

# The SDP value of random 2CSPs

Amulya Musipatla\*    Ryan O’Donnell\*†    Tselil Schramm‡    Xinyu Wu\*†

August 3, 2021

## Abstract

We consider a very wide class of models for sparse random Boolean 2CSPs; equivalently, degree-2 optimization problems over  $\{\pm 1\}^n$ . For each model  $\mathcal{M}$ , we identify the “high-probability value”  $s_{\mathcal{M}}^*$  of the natural SDP relaxation (equivalently, the quantum value). That is, for all  $\epsilon > 0$  we show that the SDP optimum of a random  $n$ -variable instance is (when normalized by  $n$ ) in the range  $(s_{\mathcal{M}}^* - \epsilon, s_{\mathcal{M}}^* + \epsilon)$  with high probability. Our class of models includes non-regular CSPs, and ones where the SDP relaxation value is strictly smaller than the spectral relaxation value.

## 1 Introduction

A large number of important algorithmic tasks can be construed as constraint satisfaction problems (CSPs): finding an assignment to Boolean variables to optimize the number of satisfied constraints. Almost every form of constraint optimization is NP-complete; thus one is led to questions of efficiently finding near-optimal solutions, or understanding the complexity of average-case rather than worst-case instances. Indeed, understanding the complexity of random sparse CSPs is of major importance not just in traditional algorithms theory, but also in, e.g., cryptography [JLS20], statistical physics [MM09], and learning theory [DLS14].

Suppose we fix the model  $\mathcal{M}$  for a random sparse CSP on  $n$  variables (e.g., random  $k$ -SAT with a certain clause density). Then it is widely believed that there should be a constant  $c_{\mathcal{M}}^*$  such that the optimal value of a random instance is  $c_{\mathcal{M}}^* \pm o_{n \rightarrow \infty}(1)$  with high probability (whp). (Here we define the optimal value to mean the maximum number of simultaneously satisfiable constraints, divided by the number of variables.) Unfortunately, it is extremely difficult to prove this sort of result; indeed, it was considered a major breakthrough when Bayati, Gamarnik, and Tetali [BGT13] established it for one of the simplest possible cases: Max-Cut on random  $d$ -regular graphs (which we will denote by  $\mathcal{MC}_d$ ). Actually “identifying” the value  $c_{\mathcal{M}}^*$  (beyond just proving its existence) is even more challenging. It is generally possible to estimate  $c_{\mathcal{M}}^*$  using heuristic methods from statistical physics, but making these estimates rigorous is beyond the reach of current methods. Taking again the example of Max-Cut on  $d$ -regular random graphs, it was only recently [DMS<sup>+</sup>17] that the value  $c_{\mathcal{MC}_d}^*$  was determined up to a factor of  $1 \pm o_{d \rightarrow \infty}(1)$ . The value for any particular  $d$ , e.g.  $c_{\mathcal{MC}_3}^*$ , has yet to be established.

Returning to algorithmic questions, we can ask about the *computational feasibility* of optimally solving sparse random CSPs. There are two complementary questions to ask: given a random

---

\*Computer Science Department, Carnegie Mellon University.

†Supported by NSF grant FET-1909310. This material is based upon work supported by the National Science Foundation under grant numbers listed above. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation (NSF).

‡Department of Statistics, Stanford University.

instance from model  $\mathcal{M}$  (with presumed optimal value  $c_{\mathcal{M}}^* \pm o_{n \rightarrow \infty}(1)$ ), can one efficiently *find* a solution achieving value  $\gtrsim c_{\mathcal{M}}^*$ , and can one efficiently *certify* that every solution achieves value  $\lesssim c_{\mathcal{M}}^*$ ? The former question is seemingly a bit more tractable; for example, a very recent breakthrough of Montanari [Mon21] gives an efficient algorithm for (whp) finding a cut in a random graph  $\mathcal{G}(n, p)$  graph of value at least  $(1 - \epsilon)c_{\mathcal{G}(n, p)}^*$ . On the other hand, we do not know any algorithm for efficiently certifying (whp) that a random instance has value at most  $(1 + \epsilon)c_{\mathcal{G}(n, p)}^*$ . Indeed, it is reasonable to conjecture that no such algorithm exists, leading to an example of a so-called “information-computation gap”.

To bring evidence for this we can consider *semidefinite programming* (SDP), which provides efficient algorithms for certifying an upper bound on the optimal value of a CSP [FL92]. Indeed, it is known [Rag09] that, under the Unique Games Conjecture, the basic SDP relaxation provides essentially optimal certificates for CSPs in the *worst case*. In this paper we in particular consider Boolean 2CSPs — more generally, optimizing a homogeneous degree-2 polynomial over the hypercube — as this is the setting where semidefinite programming is most natural. Again, for a fixed model  $\mathcal{M}$  of random sparse Boolean 2CSPs, one expects there should exist a constant  $s_{\mathcal{M}}^*$  such that the optimal SDP-value of an instance from  $\mathcal{M}$  is whp  $s_{\mathcal{M}}^* \pm o_{n \rightarrow \infty}(1)$ . Philosophically, since semidefinite programming is doable in polynomial time, one may be more optimistic about proving this and explicitly identifying  $s_{\mathcal{M}}^*$ . Indeed, some results in this direction have recently been established.

## 1.1 Prior work on identifying high-probability SDP values

Let us consider the most basic case:  $\mathcal{MC}_d$ , Max-Cut on random  $d$ -regular graphs. For ease of notation, we will consider the equivalent problem of maximizing  $\frac{1}{n}x^T(-\mathbf{A})x$  over  $x \in \{\pm 1\}^n$ , where  $\mathbf{A}$  is the adjacency matrix of a random  $n$ -vertex  $d$ -regular graph.<sup>1</sup> Although  $s_{\mathcal{MC}_d}^*$ , the high-probability SDP relaxation value, was pursued as early as 1987 [Bop87] (see also [DH73]), it was not until 2015 that Montanari and Sen [MS16] established the precise result  $s_{\mathcal{MC}_d}^* = 2\sqrt{d-1}$ . That is, in a random  $d$ -regular graph, whp the basic SDP relaxation value [Bop87, RW95, DP93, PR95] for the size of the maximum cut is  $(\frac{d}{4} + \sqrt{d-1} \pm o_{n \rightarrow \infty}(1))n$ . Here the special number  $2\sqrt{d-1}$  is the maximum eigenvalue of the  $d$ -regular infinite tree.

The proof of this result has two components: showing  $\text{SDP}(-\mathbf{A}) \geq 2\sqrt{d-1} - \epsilon$  whp, and showing  $\text{SDP-DUAL}(-\mathbf{A}) \leq 2\sqrt{d-1} + \epsilon$  whp. Here  $\text{SDP}(A) = \max\{\langle \rho, A \rangle : \rho \succeq 0, \rho_{ii} = \frac{1}{n} \forall i\}$  denotes the “primal” SDP value on matrix  $A$  (commonly associated with the Goemans–Williamson rounding algorithm [GW95]), and  $\text{SDP-DUAL}(A) = \min\{\lambda_{\max}(A + \text{diag}(\zeta)) : \text{avg}_i(\zeta_i) = 0\}$  denotes the (equal) “dual” SDP value on  $A$ . To show the latter bound, it is sufficient to observe that  $\text{SDP-DUAL}(-\mathbf{A}) \leq \lambda_{\max}(-\mathbf{A})$ , the “eigenvalue bound”, and  $\lambda_{\max}(-\mathbf{A}) \leq 2\sqrt{d-1} + o_n(1)$  whp by Friedman’s Theorem [Fri08]. As for lower-bounding  $\text{SDP}(-\mathbf{A})$ , Montanari and Sen used the “Gaussian wave” method [Elo09, CGHV15, HV15] to construct primal SDP solutions achieving at least  $2\sqrt{d-1} - \epsilon$  (whp). The idea here is essentially to build the SDP solutions using an approximate eigenvector (of finite support) of the infinite  $d$ -regular tree achieving eigenvalue  $2\sqrt{d-1} - \epsilon$ ; the fact that SDP constraint “ $\rho_{ii} = \frac{1}{n} \forall i$ ” can be satisfied relies heavily on the regularity of the graph.

**Remark 1.1.** The Montanari–Sen result in passing establishes that the (high-probability) eigenvalue and SDP bounds coincide for random regular graphs. This is consistent with a known theme, that the two bounds tend to be the same (or nearly so) for graphs where “every vertex looks similar” (in particular, for regular graphs). This theme dates back to Delorme and Poljak [DP93], who

<sup>1</sup>Throughout this work, **boldface** is used to denote random variables.

showed that  $\text{SDP-DUAL}(-A) = \lambda_{\max}(-A)$  whenever  $A$  is the adjacency matrix of a vertex-transitive graph.

Subsequently, the high-probability SDP value  $s_{\mathcal{M}}^*$  was established for a few other models of random regular 2CSPs. Deshpande, Montanari, O’Donnell, Schramm, and Sen [DMO<sup>+</sup>19] showed that for  $\mathcal{M} = \mathcal{NAE3}_c$  — meaning random regular instances of NAE-3SAT (not-all-equals 3Sat) with each variable participating in  $c$  clauses — we have  $s_{\mathcal{M}}^* = \frac{9}{8} - \frac{3}{8} \cdot \frac{\sqrt{c-1}-\sqrt{2}}{c}$ . We remark that NAE-3SAT is effectively a 2CSP, as the predicate  $\text{NAE}_3 : \{\pm 1\}^3 \rightarrow \{0, 1\}$  may be expressed as  $\frac{3}{4} - \frac{1}{4}(xy+yz+zx)$ , supported on the “triangle” formed by variables  $x, y, z$ . The analysis in this paper is somewhat similar to that in [MS16], but with the infinite graph  $\mathfrak{X} = K_3 \star K_3 \star \cdots \star K_3$  ( $c$  times) replacing the  $d$ -regular infinite tree. This  $\mathfrak{X}$  is the  $2c$ -regular infinite “tree of triangles” depicted (partly, in the case  $c = 3$ ) in Figure 1. More generally, [DMO<sup>+</sup>19] established the high-probability SDP value for large random (edge-signed) graphs that locally resemble  $K_r \star K_r \star \cdots \star K_r$ , the  $(r-1)c$ -regular infinite “tree of cliques  $K_r$ ”. (The  $r = 2$  case essentially generalizes [MS16].) As in [MS16],  $s_{\mathcal{M}}^*$  coincides with the (high-probability) eigenvalue bound. The upper bound on  $s_{\mathcal{M}}^*$  is shown by using Bordenave’s proof [Bor20] of Friedman’s Theorem for random  $(c, r)$ -biregular graphs. The lower bound on  $s_{\mathcal{M}}^*$  is shown using the Gaussian wave technique, relying on the distance-regularity of the graphs  $K_r \star K_r \star \cdots \star K_r$  (indeed, it is known that every infinite distance-regular graph is of this form).

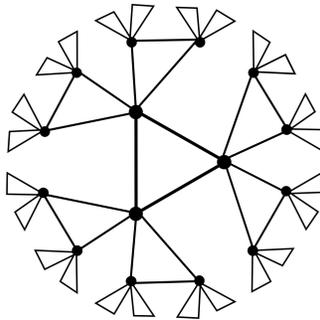


Figure 1: The 6-regular infinite graph  $K_3 \star K_3 \star K_3$ , modeling random 3-regular NAE3-SAT.

Mohanty, O’Donnell, and Paredes [MOP20] generalized the preceding two results to the case of “two-eigenvalue” 2CSPs. Roughly speaking, these are 2CSPs formed by placing copies of a small weighted “constraint graph”  $\mathcal{H}$  — required to have just two distinct eigenvalues — in a random regular fashion onto  $n$  vertices/variables. (This is indeed a generalization [DMO<sup>+</sup>19], as cliques have just two distinct eigenvalues.) As two-eigenvalue examples, [MOP20] considered CSPs with the “CHSH constraint” — and its generalizations, the “Forrelation<sub>k</sub>” constraints — which are important in quantum information theory [CHSH69, AA18]. Here the SDP value of an instance is particularly relevant as it is precisely the optimal “quantum entangled value” of the 2CSP [CHTW04]. Once again, it is shown in [MOP20] that the high-probability SDP and eigenvalue bounds coincide for these types of CSPs. The two-eigenvalue condition is used at a technical level in both the variant of Bordenave’s theorem proven for the eigenvalue upper bound, and in the Gaussian wave construction in the SDP lower bound.

