Math. Program., Ser. B (2011) 129:69–89
DOI 10.1007/s10107-011-0459-x

FULL LENGTH PAPER

Nuclear norm minimization for the planted clique and biclique problems

Brendan P. W. Ames · Stephen A. Vavasis

Received: 6 February 2009 / Accepted: 12 January 2011 / Published online: 18 May 2011

Abstract We consider the problems of finding a maximum clique in a graph and finding a maximum-edge biclique in a bipartite graph. Both problems are NP-hard. We write both problems as matrix-rank minimization and then relax them using the nuclear norm. This technique, which may be regarded as a generalization of compressive sensing, has recently been shown to be an effective way to solve rank optimization problems. In the special case that the input graph has a planted clique or biclique (i.e., a single large clique or biclique plus diversionary edges), our algorithm successfully provides an exact solution to the original instance. For each problem, we provide two analyses of when our algorithm succeeds. In the first analysis, the diversionary edges are placed by an adversary. In the second, they are placed at random. In the case of random edges for the planted clique problem, we obtain the same bound as Alon, Krivelevich and Sudakov as well as Feige and Krauthgamer, but we use different techniques.

Mathematics Subject Classification (2000) 90C25 · 65K05 · 68Q25

Supported in part by a Discovery Grant and Postgraduate Scholarship (Doctoral) from NSERC (Natural Science and Engineering Research Council of Canada) and the US Air Force Office of Scientific Research.

B. P. W. Ames (✉) · S. A. Vavasis
Department of Combinatorics and Optimization, University of Waterloo,
200 University Avenue W., Waterloo, ON N2L 3G1, Canada
e-mail: bpames@math.uwaterloo.ca

S. A. Vavasis
e-mail: vavasis@math.uwaterloo.ca

1 Introduction

Several recent papers including Recht et al. [18] and Candès and Recht [3] consider nuclear norm minimization as a convex relaxation of matrix rank minimization. *Matrix rank minimization* refers to the problem of finding a matrix $X \in \mathbf{R}^{m\times n}$ to minimize rank (X) subject to linear constraints on X. As we shall show in Sects. 4 and 5, the clique and biclique problems, both NP-hard, are easily expressed as matrix rank minimization, thus showing that matrix rank minimization is also NP-hard.

Each of the two papers mentioned in the previous paragraph has results of the following general form. Suppose an instance of matrix rank minimization is posed in which it is known *a priori* that a solution of very low rank exists. Suppose further that the constraints are random in some sense. Then the nuclear norm relaxation turns out to be exact, i.e., it recovers the (unique) solution of low rank. The *nuclear norm* of a matrix X, also called the *trace norm*, is defined to be the sum of the singular values of X.

These authors build upon recent breakthroughs in compressive sensing [4,5,10]. In compressive sensing, the problem is to recover a sparse vector that solves a set of linear equations. In the case that the equations are randomized and a very sparse solution exists, compressive sensing can be solved by relaxation to the l_1 norm. The correspondence between matrix rank minimization and compressive sensing is as follows: matrix rank (number of nonzero singular values) corresponds to vector sparsity (number of nonzero entries) and nuclear norm corresponds to l_1 norm.

Our results follow the spirit of Recht et al. but use different technical approaches. We establish results about two well known graph theoretic problems, namely maximum clique and maximum-edge biclique. The maximum clique problem takes as input an undirected graph and asks for the largest clique (i.e., induced subgraph of nodes that are completely interconnected). This problem is one of Karp's original NP-hard problems [8]. The maximum-edge biclique takes as input a bipartite graph (U, V, E) and asks for the subgraph that is a complete bipartite graph $K_{m,n}$ that maximizes the product mn. This problem was shown to be NP-hard by Peeters [17].

In Sects. 4 and 5, we relax these problems to convex optimization using the nuclear norm. For each problem, we show that convex optimization can recover the exact solution in two cases. The first case, described in Sect. 4.1, is the adversarial case: the N-node graph under consideration consists of a single n-node clique plus a number of diversionary edges chosen by an adversary. We show that the algorithm can tolerate up to $O(n^2)$ diversionary edges provided that no non-clique vertex is adjacent to more than $O(n)$ clique vertices. We argue also that these two bounds, $O(n^2)$ and $O(n)$, are the best possible. We show analogous results for the biclique problem in Sect. 5.1.

Our second analysis, described in Sects. 4.2 and 5.2, supposes that the graph contains a single clique or biclique, while the remaining nonclique edges are inserted independently at random with fixed probability p. This problem has been studied by Alon et al. [2] and by Feige and Krauthgamer [6]. In the case of clique, we obtain the same result as they do, namely, that as long as the clique has at least $O(N^{1/2})$ nodes, where N is the number of nodes in G, then our algorithm will find it. Like Feige and Krauthgamer, our algorithm also certifies that the maximum clique has been found due to a uniqueness result for convex optimization, which we present in Sect. 3. We believe that our technique is more general than Feige and Krauthgamer; for example, ours extends essentially without alteration to the biclique problem, whereas Feige and Krauthgamer rely on some special properties of the clique problem. Furthemore, our results use only Chernoff bounds and classical theorems about the norms of random matrices. The random matrix results needed for our main theorems are presented in Sect. 2.

Our interest in the planted clique and biclique problems arises from applications in data mining. In data mining, one seeks a pattern hidden in an apparently unstructured set of data. A natural question to ask is whether a data mining algorithm is able to find the hidden pattern in the case that it is actually present but obscured by noise. For example, in the realm of clustering, Ben-David [1] has shown that if the data is actually clustered, then a clustering algorithm can find the clusters. The clique and biclique problems are both simple model problems for data mining. For example, Iasemidis et al. [14] reduce a data mining problem in epilepsy prediction to a maximum clique problem. Gillis and Glineur [11] use the biclique problem as a model problem for nonnegative matrix factorization and finding features in images.

2 Results on norms of random matrices

In this section we provide a few results concerning random matrices with independently identically distributed (i.i.d.) entries of mean 0. In particular, the probability distribution Ω for an entry A_{ij} will be as follows:

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

It is easy to check that the variance of A_{ij} is $\sigma^2 = p/(1-p)$.

We start by recalling a theorem of Füredi and Komlós [7]:

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

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

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

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

Remark 1 In this theorem and for the rest of the paper, $\|A\|$ denotes $\|A\|_2$, often called the spectral norm. It is equal to the maximum singular value of A or equivalently to the square root of the maximum eigenvalue of $A^T A$.

Remark 2 The theorem is not stated exactly in this way in [7]; 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.

Remark 3 As mentioned above, the mean value of entries of A is 0. This is crucial for the theorem; a distribution with any other mean value would lead to $\|A\| = \Omega(n)$.

A similar theorem due to Geman [9] is available for unsymmetric matrices.

Theorem 2 *Let A be a $\lceil yn \rceil \times n$ matrix whose entries are chosen according to Ω 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 $c_3 > 0$ depend on p and y,*

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

for some $c_4 > 0$ also depending on p, y.

As in the case of [7], this theorem is not stated exactly this way in Geman's paper, but can be deduced from the equations on pp. 255–256 by taking $k = n^q$ for a q satisfying $(2\alpha + 4)q < 1$.

The last theorem about random matrices requires a version of the well known Chernoff bounds, which is as follows (see [16, Theorem 4.4]).

Theorem 3 (Chernoff Bounds) *Let $X_1, \ldots, X_k$ 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 $\delta > 0$*

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

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

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

The final theorem of this section is as follows.

Theorem 4 *Let A be an $n \times N$ matrix whose entries are chosen according to Ω. 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*

$$P\left(\|A - \tilde{A}\|_F^2 \leq c_1 N \right) \geq 1 - (2/3)^N - N c_2^n. \tag{3}$$

Remark 1 The notation $\|A\|_F$ denotes the Frobenius norm of A, that is, $\left(\sum_i \sum_j A_{ij}^2\right)^{1/2}$. It is well known that $\|A\|_F \geq \|A\|$ for any A.

