \times r_s$ random matrix \tilde{P} with entries i.i.d. with expected value equal to 0 by adding $r_s - r_q$ rows with entries i.i.d. such that each additional entry takes value equal to $-\theta$ with probability p and value equal to p with probability $1 - p$. Therefore, by Theorem 4.1

$$\|P\| = \|H - \theta \mathbf{e}\mathbf{e}^T\| \leq \|\tilde{P}\| \leq \gamma_1 \sqrt{r_s} \quad (73)$$

for some $\gamma_1 > 0$ depending only on p with probability at least $1 - \bar{c}_1 \exp(-\bar{c}_2 r_s^{\bar{c}_3})$ where $\bar{c}_i > 0$ depend only on p . Combining (72), (73), (65) and applying the union bound shows that

$$\|A^{-1}\|\|P\| = \frac{\gamma_1 \sqrt{r_s}}{(\theta - \delta p)r_q} < 1$$

with probability at least

$$1 - (r_q + r_s) \left(\frac{e^\delta}{(1 + \delta)^{(1 + \delta)}} \right)^{p\hat{r}} - \bar{c}_1 \exp(-\bar{c}_2 r_s^{\bar{c}_3})$$

for sufficiently large r_q . Therefore, $A + P$ is nonsingular and \mathbf{y} and \mathbf{z} are uniquely determined with probability tending exponentially to 1 as $\hat{r} \rightarrow \infty$.

For the remainder of the section we assume that $A + P$ is nonsingular. We define $Q, Q_1, Q_2, \bar{\mathbf{y}}$ and $\bar{\mathbf{z}}$ as in Section 2.1. To find an upper bound on $\|\tilde{S}_{C_q, C_s}\|$, we decompose \tilde{S}_{C_q, C_s} as

$$\tilde{S}_{C_q, C_s} = M_1 + M_2$$

where $M_1 := H \circ (\bar{\mathbf{y}}\mathbf{e}^T + \mathbf{e}\bar{\mathbf{z}}^T) - c(\mathbf{e}\mathbf{e}^T - H)$ and $M_2 := H \circ (Q_1\mathbf{b}\mathbf{e}^T + \mathbf{e}\mathbf{b}^T Q_2^T)$

We first obtain an upper bound on the norm of M_1 . We define $\mathbf{d} \in \mathbf{R}^{C_q}$ to be the vector such that the entry d_i is the difference between the number of edges added between the node i and C_s and the expected number of such edges for each $i \in C_q$. That is,

$$\mathbf{d} = \mathbf{n}_{C_q}^s - E[\mathbf{n}_{C_q}^s] = \mathbf{n}_{C_q}^s - pr_s\mathbf{e}.$$

Similarly, we let $\mathbf{f} := \mathbf{n}_{C_s}^q - pr_q\mathbf{e}$. Note that, by our choice of \mathbf{d} and \mathbf{f} , we have $r_s I - D = pr_s I + \text{Diag}(\mathbf{d})$ and $r_q I - F = pr_q I + \text{Diag}(\mathbf{f})$. Notice that for $\theta = 1 - p$ we have $D = \theta r_s I - \text{Diag}(\mathbf{d})$. Expanding (16) we have

$$\begin{aligned} \bar{\mathbf{y}} &= \left(D^{-1} - \frac{\theta D^{-1} \mathbf{e} \mathbf{e}^T D^{-1}}{1 + \theta \mathbf{e}^T D^{-1} \mathbf{e}} \right) \mathbf{b}_1 \\ &= \frac{D^{-1}}{1 + \theta \mathbf{e}^T D^{-1} \mathbf{e}} (\mathbf{b}_1 + \theta \mathbf{b}_1 \mathbf{e}^T D^{-1} \mathbf{e} - \theta \mathbf{e} \mathbf{e}^T D^{-1} \mathbf{b}_1) \\ &= \frac{D^{-1}}{1 + \theta \mathbf{e}^T D^{-1} \mathbf{e}} (\mathbf{b}_1 + \theta (\mathbf{b}_1 \mathbf{e}^T - \mathbf{e} \mathbf{b}_1^T) D^{-1} \mathbf{e}) \end{aligned}$$

since $\mathbf{e}^T D^{-1} \mathbf{b}_1 = \mathbf{b}_1^T D^{-1} \mathbf{e}$. Substituting $\mathbf{b}_1 = c(r_s \mathbf{e} - \bar{\mathbf{d}})$, where $\bar{\mathbf{d}} := \text{diag}(D)$, we have

$$\mathbf{b}_1 \mathbf{e}^T - \mathbf{e} \mathbf{b}_1^T = c(\mathbf{e} \bar{\mathbf{d}}^T - \bar{\mathbf{d}} \mathbf{e}^T)$$

and, hence,

$$\begin{aligned} \bar{\mathbf{y}} &= \frac{cD^{-1}}{1 + \theta \mathbf{e}^T D^{-1} \mathbf{e}} (r_s \mathbf{e} - \bar{\mathbf{d}} + \theta (\mathbf{e} \bar{\mathbf{d}}^T - \bar{\mathbf{d}} \mathbf{e}^T) D^{-1} \mathbf{e}) \\ &= \frac{cD^{-1}}{1 + \theta \mathbf{e}^T D^{-1} \mathbf{e}} (r_s \mathbf{e} - \bar{\mathbf{d}} + \theta \mathbf{e} \mathbf{e}^T \mathbf{e} - \theta \bar{\mathbf{d}} \mathbf{e}^T D^{-1} \mathbf{e}) \\ &= \frac{cD^{-1}}{1 + \theta \mathbf{e}^T D^{-1} \mathbf{e}} (r_s \mathbf{e} + \theta r_q \mathbf{e}) - c\mathbf{e} \\ &= \frac{c(r_s + \theta r_q) D^{-1}}{1 + \theta \mathbf{e}^T D^{-1} \mathbf{e}} \left(\frac{1}{\theta r_s} (D + \text{Diag}(\mathbf{d})) \right) \mathbf{e} - c\mathbf{e} \\ &= \left(\frac{c(r_s + \theta r_q)}{(1 + \theta \mathbf{e}^T D^{-1} \mathbf{e}) \theta r_s} - c \right) \mathbf{e} + \frac{c(r_s + \theta r_q)}{(1 + \theta \mathbf{e}^T D^{-1} \mathbf{e}) \theta r_s} D^{-1} \mathbf{d} \end{aligned}$$