Most recently, O’Donnell and Wu [OW20] used the results of Bordenave and Collins [BC19] (a substantial generalization of [Bor20]) to establish the high-probability *eigenvalue relaxation bound* for very wide range of random 2CSPs, encompassing all those previously mentioned: namely, any quadratic optimization problem defined by random “matrix polynomial lifts” over literals.

## 1.2 Our work

In this work, we establish the high-probability SDP value  $s_{\mathcal{M}}^*$  for random instances of any 2CSP model  $\mathcal{M}$  arising from lifting matrix polynomials (as in [OW20]). This generalizes all previously described work on SDP values, and covers many more cases, including random lifts of any base 2CSP and random graphs modeled on any free/additive/amalgamated product. Such graphs have seen numerous applications within theoretical computer science, for example the zig-zag product in derandomization (e.g. [RVW02]) and lifts of 2CSPs in the study of the stochastic block model (e.g [BBK<sup>+</sup>20]). See Section 2 for more details and definitions, and see [OW20] for a thorough description of the kinds of random graph/2CSP models that can arise from matrix polynomials.

Very briefly, a matrix polynomial  $p$  is a small, explicit “recipe” for producing random  $n$ -vertex edge-weighted graphs, each of which “locally resembles” an associated *infinite* graph  $\mathfrak{X}_p$ . For example,  $p_3(Y_1, Y_2, Y_3) := Y_1 + Y_2 + Y_3$  is a recipe for random (edge-signed) 3-regular  $n$ -vertex graphs, and here  $\mathfrak{X}_{p_3}$  is the infinite 3-regular tree. As another example, if  $p_{333}(Z_{1,1}, \dots, Z_{3,3})$  denotes the following matrix polynomial —

$$\begin{pmatrix} 0 & Z_{1,1}Z_{1,2}^* + Z_{2,1}Z_{2,2}^* + Z_{3,1}Z_{3,2}^* & Z_{1,1}Z_{1,3}^* + Z_{2,1}Z_{2,3}^* + Z_{3,1}Z_{3,3}^* \\ Z_{1,2}Z_{1,1}^* + Z_{2,2}Z_{2,1}^* + Z_{3,2}Z_{3,1}^* & 0 & Z_{1,2}Z_{1,3}^* + Z_{2,2}Z_{2,3}^* + Z_{3,2}Z_{3,3}^* \\ Z_{1,3}Z_{1,1}^* + Z_{2,3}Z_{2,1}^* + Z_{3,3}Z_{3,1}^* & Z_{1,3}Z_{1,2}^* + Z_{2,3}Z_{2,2}^* + Z_{3,3}Z_{3,2}^* & 0 \end{pmatrix} \quad (1)$$

— then  $p_{333}$  is a recipe for random (edge-signed) 6-regular  $n$ -vertex graphs where every vertex participates in 3 triangles. In this case,  $\mathfrak{X}_{p_{333}}$  is the infinite graph (partly) depicted in Figure 1. The Bordenave–Collins theorem [BC19] shows that if  $\mathbf{A}$  is the adjacency matrix of a random *unsigned*  $n$ -vertex graph produced from a matrix polynomial  $p$ , then whp the “nontrivial” spectrum of  $\mathbf{A}$  will be within  $\epsilon$  (in Hausdorff distance) of the spectrum of  $\mathfrak{X}_p$ . In the course of derandomizing this theorem, O’Donnell and Wu [OW20] established that for random *edge-signed* graphs, the modifier “nontrivial” should be dropped. As a consequence, in the signed case one gets  $\lambda_{\max}(\mathbf{A}) \approx \lambda_{\max}(\mathfrak{X}_p)$  up to an additive  $\epsilon$ , whp; i.e., the high-probability eigenvalue bound for CSPs derived from  $p$  is precisely  $\lambda_{\max}(\mathfrak{X}_p)$ . We remark that for simple enough  $p$  there are formulas for  $\lambda_{\max}(\mathfrak{X}_p)$ ; regarding our two example above, it is  $2\sqrt{2}$  for  $p = p_3$ , and it is 5 for  $p = p_{333}$ . In particular, if  $p$  is a linear matrix polynomial,  $\lambda_{\max}(\mathfrak{X}_p)$  may be determined numerically with the assistance of a formula of Lehner [Leh99] (see also [GVK21] for the case of standard random lifts of a fixed base graph).

In this paper we investigate the high-probability SDP value — denote it  $s_p^*$  — of a large random 2CSP (Boolean quadratic optimization problem) produced by a matrix polynomial  $p$ . Critically, our level of generality lets us consider *non-regular* random graph models, in contrast to all previous work. Because of this, see we cases in contrast to Remark 1.1, where (whp) the SDP value is strictly smaller than the eigenvalue relaxation bound. As a simple example, for random edge-signed (2, 3)-biregular graphs, the high-probability eigenvalue bound is  $\sqrt{2-1} + \sqrt{3-1} = 1 + \sqrt{2} \approx 2.414$ , but our work establishes that the high-probability SDP value is  $\sqrt{\frac{13}{4}} + 2\sqrt{2} - \frac{1}{10} \approx 2.365$ .

An essential part of our work is establishing the appropriate notion of the “SDP value” of an infinite graph  $\mathfrak{X}_p$ , with adjacency operator  $A_\infty$ . While the eigenvalue bound  $\lambda_{\max}(A_\infty)$  makes sense for the infinite-dimensional operator  $A_\infty$ , the SDP relaxation does not. The definition  $\text{SDP}(A_\infty) = \max\{\langle \rho, A_\infty \rangle : \rho \succeq 0, \rho_{ii} = \frac{1}{n} \forall i\}$  does not make sense, since “ $n$ ” is  $\infty$ . Alternatively, if one tries the normalization  $\rho_{ii} = 1$ , any such  $\rho$  will have infinite trace, and hence  $\langle \rho, A_\infty \rangle$  may be  $\infty$ . Indeed, since the only control we have on  $A_\infty$ ’s “size” will be an operator norm (“ $\infty$ -norm”) bound, the expression  $\langle \rho, A_\infty \rangle$  is only guaranteed to make sense if  $\rho$  is restricted to be trace-class (i.e., have

finite “1-norm”).

On the other hand, we know that the eigenvalue bound  $\lambda_{\max}(A_\infty)$  is too weak, intuitively because it does not properly “rebalance” graphs  $\mathfrak{X}_p$  that are not regular/vertex-transitive. The key to obtaining the correct bound is introducing a new notion, intermediate between the eigenvalue and SDP bounds, that is appropriate for graphs  $\mathfrak{X}_p$  arising from matrix polynomial recipes. Although these graphs may be irregular, their definition also allows them to be viewed as *vertex-transitive* infinite graphs with  $r \times r$  matrix edge-weights. In light of their vertex-transitivity, [Remark 1.1](#) suggests that a “maximum eigenvalue”-type quantity — suitably defined for matrix-edge-weighted graphs — might serve as the sharp replacement for SDP value. We introduce such a quantity, calling it the *partitioned SDP* bound. Let  $G$  be an  $n$ -vertex graph with  $r \times r$  matrices as edge-weights, and let  $A$  be its adjacency matrix, thought of as a Hermitian  $n \times n$  matrix whose entries are  $r \times r$  matrices. We will define

$$\text{PARTSDP}(A) = \sup\{\langle \rho, A \rangle : \rho \succeq 0, \text{tr}(\rho)_{ii} = \frac{1}{r}\}, \quad (2)$$

where here  $\text{tr}(\rho)$  refers to the  $r \times r$  matrix obtained by summing the entries on  $A$ ’s main diagonal (themselves  $r \times r$  matrices), and  $\text{tr}(\rho)_{ii}$  denotes the scalar in the  $(i, i)$ -position of  $\text{tr}(\rho)$ . This partitioned SDP bound may indeed be regarded as intermediate between the maximum eigenvalue and the SDP value. On one hand, given an scalar-edge-weighted  $n$ -vertex graph with adjacency matrix  $A$ , we may take  $r = 1$  and then it is easily seen that  $\text{PARTSDP}(A)$  coincides with  $\lambda_{\max}(A)$ . On the other hand, if we regard  $A$  as a  $1 \times 1$  matrix and take  $r = n$  (so that we have single vertex with a self-loop weighted by all of  $A$ ), then  $\text{PARTSDP}(A) = \text{SDP}(A)$ .

As we will see in [Section 3](#),  $\text{PARTSDP}(A)$  can be suitably defined even for bounded-degree infinite graphs with  $r \times r$  edge-weights. Furthermore, it has the following SDP dual:

$$\text{PARTSDP-DUAL}(A) = \inf\{\lambda_{\max}(A + \mathbb{1}_{n \times n} \otimes \text{diag}(\zeta)) : \text{avg}(\zeta_1, \dots, \zeta_r) = 0\}.$$

In the technical [Section 3.1](#), we show that there is no SDP duality gap between  $\text{PARTSDP}(A)$  and  $\text{PARTSDP-DUAL}(A)$ , even in the case of infinite graphs. It is precisely the common value of  $\text{PARTSDP}(\mathfrak{X}_p)$  and  $\text{PARTSDP-DUAL}(\mathfrak{X}_p)$  that is the high-probability SDP value of large random 2CSPs produced from  $p$ ; our main theorem is the following:

**Theorem 1.2.** *Let  $p$  be a matrix polynomial with  $r \times r$  coefficients. Let  $A_\infty$  be the adjacency operator (with  $r \times r$  entries) of the associated infinite lift  $\mathfrak{X}_p$ , and write  $s_p^* = \text{PARTSDP}(A_\infty) = \text{PARTSDP-DUAL}(A_\infty)$ . Then for any  $\epsilon, \beta > 0$  and sufficiently large  $n$ , if  $\mathbf{A}_n$  is the adjacency matrix of a random edge-signed  $n$ -lift of  $p$ , it holds that  $s_p^* - \epsilon \leq \text{SDP}(\mathbf{A}_n) \leq s_p^* + \epsilon$  except with probability at most  $\beta$ .*

Note that  $\text{PARTSDP}(A_\infty)$  is a fixed value only dependent on the polynomial  $p$ , a finitary object.

The upper bound  $\text{SDP}(\mathbf{A}_n) \leq \text{PARTSDP-DUAL}(A_\infty) + \epsilon$  in this theorem can be derived from the results of [\[BC19, OW20\]](#). Our main work is to prove the lower bound  $\text{SDP}(\mathbf{A}_n) \geq \text{PARTSDP}(A_\infty) - \epsilon$ . For this, our approach is inspired by the Gaussian Wave construction of [\[MS16, DMO<sup>+</sup>19\]](#) for  $d$ -regular graphs (in the random lifts model), which can be viewed as constructing a feasible  $\text{SDP}(\mathbf{A}_n)$  solution of value  $\lambda_{\max}(A_\infty) - \epsilon$  using a truncated eigenfunction of  $A_\infty$ . Since local neighborhoods in  $A_\infty$  look like local neighborhoods in  $\mathbf{A}_n$  with high probability, the eigenfunction can be “pasted” almost everywhere into the graph  $\mathbf{G}_n$ , which gives an SDP solution of value near  $\lambda_{\max}(A_\infty)$ .

This approach runs into clear obstacles in our setting. Indeed, the raw eigenfunctions are of no use to us, as *the SDP value may be smaller than the spectral relaxation value*. Instead, we first show that there is a  $\rho_0$  with only finitely many nonzero entries that achieves the sup in [Equation \(2\)](#)

up to  $\epsilon$ . This is effectively a finite  $r \times r$  matrix-edge-weighted graph. We then show that this  $\rho_0$  can (just as in the regular case) whp be “pasted” almost everywhere into the graph  $\mathbf{G}_n$  defined by  $\mathbf{A}_n$ , which gives an SDP solution of value close to  $\text{PARTSDP}(A_\infty)$ . The fact that  $\mathfrak{X}_p$  and  $\mathbf{G}_n$  are regarded as regular tree-like graphs with matrix edge-weights (rather than as irregular graphs with scalar edges-weights) is crucially used to show that the “pasted solution” satisfies the finite SDP’s constraints “ $\rho_{ii} = \frac{1}{n} \forall i$ ”.