Remark 2 Note that $\tilde{A}$ is undefined if there is a j such that $n_j = n$. In this case we assume that $\|A - \tilde{A}\| = \infty$, i.e., the event considered in (3) fails.

Remark 3 Observe that the column sums of A are random variables with mean zero since the mean of the entries is 0. On the other hand, the column sums of $\tilde{A}$ are identically zero deterministically; this is the rationale for the choice of $\tilde{A} = -n_j/(n - n_j)$.

The proof of Theorem 4 is included as an appendix.

3 Optimality conditions for an instance of nuclear norm minimization

In this section, we prove a theorem that gives sufficient conditions for optimality and uniqueness of a solution to the convex optimization problem

$$\begin{aligned} \min \quad & \|X\|_* \\ \text{s.t.} \quad & \sum_{i=1}^{M}\sum_{j=1}^{N} X_{i,j} \geq mn, \\ & X_{i,j} = 0 \quad \text{for } (i,j) \in \tilde{E} \end{aligned} \tag{4}$$

which will arise in our analysis of the maximum clique and maximum edge biclique problems. Here, $X \in \mathbf{R}^{M\times N}$, $\|X\|_* = \sigma_1(X)+\cdots+\sigma_N(X)$ denotes the nuclear norm, the sum of the singular values of the matrix, E is a subset of $\{1,\ldots,M\}\times\{1,\ldots,N\}$, and the complement of E is denoted $\tilde{E}$. These conditions involve multipliers λ_{ij} and μ and a matrix W. In subsequent subsections we explain how to select λ_{ij}, μ and W based on the underlying graph to show that a particular feasible solution is optimal for certain instances of the maximum clique and maximum edge biclique problems.

Recall that if $f : \mathbf{R}^n \to \mathbf{R}$ is a convex function, then a *subgradient* of f at a point $\mathbf{x}$ is defined to be a vector $\mathbf{g} \in \mathbf{R}^n$ such that for all $\mathbf{y} \in \mathbf{R}^n$, $f(\mathbf{y}) - f(\mathbf{x}) \geq \mathbf{g}^T(\mathbf{y} - \mathbf{x})$. It is a well-known theorem that for a convex f and for every $\mathbf{x} \in \mathbf{R}^n$, the set of subgradients forms a nonempty closed convex set. This set of subgradients, called the *subdifferential*, is denoted as $\partial f(\mathbf{x})$.

The following lemma characterizes the subdifferential of $\|\cdot\|_*$ (see [3, Equation 3.4] and also [20]).

Lemma 1 *Suppose* $A \in \mathbf{R}^{m\times n}$ *has rank* r *with singular value decomposition* $A = \sum_{k=1}^{r} \sigma_k \mathbf{u}_k \mathbf{v}_k^T$. *Then* ϕ *is a subgradient of* $\|\cdot\|_*$ *at* A *if and only if* ϕ *is of the form*

$$\phi = \sum_{k=1}^{r} \mathbf{u}_k \mathbf{v}_k^T + W$$

where W *satisfies* $\|W\| \leq 1$ *such that the column space of* W *is orthogonal to* $\mathbf{u}_k$ *and the row space of* W *is orthogonal to* $\mathbf{v}_k$ *for all* $k = 1, 2, \ldots, r$.

Let I be a subset of $\{1,\ldots,N\}$. We say that $\mathbf{u} \in \mathbf{R}^N$ is the *characteristic vector* of I if $u_i = 1$ for $i \in I$ while $u_i = 0$ for $i \in \{1,\ldots,N\} - I$.

Let U^* be a subset of $\{1,\ldots,M\}$ and V^* a subset of $\{1,\ldots,N\}$, and let $\bar{\mathbf{u}}, \bar{\mathbf{v}}$ be their characteristic vectors respectively. Suppose $|U^*| = m$ and $|V^*| = n$ with $m > 0, n > 0$. Let $X^* = \bar{\mathbf{u}}\bar{\mathbf{v}}^T$, an $M \times N$ matrix. Clearly X^* has rank 1. Note that Lemma 1 implies that

$$\partial\|\cdot\|_*(X^*) = \{\bar{\mathbf{u}}\bar{\mathbf{v}}^T/\sqrt{mn} + W : W\bar{\mathbf{v}} = \mathbf{0},\ W^T\bar{\mathbf{u}} = \mathbf{0},\ \|W\| \le 1\}. \tag{5}$$

This leads to the main theorem for this section.

Theorem 5 *Let U^* be a subset of $\{1, \ldots, M\}$ of cardinality m, and let V^* be a subset of $\{1, \ldots, N\}$ of cardinality n. Let $\bar{\mathbf{u}}$ and $\bar{\mathbf{v}}$ be the characteristic vectors of U^*, V^* respectively. Let $X^* = \bar{\mathbf{u}}\bar{\mathbf{v}}^T$. Suppose X^* is feasible for (4). Suppose also that there exist $W \in \mathbf{R}^{M\times N}$, $\lambda \in \mathbf{R}^{M\times N}$ and $\mu \in \mathbf{R}_+$ such that $W\bar{\mathbf{v}} = \mathbf{0}$, $\bar{\mathbf{u}}^T W = \mathbf{0}$, $\|W\| \le 1$ and*

$$\frac{\bar{\mathbf{u}}\bar{\mathbf{v}}^T}{\sqrt{mn}} + W = \mu\mathbf{e}\mathbf{e}^T + \sum_{(i,j)\in\tilde{E}} \lambda_{ij}\mathbf{e}_i\mathbf{e}_j^T. \tag{6}$$

Here, $\mathbf{e}$ denotes the vector of all 1's while $\mathbf{e}_i$ denotes the ith column of the identity matrix (either in $\mathbf{R}^M$ or $\mathbf{R}^N$). Then X^* is an optimal solution to (4). Moreover, for any $I \subset \{1, \ldots, M\}$ and $J \subset \{1, \ldots, N\}$ such that $I \times J \subset E$, $|I| \cdot |J| \le mn$.

Furthermore, if $\|W\| < 1$ and $\mu > 0$, then X^* is the unique optimizer of (4) (and hence will be found if a solver is applied to (4)).

Proof The fact that X^* is optimal is a straightforward application of the well-known Karush-Kuhn-Tucker conditions (see, for example, [19, Theorem 28.3]).

Now consider (I, J) such that $I \times J \subset E$. Then $X' = \bar{\mathbf{u}}'(\bar{\mathbf{v}}')^T \cdot mn/(|I| \cdot |J|)$, where $\bar{\mathbf{u}}'$ is the characteristic vector of I and $\bar{\mathbf{v}}'$ is the characteristic vector of J, is also a feasible solution to (4). Recall that for a matrix of the form $\mathbf{u}\mathbf{v}^T$, the unique nonzero singular value (and hence the nuclear norm) equals $\|\mathbf{u}\| \cdot \|\mathbf{v}\|$. Thus, $\|X'\|_* = mn/(|I| \cdot |J|)^{1/2}$ and $\|X^*\|_* = \sqrt{mn}$. Since X^* is optimal, $\|X'\|_* \ge \|X^*\|$, i.e., $\sqrt{mn} \le mn/(|I| \cdot |J|)^{1/2}$. Simplifying yields $|I| \cdot |J| \le mn$.