since

$$I = \frac{1}{\theta r_s} (D + \text{Diag}(\mathbf{d})).$$

Let $\bar{\mathbf{y}}_1 := \omega_1 \mathbf{e}$, $\bar{\mathbf{y}}_2 := v_1 \mathbf{e}$ where

$$\begin{aligned} \omega_1 &= \frac{c(\theta r_q + r_s)}{\theta(r_s + r_q)} - c \\ v_1 &= \frac{c(\theta r_q + r_s)}{(1 + \theta \mathbf{e}^T D^{-1} \mathbf{e}) \theta r_s} - c - \omega_1 \end{aligned}$$

and let

$$\bar{\mathbf{y}}_3 := \frac{c(\theta r_q + r_s)}{(1 + \theta \mathbf{e}^T D^{-1} \mathbf{e}) \theta r_s} D^{-1} \mathbf{d}.$$

Hence, $\bar{\mathbf{y}} = \bar{\mathbf{y}}_1 + \bar{\mathbf{y}}_2 + \bar{\mathbf{y}}_3$. Similarly, $\bar{\mathbf{z}} = \bar{\mathbf{z}}_1 + \bar{\mathbf{z}}_2 + \bar{\mathbf{z}}_3$ where $\bar{\mathbf{z}}_1 := \omega_2 \mathbf{e}$, $\bar{\mathbf{z}}_2 := v_2 \mathbf{e}$ where

$$\begin{aligned}\omega_2 &= \frac{c(r_q + \theta r_s)}{\theta(r_s + r_q)} - c \\ v_2 &= \frac{c(r_q + \theta r_s)}{(1 + \theta \mathbf{e}^T F^{-1} \mathbf{e}) \theta r_q} - c - \omega_2\end{aligned}$$

and

$$\bar{\mathbf{z}}_3 := \frac{c(r_q + \theta r_s)}{(1 + \theta \mathbf{e}^T F^{-1} \mathbf{e}) \theta r_q} F^{-1} \mathbf{f}.$$

Therefore, we can further decompose M_1 as $M_1 = \tilde{M}_1 + \tilde{M}_2 + \tilde{M}_3$ where

$$\begin{aligned}\tilde{M}_1 &:= H \circ (\bar{\mathbf{y}}_1 \mathbf{e}^T + \mathbf{e} \bar{\mathbf{z}}_1^T) - c(\mathbf{e} \mathbf{e}^T - H), \\ \tilde{M}_2 &:= H \circ (\bar{\mathbf{y}}_2 \mathbf{e}^T + \mathbf{e} \bar{\mathbf{z}}_2^T), \quad \tilde{M}_3 := H \circ (\bar{\mathbf{y}}_3 \mathbf{e}^T + \mathbf{e} \bar{\mathbf{z}}_3^T).\end{aligned}$$

Notice that the matrix \tilde{M}_1 has entries corresponding to edges equal to $-c$ and remaining entries equal to $cp/(1-p)$ since

$$\omega_1 + \omega_2 = \frac{c(1+\theta)(r_q + r_s)}{\theta(r_q + r_s)} - 2c = \frac{cp}{\theta}.$$

Therefore, each entry of the matrix \tilde{M}_1 has expected value equal to 0. Moreover, each entry of the random block matrix \hat{M} of the form

$$\hat{M} = \begin{bmatrix} \tilde{M}_1 \\ \tilde{R} \end{bmatrix}$$

has expected value equal to 0 if \tilde{R} has identically independently distributed entries such that

$$\tilde{R}_{i,j} = \begin{cases} -c, & \text{with probability } p \\ cp/(1-p), & \text{with probability } 1-p. \end{cases}$$

Therefore, there exists $c_1, c_2, c_3, c_4 > 0$ such that

$$\|\tilde{M}_1\| \leq \|\hat{M}\| \leq c_4 \sqrt{r_s} \quad (74)$$

with probability at least $1 - c_1 \exp(-c_2 r_s^{c_3})$ by Theorem 4.1.

Next, to obtain upper bounds on $\|\tilde{M}_2\|$ and $\|\tilde{M}_3\|$ we will use the following lemma.

Lemma 4.1 *There exists $B > 0$ depending only on p such that*

$$\sum_{i \in C_q} \frac{|\theta r_s - D_{ii}|^\alpha}{D_{ii}} \leq B \frac{r_q}{r_s^{1-\alpha/2}} \quad (75)$$

and

$$\sum_{i \in C_s} \frac{|\theta r_q - F_{ii}|^\alpha}{F_{ii}} \leq B \frac{r_s}{r_q^{1-\alpha/2}} \quad (76)$$

for $\alpha = 1, 2$ with probability at least

$$1 - (r_q + r_s) v_p^{\hat{r}} - 2(2/3)^{\hat{r}} \quad (77)$$

where $v_p = (e^\delta / (1 + \delta)^{(1+\delta)})^p$ and $\delta = \min\{p, \sqrt{p} - p\}$.