## 2 Preliminaries

To preface the following definitions and concepts, we remark that our “real” interest is in graphs (or 2CSPs) with real scalar weights. The introduction of matrix edge-weights facilitates two things: helping us define a wide variety of interesting scalar-weighted graphs via matrix polynomial lifts; and, facilitating the definition of  $\text{PARTSDP}(\cdot)$ , which we use to bound the SDP relaxation value of the associated 2CSPs. Our use of potentially complex matrices is also not really essential; we kept them because prior work that we utilize ([BC19], the tools in [Section 3](#)) is stated in terms of complex matrices. However the reader will not lose anything by assuming that all Hermitian matrices are in fact symmetric real matrices.

### 2.1 Matrix-weighted graphs

In the following definitions, we’ll restrict attention to graphs with at-most-countable vertex sets  $V$  and bounded degree. We also often use bra-ket notation, with  $(|v\rangle)_{v \in V}$  denoting the standard orthonormal basis for the complex vector space  $\ell_2(V)$ .

**Definition 2.1** (Matrix-weighted graph). Fix any  $r \in \mathbb{N}^+$ . A *matrix-weighted graph* will refer to a directed simple graph  $G = (V, E)$  with self-loops allowed, in which each directed edge  $e$  has an associated *weight*  $a_e \in \mathbb{C}^{r \times r}$ . If  $(v, w) \in E \implies (w, v) \in E$  and  $a_{(w,v)} = a_{(v,w)}^*$ , we say that  $G$  is an *undirected* matrix-weighted graph. The *adjacency matrix* of  $G$  is the operator  $A$ , acting on  $\ell_2(V) \otimes \mathbb{C}^r$ , given by

$$\sum_{(v,w) \in E} |w\rangle\langle v| \otimes a_{(v,w)}.$$

It can be helpful to think to think of  $A$  in matrix form, as a  $|V| \times |V|$  matrix whose entries are themselves  $r \times r$  edge-weight matrices. Note that if  $G$  is undirected if and only if  $A$  is self-adjoint,  $A = A^*$ .

**Definition 2.2** (Extension of a matrix-weighted adjacency matrix). Given a  $|V| \times |V|$  matrix  $A$  with  $r \times r$  entries, we may also view it as a  $|V|r \times |V|r$  matrix with scalar entries. When we wish to explicitly call attention to the distinction, we will call the latter matrix the *extension* of  $A$ , and denote it by  $\tilde{A}$ .

### 2.2 Matrix polynomials

**Definition 2.3** (Matrix polynomial). Let  $Y_1, \dots, Y_d$  be formal indeterminates that are their own inverses, and let  $Z_1, \dots, Z_e$  be formal indeterminates with formal inverses  $Z_1^*, \dots, Z_e^*$ . For a fixed  $r$ , we define a *matrix polynomial* to be a formal noncommutative polynomial  $p$  over the indeterminates  $Y_1, \dots, Z_e^*$ , with coefficients in  $\mathbb{C}^{r \times r}$ . In particular, we may write

$$p = \sum_w a_w w,$$

where the sum is over words  $w$  on the alphabet of indeterminates, each  $a_w$  is in  $\mathbb{C}^{r \times r}$ , and only finitely many  $a_w$  are nonzero. Here we call a word *reduced* if it has no adjacent  $Y_i Y_i$  or  $Z_i Z_i^*$  pairs. We will denote the empty word by  $\mathbf{1}$ .

As we will shortly describe, we will always be considering substituting unitary operators for the  $Z_i$ 's, and unitary involution operators for the  $Y_i$ 's. Thus we can think of  $Z_i^*$  as both the inverse and the “adjoint” of indeterminate  $Z_i$ , and similarly we think of  $Y_i^* = Y_i$ .

**Definition 2.4** (Adjoint of a polynomial). Given a matrix polynomial  $p = \sum_w a_w w$  as above, we define its *adjoint* to be

$$p^* = \sum_w a_w^* w^*,$$

where  $a^*$  is the usual adjoint of  $a \in \mathbb{C}^{r \times r}$ , and  $w^*$  is the adjointed reverse of  $w$ . That is, if  $w = w_1 \cdots w_k$  then  $w^* = w_k^* \cdots w_1^*$ , where  $\mathbf{1}^* = \mathbf{1}$ ,  $Y_i^* = Y_i$ , and  $Z_i^{**} = Z_i$ . We say  $p$  is *self-adjoint* if  $p^* = p$  formally.

Note that in any self-adjoint polynomial, some terms will be self-adjoint, and others will come in self-adjoint pairs. In this work, we will only be considering self-adjoint polynomials.

### 2.3 Lifts of matrix polynomials

**Definition 2.5** ( $n$ -lift). Given a matrix polynomial over the indeterminates  $Y_1, \dots, Z_e^*$ , we define an  $n$ -lift to be a sequence  $\mathcal{L} = (M_1, \dots, M_d, P_1, \dots, P_e)$  of  $n \times n$  matrices, where each  $P_i$  is a signed permutation matrix and each  $M_i$  is a signed matching matrix.<sup>2</sup> A *random  $n$ -lift*  $\mathcal{L} = (M_1, \dots, M_d, P_1, \dots, P_e)$  is one where the matrices are chosen independently and uniformly at random.

**Definition 2.6** (Evaluation/substitution of lifts.). Given an  $n$ -lift  $\mathcal{L}$  and a word  $w$ , we define  $\mathcal{L}^w$  to be the  $n \times n$  operator obtained by substituting appropriately into  $w$ : namely,  $Y_i = M_i$ ,  $Z_i = P_i$ , and  $Z_i^* = P_i^*$  for each  $i$  (and substituting the empty word with the  $n \times n$  identity operator). Given also a matrix polynomial  $p = \sum_w a_w w$ , we define the *evaluation of  $p$  at  $\mathcal{L}$*  to be the following operator on  $\mathbb{C}^n \otimes \mathbb{C}^r$ :<sup>3</sup>

$$p(\mathcal{L}) = \sum_w \mathcal{L}^w \otimes a_w.$$

**Remark 2.7.** Note that each  $P_i$  is unitary and each  $M_i$  a unitary involution (as promised), so  $p^*(\mathcal{L}) = p(\mathcal{L})^*$ . Thus  $p(\mathcal{L})$  is a self-adjoint operator whenever  $p^*$  is a self-adjoint polynomial. In this case we also have that  $p(\mathcal{L})$  may be viewed as the adjacency matrix of an undirected graph on vertex set  $[n]$  with  $r \times r$  edge-weights.

Note that the evaluation  $p(\mathcal{L})$  of a matrix polynomial may be viewed as the adjacency matrix of an undirected graph on  $[n]$  with  $r \times r$  edge-weights; or, its extension may be viewed as the adjacency matrix of an undirected graph on  $[n] \times [r]$  with scalar edge-weights. In this way, each fixed matrix polynomial  $p$ , when applied to a random lift, gives rise to a random (undirected, scalar-weighted) graph model.

<sup>2</sup>A signed matching matrix is the adjacency matrix of a perfect matching with  $\pm 1$  edge-signs. If  $d > 0$  then we must restrict to even  $n$ .

<sup>3</sup>Note that coefficients  $a_w$  are written on the left in  $p$ , as is conventional, but we take the tensor/Kronecker product on the right so that the matrix form of  $p(\mathcal{L})$  may be more naturally regarded as an  $n \times n$  matrix with  $r \times r$  entries.

**Example 2.8.** A simple example is the matrix polynomial

$$p(Y_1, Y_2, Y_3) = Y_1 + Y_2 + Y_3.$$

Here  $r = 1$  and each coefficient is just the scalar 1. This  $p$  gives rise to a model of random edge-signed 3-regular graphs on  $[n]$ .

By moving to actual matrix coefficients with  $r > 1$ , one can get the random (signed) graph model given by randomly  $n$ -lifting any base  $r$ -vertex graph  $H$ .

**Example 2.9.** As a simple example,

$$p(Z_1, Z_2, Z_3) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} Z_1 + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} Z_1^* + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} Z_2 + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} Z_2^* + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} Z_3 + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} Z_3^*$$

is the recipe for random 3-regular (edge-signed)  $(n + n)$ -vertex bipartite graphs. The reader may like to view this as a  $2 \times 2$  matrix of polynomials,

$$p(Z_1, Z_2, Z_3) = \begin{pmatrix} 0 & Z_1 + Z_2 + Z_3 \\ Z_1^* + Z_2^* + Z_3^* & 0 \end{pmatrix},$$

but recall that we actually Kronecker-product the coefficient matrices “on the other side”. So rather than as a  $2 \times 2$  block-matrix with  $n \times n$  blocks, we think of the resulting adjacency matrix as an  $n \times n$  block-matrix with  $2 \times 2$  blocks; equivalently, an  $n$ -vertex graph with  $2 \times 2$  matrix edge-weights.

**Example 2.10.** The matrix polynomial  $p_{333}$  mentioned in (1) gives an example of a *nonlinear* polynomial with matrix coefficients. Again, we wrote it there as a  $3 \times 3$  matrix of polynomials for compactness, but for analysis purposes we will view it as a degree-2 polynomial with  $3 \times 3$  coefficients.

**Definition 2.11** ( $\infty$ -lift). Formally, we extend [Definition 2.5](#) to the case of  $n = \infty$  as follows. Let  $V_\infty$  denote the free product of groups  $\mathbb{Z}_2^{*d} \star \mathbb{Z}^{*e}$ , with its components generated by  $g_1, \dots, g_d, h_1, h_1^{-1}, \dots, h_e, h_e^{-1}$ . Thus the elements of  $V_\infty$  are in one-to-one correspondence with the reduced words over indeterminates  $Y_1, \dots, Z_e^*$ . The generators  $g_1, \dots, g_d, h_1, \dots, h_e$  act as permutations on  $V_\infty$  by left-multiplication, with the first  $d$  in fact being matchings. We write  $\sigma_1, \dots, \sigma_{d+e}$  for these permutations, and we also identify them with their associated permutation operators on  $\ell_2(V_\infty)$ . Finally, we write  $\mathcal{L}_\infty = (\sigma_0, \dots, \sigma_{d+2e})$  for “the”  $\infty$ -lift associated to  $p$ . (Note that this lift is “unsigned”.)

**Definition 2.12** (Evaluation at the  $\infty$ -lift, and  $\mathfrak{X}_p$ ). The evaluation of a matrix polynomial  $p$  at the infinity lift  $\mathcal{L}_\infty$  is now defined just as in [Definition 2.6](#); the resulting operator  $p(\mathcal{L}_\infty)$  operates on  $\ell_2(V_\infty) \otimes \mathbb{C}^r$ . We may think of the result as a matrix-weighted graph on vertex set  $V_\infty$ , and we will sometimes denote this graph by  $\mathfrak{X}_p$ . When  $p$  is understood, we often write  $A_\infty = p(\mathcal{L}_\infty)$  for the adjacency operator of  $\mathfrak{X}_p$ , which can be thought of as an infinite matrix with rows/columns indexed by  $V_\infty$  and entries from  $\mathbb{C}^{r \times r}$ , or as its “extension”  $\tilde{A}_\infty$ , an infinite matrix with rows/columns indexed by  $V_\infty \times [r]$  and scalar entries.

**Example 2.13.** For the polynomial  $p = Y_1 + \dots + Y_d$ , the corresponding graph  $\mathfrak{X}_p$  is the infinite (unweighted)  $d$ -regular tree.

We may now state a theorem which is essentially the main result (“Theorem 2”) of [BC19]. The small difference is that our notion of random  $n$ -lifts, which includes  $\pm 1$  signs on the matchings/permutations, lets us eliminate mention of “trivial” eigenvalues (see [OW20, Thms. 1.9, 10.10]).