Now finally we turn to the uniqueness of X^*. The optimization problem (4) can be formulated as the semidefinite program

$$\begin{array}{ll} \min & \frac{1}{2}(\mathrm{tr}\,(Z_1) + \mathrm{tr}\,(Z_2)) \\ \text{s.t.} & Z = \begin{bmatrix} Z_1 & X \\ X^T & Z_2 \end{bmatrix} \succeq 0 \\ & \sum_{i=1}^{M}\sum_{j=1}^{N} X_{ij} \ge mn \\ & X_{ij} = 0\ \forall (i,j) \in \tilde{E} \end{array} \tag{7}$$

(see [18, Equation 2.8]). This problem is strictly feasible, and, hence, strong duality holds. The dual problem of (7) is

$$\begin{array}{ll} \max\ mn\mu & \\ \text{s.t.}\quad Q = \begin{bmatrix} I & -\sum_{(i,j)\in\tilde{E}} \lambda_{ij}\mathbf{e}_i\mathbf{e}_j^T - \mu\mathbf{e}\mathbf{e}^T \\ -\sum_{(i,j)\in\tilde{E}} \lambda_{ij}\mathbf{e}_j\mathbf{e}_i^T - \mu\mathbf{e}\mathbf{e}^T & I \end{bmatrix} \succeq 0 & \\ \quad\ \mu \ge 0. & \end{array} \tag{8}$$

Now suppose that there exists $W, \lambda \in \mathbf{R}^{M \times N}$ and $\mu > 0$ satisfying (6) such that $W\bar{\mathbf{v}} = \mathbf{0}$, $\bar{\mathbf{u}}^T W = \mathbf{0}$ and $\|W\| < 1$. Notice that

$$X^* = \bar{\mathbf{u}}\bar{\mathbf{v}}^T, \quad Z_1^* = \frac{n}{\sqrt{mn}}\bar{\mathbf{u}}\bar{\mathbf{u}}^T, \quad Z_2^* = \frac{m}{\sqrt{mn}}\bar{\mathbf{v}}\bar{\mathbf{v}}^T$$

forms a primal feasible triple for (7) and the matrix Q^* as defined by μ and λ is dual feasible. Moreover, $Z^* Q^* = 0$, and, thus, by complementary slackness, Z^* is optimal for (7), Q^* is the corresponding dual optimal solution for (8) and X^* is optimal for (4). Also, note that $\bar{\mathbf{u}}\bar{\mathbf{v}}^T/\sqrt{mn} + W$ has maximum singular value equal to 1 with multiplicity 1 since $\|W\| < 1$. Therefore, since υ is an eigenvalue of $Q - I$ if and only υ is an eigenvalue of $\bar{\mathbf{u}}\bar{\mathbf{v}}^T/\sqrt{mn} + W$ or $-\upsilon$ is an eigenvalue of $\bar{\mathbf{u}}\bar{\mathbf{v}}^T/\sqrt{mn} + W$, Q has exactly one singular value equal to 0. It follows immediately that Q^* has rank equal to $M + N - 1$.

Now suppose that

$$\hat{Z} = \begin{bmatrix} \hat{Z}_1 & \hat{X} \\ \hat{X}^T & \hat{Z}_2 \end{bmatrix}$$

is optimal for (7) and, hence, $\hat{X}$ is optimal for (4). Moreover, since Q^* is dual optimal, $\hat{Z}Q^* = 0$ by complementary slackness. Therefore $\hat{Z}$ has rank 1 and must be equal to tZ^* for some nonnegative scalar t. It follows immediately that $\hat{X} = (1/t)\bar{\mathbf{u}}\bar{\mathbf{v}}^T$. However, since $\hat{X}$ is optimal for (4),

$$\sqrt{mn} = \|\hat{X}\| = \|\bar{\mathbf{u}}\bar{\mathbf{v}}^T\|/t = \sqrt{mn}/t,$$

and, hence, $t = 1$. Therefore, $X^* = \bar{\mathbf{u}}\bar{\mathbf{v}}^T$ is the unique minimizer of (4). □

4 The maximum clique problem

Let $G = (V, E)$ be a simple graph. A *clique* C of G is a subset of V such that the subgraph of G induced by C, denoted $G(C)$, is complete. The **maximum clique problem** focuses on finding the largest clique of graph G. The problem of determining the size of the maximum clique in a graph G, called the **clique number**, $\omega(G)$, is one of Karp's original NP-hard problems [8]. Moreover, $\omega(G)$ is hard to approximate. That is, for any fixed $\epsilon > 0$, $\omega(G)$ cannot be approximated within a ratio of $n^{1-\epsilon}$ unless NP has randomized polynomial time algorithms (see [12]). This intractibility suggests studying this problem for randomly generated input graphs. One particular set of such graphs are the planted or hidden cliques in which a random graph on n vertices is generated by first placing a clique of size k in the graph and then choosing each remaining pair of vertices to be adjacent independently at random with fixed probability p. Alon, Krivelevich, and Sudakov [2] and Feige and Krauthgamer [6] showed that in the case that $k = \Omega(\sqrt{n})$ then the planted clique can be found, with extremely high probability, in polynomial time. The algorithm of Feige and Krauthgamer uses the relaxation of the independent set problem on the complement $\overline{G}$ of G given by the Lovász theta function $\vartheta(\overline{G})$ and relies on the fact that, with extremely high probability,

$$\vartheta(\overline{G \setminus v}) = \begin{cases} k-1 & \text{if } v \text{ is in the planted clique} \\ k & \text{otherwise} \end{cases}$$

for every $v \in V(G)$ if G contains a planted clique of size $\Omega(\sqrt{n})$. On the other hand, Alon et al. exploit certain properties of the spectrum of the adjacency matrix A of G. In particular, the algorithm of Alon et al. relies on the fact that if G contains a planted clique containing $k = \Omega(\sqrt{n})$ vertices, then, if W is the set of indices corresponding to the k largest (in absolute value) entries of the eigenvector corresponding to the second largest eigenvalue of A, any vertex not in the planted clique is adjacent to less than $3k/4$ vertices in W with extremely high probability.

In this section, we will show that the same result can be obtained by relaxing the maximum clique problem to convex optimization using the nuclear norm. We believe our technique improves over the algorithms in [2,6] in several ways. First, our technique can be applied to find planted cliques in graphs that are constructed deterministically by having an adversary add noise to a graph containing a sufficiently large clique. Moreover, our technique extends to the maximum edge biclique problem without modification, whereas the algorithms of Alon et al. and Feige and Krauthgamer exploit special properties of the clique problem. Indeed, the complement of a biclique is not an independent set, and, hence, the relaxation of Feige and Krauthgamer using the Lovász theta number cannot be used to find a biclique without alteration. Similarly, the spectral properties of the adjacency matrix of a planted clique used by Alon et al. do not hold for the spectrum of a planted biclique.

For any clique K of G, the adjacency matrix of the graph K' obtained by taking the union of $G(K)$ and the set of loops for each $v \in K$ is a rank-one matrix with 1's in the entries indexed by $K \times K$ and 0's everywhere else. Therefore, a clique K of G containing n vertices can be found by solving the rank minimization problem

$$\begin{aligned} \min \quad & \text{rank}\,(X) \\ \text{s.t.} \quad & \sum_{i \in V} \sum_{j \in V} X_{ij} \geq n^2, && (9) \\ & X_{ij} = 0 \quad \text{if } ij \notin E \text{ and } i \neq j, && (10) \\ & X = X^T && (11) \\ & X \in [0,1]^{V \times V}. && (12) \end{aligned}$$

Unfortunately, this rank minimization problem is also NP-hard. We consider the relaxation obtained by replacing the objective function with the nuclear norm, the sum of the singular values of the matrix: $\|X\|_* = \sigma_1(X) + \cdots + \sigma_N(X)$.

Underestimating rank (X) with $\|X\|_*$, we obtain the following convex optimization problem:

$$\begin{aligned} \min \quad & \|X\|_* \\ \text{s.t.} \quad & \sum_{i\in V}\sum_{j\in V} X_{ij} \geq n^2, \\ & X_{ij} = 0 \quad \text{if } ij \notin E \text{ and } i \neq j. \end{aligned} \tag{13}$$

Notice that the relaxation has dropped the constraint $0 \leq X_{ij} \leq 1$ that was present in the original formulation. This constraint turns out to be superfluous [and, in fact, unhelpful—see the remark following (14)] for our approach. Notice that (13) is exactly the convex optimization problem given by (4) when $m = n$, $M = N$ and the subset E of $\{1, \dots, N\} \times \{1, \dots, N\}$ is equal to the union of the set of edges of G and the set of loops $\{ii : i \in V\}$. Hence, Theorem 5 immediately specializes to the following theorem.