Proof: We first prove (75). For each $j \in C_q$, let $n_j := n_j^s$. The random numbers $\{n_j : j \in C_q\}$ are independent, and each is the result of r_s Bernoulli trials, each with probability of success equal to p . We define Ψ to be the event that at least one n_j is very far from its expected value. That is, Ψ is the event that there exists $j \in C_q$ such that $n_j > tr_s$, where $t = \min\{\sqrt{p}, 2p\}$. Moreover, we define $\tilde{\Psi}$ to be its complement, and let $\tilde{\psi}(n_j)$ be the indicator function such that

$$\tilde{\psi}(n_j) = \begin{cases} 1, & \text{if } n_j \leq tr_s \\ 0, & \text{otherwise.} \end{cases}$$

Let B be a positive scalar depending on p to be determined later. Then

$$\begin{aligned} P\left(\sum_{i \in C_q} \frac{|\theta r_s - D_{ii}|^\alpha}{D_{ii}} \geq B \frac{r_q}{r_s^{1-\alpha/2}}\right) &= P\left(\sum_{i \in C_q} \frac{|n_i - pr_s|^\alpha}{r_s - n_i} \geq B \frac{r_q}{r_s^{1-\alpha/2}}\right) \\ &\leq P\left(\sum_{i \in C_q} \frac{|n_i - pr_s|^\alpha}{r_s - n_i} \geq B \frac{r_q}{r_s^{1-\alpha/2}} \wedge \tilde{\Psi}\right) + P(\Psi). \end{aligned} \quad (78)$$

We will analyze the two terms separately. For the first term we use a technique of Bernstein (see [13]). Let ϕ be the indicator function of the nonnegative reals. Then,

$$\begin{aligned} P\left(\sum_{j \in C_q} \frac{|n_j - pr_s|^\alpha}{r_s - n_j} \geq B \frac{r_q}{r_s^{1-\alpha/2}} \wedge \tilde{\Psi}\right) &= P\left(\sum_{j \in C_q} \frac{|n_j - pr_s|^\alpha}{r_s - n_j} - B \frac{r_q}{r_s^{1-\alpha/2}} \geq 0 \wedge \tilde{\psi}(n_j) = 1 \ \forall j \in C_q\right) \\ &= P\left(\sum_{j \in C_q} \frac{r_s^{1-\alpha/2} |n_j - pr_s|^\alpha}{r_s - n_j} - Br_q \geq 0 \wedge \tilde{\psi}(n_j) = 1 \ \forall j \in C_q\right) \\ &= E\left(\phi\left(\sum_{j \in C_q} \frac{r_s^{1-\alpha/2} |n_j - pr_s|^\alpha}{r_s - n_j} - Br_q\right) \cdot \prod_{j \in C_q} \tilde{\psi}(n_j)\right). \end{aligned}$$

Let h be a positive scalar depending p to be determined later. Notice that for any $h > 0$ and all $x \in \mathbf{R}$, $\phi(x) \leq \exp(hx)$. Thus, by the independence of the n_j 's,

$$\begin{aligned} P\left(\sum_{i \in C_q} \frac{|n_i - pr_s|^\alpha}{r_s - n_i} \geq B \frac{r_q}{r_s^{1-\alpha/2}} \wedge \tilde{\Psi}\right) &\leq E\left(\exp\left(h \sum_{j \in C_q} \left(\frac{|n_j - pr_s|^\alpha}{r_s^{\alpha/2-1}(r_s - n_j)} - B\right)\right) \cdot \prod_{j \in C_q} \tilde{\psi}(n_j)\right) \\ &= \prod_{j \in C_q} E\left(\exp\left(h \left(\frac{|n_j - pr_s|^\alpha}{r_s^{\alpha/2-1}(r_s - n_j)} - B\right)\right) \tilde{\psi}(n_j)\right) \\ &= f_1 \cdots f_{r_q} \end{aligned}$$

where

$$f_j = E \left(\exp \left(h \left(\frac{|n_j - pr_s|^\alpha}{r_s^{\alpha/2-1}(r_s - n_j)} - B \right) \right) \tilde{\psi}(n_j) \right).$$

We now analyze each f_j individually. Fix $j \in C_q$. Then

$$\begin{aligned} f_j &= \sum_{i=0}^{\lfloor tr_s \rfloor} \exp \left(h \left(\frac{|n_j - pr_s|^\alpha}{r_s^{\alpha/2-1}(r_s - n_j)} - B \right) \right) P(n_j = i) \\ &\leq \sum_{i=0}^{\lfloor tr_s \rfloor} \exp \left(h \left(\frac{|i - pr_s|^\alpha}{(1 - \sqrt{p})r_s^{\alpha/2}} - B \right) \right) P(n_j = i) \end{aligned}$$

since $i \leq tr_s$ and, hence, $i \leq \sqrt{p}r_s$. We now reorganize this summation by considering i such that $|i - pr_s| < \sqrt{r_s}$, then i such that $\sqrt{r_s} \leq |i - pr_s| < 2\sqrt{r_s}$ and so on. Notice that, since $i \leq tr_s \leq 2pr_s$, we need only to consider intervals until $|i - pr_s|$ reaches pr_s . Hence,

$$\begin{aligned} f_j &\leq \sum_{k=0}^{\lfloor p\sqrt{r_s} \rfloor} \sum_{i: |i - pr_s| \in [k\sqrt{r_s}, (k+1)\sqrt{r_s})} \exp \left(h \left(\frac{|i - pr_s|^\alpha}{(1 - \sqrt{p})r_s^{\alpha/2}} - B \right) \right) P(n_j = i) \\ &\leq \sum_{k=0}^{\lfloor p\sqrt{r_s} \rfloor} \sum_{i: |i - pr_s| \in [k\sqrt{r_s}, (k+1)\sqrt{r_s})} \exp \left(h \left(\frac{(k+1)^\alpha}{1 - \sqrt{p}} - B \right) \right) P(n_j = i) \\ &\leq 2 \sum_{k=0}^{\lfloor p\sqrt{r_s} \rfloor} \exp \left(h \left(\frac{(k+1)^\alpha}{1 - \sqrt{p}} - B \right) \right) \exp(-k^2/p) \end{aligned}$$

by (60). Overestimating the finite sum with an infinite sum, we have

$$f_j \leq 2 \exp(-hB) \cdot \sum_{k=0}^{\infty} \exp \left(\frac{h(k+1)^\alpha}{1 - \sqrt{p}} - k^2/p \right).$$