**Theorem 2.14.** *Let  $p$  be a self-adjoint matrix polynomial with coefficients from  $\mathbb{C}^{r \times r}$  on indeterminates  $Y_1, \dots, Z_e^*$ . Then for all  $\epsilon, \beta > 0$  and sufficiently large  $n$ , the following holds:*

*Let  $\mathbf{A}_n = p(\mathcal{L}_n)$ , where  $\mathcal{L}_n$  is a random  $n$ -lift, and let  $A_\infty = p(\mathcal{L}_\infty)$ . Then except with probability at most  $\beta$ , the spectra  $\text{spec}(\mathbf{A}_n)$  and  $\text{spec}(A_\infty)$  are at Hausdorff distance at most  $\epsilon$*

## 2.4 Random lifts as optimization problems

Given a Hermitian (i.e., self-adjoint) matrix  $A \in \mathbb{C}^{n \times n}$ , we are interested in the task of maximizing  $x^\top A x$  over all Boolean vectors  $x \in \{\pm 1\}^n$ . (Since  $A$  is Hermitian, the quantity  $x^\top A x$  is always real, so this maximization problem makes sense.) This is the same as maximizing the homogeneous degree-2 (commutative) polynomial  $\sum_{i,j} A_{ij} x_i x_j$  over  $x \in \{\pm 1\}^n$ , and it is also essentially the same task as the Max-Cut problem on (scalar-)weighted undirected graphs. More precisely, if  $G$  is a weighted graph on vertex set  $[n]$  with adjacency matrix  $A$ , then  $G$ 's maximum cut is indicated by the  $x \in \{\pm 1\}^n$  that maximizes  $x^\top (-A)x$ . For the sake of scaling we will also include a factor of  $\frac{1}{n}$  in this optimization problem, leading to the following definition:

**Definition 2.15** (Optimal value). Given a Hermitian matrix  $A \in \mathbb{C}^{n \times n}$ , we define

$$\text{OPT}(A) = \sup_{x \in \{\pm 1\}^n} \left\{ \frac{1}{n} x^\top A x \right\} = \sup_{x \in \left\{ \pm \frac{1}{\sqrt{n}} \right\}^n} \left\{ x^\top A x \right\}.$$

(For finite-dimensional  $A$ , the sups and infs mentioned in this section are all achieved.)

We remark that

$$x^\top A x = \text{tr}(x^\top A x) = \text{tr}(x x^\top A) = \langle x x^\top, A \rangle,$$

where we use the notation  $\langle B, C \rangle = \text{tr}(BC)$ . Thus we also have

$$\text{OPT}(A) = \sup_{\rho \in \text{Cut}_n} \left\{ \frac{1}{n} \langle \rho, A \rangle \right\},$$

where  $\text{Cut}_n$  is the “cut polytope”, the convex hull of all matrices of the form  $x x^\top$  for  $x \in \{\pm 1\}^n$ . (Since  $\langle \rho, A \rangle$  is linear in  $\rho$ , maximizing over the convex hull is the same as maximizing over the extreme points, which are just those matrices of the form  $x x^\top$ .)

The above optimization problem has a natural relaxation: maximizing  $\frac{1}{n} x^\top A x$  over all *unit* vectors  $x$ . This leads to the following efficiently computable upper bound on  $\text{OPT}(A)$ :

**Definition 2.16** (Eigenvalue bound). Given a Hermitian matrix  $A \in \mathbb{C}^{n \times n}$ , we define the *eigenvalue bound* to be

$$\text{EIG}(A) = \sup \{ \langle \rho, A \rangle : \rho \succeq 0, \text{tr}(\rho) = 1 \},$$

where here  $\rho \succeq 0$  denotes that  $\rho$  is (Hermitian and) positive semidefinite.

The matrices  $\rho$  being optimized over in  $\text{EIG}(A)$  are known as *density matrices*; i.e.,  $\text{EIG}(A)$  is the maximal inner product between  $A$  and any density matrix. Note that if  $\varrho \in \text{Cut}_n$ , then  $\rho = \frac{1}{n} \varrho$  is a density matrix. Thus,  $\text{EIG}(A)$  is a relaxation of  $\text{OPT}(A)$ , or in other words,  $\text{OPT}(A) \leq \text{EIG}(A)$ .

The set of density matrices is convex, and it's well known that its extreme points are all the rank-1 density matrices; i.e., those  $\rho$  of the form  $xx^\top$  for  $x \in \mathbb{C}^n$  with  $\|x\|_2^2 = 1$ . Thus in  $\text{EIG}(A)$  it is equivalent to just maximize over these extreme points:

$$\text{EIG}(A) = \sup_{\substack{x \in \mathbb{C}^n \\ \|x\|_2^2=1}} \left\{ \langle xx^\top, A \rangle \right\} = \sup_{\substack{x \in \mathbb{C}^n \\ \|x\|_2^2=1}} \left\{ x^\top A x \right\}.$$

From this formula we see that  $\text{EIG}(A)$  is also equal to  $\lambda_{\max}(A)$ , the maximum eigenvalue of  $A$ ; hence the terminology ‘‘eigenvalue bound’’. One may also think of  $\text{EIG}(A)$  and  $\lambda_{\max}(A)$  as SDP duals of one another.

We now mention another well known, tighter, upper bound on  $\text{OPT}(A)$ .

**Definition 2.17** (Basic SDP bound). Given a Hermitian matrix  $A \in \mathbb{C}^{n \times n}$ , the *basic SDP bound* is defined to be

$$\text{SDP}(A) = \sup \{ \langle \rho, A \rangle : \rho \succeq 0, \rho_{ii} = \frac{1}{n}, \forall i \}.$$

Recall that an  $n \times n$  matrix  $\rho$  is a *correlation matrix* [Sty73] if it is PSD and has all diagonal entries equal to 1. Thus  $\text{SDP}(A)$  is equivalently maximizing  $\frac{1}{n} \langle \rho, A \rangle$  over all correlation matrices  $\rho$ . We also note that any cut matrix is a correlation matrix, and any correlation matrix is a density matrix, hence so

$$\text{OPT}(A) \leq \text{SDP}(A) \leq \text{EIG}(A).$$

**Definition 2.18** (Dual SDP bound). The semidefinite dual of  $\text{SDP}(A)$  is the following [DP93]:

$$\text{SDP-DUAL}(A) = \inf_{\substack{\zeta \in \mathbb{R}^n \\ \text{avg}(\zeta_1, \dots, \zeta_n) = 0}} \{ \lambda_{\max}(A + \text{diag}(\zeta)) \}.$$

Despite the fact that the usual ‘‘Slater condition’’ for strong SDP duality fails in this case (because the set of correlation matrices isn't full-dimensional), one can still show [PR95] that  $\text{SDP}(A) = \text{SDP-DUAL}(A)$  indeed holds for finite-dimensional  $A$ .

**Remark 2.19.** In this work we frequently consider matrix-weighted graphs with adjacency matrices  $A$ , thought of as  $n \times n$  matrices with entries from  $\mathbb{C}^{r \times r}$ . For such matrices, whenever we write  $\text{OPT}(A)$ , we mean  $\text{OPT}(\tilde{A})$  for the  $nr \times nr$  ‘‘extension’’ matrix  $\tilde{A}$  (see Definition 2.2), and similarly for  $\text{EIG}(A)$ ,  $\lambda_{\max}(A)$ ,  $\text{SDP}(A)$ ,  $\text{SDP-DUAL}(A)$ .

As mentioned in Section 1, the eigenvalue bound  $\lambda_{\max}(A)$  makes sense when  $A$  is the adjacency matrix of an infinite graph (with bounded degree). However  $\text{SDP}(A)$  does not extend to the infinite case, as the number ‘‘ $n$ ’’ appearing in its definition is not finite. On the other hand, we now introduce a new, intermediate, ‘‘maximum eigenvalue-like’’ bound that is appropriate for matrix-weighted graphs. This is the ‘‘partitioned SDP bound’’ appearing in the statement of our main Theorem 1.2. In the following Section 3, we will show that it generalizes well to the case of infinite graphs.

**Definition 2.20** (Partitioned SDP bound). Let  $A$  be an  $n \times n$  Hermitian matrix with entries from  $\mathbb{C}^{r \times r}$ . We define its *partitioned SDP bound* to be

$$\text{PARTSDP}(A) = \sup \{ \langle \rho, A \rangle : \rho \succeq 0, \text{tr}(\rho) \in \frac{1}{r} \text{Corr}_r \},$$

where:

- the matrices  $\rho$  are also thought of as  $n \times n$  matrices with entries from  $\mathbb{C}^{r \times r}$ ;

- $\langle \rho, A \rangle$  is interpreted as  $\langle \tilde{\rho}, \tilde{A} \rangle$ ;
- $\text{tr}(\rho)$  denotes the sum of the diagonal entries of  $\rho$ , which is an  $r \times r$  matrix;
- $\text{Corr}_r$  is the set of  $r \times r$  correlation matrices;
- in other words, the final condition is that  $\text{tr}(\rho)_{ii} = \frac{1}{r}$  for all  $i \in [r]$ .

**Remark 2.21.** As mentioned, the partitioned SDP bound can be viewed as “intermediate” between the eigenvalue bound and the SDP bound. To explain this, suppose  $A$  is an  $n \times n$  Hermitian matrix. On one hand, we can regard  $A$  as an  $n \times n$  matrix with  $1 \times 1$  matrix entries ( $r = 1$ ); in this viewpoint,  $\text{PARTSDP}(A) = \text{EIG}(A)$ . On the other hand, we can regard  $A$  as a  $1 \times 1$  matrix with a single  $n \times n$  matrix entry ( $r = n$ ); in this viewpoint,  $\text{PARTSDP}(A) = \text{SDP}(A)$ .

It is easy to see that the partitioned SDP bound indeed has an SDP formulation, and we now state its SDP dual:

**Definition 2.22.** The SDP dual of  $\text{PARTSDP}(A)$  is the following:

$$\text{PARTSDP-DUAL}(A) = \inf_{\substack{\zeta \in \mathbb{R}^r \\ \text{avg}(\zeta_1, \dots, \zeta_r) = 0}} \{ \lambda_{\max}(A + \mathbb{1}_{n \times n} \otimes \text{diag}(\zeta)) \}.$$

Weak SDP duality,  $\text{PARTSDP}(A) \leq \text{PARTSDP-DUAL}(A)$ , holds as always, but again it is not obvious that strong SDP duality holds. In fact, not only does strong duality hold, it even holds in the case of *infinite* matrices  $A$ . This fact is crucial for our work, and proving it the subject of the upcoming technical section.

### 3 The infinite SDPs

This technical section has two goals. First, in [Section 3.1](#) we show that strong duality holds with  $\text{PARTSDP}(A) = \text{PARTSDP-DUAL}(A)$ , even for infinite matrices  $A$  with  $r \times r$  entries. Even in the finite case this is not trivial, as the feasible region for the SDP  $\text{PARTSDP}(A)$  is not full-dimensional, and hence the Slater condition ensuring strong duality does not apply. The infinite case involves some additional technical considerations, needed so that we may eventually apply the strong duality theorem for conic linear programming of Bonnans and Shapiro [[BS00](#), Thm. 2.187]. Second, in [Section 3.2](#), we show that in the optimization problem  $\text{PARTSDP}(A)$ , values arbitrarily close to the optimum can be achieved by matrices  $\rho$  of finite support (i.e., with only finitely many nonzero entries). Indeed (though we don’t need this fact), these finite-rank  $\rho$  need only have rank at most  $r$ . This fact is familiar from the case of  $r = 1$ , where the optimizer in the eigenvalue bound [Definition 2.16](#) is achieved by a  $\rho$  of rank 1 (namely  $|\psi\rangle\langle\psi|$  for any maximum eigenvector  $|\psi\rangle$ ). Finally, in [Section 3.3](#) we consolidate all these results into a theorem statement suitable for use with graphs produced by infinite lifts of matrix polynomials.

#### 3.1 SDP duality

Let  $V$  be a countable set and write  $\mathcal{H} = \ell_2(V)$  for the associated (complex, separable) Hilbert space of square-summable functions  $f : V \rightarrow \mathbb{C}$ . We write  $B_{00}(\mathcal{H})$ ,  $B_1(\mathcal{H})$ ,  $B(\mathcal{H})$  for the spaces of finite-rank, trace-class, and bounded operators on  $\mathcal{H}$ , respectively. Focusing on  $B_1(\mathcal{H})$  (with the weak topology) and  $B(\mathcal{H})$  (with the  $\sigma$ -weak topology), these are locally convex topological vector spaces forming a dual pair with bilinear map  $\langle \cdot, \cdot \rangle : B_1(\mathcal{H}) \times B(\mathcal{H}) \rightarrow \mathbb{C}$  defined by  $\langle \rho, a \rangle = \text{tr}(\rho a)$

(see [RS80, Thm. VI.26], [Lan17, Thm. B.146], or [BR02, Prop. 2.4.3]). Recall [Lan17, Lem. B.142] that  $\langle \rho, a \rangle \leq \|\rho\|_1 \|a\|$ , where  $\|\rho\|_1 = \text{tr} \sqrt{\rho^* \rho}$  is the trace-norm.