Theorem 6 *Let V^* be an n-node clique contained in an N-node undirected graph $G = (V, E)$. Let $\bar{\mathbf{v}} \in \mathbf{R}^V$ be the characteristic vector of V^*. Let $X^* = \bar{\mathbf{v}}\bar{\mathbf{v}}^T$. (Clearly X^* is feasible for (13). Suppose also that there exist $W \in \mathbf{R}^{V\times V}$, $\lambda \in \mathbf{R}^{V\times V}$ and $\mu \in \mathbf{R}_+$ such that $W\bar{\mathbf{v}} = \mathbf{0}$, $\bar{\mathbf{v}}^T W = \mathbf{0}$, $\|W\| \leq 1$ and*

$$\frac{\bar{\mathbf{v}}\bar{\mathbf{v}}^T}{n} + W = \mu \mathbf{e}\mathbf{e}^T + \sum_{(i,j)\in\tilde{E}} \lambda_{ij}\mathbf{e}_i\mathbf{e}_j^T. \tag{14}$$

Then X^* is an optimal solution to (13). Moreover, V^* is a maximum clique of G. Furthermore, if $\|W\| < 1$ and $\mu > 0$, then X^* is the unique optimizer of (13), and V^* is the unique maximum clique of G.

Remark It may appear that we need to know the value of n prior to applying the theorem since n is present in the statement of (13). In fact, this is not the case: we observe that the factor n^2 appearing in (13) is the sole inhomogeneity in the problem. This means that we obtain the same solution, rescaled in the appropriate way, if we replace n^2 by 1 in (13). Thus, n does not need to be known in advance to apply this theorem.

For the next two subsections, we consider two scenarios for constructing G and try to find X^*, W and values for the multipliers to satisfy the conditions of the previous theorem. For both subsections, we use the following choices. We take $\mu = 1/n$ where $n = |V^*|$. We define W and λ by considering the following cases:

- (ω_1) If $(i, j) \in V^* \times V^*$, we choose $W_{ij} = 0$ and $\lambda_{ij} = 0$. In this case, the entries on other side of (14) corresponding to this case become $1/n + 0 = 1/n + 0$.
- (ω_2) If $(i, j) \in E - (V^* \times V^*)$ such that $i \neq j$, then we choose $W_{ij} = 1/n$ and $\lambda_{ij} = 0$. Then the two sides of (14) become $0 + 1/n = 1/n + 0$.
- (ω_3) If $i \notin V^*$, we set $W_{ii} = 1/n$. Again the two sides of (14) become $0 + 1/n = 1/n + 0$.
- (ω_4) If $(i, j) \notin E, i \notin V^*, j \notin V^*$, then we choose $W_{ij} = -\gamma/n$ and $\lambda_{ij} = -(1+\gamma)/n$ for some constant $\gamma \in \mathbf{R}$. The two sides of (14) become $0 - \gamma/n = 1/n - (1 + \gamma)/n$. The value of γ is specified below.

(ω_5) If $(i, j) \notin E, i \in V^*, j \notin V^*$, then we choose

$$W_{ij} = -\frac{p_j}{n(n-p_j)}, \quad \lambda_{ij} = -\frac{1}{n} - \frac{p_j}{n(n-p_j)}$$

where p_j is equal to the number of edges in E from j to V^*.

(ω_6) If $(i, j) \notin E, i \notin V^*, j \in V^*$ then choose W_{ij}, λ_{ij} symmetrically with the previous case.

First, observe that $W\bar{\mathbf{v}} = \mathbf{0}$. Indeed, for entries $i \in V^*, W(i, :)\bar{\mathbf{v}} = 0$ since $W(i, V^*) = 0$ for such entries. For entries $i \in V - V^*$,

$$W(i, :)\bar{\mathbf{v}} = p_i \frac{1}{n} - (n - p_i)\frac{p_i}{n(n-p_i)} = 0$$

by our special choice of $W(i, j)$ in cases 5 and 6.

It remains to determine which graphs G yield W as defined by (ω_1)–(ω_6) such that $\|W\| < 1$. We present two different analyses.

4.1 The adversarial case

Suppose that the edge set of the graph $G = (V, E)$ is generated as follows. We first add a complete subgraph K_{V^*} with vertex set V^* of size n. Then, an adversary is allowed to add a number of the remaining $|V|(|V| - 1)/2 - n(n-1)/2$ potential edges to the graph. We will show that, under certain conditions, our adversary can add up to $O(n^2)$ edges to the graph and V^* will still be the unique maximum clique of G.

We first introduce the following notation. Let $W^D \in \mathbf{R}^{V \times V}$ denote the matrix with diagonal entries equal to the diagonal entries of W and all other entries equal to 0. Let W^{ND} be the matrix whose nondiagonal entries are equal to the corresponding nondiagonal entries of W and whose diagonal entries are equal to 0. So $W = W^D + W^{ND}$.

Now suppose $G = (V, E)$ contains a clique V^* of size n with vertices indexed by $V^* \in \mathbf{R}^V$. Moreover, suppose that G contains at most r edges not in K_{V^*} and each vertex in $V - V^*$ is adjacent to at most δn vertices in V^* for some $\delta \in (0, 1)$. Consider W as defined by (ω_1)–(ω_6) with $\gamma = 0$. By the triangle inequality,

$$\|W\|^2 \le (\|W^D\| + \|W^{ND}\|)^2 \le 2(\|W^D\|^2 + \|W^{ND}\|^2) = 2(1/n^2 + \|W^{ND}\|^2)$$

since $\|W^D\| = 1/n$. Applying the bound $\|W\| \le \|W\|_F$, it suffices to determine which values of r yield

$$\|W^{ND}\|_F^2 = 2\|W(V^*, V - V^*)\|_F^2 + \|W^{ND}(V - V^*, V - V^*)\|_F^2 < (n^2 - 2)/(2n^2)$$

since, by the symmetry of W,

$$W^{ND}(V^*, V - V^*) = W(V^*, V - V^*) = W(V - V^*, V^*).$$

The diagonal entries of $W^{ND}(V - V^*, V - V^*)$ are equal to 0 and at most $2r$ of the remaining entries are equal to $1/n$. Therefore,

$$\|W^{ND}(V - V^*, V - V^*)\|_F^2 \leq 2r/n^2.$$

Moreover, since $n - p_j \geq (1 - \delta)n$,

$$\begin{aligned}
\|W(V^*, V - V^*)\|_F^2 &= \sum_{j \in V - V^*} \left(p_j \cdot \frac{1}{n^2} + (n - p_j) \cdot \frac{p_j^2}{(n - p_j)^2 n^2} \right) \\
&= \sum_{j \in V - V^*} \left(\frac{p_j}{n^2} + \frac{p_j^2}{(n - p_j) n^2} \right) \\
&\leq \sum_{j \in V - V^*} \left(\frac{p_j}{n^2} + \frac{\delta n p_j}{(1 - \delta) n^3} \right) \\
&= \left(\frac{1}{1 - \delta} \right) \sum_{j \in V - V^*} \frac{p_j}{n^2} \\
&\leq \left(\frac{1}{1 - \delta} \right) \frac{r}{n^2}.
\end{aligned}$$

Thus, the optimality and uniqueness conditions given by Theorem 5 are satisfied by X^* if

$$\left(1 + \frac{1}{1 - \delta} \right) r < (n^2 - 2)/4.$$

Equivalently,

$$r < \frac{1 - \delta}{4(2 - \delta)} (n^2 - 2).$$

Therefore, G can contain up to $O(n^2)$ edges other than those in $V^* \times V^*$, and yet V^* will remain the unique maximum clique of G.