Choosing h such that $h \leq (1 - \sqrt{p})/(8p)$ ensures that

$$\frac{h(k+1)^\alpha}{1 - \sqrt{p}} - k^2/p \leq -k^2/(2p)$$

for all r_q, r_s and $k \geq 1$. Hence, splitting off the $k = 0$ term, we have

$$f_j \leq 2 \exp \left(\frac{h}{1 - \sqrt{p}} - hB \right) + 2 \exp(-hB) \cdot \sum_{k=1}^{\infty} \exp(-k^2/(2p)). \quad (79)$$

Since $\sum_{k=1}^{\infty} \exp(-k^2/(2p))$ is dominated by a geometric series, the summation in (79) is a finite number depending on p . Therefore, once h is chosen, it is possible to choose B , depending only on p and h , sufficiently large so that each of the two terms in (79) is at most $1/3$. Therefore, we can choose h and B so that $f_j \leq 2/3$ for all $j \in C_q$. It follows immediately that

$$P \left(\sum_{i \in C_q} \frac{|\theta r_s - D_{ii}|^\alpha}{D_{ii}} \geq B \frac{r_q}{r_s^{1-\alpha/2}} \wedge \tilde{\Psi} \right) \leq (2/3)^{r_q} \leq (2/3)^{\hat{r}}. \quad (80)$$

To obtain a bound on the second term in (78), notice that the probability that $n_j > tr_s$ is at most $v_p^{r_s} \leq v_p^{\hat{r}}$ where $v_p = (e^\delta / (1 + \delta)^{(1+\delta)})^p$ by Theorem 4.3, where $\delta = t/p - 1 = \min\{p, \sqrt{p} - p\}$. Thus, applying the union bound shows that

$$P \left(\sum_{i \in C_q} \frac{|\theta r_s - D_{ii}|^\alpha}{D_{ii}} \geq B \frac{r_q}{r_s^{1-\alpha/2}} \right) \leq (2/3)^{\hat{r}} + r_q v_p^{\hat{r}}.$$

By an identical argument

$$P \left(\sum_{i \in C_s} \frac{|\theta r_q - F_{ii}|^\alpha}{F_{ii}} \geq B \frac{r_s}{r_q^{1-\alpha/2}} \right) \leq (2/3)^{\hat{r}} + r_s v_p^{\hat{r}}.$$

Applying the union bound one last time shows that (75) and (76) hold simultaneously with probability at least $1 - (r_q + r_s)v_p^{\hat{r}} - 2(2/3)^{\hat{r}}$ as required. \blacksquare

As an immediate corollary of Lemma 4.1, we have the following bound on $|v_1|$ and $|v_2|$.

Corollary 4.1 *There exists $B_1 > 0$ depending only on p such that*

$$|v_1| + |v_2| \leq B_1 \frac{r_q^{3/2} + r_s^{3/2}}{(r_q + r_s)(r_q r_s)^{1/2}}$$

with probability at least $1 - (r_q + r_s)v_p^{\hat{r}} - 2(2/3)^{\hat{r}}$.

Proof: We begin with v_1 . Notice that

$$\begin{aligned} v_1 &= \frac{c(\theta r_q + r_s)}{\theta} \left(\frac{1}{(1 + \theta \mathbf{e}^T D^{-1} \mathbf{e}) r_s} - \frac{1}{r_q + r_s} \right) \\ &= \frac{c(\theta r_q + r_s)(r_q - \theta r_s \mathbf{e}^T D^{-1} \mathbf{e})}{\theta r_s (r_q + r_s)(1 + \theta \mathbf{e}^T D^{-1} \mathbf{e})}. \end{aligned}$$

Moreover,

$$|\theta r_s \mathbf{e}^T D^{-1} \mathbf{e} - r_q| = \left| \sum_{i \in C_q} \frac{\theta r_s}{D_{ii}} - \sum_{i \in C_q} 1 \right| \leq \sum_{i \in C_q} \left| \frac{\theta r_s}{D_{ii}} - 1 \right| = \sum_{i \in C_q} \frac{|\theta r_s - D_{ii}|}{D_{ii}}$$

and, since $D_{ii} \leq r_s$ for all $i \in C_q$, we have

$$r_s(1 + \theta \mathbf{e}^T D^{-1} \mathbf{e}) \geq r_s \left(1 + \frac{\theta r_q}{r_s} \right) = \theta r_q + r_s. \quad (81)$$

Therefore, setting $\alpha = 1$ in (75) shows that

$$\begin{aligned} |v_1| &\leq \frac{c(\theta r_q + r_s) B r_q}{\theta r_s^{3/2} (r_q + r_s) (1 + \theta \mathbf{e}^T D^{-1} \mathbf{e})} \\ &\leq \frac{c B (\theta r_q + r_s) r_q}{\theta \sqrt{r_s} (r_q + r_s) (\theta r_q + r_s)} \\ &\leq B_1 \frac{r_q}{\sqrt{r_s} (r_q + r_s)} \end{aligned} \quad (82)$$

where $B_1 := B/\theta$ and where (82) holds with probability at least $1 - (2/3)^{\hat{r}} - r_q v_p^{\hat{r}}$. By an identical calculation

$$|v_2| \leq \frac{B_1 r_s}{\sqrt{r_q}(r_q + r_s)} \quad (83)$$

with probability at least $1 - (2/3)^{\hat{r}} - r_s v_p^{\hat{r}}$. Applying the union bound completes the proof. \blacksquare