Write  $B_1(\mathcal{H})_{\text{sa}}$  (respectively,  $B(\mathcal{H})_{\text{sa}}$ ) for the (closed) real subspace of self-adjoint operators in  $B_1(\mathcal{H})$  (respectively,  $B(\mathcal{H})$ ); note that  $B_1(\mathcal{H})_{\text{sa}}$  and  $B(\mathcal{H})_{\text{sa}}$  continue to form a dual pair (see, e.g., [Mey06, p. 212]). Recall that  $a \in B(\mathcal{H})_{\text{sa}}$  is positive semidefinite if and only if  $\langle \varphi | a | \varphi \rangle \geq 0$  for all  $|\varphi\rangle \in \mathcal{H}$  (for such  $a$  we have  $\sqrt{a^* a} = a$ ); as usual we write  $b \succeq a$  to mean that  $b - a$  is positive semidefinite. Letting  $B_1(\mathcal{H})_+$  (respectively,  $B(\mathcal{H})_+$ ) denote the positive semidefinite operators in  $B_1(\mathcal{H})_{\text{sa}}$  (respectively,  $B(\mathcal{H})_{\text{sa}}$ ), we have that these are both (nonempty) closed, convex cones (for  $B(\mathcal{H})_+$ , see [Con90, Prop. 3.7], [Lan17, Prop. C.51]; for  $B_1(\mathcal{H})_+$  see [Fri19, Sec. 4], [vDE20, App. A.2], [SH08, Sec. 2]); further, they are topologically dual cones [vDE20, App. A.2].

**Our SDP.** We now introduce an SDP that is equivalent to our partitioned SDP SDP; however, we will express it with scalar entries (rather than matrix entries). Fix any  $a \in B(\mathcal{H})_{\text{sa}}$ . Let  $V = V_1 \sqcup V_2 \sqcup \cdots \sqcup V_r$  be a partition of  $V$  into  $r \in \mathbb{N}_+$  nonempty parts, and for  $1 \leq j \leq r$  let  $I_j \in B(\mathcal{H})_{\text{sa}}$  be the operator  $\sum_{v \in V_j} |v\rangle\langle v|$ . Consider the following semidefinite program:

$$\sup_{\rho \in B_1(\mathcal{H})_+} \langle \rho, a \rangle \quad \text{subject to} \quad \langle \rho, I_j \rangle = \frac{1}{r}, \quad j = 1 \dots r. \quad (\text{SDP-P})$$

To relate this back to our definition of  $\text{PARTSDP}(A)$  (for infinite matrices), suppose that  $A$  is a self-adjoint matrix, indexed by countable vertex  $V_\infty$ , with  $r \times r$  entries (and only finitely many nonzero entries per row/column). Let  $a = \tilde{A}$  be its “extension”, with rows/columns indexed by  $V = V_\infty \times [r]$ , and let  $V_j = V_\infty \times \{j\}$ . Then (SDP-P) is precisely how we define  $\text{PARTSDP}(A)$ .

We remark that (SDP-P) is always feasible. By summing the constraints on  $\rho$  we get  $\text{tr}(\rho) = 1$ ; i.e.,  $\rho$  is required to be a density operator. In particular,  $\langle \rho, a \rangle \leq \text{tr}(\rho) \|a\|$  always so the optimum value of (SDP-P) is finite.

We would like to show that the optimum value of (SDP-P) is equal to that of the following dual semidefinite program which, in the setup mentioned above, is equivalent to our definition of  $\text{PARTSDP-DUAL}(A)$ :

$$\inf_{\zeta \in \mathbb{R}^r} \lambda_{\max}(a + \zeta_1 I_1 + \cdots + \zeta_r I_r) \quad \text{subject to} \quad \text{avg}(\zeta_1, \dots, \zeta_r) = 0. \quad (\text{SDP-D})$$

Showing this will take a couple of steps. The first is to mechanically write down the Lagrangian dual of (SDP-P), which is:

$$\inf_{y \in \mathbb{R}^r} \frac{1}{r} \sum_j y_j \quad \text{subject to} \quad a \preceq y_1 I_1 + \cdots + y_r I_r. \quad (\text{D1})$$

We first show that (D1) is equivalent to (SDP-D). We may reparameterize all  $y \in \mathbb{R}^r$  by defining  $\bar{y} = \frac{1}{r} \sum_j y_j$  and writing  $y_j = \bar{y} - \frac{1}{r}$ ; in this way, every  $y \in \mathbb{R}^r$  corresponds to a pair  $\bar{y} \in \mathbb{R}$  and  $\zeta \in \mathbb{R}^r$  satisfying  $\text{avg}(\zeta_1, \dots, \zeta_r) = 0$ , and vice versa. Under this reparameterization, (D1) becomes

$$\inf_{\substack{\bar{y} \in \mathbb{R} \\ \zeta \in \mathbb{R}^r}} \bar{y} \quad \text{subject to} \quad \text{avg}(\zeta_1, \dots, \zeta_r) = 0 \quad \text{and} \quad a \preceq \bar{y} I - \sum_j \zeta_j I_j. \quad (\text{D2})$$

The second constraint here is equivalent to  $\lambda_{\max}(a + \sum_j \zeta_j I_j) \leq \bar{y}$ , and hence the optimal choice of  $\bar{y}$  given  $\zeta$  is achieved by  $\lambda_{\max}(a + \sum_j \zeta_j I_j)$ . Thus (D1) is equivalent to (D2) is equivalent to (SDP-D), as claimed.

We would next like to claim that strong duality holds for the dual pair (SDP-P) and (D1); however, there is a difficulty because the feasible region of (SDP-P) only has nonempty *relative*

interior, not interior. Alternatively, one might say the difficulty is that the feasible region for (D1) is not bounded. However we can fix this by introducing the following equivalent bounded variant:

$$\inf_{y \in \mathbb{R}^r} \frac{1}{r} \sum_j y_j \quad \text{subject to} \quad a \preceq y_1 I_1 + \cdots + y_r I_r \text{ and } |y_j| \leq C, \quad j = 1 \dots r, \quad (\text{D3})$$

where  $C = 2r\|a\|$ .

**Claim 3.1.** (D3) is equivalent to (D1).

*Proof.* Suppose we have any feasible solution  $y$  for (D1). First observe that for any  $1 \leq j \leq r$ , if we take an arbitrary  $v_j \in V_j$  we have

$$y_1 I_1 + \cdots + y_r I_r \succeq a \implies y_j \geq \langle v_j | a | v_j \rangle \geq -\|a\|. \quad (3)$$

Next observe that the optimum value of (D1) is at most  $\|a\|$  (since we could take  $y_j = \|a\|$  for all  $j$ ); thus we may restrict attention to  $y$ 's that achieve objective value at most  $\|a\|$ . But now we claim any such feasible  $y$  will have  $y_k \leq 2r\|a\|$  for any  $1 \leq k \leq r$ . To see this, observe (using [Inequality \(3\)](#)) that

$$\frac{1}{r} \sum_j y_i \geq \frac{1}{r} (y_k - \sum_{j \neq k} \|a\|) \geq y_k/r - \|a\|.$$

Thus if  $y_k > 2r\|a\|$ , it is strictly not optimal. We conclude that we may add the constraints  $-\|a\| \leq y_j \leq 2r\|a\|$  for  $j = 1 \dots r$  to (D1) without changing it. Clearly it doesn't hurt to widen this interval to  $-C \leq y_j \leq C$  (for the sake of symmetry), and so we conclude that (D3) is indeed equivalent to (D1).  $\square$

We can now write the Lagrangian dual of (D3), which is

$$\sup_{\rho \in B_1(\mathcal{H})_+} \langle \rho, a \rangle - C \sum_{j=1}^r |\langle \rho, I_j \rangle - \frac{1}{r}|. \quad (\text{P3})$$

**Claim 3.2.** (P3) is equivalent to (SDP-P).

*Proof.* Clearly the objective value of (SDP-P) is at most that of (P3), so it suffices to show the converse inequality. To that end, let  $\rho$  be any feasible solution for (P3), and write the objective value as  $o(\rho) - p(\rho)$ , where  $o(\rho) = \langle \rho, a \rangle$  ("objective") and  $p(\rho) = C \sum_j |\langle \rho, I_j \rangle - \frac{1}{r}|$  ("penalty").

Let  $\epsilon = (\max_j \{r\langle \rho, I_j \rangle\} - 1)_+$ , so that  $\langle \rho, I_j \rangle \leq \frac{1}{r}(1 + \epsilon)$  for all  $j$ , and write  $\tilde{\rho} = \rho/(1 + \epsilon) \in B_1(\mathcal{H})_+$ . Note that  $p(\tilde{\rho}) = C \sum_j \delta_j$ , where  $\delta_j = \frac{1}{r} - \langle \tilde{\rho}, I_j \rangle \geq 0$ . Finally, define  $\rho' = \tilde{\rho} + \sum_j \delta_j |v_j\rangle\langle v_j| \in B_1(\mathcal{H})_+$ , where each  $v_j$  is an arbitrary element of  $V_j$ . By construction,  $p(\rho') = 0$ . We now want to show

$$o(\rho') \geq o(\rho) - p(\rho) \iff p(\rho) \geq o(\rho) - o(\rho'). \quad (4)$$

We will then have that  $\rho'$  is feasible for (SDP-P), with an objective value at least that of  $\rho$ 's in (P3); this will complete the proof.

First, we have

$$p(\rho) = C \sum_j |\langle \rho, I_j \rangle - \frac{1}{r}| = C \sum_j |(1 + \epsilon)\langle \tilde{\rho}, I_j \rangle - \frac{1}{r}| = 2r\|a\| \sum_j |\epsilon/r - (1 + \epsilon)\delta_j|.$$

On the other hand, we have

$$o(\rho) - o(\rho') = \langle \rho, a \rangle - \frac{1}{1+\epsilon} \langle \rho, a \rangle - \delta_j \sum_j \langle v_j | a | v_j \rangle \leq \frac{\epsilon}{1+\epsilon} \langle \rho, a \rangle + \|a\| \sum_j \delta_j \leq \|a\| \left( \frac{\epsilon}{1+\epsilon} \text{tr } \rho + \sum_j \delta_j \right).$$

Summing the inequality  $\langle \rho, I_j \rangle \leq \frac{1}{r}(1+\epsilon)$  yields  $\text{tr } \rho \leq 1+\epsilon$ , so to establish [Inequality \(4\)](#) it remains to show

$$2r \sum_j |r\epsilon - (1+\epsilon)\delta_j| \geq \epsilon + \sum_j \delta_j \iff \sum_j (2r|r\epsilon - (1+\epsilon)\delta_j| - \delta_j) \geq \epsilon. \quad (5)$$

This is certainly true if  $\epsilon = 0$ , so it remains to assume  $\epsilon > 0$ . Then writing  $j_0$  for the  $j$  achieving  $\max_j \{r\langle \rho, I_j \rangle\}$ , we have  $\delta_{j_0} = 0$  by construction. This gives us a contribution of  $2\epsilon$  to the sum on the right in [Inequality \(5\)](#). On the other hand, for  $j \neq j_0$ , the worst possible value for  $\delta_j$  is  $\frac{\epsilon}{r(1+\epsilon)}$ , in which case the contribution to the sum is  $-\frac{\epsilon}{r(1+\epsilon)}$ ; summing this over all  $j \neq j_0$  yields at worst  $-\frac{\epsilon}{1+\epsilon}$ . Thus we have shown the sum in [Inequality \(5\)](#) is at least  $2\epsilon - \frac{\epsilon}{1+\epsilon} \geq \epsilon$ , as needed.  $\square$

By inspecting the above proof we may extract the following lemma, which will be useful later:

**Lemma 3.3.** *Suppose  $\rho \in B_1(\mathcal{H})_+$  has  $|\langle \rho, I_j \rangle - \frac{1}{r}| \leq \eta$  for each  $1 \leq j \leq r$ . Then there is  $\rho' \in B_1(\mathcal{H})_+$  satisfying  $\langle \rho', I_j \rangle = \frac{1}{r}$  for each  $j$  and with  $|\langle \rho, a \rangle - \langle \rho', a \rangle| \leq Cr\eta$ . Furthermore, if  $\rho$  has finite rank (respectively, support) then so too does  $\rho'$ .*

Returning to (P3), formally speaking we are regarding it as the conic linear program

$$\sup_{(\rho, \lambda^+, \lambda^-) \in B_1(\mathcal{H})_+ \oplus \mathbb{R}_{\geq 0}^r \oplus \mathbb{R}_{\geq 0}^r} \langle \rho, a \rangle - C \sum_{j=1}^r (\lambda_j^+ + \lambda_j^-) \quad \text{subject to} \quad \langle \rho, I_j \rangle - \frac{1}{r} = \lambda_j^+ - \lambda_j^-, \quad j = 1 \dots r. \quad (\text{P3})$$

Not only is this feasible, meaning the vector  $(\frac{1}{r}, \dots, \frac{1}{r})$  is in the region

$$D = \left\{ (\langle \sigma, I_1 \rangle - (\lambda_1^+ - \lambda_1^-), \dots, \langle \sigma, I_r \rangle - (\lambda_r^+ - \lambda_r^-)) : \sigma \in B_1(\mathcal{H})_+, \lambda^+ \in \mathbb{R}_{\geq 0}^r, \lambda^- \in \mathbb{R}_{\geq 0}^r \right\} \subseteq \mathbb{R}^r,$$

but it is even in the interior of the region, since  $D$  is in fact all of  $\mathbb{R}^r$ . We are therefore finally in a position to apply a strong duality theorem for conic linear programming [[BS00](#), Thm. 2.187] to the pair (P3), (D3). We conclude:

**Theorem 3.4.** (SDP-P) and (SDP-D) have a common value  $s^*$ , and  $s^*$  is achieved in (SDP-D) by some  $\hat{\zeta} \in \mathbb{R}^r$ . We may therefore write  $s^* = \lambda_{\max}(\hat{a})$ , where  $\hat{a} = a + \sum_j \hat{\zeta}_j I_j$ .