Note that these bounds are the best possible up to the constant factors. In particular, if the adversary were able to insert $(n + 1)(n + 2)/2$ edges, then a new clique could be created larger than the planted clique. Thus, the adversary must be limited to const $\cdot\, n^2$ edges for const $< 1/2$. Similarly, if the adversary could join a nonclique vertex to n clique vertices, then the adversary would have enlarged the clique. Thus, the restriction that a nonclique vertex is adjacent to at most const $\cdot\, n$ clique vertices is the best possible.

4.2 The randomized case

Let V be a set of vertices with $|V| = N$ and consider a subset $V^* \subseteq V$ such that $|V^*| = n$. We construct the edge set E of the graph $G = (V, E)$ as follows:

(Γ_1) For all $(i, j) \in V^* \times V^*$, $ij \in E$.
(Γ_2) Each of the remaining $N(N-1)/2 - n(n-1)/2$ possible edges is added to E independently at random with probability $p \in [0, 1)$.

Notice that, by our construction of E, V^* is a clique of G of size n. We wish to determine which n, N yield G as constructed by (Γ_1) and (Γ_2) such that with high probability $X^* = \bar{\mathbf{v}}\bar{\mathbf{v}}^T$ is optimal for the convex relaxation of the clique problem given by (13). The following theorem states the desired result.

Theorem 7 *There exists an $\alpha > 0$ depending on p such that for all G constructed via (Γ_1), (Γ_2) with $n \geq \alpha\sqrt{N}$, the clique defined by $V^* \times V^*$ is the unique maximum clique of G and will correspond to the unique solution of (13) with probability tending exponentially to 1 as $N \to \infty$.*

Proof Consider the matrix W constructed as in (ω_1)–(ω_6) with $\gamma = -p/(1-p)$. By Theorem 6, X^* is the unique optimum if

$$\|W\| < 1 \quad \text{and} \quad p_j < n \quad \text{for all } j \in V - V^*.$$

We first show that $\|W\| < 1$ with probability tending exponentially to 1 as $N \to \infty$ in the case that $n = \Omega(\sqrt{N})$. We write $W = W_1 + W_2 + W_3 + W_4 + W_5$, where each of the five terms is defined as follows.

We first define W_1. For cases (ω_2) and (ω_4), choose $W_1(i, j) = W(i, j)$. For cases (ω_5) and (ω_6), take $W_1(i, j) = -p/((1-p)n)$. For case (ω_1), choose $W_1(i, j)$ randomly such that $W_1(i, j)$ is equal to $1/n$ with probability p and equal to $-p/((1-p)n)$ otherwise. Similarly, in case (ω_3), take $W_1(i, i)$ to be equal to $1/n$ with probability p and equal to $-p/((1-p)n)$ otherwise. By construction, each entry of W_1 is an independent random variable with the distribution

$$W_1(i, j) = \begin{cases} 1/n & \text{with probability } p, \\ -p/((1-p)n) & \text{with probability } 1-p. \end{cases}$$

Therefore, applying Lemma 1 shows that there exists constant $c_1 > 0$ such that

$$\|W_1\| \leq 3\left(\frac{p}{1-p}\right)^{1/2} \frac{\sqrt{N}}{n} \tag{15}$$

with probability at least to $1 - \exp(c_1 N^{1/6})$ for some constant $c_1 > 0$.

Next, W_2 is the correction matrix to W_1 in case (ω_1). That is, $W_2(i, j)$ is chosen such that

$$W_2(i, j) + W_1(i, j) = W(i, j) = 0$$

for all $(i, j) \in V^* \times V^*$ and is zero everywhere else. As before, applying Lemma 1 shows that

$$\|W_2\| \leq 3\left(\frac{p}{1-p}\right)^{1/2} \frac{1}{\sqrt{n}} \tag{16}$$

with probability at least $1 - \exp(c_1 n^{1/6})$ for some constant $c_1 > 0$. Similarly, W_3 is the correction to W_3 in case (ω_3), that is

$$W_3(i, i) = W(i, i) - W_1(i, i)$$

for all $i \in V - V^*$ and all other entries are equal 0. Therefore, W_3 is a diagonal matrix with diagonal entries bounded by $2/n$. It follows that

$$\|W_3\| \leq \frac{2}{n}. \tag{17}$$

Finally, W_4 and W_5 are the corrections for cases (ω_5) and (ω_6) respectively. These are exactly of the form $(A - \tilde{A})/n$ as in Theorem 4, in which N in the theorem stands for $N - n$ in the present context. Examining each term of (3) shows that in the case $n = \Omega(N^{1/2})$, the probability that $\|A - \tilde{A}\|_F^2 = O(N)$ is at least $1 - k_1 \exp(-k_2 N^{k_3})$ for some $k_1, k_2, k_3 > 0$ depending on p. It follows that there exists constant $\alpha_4 > 0$ such that

$$\|W_4\|^2 \leq \|W_4\|_F^2 < \alpha_4^2 N n^{-2}$$

with probability tending to 1 as $N \to \infty$. Moreover, in the case that $\|W_4\|^2 < \alpha_4^2 N n^{-2}$, the Condition Ψ in the proof of Theorem 4 is not violated; here n_j in the proof stands for p_j, for each $j \in V - V^*$. This implies that $p_j < n$ in the case that $\|W_4\|^2 < \alpha_4^2 N n^{-2}$. Notice that, by symmetry, $W_4 = W_5^T$. Thus, since each of $W_1, W_2, \ldots, W_5$ is bounded by an arbitrarily small constant if $n = \Omega(\sqrt{N})$, there exists constant $\alpha > 0$ such that $\|W\| < 1$ with probability tending exponentially to 1 as $N \to \infty$ as required. □

5 The maximum edge biclique problem

Consider a bipartite graph $G = ((U, V), E)$ where $|U| = M$, $|V| = N$. A pair of disjoint independent subsets $U' \subseteq U$, $V' \subseteq V$ is a *biclique* of G if $uv \in E$ for all $u \in U'$, $v \in V'$. That is, (U', V') is a biclique of G if the subgraph of G induced by (U', V') is complete bipartite. The adjacency matrix of the subgraph of a biclique (U', V') of G is a rank-one matrix $X \in \mathbf{R}^{M \times N}$. This matrix X has the property that $X_{ij} = 0$ for all $i \in U$, $j \in V$ such that $ij \notin E$. It follows that a biclique of G of size

mn can be found (if one exists) by solving the rank minimization problem

$$\begin{aligned} \min \quad & \text{rank}\,(X) \\ \text{s.t.} \quad & \sum_{i \in U} \sum_{j \in V} X_{ij} \geq mn, && (18) \\ & X_{ij} = 0 \quad \forall\, ij \in (U \times V) - E, && (19) \\ & X \in [0, 1]^{U \times V}. && (20) \end{aligned}$$

A rank-one solution X^* to this problem corresponds to the adjacency matrix of the subgraph induced by a biclique of G containing at least mn edges. As with the maximum clique problem, this rank minimization problem is still NP-hard. As before, we underestimate rank (X) with $\|X\|_*$. We obtain the following convex optimization problem:

$$\begin{aligned} \min \quad & \|X\|_* && (21) \\ \text{s.t.} \quad & \sum_{i \in U} \sum_{j \in V} X_{ij} \geq mn, \\ & X_{ij} = 0 \quad \text{if } ij \notin E. \end{aligned}$$

Using the Karush–Kuhn–Tucker conditions, we derive conditions for which the adjacency matrix of a subgraph induced by a biclique of G is optimal for this relaxation. Indeed, the following is an immediate consequence (essentially a restatement) of Theorem 5.