Observe that, as an immediate consequence of Corollary 4.1 and the facts that $H \circ \mathbf{e}\mathbf{e}^T = H$ and $\|H\|_F \leq \sqrt{r_q r_s}$, we have

$$\begin{aligned} \|\tilde{M}_2\| &= \|H \circ (\bar{\mathbf{y}}_2 \mathbf{e}^T + \mathbf{e} \bar{\mathbf{z}}_2^T)\| \leq (|v_1| + |v_2|) \|H\|_F \\ &\leq B_1 \frac{r_q^{3/2} + r_s^{3/2}}{r_q + r_s} \leq 2B_1 \sqrt{r_s} \end{aligned} \quad (84)$$

with probability at least $1 - (r_q + r_s) v_p^{\hat{r}} - 2(2/3)^{\hat{r}}$.

The following corollary of Lemma 4.1 provides an upper bound on $\|\tilde{M}_3\|$.

Corollary 4.2 *There exists B_2 depending only on p such that*

$$\|\tilde{M}_3\| \leq \|H \circ (\bar{\mathbf{y}}_3 \mathbf{e}^T + \mathbf{e} \bar{\mathbf{z}}_3^T)\| \leq B_2 (\sqrt{r_q} + \sqrt{r_s}) \quad (85)$$

with probability at least $1 - (r_q + r_s) v_p^{\hat{r}} - 2(2/3)^{\hat{r}}$.

Proof: To obtain an upper bound on $\|\tilde{M}_3\|$, we first obtain upper bounds on $\|H \circ (\bar{\mathbf{y}}_3 \mathbf{e}^T)\|$ and $\|H \circ (\mathbf{e} \bar{\mathbf{z}}_3^T)\|$. We begin with $\|H \circ (\bar{\mathbf{y}}_3 \mathbf{e}^T)\|$. Since

$$\sum_{i \in C_q} D_{ii} (D^{-1} \mathbf{d})_i^2 = \sum_{i \in C_q} \frac{|\theta r_s - D_{ii}|^2}{D_{ii}}$$

applying (75) with $\alpha = 2$ and (39) with $W = H$, $\mathbf{u} = \bar{\mathbf{y}}_3$, and $\mathbf{v} = \mathbf{e}$ shows that

$$\begin{aligned} \|H \circ (\bar{\mathbf{y}}_3 \mathbf{e}^T)\| &\leq \left(\sum_{i \in C_q} \bar{\mathbf{y}}_3(i)^2 \|H(i, :)\|^2 \right)^{1/2} \\ &= \frac{c(\theta r_q + r_s)}{(1 + \theta \mathbf{e}^T D^{-1} \mathbf{e}) \theta r_s} \cdot \left(\sum_{i \in C_q} D_{ii} [D^{-1} \mathbf{d}]_i^2 \right)^{1/2} \\ &\leq \frac{c(\theta r_q + r_s)}{(1 + \theta \mathbf{e}^T D^{-1} \mathbf{e}) \theta r_s} (B r_q)^{1/2} \end{aligned} \quad (86)$$

$$\leq B_2 \sqrt{r_q} \quad (87)$$

where $B_2 := \sqrt{B}/\theta$, (87) follows from (81) and (86) holds with probability at least $1 - (2/3)^{\hat{r}} - r_q v_p^{\hat{r}}$. Similarly,

$$\|H \circ (\mathbf{e} \bar{\mathbf{z}}_3^T)\| \leq B_2 \sqrt{r_s} \quad (88)$$

with probability at least $1 - ((2/3)^{\hat{r}} + r_s v_p^{\hat{r}})$. Applying the union bound shows that

$$\|\tilde{M}_3\| \leq \|H \circ (\bar{\mathbf{y}}_3 \mathbf{e}^T)\| + \|H \circ (\mathbf{e} \bar{\mathbf{z}}_3^T)\| \leq B_2(\sqrt{r_q} + \sqrt{r_s}) \quad (89)$$

with probability at least $1 - (r_q + r_s)v_p^{\hat{r}} - 2(2/3)^{\hat{r}}$ as required. \blacksquare

We complete the proof of Theorem 4.6 by showing that $M_2 = H \circ (Q_1 \mathbf{b} \mathbf{e}^T + \mathbf{e} \mathbf{b}^T Q_2^T)$ has norm at most a constant multiple of $r_s/\sqrt{r_q}$ with high probability. The following lemma provides an upper bound on $\|Q_1 \mathbf{b}\|$ and $\|Q_2 \mathbf{b}\|$.

Lemma 4.2 *There exists B_3, B_4 and $\bar{c}_i > 0, i = 1, 2, 3$, depending only on p such that*

$$\|Q_1 \mathbf{b}\| \leq B_3 \frac{r_s^{1/2}}{r_q^{1/2}} \quad (90)$$

$$\|Q_2 \mathbf{b}\| \leq B_4 \frac{r_s^{1/2}(r_q + r_s^{1/2})}{r_q^{3/2}} \quad (91)$$

with probability at least

$$1 - \bar{c}_1 \exp(-\bar{c}_2 \hat{r}^{\bar{c}_3}) - (r_q + r_s) \left(\frac{e^\delta}{(1 + \delta)^{(1 + \delta)}} \right)^{p\hat{r}} \quad (92)$$

where $\delta = (1 - p)/(2p)$.