### 3.2 Nearly optimal finite, rank- $r$ solutions

Although the optimal value  $s^*$  is achieved in (SDP-D) (that is, its ‘‘inf’’ may be written ‘‘min’’), the value  $s^*$  might not be achieved in (SDP-P). However for any  $\delta > 0$ , we can find  $\rho \in B_1(\mathcal{H})_+$  satisfying  $\langle \rho, I_j \rangle = \frac{1}{r}$  for all  $j$  and with  $\langle \rho, a \rangle \geq s^* - \delta$ . Let us further simplify this  $\rho$ .

**Finite rank.** First, recall that  $B_{00}(\mathcal{H})_+$  (the finite-rank positive semidefinite operators on  $\mathcal{H}$ ) is dense in  $B_1(\mathcal{H})_+$  with respect to the trace-norm (see [[Con85](#), Thm. 1.11(d)], with the proof clearly holding under the positive semidefinite restriction). Thus we can find a finite-rank  $\rho_0 \in B_{00}(\mathcal{H})_+$  satisfying  $|\langle \rho_0, I_j \rangle - \frac{1}{r}| \leq \delta$  for all  $j$  and with  $\langle \rho_0, a \rangle \geq s^* - \delta - \|a\|\delta$ . Next, using [Lemma 3.3](#) we can convert this to another finite-rank  $\rho_1 \in B_{00}(\mathcal{H})_+$  satisfying  $\langle \rho_1, I_j \rangle = \frac{1}{r}$  for all  $j$  and with  $\langle \rho_1, a \rangle \geq s^* - \delta_1$ , where  $\delta_1 = \delta + \|a\|\delta + C\delta \rightarrow 0$  as  $\delta \rightarrow 0$ .

**Finite support.** Since  $\rho_1 \in B_{00}(\mathcal{H})_+$  has trace 1 we can write  $\rho_1 = \sum_{i=1}^{k_1} \lambda_i |\varphi_i\rangle\langle\varphi_i|$  for some  $\lambda = (\lambda_1, \dots, \lambda_{k_1}) \in \mathbb{R}_+^{k_1}$  forming a probability distribution on  $[k_1]$  and some orthonormal vectors  $|\varphi_i\rangle \in \ell_2(V)$ . We can approximate each  $|\varphi_i\rangle$  by a unit vector  $|\varphi'_i\rangle$  of finite support satisfying  $|\langle\varphi_i|\varphi'_i\rangle|^2 \geq 1 - \delta_1$ , from which it follows (an easy calculation [NC10, (9.60), (9.99)]) that  $\|e_i\|_1 \leq 2\sqrt{\delta_1}$ , where  $e_i = |\varphi_i\rangle\langle\varphi_i| - |\varphi'_i\rangle\langle\varphi'_i|$ . Now let  $\rho_2 = \sum_{i=1}^{k_1} \lambda_i |\varphi'_i\rangle\langle\varphi'_i|$ , a positive semidefinite matrix of finite support. Then

$$\langle\rho_2, a\rangle = \langle\rho_1, a\rangle - \sum_{i=1}^{k_1} \lambda_i \langle e_i, a\rangle \geq s^* - \delta_1 - \sum_{i=1}^{k_1} \lambda_i \cdot 2\sqrt{\delta_1}\|a\| = s^* - \delta_1 - 2\sqrt{\delta_1}\|a\|,$$

and one can similarly check that  $|\langle\rho_2, I_j\rangle - \frac{1}{r}| \leq 2\sqrt{\delta_1}$  for all  $j$ . Again applying Lemma 3.3, we can convert this to another finite-support  $\hat{\rho} \in B_{00}(\mathcal{H})_+$  satisfying  $\langle\hat{\rho}, I_j\rangle = \frac{1}{r}$  for all  $j$  and with  $\langle\hat{\rho}, a\rangle \geq s^* - \delta_2$ , where  $\delta_2 = \delta_1 + 2\sqrt{\delta_1}\|a\| + 2C\sqrt{\delta_1} \rightarrow 0$  as  $\delta \rightarrow 0$ . Since  $\delta_2$  can be made arbitrarily small by taking  $\delta \rightarrow 0$ , we finally conclude:

**Proposition 3.5.** *For any  $\epsilon > 0$ , there is a feasible finite-support solution  $\hat{\rho}$  to (SDP-P) achieving objective value at least  $s^* - \epsilon = \lambda_{\max}(\hat{a}) - \epsilon$ .*

**Rank r.** In general, any solution  $\rho$  to (SDP-P) need not have rank more than  $r$ . Although we don't strictly need this fact, it is particularly easy to show for  $\rho$  of finite support, and we do so now.

**Proposition 3.6.** *Let  $\hat{\rho}$  be a feasible solution to (SDP-P), with nonzero entries only in rows or columns indexed by a finite set  $F \subseteq V$ . Then there is another  $\tilde{\rho}$  feasible for (SDP-P), supported on  $F$  and with rank at most  $r$  such that  $\langle\tilde{\rho}, a\rangle \geq \langle\hat{\rho}, a\rangle$ .*

*Proof.* We think of  $\hat{\rho}$  as a matrix indexed just by  $\hat{F}$ , and write  $a_{\hat{F}}$  for the submatrix of  $a$  on rows/columns indexed by  $\hat{F}$ . Suppose  $\hat{\rho}$  has eigenvalues  $\mu_1, \dots, \mu_{|\hat{F}|} \geq 0$  (nonnegative, since  $\hat{\rho} \geq 0$ ), with corresponding orthonormal eigenvectors  $|\psi_1\rangle, \dots, |\psi_{|\hat{F}|}\rangle$ . Then  $\hat{\rho}$  achieves objective value

$$c := \langle\hat{\rho}, a\rangle = \langle\hat{\rho}, a_{\hat{F}}\rangle = \text{tr}(\hat{\rho}a_{\hat{F}}) = \text{tr}\left(\left(\sum_i \mu_i |\psi_i\rangle\langle\psi_i|\right)a_{\hat{F}}\right) = \sum_i \mu_i \langle\psi_i|a_{\hat{F}}|\psi_i\rangle.$$

Also, writing  $I_j$  for the identity matrix restricted to  $V_j \cap \hat{F}$ , feasibility of  $\rho$  implies

$$\frac{1}{r} = \langle\hat{\rho}, I_j\rangle = \sum_i \mu_i \langle\psi_i|I_j|\psi_i\rangle, \quad j = 1 \dots r.$$

Now if we imagine  $|\psi_1\rangle, \dots, |\psi_{|\hat{F}|}\rangle$  are fixed, and  $\lambda_1, \dots, \lambda_{|\hat{F}|}$  are real variables, the above tells us that the following linear program —

$$\max_{\lambda \in \mathbb{R}_{\geq 0}^{|\hat{F}|}} \left\{ \sum_i \lambda_i \langle\psi_i|a_{\hat{F}}|\psi_i\rangle : \sum_i \lambda_i \langle\psi_i|I_j|\psi_i\rangle = \frac{1}{r}, \quad j = 1 \dots r \right\} \quad (\text{LP})$$

— has optimal value at least  $c$ , since we may take  $\lambda = \mu$ . By standard theory of finite linear programs, the optimal value occurs at a vertex where  $|\hat{F}|$  linearly independent constraints are tight. Since there are only  $r$  equality constraints in (LP), there is an optimal solution  $\lambda^*$  with at least  $|\hat{F}| - r$  zero entries. Thus  $\tilde{\rho} = \sum_i \lambda_i^* |\psi_i\rangle\langle\psi_i|$  has rank at most  $r$ , achieves objective value at least  $c$ , and is feasible for (SDP-P).  $\square$

### 3.3 Conclusion for matrix edge-weighted graphs

Let  $V_\infty$  be a countable set of nodes and let  $G_\infty$  be a bounded-degree graph on  $V_\infty$  with matrix edge-weights from  $\mathbb{C}^{r \times r}$ . Let  $A_\infty$  be the adjacency operator for  $G_\infty$ , acting on  $\ell_2(V_\infty) \otimes \mathbb{C}^r$  and assumed self-adjoint; we may think of it as an infinite matrix with rows and columns indexed by  $V_\infty$ , and with entries from  $\mathbb{C}^{r \times r}$ . As in [Definition 2.2](#), we write  $\tilde{A}_\infty$  for its “extension”; this is a self-adjoint bounded operator on  $\ell_2(V_\infty \times [r])$ . Conversely, given any solution  $\hat{\rho}$  for (SDP-P) from [Section 3](#), we may “unextend” it and think of it as an infinite matrix  $\rho$  with rows and columns indexed by  $V_\infty$ , and entries from  $\mathbb{C}^{r \times r}$ .

We apply the the SDP duality theory from [Sections 3](#) and [3.2](#), with “ $V$ ” being  $V_\infty \times [r]$ , “ $a$ ” being  $\tilde{A}_\infty$ , and “ $V_j$ ” being  $V_\infty \times \{j\}$  for all  $1 \leq j \leq r$ .

The conclusion is that for any  $\epsilon > 0$ , there exists:

- $\hat{\zeta} \in \mathbb{R}^r$  with  $\text{avg}_j(\hat{\zeta}_j) = 0$ ;
- a finite subset  $F \subset V_\infty$ ;
- a PSD matrix  $\rho$  (the “unextension” of  $\tilde{\rho}$  of rank at most  $r$  from [Proposition 3.6](#)) with rows/columns indexed by  $V_\infty$  and entries from  $\mathbb{C}^{r \times r}$ , supported on the rows/columns  $F$ , with

$$\text{tr}(\rho)_{jj} = \frac{1}{r}, \quad j = 1 \dots r;$$

such that for

$$\hat{A} = \tilde{A}_\infty + \mathbb{1}_{V_\infty} \otimes \text{diag}(\hat{\zeta}),$$

we have

$$s^* := \lambda_{\max}(\hat{A}) = \text{PARTSDP-DUAL}(A_\infty) = \text{PARTSDP}(A_\infty) \geq \langle \rho, A_\infty \rangle = \langle \rho_F, A_F \rangle \geq s^* - \epsilon, \quad (6)$$

where  $\rho_F, A_F$  denote  $\rho, A_\infty$  (respectively) restricted to the rows/columns  $F$ .