Theorem 8 *Let (U^*, V^*) be a biclique in G in which $|U^*| = m$ and $|V^*| = n$. Let $\bar{\mathbf{u}} \in \mathbf{R}^M$ be the characteristic vector of U^*, and let $\bar{\mathbf{v}} \in \mathbf{R}^N$ be the characteristic vector of V^*. Let $X^* = \bar{\mathbf{u}}\bar{\mathbf{v}}^T$. (Clearly X^* is feasible for (21)). Let $E = E(G)$ and let $\tilde{E}$ be its complement. Suppose also that there exist $W \in \mathbf{R}^{M \times N}$, $\lambda \in \mathbf{R}^{M \times N}$ and $\mu \in \mathbf{R}_+$ such that $W\bar{\mathbf{v}} = \mathbf{0}$, $\bar{\mathbf{u}}^T W = \mathbf{0}$, $\|W\| \leq 1$ and*

$$\frac{\bar{\mathbf{u}}\bar{\mathbf{v}}^T}{\sqrt{mn}} + W = \mu \mathbf{e}\mathbf{e}^T + \sum_{i,j) \in \tilde{E}} \lambda_{ij} e_i e_j^T. \tag{22}$$

Then X^* is an optimal solution to (21). Moreover, G does not contain any biclique with more than mn edges. Furthermore, if $\|W\| < 1$ and $\mu > 0$, then X^* is the unique optimizer of (21) and (U^*, V^*) is the unique optimal biclique of G.

In the next two subsections, we consider two scenarios for how to construct a bipartite graph G and biclique that satisfy the conditions of the theorem.

In both scenarios, we will take $\mu = 1/\sqrt{mn}$ and consider W and λ defined according to the following cases.

(ψ_1) For $ij \in U^* \times V^*$, taking $W_{ij} = 0$ and $\lambda_{ij} = 0$ ensures the ij-entries of both sides of (22) are equal to $1/\sqrt{mn}$.

(ψ_2) For $ij \in E - (U^* \times V^*)$, we take $W_{ij} = 1/\sqrt{mn}$ and $\lambda_{ij} = 0$. Again, the ij-entries of both sides of (22) are equal to $1/\sqrt{mn}$.

(ψ_3) For $ij \notin E$ such that $i \notin U^*$ and $j \notin V^*$, we select $W_{ij} = -\gamma/\sqrt{mn}$ and $\lambda_{ij} = -(1+\gamma)/\sqrt{mn}$ where γ will be defined below. In this case, the ij-entries of each side of (22) are 0.

(ψ_4) For $ij \notin E$ such that $i \notin U^*$ and $j \in V^*$, we choose

$$W_{ij} = -\frac{p_i}{(n - p_i)\sqrt{mn}} \quad \text{and} \quad \lambda_{ij} = \frac{1}{\sqrt{mn}}\left(\frac{-p_i}{n - p_i} - 1\right)$$

where p_i is equal to the number of edges with left endpoint equal to i and right endpoint in V^*. Note that if $n = p_i$ then i is connected to every vertex of V^* and thus the KKT condition cannot possibly be satisfied. If $p_i < n$, both sides of (22) are equal to $-p_i/((n - p_i)\sqrt{mn})$.

(ψ_5) For $ij \notin E$ such that $i \in U^*$ and $j \notin V^*$, we choose

$$W_{ij} = -\frac{q_j}{(m - q_j)\sqrt{mn}} \quad \text{and} \quad \lambda_{ij} = \frac{1}{\sqrt{mn}}\left(\frac{-q_j}{m - q_j} - 1\right)$$

where q_j is equal to the number of edges with right endpoint equal to j and left endpoint in U^*. As before, this is appropriate only if $q_j < m$.

We next check that W satisfies the requirements for ϕ to be a subgradient of $\bar{\mathbf{u}}\bar{\mathbf{v}}^T$: $W\bar{\mathbf{v}} = \mathbf{0}$, $W^T\bar{\mathbf{u}} = \mathbf{0}$, and $\|W\| \leq 1$. To show that $W\bar{\mathbf{v}} = \mathbf{0}$, choose row i of W and consider $W(i,:)\bar{\mathbf{v}} = \sum_{j \in V^*} W_{ij}$. If $i \in U^*$ then $W_{ij} = 0$ for all $j \in V^*$, so $W(i,:)\bar{\mathbf{v}} = 0$. In the case $i \notin U^*$, consider each $j \in V^*$. If $ij \in E$ then, by Case 2, $W_{ij} = 1/\sqrt{mn}$. There are p_i such entries, with sum $p_i/\sqrt{mn}$. If $ij \notin E$, then $W_{ij} = -p_i/((n - p_i)\sqrt{mn})$. There are $n - p_i$ such entries, with sum $-p_i/\sqrt{mn}$. It follows that $W(i,:)\bar{\mathbf{v}} = 0$ as required.

The proof that $W^T\bar{\mathbf{u}} = \mathbf{0}$ follows is symmetric. It remains to determine which bipartite graphs G yield W as defined above such that $\|W\| < 1$. As in the maximum clique case, we present two different analyses.

5.1 The adversarial case

Suppose that the edge set of the bipartite graph $G = ((U, V), E)$ is generated as follows. We first add a complete bipartite subgraph K_{U^*,V^*} with vertex sets U^*, V^* of sizes $|U^*| = m$, $|V^*| = n$ respectively. Then, as in the adversarial case for the maximum clique problem, an adversary is allowed to add a number of the remaining $|U||V| - mn$ potential edges to the graph. We will show that, under certain conditions, our adversary can add up to $O(mn)$ edges to the graph and (U^*, V^*) will still be a maximum edge biclique of G.

We make the following assumptions on the structure of G:

1. G contains at most r edges aside from those of K_{U^*,V^*}.
2. Each vertex of $V - V^*$ is adjacent to at most αm vertices of U^* for some $\alpha \in (0, 1)$.
3. Each vertex of $U - U^*$ is adjacent to at most βn vertices of V^* for some $\beta \in (0, 1)$.

Consider W as defined by (ψ_1)–(ψ_5) with $\gamma = 0$. As before, we use the bound $\|W\| \leq \|W\|_F$. Notice that at most r entries of $W(U - U^*, V - V^*)$ are equal to $1/\sqrt{mn}$ and the remainder are equal to 0. Therefore,

$$\|W(U - U^*, V - V^*)\|_F^2 \leq \frac{r}{mn}.$$

Moreover, for each $j \in V - V^*$, $q_j \leq \alpha m$. It follows that

$$\begin{aligned}
\|W(U^*, V - V^*)\|_F^2 &= \sum_{v \in V - V*} \left(\frac{q_v}{mn} + (m - q_v) \frac{q_v^2}{mn(m - q_v)^2} \right) \\
&= \sum_{v \in V^*} \frac{q_v}{mn} \left(1 + \frac{q_v}{m - q_v} \right) \\
&\leq \sum_{v \in V^*} \frac{q_v}{mn} \left(1 + \frac{\alpha}{1 - \alpha} \right) \\
&= \sum_{v \in V^*} \frac{q_v}{mn(1 - \alpha)} \leq \frac{r}{mn(1 - \alpha)}.
\end{aligned}$$

Similarly,

$$\|W(U - U^*, V^*)\|_F^2 \leq \frac{r}{(1 - \beta)mn}.$$

Therefore, $\|W\| < 1$ if

$$r \left(1 + \frac{1}{1 - \alpha} + \frac{1}{1 - \beta} \right) < mn.$$

Thus, the graph can contain up to $O(mn)$ diversionary edges, yet the optimality and uniqueness conditions given by Theorem 8 are still satisfied. This result is the best possible up to constants for the same reasons explained at the end of Sect. 4.1.