Proof: We first derive a bound on each of $\|Q_1\|$, $\|Q_2\|$ and $\|\mathbf{b}\|$ and consequently a bound on each of $\|Q_1 \mathbf{b}\|$ and $\|Q_2 \mathbf{b}\|$ by applying the inequalities

$$\|Q_1 \mathbf{b}\| \leq \|Q_1\| \|\mathbf{b}\| \quad \text{and} \quad \|Q_2 \mathbf{b}\| \leq \|Q_2\| \|\mathbf{b}\|.$$

Recall that

$$\|Q_1\| \leq \|(D + \theta \mathbf{e} \mathbf{e}^T)^{-1}\| \sum_{\ell=1}^{\infty} \|P_1 P_2\|^\ell + \|P_1\| \|(F + \theta \mathbf{e} \mathbf{e}^T)^{-1}\| \sum_{\ell=0}^{\infty} \|P_1 P_2\|^\ell$$

and

$$\|Q_2\| \leq \|(F + \theta \mathbf{e} \mathbf{e}^T)^{-1}\| \sum_{\ell=1}^{\infty} \|P_1 P_2\|^\ell + \|P_2\| \|(D + \theta \mathbf{e} \mathbf{e}^T)^{-1}\| \sum_{\ell=0}^{\infty} \|P_1 P_2\|^\ell$$

where

$$P_1 = (D + \theta \mathbf{e} \mathbf{e}^T)^{-1}(H - \theta \mathbf{e} \mathbf{e}^T), \quad P_2 = (F + \theta \mathbf{e} \mathbf{e}^T)^{-1}(H^T - \theta \mathbf{e} \mathbf{e}^T).$$

Applying the upper bounds on $\|(D + \theta \mathbf{e} \mathbf{e}^T)^{-1}\|$, $\|(F + \theta \mathbf{e} \mathbf{e}^T)^{-1}\|$, and $\|H - \theta \mathbf{e} \mathbf{e}^T\|$ given by (70), (71), and (73) shows that

$$\|P_1 P_2\| \leq \frac{\|H - \theta \mathbf{e} \mathbf{e}^T\|^2}{(\min_{i \in C_q} D_{ii})(\min_{i \in C_s} F_{ii})} \leq \frac{\gamma_1^2}{(\theta - \delta p)^2 r_q} \quad (93)$$

with probability at least (92). Therefore, there exists $\gamma_2 > 0$ depending only on p such that

$$\begin{aligned}\|Q_1\| &\leq \frac{1}{(\theta - \delta p)r_s} \sum_{\ell=1}^{\infty} \left(\frac{\gamma_1^2}{(\theta - \delta p)^2 r_q} \right)^{\ell} + \frac{\gamma_1}{(1 - \delta p)^2 r_q \sqrt{r_s}} \sum_{\ell=0}^{\infty} \left(\frac{\gamma_1^2}{(\theta - \delta p)^2 r_q} \right)^{\ell} \\ &\leq \frac{\gamma_2}{r_q \sqrt{r_s}}\end{aligned}\tag{94}$$

with probability at least (92) since

$$\sum_{\ell=0}^{\infty} \left(\frac{\gamma_1^2}{(\theta - \delta p)^2 r_q} \right)^{\ell} \leq O(1)$$

and

$$\sum_{\ell=1}^{\infty} \left(\frac{\gamma_1^2}{(\theta - \delta p)^2 r_q} \right)^{\ell} \leq O(r_q^{-1}).$$

with probability at least (92) in the case that $r_q > (\gamma_1/(\theta - \delta p))^2$. Similarly, there exists $\gamma_3 > 0$ depending only on p such that

$$\|Q_2\| \leq \frac{\gamma_3}{r_q} (r_q^{-1} + r_s^{-1/2}) = \frac{\gamma_3(r_q + r_s^{1/2})}{r_q^2 r_s^{1/2}}\tag{95}$$

with probability at least (92). Finally, recall that

$$[\mathbf{b}]_i = c \cdot \begin{cases} n_i^s, & \text{if } i \in C_q \\ n_i^q, & \text{if } i \in C_s. \end{cases}$$

Therefore, by (67) and (68)

$$\|\mathbf{b}\| = c \left(\sum_{i \in C_q} (n_i^s)^2 + \sum_{i \in C_s} (n_i^q)^2 \right)^{1/2} \leq (1 + \delta) p c (r_q r_s)^{1/2} (r_q + r_s)^{1/2}$$

with probability at least $1 - (r_q + r_s) (e^{\delta}/(1 + \delta)^{(1+\delta)})^{p\hat{r}}$. Thus, applying the union bound shows that there exists B_3, B_4 depending only on p such that

$$\begin{aligned}\|Q_1 \mathbf{b}\| &\leq \frac{\gamma_2 (1 + \delta) p c (r_q r_s)^{1/2} (r_q + r_s)^{1/2}}{r_q r_s^{1/2}} \leq B_3 \frac{r_s^{1/2}}{r_q^{1/2}} \\ \|Q_2 \mathbf{b}\| &\leq \frac{\gamma_3 (1 + \delta) p c (r_q r_s)^{1/2} (r_q + r_s)^{1/2} (r_q + r_s^{1/2})}{r_q^2 r_s^{1/2}} \leq B_4 \frac{r_s^{1/2} (r_q + r_s^{1/2})}{r_q^{3/2}}\end{aligned}$$

with probability at least (92) since $r_q \leq r_s$ by Assumption (65). ■

Finally, to obtain an upper bound on $\|M_2\|$ we decompose M_2 as

$$M_2 = (H - \theta \mathbf{e} \mathbf{e}^T) \circ (Q_1 \mathbf{b} \mathbf{e}^T) + \theta Q_1 \mathbf{b} \mathbf{e}^T + (H - \theta \mathbf{e} \mathbf{e}^T) \circ (\mathbf{e} (Q_2 \mathbf{b})^T) + \theta \mathbf{e} (Q_2 \mathbf{b})^T.$$