## 4 The SDP value of random matrix polynomial lifts

In this section we prove our main [Theorem 1.2](#). To that end, let  $p$  be any self-adjoint matrix polynomial over indeterminates  $Y_1, \dots, Y_d, Z_1, \dots, Z_e^*$  with  $r \times r$  coefficients. Let  $A_\infty = p(\mathcal{L}_\infty)$  denote the adjacency operator of the infinite lift  $\mathfrak{X}_p$ , and write  $s^* = \text{PARTSDP}(A_\infty) = \text{PARTSDP-DUAL}(A_\infty)$  as in [Equation \(6\)](#). Fix any  $\epsilon, \beta > 0$ , and let  $\mathbf{A}_n = p(\mathcal{L})$  denote the adjacency matrix of a corresponding  $n$ -lift, formed from  $\mathcal{L} = (\mathbf{M}_1, \dots, \mathbf{M}_d, \mathbf{P}_1, \dots, \mathbf{P}_e)$ . Our goal is to show that except with probability at most  $\beta$  (assuming  $n$  is sufficiently large),

$$s^* - \epsilon \leq \text{SDP}(\mathbf{A}_n) = \text{SDP-DUAL}(\mathbf{A}_n) \leq s^* + \epsilon.$$

Given our setup, the upper bound follows easily from prior work, namely [Theorem 2.14](#). Let  $\hat{\zeta}$  and  $\hat{A}$  be as in [Section 3.3](#), and consider the matrix polynomial  $p'$  defined by

$$p' = p + \text{diag}(\hat{\zeta})\mathbb{1}.$$

Then on one hand, the  $\infty$ -lift of  $p'$  has adjacency operator precisely  $\hat{A}$ ; on the other hand,

$$\mathbf{A}'_n := p'(\mathcal{L}) = \mathbf{A}_n + \mathbb{1}_{n \times n} \otimes \text{diag}(\hat{\zeta}).$$

Thus [Theorem 2.14](#) tells us that except with probability  $\beta/2$  (provided  $n$  is large enough), the spectra  $\text{spec}(\mathbf{A}'_n)$  and  $\text{spec}(\hat{A})$  are at Hausdorff distance at most  $\epsilon$ , from which it follows that

$$\lambda_{\max}(\mathbf{A}'_n) \leq \lambda_{\max}(\hat{A}) + \epsilon = s^* + \epsilon.$$

But this indeed proves  $\text{SDP-DUAL}(\mathbf{A}_n) \leq s^* + \epsilon$ , because  $\hat{\zeta}$  has  $\text{avg}(\hat{\zeta}) = 0$  and hence is feasible for  $\text{SDP-DUAL}(\mathbf{A}_n)$ .

It therefore remains to prove  $\text{SDP}(\mathbf{A}_n) \geq s^* - \epsilon$ .

#### 4.1 A lower bound on the basic SDP value

In this section we complete the proof of our main theorem by showing that  $\text{SDP}(\mathbf{A}_n) \geq s^* - \epsilon$  except with probability at most  $o(1) = o_{n \rightarrow \infty}(1)$  (which is at most  $\beta/2$  as needed, provided  $n$  is large enough).

Let  $F, \rho, \rho_F, A_F$  be as in [Section 3.3](#), except with that section's " $\epsilon$ " replaced by  $\epsilon/2$ , so that  $\langle \rho_F, A_F \rangle \geq s^* - \epsilon/2$ . Adding finitely many vertices to  $F$  if necessary, we may assume that it consists of all reduced words over  $Y_1, \dots, Y_d, Z_1, \dots, Z_e^*$  of length at most some finite  $f_0$ . We also make the following definition:

**Definition 4.1** (Cycle in a lift). Given an  $n$ -lift  $\mathcal{L}$ , a *cycle of length  $\ell > 0$*  is a pair  $(i, w)$ , where  $i \in [n]$  and  $w$  is a reduced word of length  $\ell$  such that  $\mathcal{L}^w |i\rangle = \pm |i\rangle$ , and  $\langle i | \mathcal{L}^{w'} |i\rangle = 0$  (i.e.,  $\mathcal{L}^{w'} |i\rangle \neq \pm |i\rangle$ ) for all proper prefixes  $w'$  of  $w$ .

We will employ the following basic random graph result, [[BC19](#), Lem. 23], stated in our language:

**Lemma 4.2.** *For the random  $n$ -lift  $\mathcal{L}$ , the expected number of cycles of length  $\ell$  is  $O(\ell(d+2e-1)^\ell)$ .*

Applying this for all  $\ell \leq f := 2f_0 + \text{deg}(p)$  and using Markov's inequality, we conclude:

**Corollary 4.3.** *Except with probability at most  $n^{-.99}$ , the random  $n$ -lift  $\mathcal{L}$  has at most  $O(n^{.99})$  cycles of length at most  $f$ . In this case, we can exclude a set of "bad" vertices  $B \subseteq [n]$  with  $|B|/n \leq O(n^{-.01}) = o(1)$  so that:*

$$\forall i \notin B, \quad \forall \text{ reduced words } w \text{ with } 0 < |w| \leq 2f_0 + \text{deg}(p), \quad \langle i | \mathcal{L}^w |i\rangle = 0.$$

We henceforth fix an outcome  $\mathcal{L} = \mathcal{L}$  (and hence  $\mathbf{A}_n = A_n$ ) such that the conclusion of [Corollary 4.3](#) holds, accruing our  $o(1)$  probability of failure. Under this assumption, we will show  $\text{SDP}(A_n) \geq s^* - \epsilon/2 - o(1)$ , which is sufficient to complete the proof.

Our plan will be to first construct a "provisional" near-feasible PSD solution  $\sigma$  for  $\text{SDP}(A_n)$ , with rows/columns indexed by  $[n]$  and with  $r \times r$  entries, such that:

- $|\langle \sigma, A_n \rangle - \langle \rho_F, A_F \rangle| \leq o(1)$ , and hence  $\langle \sigma, A_n \rangle \geq s^* - \epsilon/2 - o(1)$ ;
- for  $i \notin B$ , the  $r \times r$  matrix  $\sigma_{ii}$  has diagonal entries  $\frac{1}{nr}$ .

Then, we will show how to "fix"  $\sigma$  to a some  $\sigma'$  that is truly feasible for  $\text{SDP}(A_n)$ , while still having  $\langle \sigma', A_n \rangle \geq \langle \sigma, A_n \rangle - o(1) \geq s^* - \epsilon/2 - o(1)$ .

### 4.1.1 Constructing a near-feasible solution

**Definition 4.4.** For each  $i \in [n]$ , we define a linear operator  $\Phi_i : \mathbb{C}^F \rightarrow \mathbb{C}^n$  by

$$\Phi_i = \sum_{v \in F} \mathcal{L}^v |i\rangle \langle v|,$$

and also  $\tilde{\Phi}_i = \Phi_i \otimes \mathbb{1}_{r \times r}$ . We furthermore define  $\sigma_i$  to be the  $n \times n$  matrix with  $r \times r$  entries whose extension  $\tilde{\sigma}_i$  is

$$\tilde{\sigma}_i = \tilde{\Phi}_i \cdot \tilde{\rho}_F \cdot \tilde{\Phi}_i^\top.$$

**Remark 4.5.**  $\tilde{\sigma}_i$  is PSD, being the conjugation by  $\tilde{\Phi}_i$  of the PSD operator  $\tilde{\rho}_F$ .

**Proposition 4.6.** *If  $i \notin B$ , then  $\tilde{\Phi}_i^\top \tilde{A}_n \tilde{\Phi}_i = \tilde{A}_F$ .*

*Proof.* Thinking of  $\tilde{\Phi}_i^\top \tilde{A}_n \tilde{\Phi}_i$  as an  $F \times F$  matrix of  $r \times r$  matrices, it follows that its  $(u, v)$  entry is given by

$$\sum_{\text{term } a_w w \text{ in } p} \langle i | \mathcal{L}^{v^* w u} | i \rangle a_w.$$

On the other hand, the  $(u, v)$  entry of  $A_F$  is by definition

$$\sum_{\text{term } a_w w \text{ in } p} \mathbb{1}[v^* w u = \mathbb{1}] a_w,$$

where “ $v^* w u = \mathbb{1}$ ” denotes that the reduced form of word  $v^* w u$  is the empty word. We therefore have equality for all  $u, v$  provided  $\langle i | \mathcal{L}^{v^* w u} | i \rangle = 0$  whenever  $v^* w u \neq \emptyset$ . But [Corollary 4.3](#) tells us this indeed holds for  $i \notin B$ , because  $|v^* w u| \leq 2f_0 + \deg(p)$ .  $\square$

**Corollary 4.7.** *For  $i \notin B$  we have  $\langle \sigma_i, A_n \rangle = \langle \rho_F, A_F \rangle$ .*

*Proof.* When  $i \notin B$ ,

$$\langle \sigma_i, A_n \rangle = \text{tr}(\tilde{\sigma}_i \tilde{A}_n) = \text{tr}(\tilde{\Phi}_i \tilde{\rho}_F \tilde{\Phi}_i^\top \tilde{A}_n) = \text{tr}(\tilde{\rho}_F \tilde{\Phi}_i^\top \tilde{A}_n \tilde{\Phi}_i) = \text{tr}(\tilde{\rho}_F \tilde{A}_F) = \langle \rho_F, A_F \rangle,$$

where the last equality used [Proposition 4.6](#)  $\square$

We now define our “provisional” SDP solution  $\sigma$  via

$$\sigma = \text{avg}_{i \in [n]} \{\sigma_i\};$$

this is indeed PSD, being the average of PSD operators. Using [Corollary 4.7](#),  $|B|/n = o(1)$ , and the fact that  $|\langle \sigma_i, A_n \rangle| \leq O(1)$  for every  $i$  (since  $\sigma_i$  only has  $O(1)$  nonzero entries, each bounded in magnitude by  $O(1)$ ), we conclude:

**Proposition 4.8.**  $|\langle \sigma, A_n \rangle - \langle \rho_F, A_F \rangle| \leq o(1)$ , and hence  $\langle \sigma, A_n \rangle \geq s^* - \epsilon/2 - o(1)$ .

Now similar to [Proposition 4.9](#) we have the following:

**Proposition 4.9.** *If  $j \notin B$ , then the  $(j, j)$  entry of  $\sigma$  is  $\frac{1}{n} \text{tr}(\rho_F)$  (and hence is an  $r \times r$  matrix with diagonal entries equal to  $\frac{1}{nr}$ ).*

*Proof.* By definition, the  $(j, j)$  entry of  $\sigma$  is

$$\begin{aligned}
&= \text{avg}_{i \in [n]} \sum_{u, v \in F} \langle j | \mathcal{L}^u | i \rangle \langle u | \rho_F | v \rangle \langle i | \mathcal{L}^{v*} | j \rangle \\
&= \frac{1}{n} \sum_{u, v \in F} \langle u | \rho_F | v \rangle \cdot \sum_{i \in [n]} \langle j | \mathcal{L}^u | i \rangle \langle i | \mathcal{L}^{v*} | j \rangle \\
&= \frac{1}{n} \sum_{u, v \in F} \langle u | \rho_F | v \rangle \cdot \langle j | \mathcal{L}^{uv*} | j \rangle \quad (\text{since } \sum_i |i\rangle\langle i| = \mathbb{1})
\end{aligned}$$

where we are writing  $\langle u | \rho_F | v \rangle$  for the  $r \times r$  matrix at the  $(u, v)$  entry of  $\rho_F$ . Now when  $j \notin B$ , we have that  $\langle j | \mathcal{L}^{vw*} | j \rangle = \mathbb{1}[vw^* = \emptyset]$  by [Corollary 4.3](#), since  $|vw^*| \leq 2f_0$ . Thus all summands above drop out, except for the ones with  $u = v$ ; this indeed gives  $\frac{1}{n} \text{tr}(\rho_F)$ .  $\square$

#### 4.1.2 Fixing $\sigma$

Finally, we slightly fix  $\sigma$  to make it truly feasible for  $\text{SDP}(A)$ . Let  $\sigma'$  be the  $n \times n$  matrix, with entries from  $\mathbb{C}^{r \times r}$ , defined as follows:

$$\sigma'_{ij} = \begin{cases} \sigma_{ij} & \text{if } i, j \notin B, \\ \frac{1}{nr} \mathbb{1}_{r \times r} & \text{if } i = j \in B, \\ 0 & \text{else.} \end{cases}$$

This  $\sigma'$  is easily seen to be PSD, being a principal submatrix of the PSD matrix  $\sigma$ , direct-summed with the PSD matrix  $\frac{1}{nr} \mathbb{1}_{r \times r}$ . As well, the  $nr \times nr$  extension matrix  $\tilde{\sigma}'$  has all diagonal entries equal to  $\frac{1}{nr}$ , by [Proposition 4.9](#). Thus  $\tilde{\sigma}'$  is feasible for  $\text{SDP}(A_n) = \text{SDP}(\tilde{A}_n)$ , and it remains for us to show that

$$\langle \sigma, A_n \rangle - \langle \sigma', A_n \rangle \leq o(1); \tag{7}$$

this will imply  $\langle \sigma', A_n \rangle \geq s^* - \epsilon/2 - o(1)$  by [Proposition 4.8](#), and hence  $\text{SDP}(A_n) \geq s^* - \epsilon$  (for sufficiently large  $n$ ), as desired.