5.2 The random case

Let y, z be fixed positive scalars. Let U, V be two disjoint vertex sets with $|V| = N$ and $|U| = \lceil yN \rceil$. Consider $U^* \subseteq U$ and $V^* \subseteq V$ such that $|V^*| = n$ and $|U^*| = m = \lceil zn \rceil$. Suppose the edges of the bipartite graph $G = ((U, V), E)$ are determined as follows:

- (β_1) For all $(i, j) \in U^* \times V^*$, $ij \in E$.
- (β_2) For each of the remaining potential edges $(i, j) \in U \times V$, we add edge ij to E with probability p (independently).

Notice G contains the biclique (U^*, V^*). As in the maximum clique problem, if $n = \Omega(\sqrt{N})$ and G is constructed as in (β_1), (β_2) then $U^* \times V^*$ is optimal for the convex problem (21). We have the following theorem.

Theorem 9 *There exists $\alpha > 0$ depending on p, y, z such that for each bipartite graph G constructed via (β_1), (β_2) with $n \geq \alpha\sqrt{N}$ the biclique (U^*, V^*) is the maximum edge biclique of G with probability tending exponentially to 1 as $N \to \infty$ and is found as the unique solution to the convex relaxation (21).*

Let W be constructed as in (ψ_1)–(ψ_5) with $\gamma = -p/(1-p)$. Then $X^* = \bar{\mathbf{u}}\bar{\mathbf{v}}^T$ is the unique optimal solution of (21) if

$$\|W\| < 1, \quad q_j < \lceil zn \rceil \; \forall\, j \in V - V^*, \quad \text{and} \quad p_j < n \; \forall\, j \in U - U^*.$$

To prove that $\|W\| < 1$ with high probability as $N \to \infty$ in the case that $n = \Omega(\sqrt{N})$, we write

$$W = W_1 + W_2 + W_3 + W_4$$

where each of the summands is defined as follows. We first define W_1. If $(i, j) \in U^* \times V^*$, then we set $W_1(i, j) = 1/\sqrt{mn}$ with probability p and equal to $\gamma/\sqrt{mn}$ with probability $(1-p)$. For $(i, j) \in (U \times V) - (U^* \times V^*)$, we set $W_1(i, j) = 1/\sqrt{mn}$ if $ij \in E$ and set $W_1(i, j) = \gamma/\sqrt{mn}$ otherwise. In order to bound $\|W_1\|$, observe that Theorem 2 implies that $\|W_1\| \leq \text{const}\sqrt{N}/\sqrt{mn}$ with probability exponentially close to 1. Since $\sqrt{mn}$ equals $\sqrt{\lceil zn \rceil n}$ and hence is proportional to n, we see that $\|W_1\| \leq$ const with probability exponentially close to 1 provided $n = \Omega(N^{1/2})$.

Next, set W_2 to be the correction matrix for W_1 for $U^* \times V^*$, that is,

$$W_2(i, j) = \begin{cases} -W_1(i, j) & \text{if } (i, j) \in U^* \times V^* \\ 0 & \text{otherwise,} \end{cases}$$

Again, by Theorem 2 we conclude that

$$\|W_2\| \leq \alpha \frac{1}{\sqrt{n}}$$

with probability at least $1 - c_1' \exp(-c_2' n^{c_3'})$ for some $c_1', c_2', c_3' > 0$.

It remains to derive bounds for $\|W_3\|$ and $\|W_4\|$. Notice that the construction of $W(U^*, V - V^*)$ and $W(U - U^*, V^*)$ is identical to that in Case (ω_5) for the maximum clique problem. Thus, we can again apply Theorem 4, first to W_3 (in which case (n, N) in the theorem stand for $(\lceil zn \rceil, N - n)$) and second to W_4^T (in which case (n, N) in the theorem stand for $(n, \lceil yN \rceil - \lceil zn \rceil)$ to conclude that $\|W_3\|$ and $\|W_4\|$ are both strictly bounded above by constants provided $n = \Omega(N^{1/2})$ with probability tending to 1 exponentially fast. Moreover, as before, the application of Theorem 4 to bound W_3 and W_4, implies that $q_j < \lceil zn \rceil \; \forall\, j \in V - V^*$ and $p_j < n \; \forall\, j \in U - U^*$ with probability tending exponentially to 1 as $N \to \infty$ as required. □

6 Conclusions

We have shown that the maximum clique and maximum biclique problems can be solved in polynomial time using nuclear norm minimization, a technique recently proposed in the compressive sensing literature, provided that the input graph consists of a single clique or biclique plus diversionary edges. The spectral technique used by Alon et al. [2] for the planted clique problem has been extended to other problems; see, e.g., McSherry [15]. It would be interesting to extend the nuclear norm approach to other NP-hard problems as well.

Acknowledgments The current proof of the uniqueness result in Theorem 5 is based on an idea provided by an anonymous referee. This suggestion allowed us to significantly simplify the original proof. We also received helpful comments from our colleagues C. Swamy and D. Chakrabarty.

7 Appendix

7.1 Proof of Theorem 4

From the definition of $\tilde{A}$, for column j, there are exactly $n - n_j$ entries of $\tilde{A}$ that differ from those of A. Furthermore, the difference of these entries is exactly $(n_j - pn)/((1-p)(n-n_j))$. Therefore, for each $j = 1, \ldots, N$, the contribution of column j to the square norm difference $\|A - \tilde{A}\|_F^2$ is given by

$$\|A(:,j) - \tilde{A}(:,j)\|_F^2 = \frac{(n_j - pn)^2}{(1-p)^2(n-n_j)}.$$

Recall that the numbers $n_1, \ldots, n_N$ are independent, and each is the result of n Bernoulli trials done with probability p.

We now define Ψ to be the event that at least one n_j is very far from the mean. In particular, Ψ is the event that there exists a $j \in \{1, \ldots, N\}$ such that $n_j > qn$, where $q = \min(\sqrt{p}, 2p)$. Let $\tilde{\Psi}$ be its complement, and let $\tilde{\psi}(n_j)$ be the indicator of this complement (i.e., $\tilde{\psi}(n_j) = 1$ if $n_j \le qn$ else $\tilde{\psi}(n_j) = 0$). Let c be a positive scalar depending on p to be determined later. Observe that

$$\begin{aligned} P\left(\|A - \tilde{A}\|_F^2 \ge cN\right) &= P\left(\|A - \tilde{A}\|_F^2 \ge cN \;\wedge\; \tilde{\Psi}\right) + P\left(\|A - \tilde{A}\|_F^2 \ge cN \;\wedge\; \Psi\right) \\ &\le P\left(\|A - \tilde{A}\|_F^2 \ge cN \;\wedge\; \tilde{\Psi}\right) + P(\Psi). \end{aligned} \tag{23}$$

We now analyze the two terms separately. For the first term we use a technique attributed to S. Bernstein (see Hoeffding [13]). Let ϕ be the indicator function of nonnegative reals, i.e., $\phi(x) = 1$ for $x \ge 0$ while $\phi(x) = 0$ for $x < 0$. Then, in general, $P(u \ge 0) \equiv E(\phi(u))$. Thus,

$$\begin{aligned} P\left(\|A - \tilde{A}\|_F^2 \ge cN \;\wedge\; \tilde{\Psi}\right) &= P\left(\|A - \tilde{A}\|_F^2 - cN \ge 0 \;\wedge\; \tilde{\psi}(n_1) = 1 \;\wedge\; \cdots \;\wedge\; \tilde{\psi}(n_N) = 1\right) \\ &= E\left(\phi\left(\|A - \tilde{A}\|_F^2 - cN\right) \cdot \tilde{\psi}(n_1) \cdots \tilde{\psi}(n_N)\right). \end{aligned}$$