As an immediate corollary of Lemma 4.2 we have

$$\|(Q_1 \mathbf{b}) \mathbf{e}^T\| \leq \|Q_1 \mathbf{b}\| \|\mathbf{e}\| = \sqrt{r_s} \|Q_1 \mathbf{b}\| \leq B_3 \frac{r_s}{r_q^{1/2}} \quad (96)$$

and

$$\|\mathbf{e}(Q_2 \mathbf{b})^T\| \leq \|\mathbf{e}\| \|Q_2 \mathbf{b}\| = \sqrt{r_q} \|Q_2 \mathbf{b}\| \leq B_4 \frac{r_s^{1/2}(r_q + r_s^{1/2})}{r_q} \quad (97)$$

with probability at least (92). Moreover, applying (38) with $W = H - \theta \mathbf{e} \mathbf{e}^T$, $\mathbf{u} = Q_1 \mathbf{b}$, and $\mathbf{v} = \mathbf{e}$ we have

$$\|(H - \theta \mathbf{e} \mathbf{e}^T) \circ (Q_1 \mathbf{b} \mathbf{e}^T)\| \leq \|H - \theta \mathbf{e} \mathbf{e}^T\| \|Q_1 \mathbf{b}\|_\infty \leq \|H - \theta \mathbf{e} \mathbf{e}^T\| \|Q_1 \mathbf{b}\|. \quad (98)$$

Thus, combining (98), (73), and (90) we have

$$\|(H - \theta \mathbf{e} \mathbf{e}^T) \circ (Q_1 \mathbf{b} \mathbf{e}^T)\| \leq B_3 \gamma_1 \frac{r_s}{r_q^{1/2}} \quad (99)$$

with probability at least (92). Similarly,

$$\|(H - \theta \mathbf{e} \mathbf{e}^T) \circ (\mathbf{e}(Q_2 \mathbf{b})^T)\| \leq B_4 \gamma_1 \frac{r_s(r_q + r_s^{1/2})}{r_q^{3/2}} \quad (100)$$

with probability at least (92). Therefore, there exists \hat{c} depending only on p such that

$$\|M_2\| \leq \hat{c} \frac{r_s}{r_q^{1/2}} \quad (101)$$

with probability at least (92) since $r_s \leq r_q^{3/2}$ by (65), and, hence $(r_q + r_s^{1/2})/r_q \leq O(1)$.

Combining (74), (84), (89), and (101), there exists some \tilde{B}_1 depending only on p such that

$$\|\tilde{S}_{C_q, C_s}\| \leq \tilde{B}_1 \frac{r_s}{r_q^{1/2}} \quad (102)$$

for all $q, s \in \{1, \dots, k\}$, $q \neq s$ with probability tending exponentially to 1 as \hat{r} approaches ∞ . The proof of Theorem 4.6 is now complete. Theorem 4.5 is an immediate consequence: from (102) it follows that

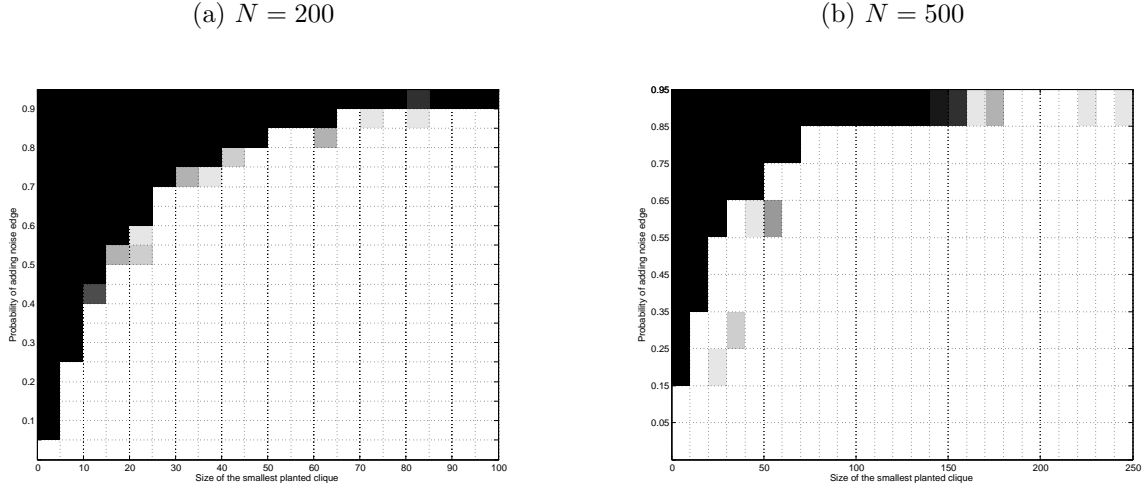
$$\sum_{q=1}^k \sum_{s=1}^k \|\tilde{S}_{C_q, C_s}\|^2 \leq 2\tilde{B}_1^2 \sum_{q=1}^k \sum_{s=1}^k \frac{r_s^2}{r_q} = 2\tilde{B}_1^2 \left(\sum_{s=1}^k r_s^2 \right) \left(\sum_{q=1}^k \frac{1}{r_q} \right)$$

and, hence, there exists some β_1 depending only on p such that

$$\|\tilde{S}_1\| \leq \beta_1 \left(\sum_{s=1}^k r_s^2 \right)^{1/2} \left(\sum_{q=1}^k \frac{1}{r_q} \right)^{1/2}$$

as required.