We have

$$\langle \sigma, A_n \rangle - \langle \sigma', A_n \rangle = \langle \sigma - \sigma', A_n \rangle \leq \|\tilde{\sigma} - \tilde{\sigma}'\|_1 \|\tilde{A}_n\|_\infty.$$

Next,

$$\|\tilde{A}_n\|_\infty = \|p(\mathcal{L})\|_\infty = \left\| \sum_w \mathcal{L}^w \otimes a_w \right\|_\infty \leq \sum_w \|\mathcal{L}^w\|_\infty \cdot \|a_w\|_\infty = \sum_w \|a_w\|_\infty \leq O(1).$$

Here the final equality is because each  $\mathcal{L}^w$  is a signed permutation matrix (hence has  $\|\mathcal{L}^w\|_\infty = 1$ ), and the final inequality is because  $p$  has only constantly many coefficients, of constant size. Thus to establish [Inequality \(7\)](#), it remains to show  $\|\tilde{\sigma} - \tilde{\sigma}'\|_1 \leq o(1)$ .

Define the orthogonal projection matrices  $\Pi_B = \sum_{i \in B} |i\rangle\langle i| \otimes \mathbb{1}_{r \times r}$  and similarly  $\Pi_{\bar{B}}$ , where  $\bar{B} = [n] \setminus B$ . Observe that

$$\sigma' = \frac{1}{nr} \Pi_B + \Pi_{\bar{B}} \sigma \Pi_{\bar{B}},$$

and thus

$$\|\tilde{\sigma} - \tilde{\sigma}'\|_1 = \|\tilde{\sigma} - \tilde{\Pi}_{\bar{B}} \tilde{\sigma} \tilde{\Pi}_{\bar{B}} - \frac{1}{nr} \tilde{\Pi}_B\|_1 \leq \|\tilde{\sigma} - \tilde{\Pi}_{\bar{B}} \tilde{\sigma} \tilde{\Pi}_{\bar{B}}\|_1 + \frac{1}{nr} \|\tilde{\Pi}_B\|_1.$$

But  $\frac{1}{nr} \|\tilde{\Pi}_B\|_1 = \frac{|B|}{n} = o(1)$ , so it remains to show

$$\|\tilde{\sigma} - \tilde{\Pi}_{\bar{B}} \tilde{\sigma} \tilde{\Pi}_{\bar{B}}\|_1 \leq o(1).$$

Note that  $\tilde{\sigma} \in \mathbb{C}^{nr \times nr}$  is nearly a density matrix: it is PSD, and all but a  $1 - o(1)$  fraction of its diagonal entries are  $\frac{1}{nr}$ , with the remaining ones being bounded in magnitude by  $O(\frac{1}{n})$ . Thus  $\text{tr}(\tilde{\sigma}) = 1 \pm o(1)$ , and we can therefore scale  $\tilde{\sigma}$  by a  $1 \pm o(1)$  factor to produce a true density matrix  $\hat{\sigma}$ . Clearly it now suffices to show

$$\|\hat{\sigma} - \tilde{\Pi}_{\bar{B}} \hat{\sigma} \tilde{\Pi}_{\bar{B}}\|_1 \leq o(1).$$

But this follows from Winter’s Gentle Measurement Lemma [Win99, Lem. 9], which bounds the quantity on the left by  $\sqrt{8\lambda}$ , where  $\lambda = 1 - \text{tr}(\hat{\sigma} \tilde{\Pi}_{\bar{B}}) = o(1)$ . This completes the proof.

## References

- [AA18] Scott Aaronson and Andris Ambainis. Forrelation: a problem that optimally separates quantum from classical computing. *SIAM J. Comput.*, 47(3):982–1038, 2018. [1.1](#)
- [BBK<sup>+</sup>20] Afonso S Bandeira, Jess Banks, Dmitriy Kunisky, Cristopher Moore, and Alexander S Wein. Spectral planting and the hardness of refuting cuts, colorability, and communities in random graphs. Technical Report 2008.12237, arXiv, 2020. [1.2](#)
- [BC19] Charles Bordenave and Benoît Collins. Eigenvalues of random lifts and polynomials of random permutation matrices. *Annals of Mathematics*, 190(3):811–875, 2019. [1.1](#), [1.2](#), [1.2](#), [2](#), [2.3](#), [4.1](#)
- [BGT13] Mohsen Bayati, David Gamarnik, and Prasad Tetali. Combinatorial approach to the interpolation method and scaling limits in sparse random graphs. *Annals of Probability*, 41(6):4080–4115, 2013. [1](#)
- [Bop87] Ravi Boppana. Eigenvalues and graph bisection: An average-case analysis. In *Proceedings of the 28th Annual IEEE Symposium on Foundations of Computer Science*, pages 280–285, 1987. [1.1](#)
- [Bor20] Charles Bordenave. A new proof of Friedman’s second eigenvalue theorem and its extension to random lifts. *Ann. Sci. Éc. Norm. Supér. (4)*, 53(6):1393–1439, 2020. [1.1](#), [1.1](#)
- [BR02] Ola Bratelli and Derek Robinson. *Operator algebras and quantum statistical mechanics*. Springer, 2002. [3.1](#)
- [BS00] J. Frédéric Bonnans and Alexander Shapiro. *Perturbation Analysis of Optimization Problems*. Springer Series in Operations Research. Springer-Verlag, New York, 2000. [3](#), [3.1](#)
- [CGHV15] Endre Csóka, Balázs Gerencsér, Viktor Harangi, and Bálint Virág. Invariant Gaussian processes and independent sets on regular graphs of large girth. *Random Structures Algorithms*, 47(2):284–303, 2015. [1.1](#)
- [CHSH69] John Clauser, Michael Horne, Abner Shimony, and Richard Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23(15):880–884, 1969. [1.1](#)
- [CHTW04] Richard Cleve, Peter Høyer, Benjamin Toner, and John Watrous. Consequences and limits of nonlocal strategies. In *Proceedings of the 19th Annual Computational Complexity Conference*, pages 236–249. IEEE, 2004. [1.1](#)

- [Con85] John B. Conway. *The Theory of Subnormal Operators*. American Mathematical Society, 1985. [3.2](#)
- [Con90] John B. Conway. *A Course in Functional Analysis*. Springer, 1990. [3.1](#)
- [DH73] Wilm Donath and Alan Hoffman. Lower bounds for the partitioning of graphs. *IBM J. Res. Develop.*, 17:420–425, 1973. [1.1](#)
- [DLS14] Amit Daniely, Nati Linial, and Shai Shalev-Shwartz. From average case complexity to improper learning complexity. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 441–448, 2014. [1](#)
- [DMO<sup>+</sup>19] Yash Deshpande, Andrea Montanari, Ryan O’Donnell, Tselil Schramm, and Subhabrata Sen. The threshold for SDP-refutation of random regular NAE-3SAT. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2305–2321. SIAM, 2019. [1.1](#), [1.1](#), [1.2](#)
- [DMS<sup>+</sup>17] Amir Dembo, Andrea Montanari, Subhabrata Sen, et al. Extremal cuts of sparse random graphs. *Annals of Probability*, 45(2):1190–1217, 2017. [1](#)
- [DP93] Charles Delorme and Svatopluk Poljak. Laplacian eigenvalues and the maximum cut problem. *Mathematical Programming*, 62:557–574, 1993. [1.1](#), [1.1](#), [2.18](#)
- [Elo09] Yehonatan Elon. Gaussian waves on the regular tree. Technical Report 0907.5065, arXiv, 2009. [1.1](#)
- [FL92] Uriel Feige and László Lovász. Two-prover one-round proof systems: Their power and their problems. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing*, pages 733–744, 1992. [1](#)
- [Fri08] Joel Friedman. A proof of Alon’s second eigenvalue conjecture and related problems. *Memoirs of the American Mathematical Society*, 195(910):viii+100, 2008. [1.1](#)
- [Fri19] Shmuel Friedland. Infinite dimensional generalizations of Choi’s theorem. *Spec. Matrices*, 7:67–77, 2019. [3.1](#)
- [GVK21] Jorge Garza-Vargas and Archit Kulkarni. Spectra of infinite graphs via freeness with amalgamation. Technical Report 1912.10137, arXiv, 2021. [1.2](#)
- [GW95] Michel Goemans and David Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. Assoc. Comput. Mach.*, 42(6):1115–1145, 1995. [1.1](#)
- [HV15] Viktor Harangi and Bálint Virág. Independence ratio and random eigenvectors in transitive graphs. *Ann. Probab.*, 43(5):2810–2840, 2015. [1.1](#)
- [JLS20] Aayush Jain, Huijia Lin, and Amit Sahai. Indistinguishability Obfuscation from well-founded assumptions. Technical Report 2008.09317, arXiv, 2020. [1](#)
- [Lan17] Klaas Landsman. *Foundations of Quantum Theory*. Springer Open, 2017. [3.1](#)
- [Leh99] Franz Lehner. Computing norms of free operators with matrix coefficients. *Amer. J. Math.*, 121(3):453–486, 1999. [1.2](#)

- [Mey06] Paul-André Meyer. *Quantum probability for probabilists*. Springer, second edition, 2006. [3.1](#)
- [MM09] Marc Mézard and Andrea Montanari. *Information, physics, and computation*. Oxford Graduate Texts. Oxford University Press, Oxford, 2009. [1](#)
- [Mon21] Andrea Montanari. Optimization of the Sherrington–Kirkpatrick hamiltonian. *SIAM Journal on Computing*, FOCS19, 2021. [1](#)
- [MOP20] Sidhanth Mohanty, Ryan O’Donnell, and Pedro Paredes. The SDP value for random two-eigenvalue CSPs. In Christophe Paul and Markus Bläser, editors, *Proceedings of the 37th Annual Symposium on Theoretical Aspects of Computer Science*, volume 154 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 50:1–50:45, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. [1.1](#)
- [MS16] Andrea Montanari and Subhabrata Sen. Semidefinite programs on sparse random graphs and their application to community detection. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*, pages 814–827. ACM, New York, 2016. [1.1](#), [1.1](#), [1.2](#)
- [NC10] Michael Nielsen and Isaac Chung. *Quantum Computation and Quantum Information*. Cambridge University Press, 10th anniversary edition edition, 2010. [3.2](#)
- [OW20] Ryan O’Donnell and Xinyu Wu. Explicit near-fully X-Ramanujan graphs. In *Proceedings of the 61st Annual IEEE Symposium on Foundations of Computer Science*, pages 1045–1056, 2020. [1.1](#), [1.2](#), [1.2](#), [1.2](#), [2.3](#)
- [PR95] Svatopluk Poljak and Franz Rendl. Nonpolyhedral relaxations of graph-bisection problems. *SIAM Journal on Optimization*, 5(3):467–487, 1995. [1.1](#), [2.4](#)
- [Rag09] Prasad Raghavendra. *Approximating NP-hard problems: efficient algorithms and their limits*. PhD thesis, University of Washington, 2009. [1](#)
- [RS80] Michael Reed and Barry Simon. *Functional Analysis*. Academic Press, 1980. [3.1](#)
- [RVW02] Omer Reingold, Salil Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Ann. of Math. (2)*, 155(1):157–187, 2002. [1.2](#)
- [RW95] Franz Rendl and Henry Wolkowicz. A projection technique for partitioning the nodes of a graph. *Ann. Oper. Res.*, 58:155–179, 1995. Applied mathematical programming and modeling, II (APMOD 93) (Budapest, 1993). [1.1](#)
- [SH08] Maksim Shirokov and Alexander Holevo. On the approximation of infinite-dimensional quantum channels. *Problemy Peredachi Informatsii*, 44(2):3–22, 2008. [3.1](#)
- [Sty73] George Styan. Hadamard products and multivariate statistical analysis. *Linear Algebra Appl.*, 6:217–240, 1973. [2.4](#)
- [vDE20] Anne van Delft and Michael Eichler. A note on Herglotz’s theorem for time series on function spaces. *Stochastic Process. Appl.*, 130(6):3687–3710, 2020. [3.1](#)

- [Win99] Andreas Winter. Coding theorem and strong converse for quantum channels. *Institute of Electrical and Electronics Engineers. Transactions on Information Theory*, 45(7):2481–2485, 1999. [4.1.2](#)