Let h be a positive scalar depending on p to be determined later. Observe that for any such h and for all $x \in \mathbf{R}$, $\phi(x) \leq \exp(hx)$. Thus,

$$\begin{aligned}
P\left(\|A-\tilde{A}\|_F^2 \geq cN \wedge \tilde{\Psi}\right) &\leq E\left(\exp\left(h\|A-\tilde{A}\|_F^2 - hcN\right)\cdot\tilde{\psi}(n_1)\cdots\tilde{\psi}(n_N)\right)\\
&= E\left(\exp\left(h\sum_{j=1}^{N}\left(\|A(:,j)-\tilde{A}(:,j)\|_F^2 - c\right)\right)\cdot\tilde{\psi}(n_1)\cdots\tilde{\psi}(n_N)\right)\\
&= E\left(\exp\left(h\sum_{j=1}^{N}\left(\frac{(n_j-pn)^2}{(1-p)^2(n-n_j)}-c\right)\right)\cdot\tilde{\psi}(n_1)\cdots\tilde{\psi}(n_N)\right)\\
&= E\left(\prod_{j=1}^{N}\exp\left(h\left(\frac{(n_j-pn)^2}{(1-p)^2(n-n_j)}-c\right)\right)\tilde{\psi}(n_j)\right)\\
&= \prod_{j=1}^{N}E\left(\exp\left(h\left(\frac{(n_j-pn)^2}{(1-p)^2(n-n_j)}-c\right)\right)\tilde{\psi}(n_j)\right) \qquad (24)\\
&= f_1\cdots f_N, \qquad (25)
\end{aligned}$$

where

$$f_j = E\left(\exp\left(h\left(\frac{(n_j-pn)^2}{(1-p)^2(n-n_j)}-c\right)\right)\tilde{\psi}(n_j)\right).$$

To obtain (24), we used the independence of the n_j's. Let us now analyze f_j in isolation.

$$\begin{aligned}
f_j &= \sum_{i=0}^{n}\exp\left(h\left(\frac{(i-pn)^2}{(1-p)^2(n-i)}-c\right)\right)\tilde{\psi}(n_j)P(n_j=i)\\
&= \sum_{i=0}^{\lfloor qn\rfloor}\exp\left(h\left(\frac{(i-pn)^2}{(1-p)^2(n-i)}-c\right)\right)P(n_j=i)\\
&\leq \sum_{i=0}^{\lfloor qn\rfloor}\exp\left(h\left(\frac{(i-pn)^2}{(1-p)^2(n-\sqrt{p}n)}-c\right)\right)P(n_j=i).
\end{aligned}$$

To derive the last line, we used the fact that $i \leq \sqrt{p}n$ since $i \leq qn$. Now let us reorganize this summation by considering first i such that $|i-pn| < \sqrt{n}$, and next i such that $|i-pn| \in [\sqrt{n}, 2\sqrt{n})$, etc. Notice that, since $i \leq qn \leq 2pn$, we need consider intervals only until $|i-pn|$ reaches pn.

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

where, for the last line, we have applied (2). The theorem is valid since $k \le p\sqrt{n}$.

Continuing this derivation and overestimating the finite sum with an infinite sum,

$$\begin{aligned}
f_j &\le 2\exp(-hc)\cdot\sum_{k=0}^{\infty}\exp\left(\frac{h(k+1)^2}{(1-p)^2(1-\sqrt{p})}-k^2/p\right) \\
&= 2\exp\left(\frac{h}{(1-p)^2(1-\sqrt{p})}-hc\right) \\
&\quad +2\exp(-hc)\cdot\sum_{k=1}^{\infty}\exp\left[\frac{h(k+1)^2}{(1-p)^2(1-\sqrt{p})}-k^2/p\right].
\end{aligned}$$

Choose h so that $h/((1-p)^2)(1-\sqrt{p}) < 1/(8p)$, i.e., $h < (1-p)^2(1-\sqrt{p})/(8p)$. Then the second term in the square-bracket expression at least twice the first term for all $k \ge 1$, hence

$$f_j \le 2\exp\left(\frac{h}{(1-p)^2(1-\sqrt{p})}-hc\right) + 2\exp(-hc)\cdot\sum_{k=1}^{\infty}\exp\left(-k^2/(2p)\right). \tag{26}$$

Observe that $\sum_{k=1}^{\infty}\exp(-k^2/(2p))$ is dominated by a geometric series and hence is a finite number depending on p. Thus, once h is selected, it is possible to choose c sufficiently large so that each of the two terms in (26) is at most $1/3$. Thus, with appropriate choices of h and c, we conclude that $f_j \le 2/3$. Thus, substituting this into (25) shows that

$$P\left(\|A-\tilde{A}\|_F^2 \ge cN \wedge \tilde{\Psi}\right) \le (2/3)^N. \tag{27}$$

We now turn to the second term in (23). For a particular j, the probability that $n_j > qn$ is bounded using (1) by v_p^n where $v_p = (e^\delta/(1+\delta)^{(1+\delta)})^p$, where $\delta = q/p - 1$, i.e., $\delta = \min(p, \sqrt{p} - p)$. Then the union bound asserts that the probability that any j satisfies $n_j > qn$ is at most $N v_p^n$. Thus,

$$P\left(\|A - \tilde{A}\|_F^2 \geq cN\right) \leq (2/3)^N + N v_p^n.$$

This concludes the proof.

References

1. Ackerman, M., Ben-David, S.: Which data sets are 'clusterable'?—A theoretical study of clusterability (2008)
2. Alon, N., Krivelevich, M., Sudakov, B.: Finding a large hidden clique in a random graph. Random Struct. Algorithms **13**, 457–466 (1998)
3. Candès, E. J., Recht, B.: Exact matrix completion via convex optimization. Available from http://arxiv.org/abs/0805.4471, May 2008
4. Candès, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inf. Theory **52**(2), 489–509 (2006)
5. Donoho, D.L.: Compressed Sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006)
6. Feige, U., Krauthgamer, R.: Finding and certifying a large hidden clique in a semirandom graph. Random Struct. Algorithms **16**(2), 195–208 (2000)
7. Füredi, Z., Komlós, J.: The eigenvalues of random symmetric matrices. Combinatorica **1**(3), 233–241 (1981)
8. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman, New York (1979)
9. Geman, S.: A limit theorem for the norm of random matrices. Ann. Probab. **8**(2), 252–261 (1980)
10. Gilbert, A. C., Guha, S., Indyk, P., Muthukrishnan, S., Strauss, M.: Near-optimal sparse fourier representations via sampling. In: STOC '02: Proceedings of the Thiry-Fourth Annual ACM Symposium on Theory of Computing, pp. 152–161. ACM, New York, NY, USA (2002)
11. Gillis, N., Glineur, F.: Nonnegative factorization and the maximum edge biclique problem. Available from http://arxiv.org/abs/0810.4225, (2008)
12. Håstad, J.: Clique is hard to approximate within $n^{1-\epsilon}$. In: 37th Annual Symposium on Foundations of Computer Science (Burlington, VT, 1996), pp. 627–636. IEEE Computer Society Press, Los Alamitos, CA (1996)
13. Hoeffding, W.: Probability inequalities for sums of bounded random variables. J. Am. Stat. Assoc. **58**, 13–30 (1962)
14. Iasemidis, L.D., Pardalos, P., Sackellares, J.C., Shiau, D.-S.: Quadratic binary programming and dynamical system approach to determine the predictability of epileptic seizures. J. Comb. Optim. **5**(1), 9–26 (2001)
15. McSherry, F.: Spectral partitioning of random graphs. In: Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science, pp. 529–537. IEEE Computer Society (2001)
16. Mitzenmacher, M., Upfal, E.: Probability and Computing. Cambridge University Press, Cambridge (2005)
17. Peeters, R.: The maximum edge biclique problem is NP-complete. Discrete Appl. Math. **131**, 651–654 (2003)
18. Recht, B., Fazel, M., Parrilo, P.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. Available from http://arxiv.org/abs/0706.4138 (2007)
19. Rockafellar, R.T.: Convex Analysis. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ (1997) (Reprint of the 1970 original, Princeton Paperbacks)
20. Watson, G.A.: Characterization of the subdifferential of some matrix norms. Linear Algebra Appl. **170**, 33–45 (1992)