Figure 1: Simulation results for N -node graph with k planted cliques of roughly identical size. We plot the average number of recoveries of the planted cliques per set of 10 trials for different minimum cluster sizes \hat{r} and different probabilities of adding noise edges. Brighter colours indicate a higher rate of recovery.



5 Experimental Results

In this section, we evaluate the performance of our relaxation of the k -disjoint-clique problem for a variety of program inputs via simulation.

We consider N -node random graphs G constructed according to (Ω_1) and (Ω_2) containing planted cliques C_1, \dots, C_k as follows. For fixed minimum clique size \hat{r} , we choose $k = \lfloor N/\hat{r} \rfloor$ and choose the clique sizes such that $r_i \in [\hat{r}, 2\hat{r}]$. Every node is assigned to exactly one of the k planted cliques. We set each entry of the adjacency matrix A_G in the diagonal blocks indexed by C_1, \dots, C_k equal to 1 and independently choose each remaining upper triangular entry of A_G to be equal to 1 with probability p . The lower triangular entries of A_G are chosen by symmetry. We solve the semidefinite program

$$\max\{\text{tr}(X\mathbf{e}\mathbf{e}^T) : X \succeq 0, X\mathbf{e} \leq \mathbf{e}, \text{tr}(X) = k, X_{ij} = 0 \text{ if } [A_G]_{ij} = 0, i \neq j\}$$

using SDPNAL [19] in Matlab. We declare the planted k -disjoint-clique subgraph to be recovered if the optimal solution X^* returned by SDPNAL satisfies $\|X^* - X_0\|_F / \|X_0\|_F < 10^{-4}$, where X_0 is the predicted optimal solution given by (6). This experiment was repeated 10 times each for different values of p and \hat{r} for $N = 200$ and $N = 500$. We plot the empirical probability of recovery of the planted k -disjoint clique subgraph for each set of trials in Figure 1.

The performance of our heuristic closely matches that predicted by Theorem 4.5, although the constants in the bounds on the sizes of planted cliques may be conservative. For example, the bounds ensuring recovery of X_0 given by Theorem 4.5 cannot be satisfied if the minimum clique size \hat{r} is less than $3(p/1-p)\sqrt{N}$, however perfect recovery occurs for smaller values of \hat{r} in almost all trials.

6 Conclusions

We have considered an NP-hard combinatorial version of the clustering problem called the k -disjoint-clique problem in which input data is an undirected graph. We have shown that a convex relaxation of the problem can exactly solve the problem for input instances constructed in a certain way. The construction of the instance is that k disjoint cliques are first placed in the input graph, and then many ‘noise’ vertices and edges are placed that obscure the k disjoint cliques. We have shown that the algorithm exactly recovers the clique for noise edges placed by an adversary provided the conditions stated in Theorem 3.1 on the number of noise edges are satisfied. In the case of random noise, many more noise edges and nodes can be tolerated compared to the adversary case; in particular, if the quantity on the right-hand side of Theorem 4.5 is at most \hat{r} , then the algorithm recovers the planted cliques with probability exponentially close to 1.

This work raises several open questions. First, as already noted in the text, our bounds may not be the best possible. Particularly in the random case, there is nothing in the way of matching lower bounds.

Another open question is whether the techniques developed herein can be applied to other formulations of clustering. For example, if clustering is posed as an optimization problem with a distance function, then can an approach like the one described here find the optimal solution for input instances constructed in a certain way?

References

- [1] N. Alon, M. Krivelevich, and B. Sudakov. Finding a large hidden clique in a random graph. *Random Structures and Algorithms*, 13:457–466, 1998.
- [2] B. Ames and S. Vavasis. Nuclear norm minimization for the planted clique and biclique problems. *Mathematical Programming*, 129(1):1–21, 2011.
- [3] P. Berkhin. A survey of clustering data mining techniques. In Jacob Kogan, Charles Nicholas, and Marc Teboulle, editors, *Grouping Multidimensional Data*, chapter 2, pages 25–71. Springer-Verlag, Berlin/Heidelberg, 2006.
- [4] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, UK, 2004.
- [5] E.J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3), 2011.
- [6] E.J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.
- [7] V. Chandrasekaran, S. Sanghavi, P.A. Parrilo, and A.S. Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011.

- [8] U. Feige and R. Krauthgamer. Finding and certifying a large hidden clique in a semi-random raph. *Random Structures Algorithms*, 16(2):195–208, 2000.
- [9] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.
- [10] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.
- [11] S. Geman. A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2):252–261, 1980.
- [12] G. Golub and C. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 1996. Third Edition.
- [13] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. American Statistical Assoc.*, 58:13–30, 1962.
- [14] A. Jalali, Y. Chen, S. Sanghavi, and H. Xu. Clustering partially observed graphs via convex optimization. *Arxiv preprint arXiv:1104.4803*, 2011.
- [15] M. Mitzenmacher and E. Upfal. *Probability and computing*. Cambridge University Press, Cambridge, 2005. Randomized algorithms and probabilistic analysis.
- [16] R. Ostrovsky, Y. Rabani, L. Schulman, and C. Swamy. The effectiveness of Lloyd-type methods for the k-means problem. In *Proceedings of 47th Annual IEEE Symposium on the Foundations of Computer Science*, 2006.
- [17] S. Oymak and B. Hassibi. Finding dense clusters via “low rank + sparse” decomposition. *Arxiv preprint arXiv:1104.5186*, 2011.
- [18] B. Recht, M. Fazel, and P.A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.
- [19] X.Y. Zhao, D. Sun, and K.C. Toh. A newton-cg augmented lagrangian method for semidefinite programming. *SIAM Journal on Optimization*, 20(4):1737–1765, 2010.