QUASI-RANDOMNESS AND ALGORITHMIC REGULARITY FOR GRAPHS WITH GENERAL DEGREE DISTRIBUTIONS*

NOGA ALON†, AMIN COJA-OGHLAN‡, HIỆP HÀN§, MIHYUN KANG¶, VOJTĚCH RÖDL‖, AND MATHIAS SCHACHT**

Abstract. We deal with two intimately related subjects: quasi-randomness and regular partitions. The purpose of the concept of quasi-randomness is to measure how much a given graph "resembles" a random one. Moreover, a regular partition approximates a given graph by a bounded number of quasi-random graphs. Regarding quasi-randomness, we present a new spectral characterization of low discrepancy, which extends to sparse graphs. Concerning regular partitions, we introduce a concept of regularity that takes into account vertex weights, and show that if $G = (V, E)$ satisfies a certain boundedness condition, then G admits a regular partition. In addition, building on the work of Alon and Naor [Proc. 36th ACM STOC (2004) 72–80], we provide an algorithm that computes a regular partition of a given (possibly sparse) graph G in polynomial time. As an application, we present a polynomial time approximation scheme for MAX CUT on (sparse) graphs without "dense spots".

Key words: quasi-random graphs, Laplacian eigenvalues, regularity lemma, Grothendieck's inequality.

AMS subject classification: 05C85, 05C50.

1. Introduction and Results. This paper deals with quasi-randomness and regular partitions. Loosely speaking, a graph is quasi-random if the global distribution of the edges resembles the expected edge distribution of a random graph. Furthermore, a regular partition approximates a given graph by a constant number of quasi-random graphs. Such partitions are of algorithmic importance, because a number of NP-hard problems can be solved in polynomial time on graphs that come with regular partitions. In this section we present our main results and discuss related work. The remaining sections contain the proofs and detailed descriptions of the algorithms.

1.1. Quasi-Randomness: Discrepancy and Eigenvalues. Random graphs are well known to have a number of remarkable properties (e.g., excellent expansion). Therefore, quantifying how much a given graph "resembles" a random one is an important problem, both from a structural and an algorithmic point of view. Providing such measures is the purpose of the notion of *quasi-randomness*. While this concept is rather well developed for dense graphs (i.e., graphs $G = (V, E)$ with $|E| = \Omega(|V|^2)$), less is known in the sparse case, which we deal with in the present work. In fact, we

---

*An extended abstract version of this work appeared in the Proceedings of ICALP 2007.

†School of Mathematics and Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel, noga@math.tau.ac.il. Supported by an ISF grant, an ERC Advanced grant, and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University.

‡University of Warwick, DIMAP, Coventry CV4 7AL, UK, A.Coja-Oghlan@warwick.ac.uk. Supported by DFG CO 646.

§Humboldt-Universität zu Berlin, Institut für Informatik, Unter den Linden 6, 10099 Berlin, Germany, hhan@informatik.hu-berlin.de. Supported by DFG within the research training group "Methods for Discrete Structures".

¶Technische Universität Berlin, Institut für Mathematik, Straße des 17. Juni 136, 10623 Berlin, Germany, kang@math.tu-berlin.de. Supported by DFG grants PR 296/7-3 and KA 2748/2-1.

‖Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA, rodl@mathcs.emory.edu. Supported by NSF Grants DMS 0300529 and DMS 0800070.

**University of Hamburg, Department of Mathematics, Bundesstraße 55, 20146 Hamburg, Germany, schacht@math.uni-hamburg.de. Supported by DFG grant SCHA 1263/1-1 and GIF Grant 889/05.

shall actually deal with (sparse) graphs with *general degree distributions*, including but not limited to the ubiquitous power-law degree distributions (cf. [1]).

We will mainly consider two types of quasi-random properties: low discrepancy and eigenvalue separation. The low discrepancy property concerns the global edge distribution and basically states that *every* set S of vertices approximately spans as many edges as we would expect in a random graph with the same degree distribution. More precisely, if $G = (V, E)$ is a graph, then we let d_v signify the degree of $v \in V$. Furthermore, the *volume* of a set $S \subset V$ is $\mathrm{vol}(S) = \sum_{v \in S} d_v$. In addition, if $S, T \subset V$ are disjoint sets, then $e(S, T)$ denotes the number of S-T-edges in G and $e(S)$ is two times the number of edges spanned by the set S. For not necessarily disjoint sets $S, T \subset V$ we let $e(S, T) = e(S \setminus T, T \setminus S) + e(S \cap T)$.

Disc(ε): We say that G *has discrepancy at most* ε ("G has Disc(ε)" for short) if

$$\forall S \subset V : \left| e(S) - \frac{\mathrm{vol}(S)^2}{\mathrm{vol}(V)} \right| < 2\varepsilon \cdot \mathrm{vol}(V). \tag{1.1}$$

To explain (1.1), let $\boldsymbol{d} = (d_v)_{v \in V}$, and let $G(\boldsymbol{d})$ signify a random graph with expected degree distribution $\boldsymbol{d}$; that is, any two vertices v, w are adjacent with probability $p_{vw} = d_v d_w / \mathrm{vol}(V)$ independently. Then in $G(\boldsymbol{d})$ the *expected* number of edges inside of $S \subset V$ equals $\frac{1}{2} \sum_{(v,w) \in S^2} p_{vw} = \frac{1}{2} \mathrm{vol}(S)^2 / \mathrm{vol}(V)$. Consequently, (1.1) just says that for *any* set S the actual number of edges inside of S must not deviate from what we expect in $G(\boldsymbol{d})$ by more than an ε-fraction of the total volume.

An obvious problem with the bounded discrepancy property (1.1) is that it seems quite difficult to check whether $G = (V, E)$ satisfies this condition. This is because apparently one would have to inspect an exponential number of subsets $S \subset V$. Therefore, we consider a second property that refers to the eigenvalues of a certain matrix representing G. More precisely, we will deal with the *normalized Laplacian* $L(G)$, whose entries $(\ell_{vw})_{v,w \in V}$ are defined as

$$\ell_{vw} = \begin{cases} 1 & \text{if } v = w \text{ and } d_v \geq 1, \\ -(d_v d_w)^{-\frac{1}{2}} & \text{if } v, w \text{ are adjacent}, \\ 0 & \text{otherwise;} \end{cases}$$

Due to the normalization by the geometric mean $\sqrt{d_v d_w}$ of the vertex degrees, $L(G)$ turns out to be appropriate for representing graphs with general degree distributions. Moreover, $L(G)$ is well known to be positive semidefinite, and the multiplicity of the eigenvalue 0 equals the number of connected components of G (cf. [9]).

Eig(δ): Letting $0 = \lambda_1 [L(G)] \leq \cdots \leq \lambda_{|V|} [L(G)]$ denote the eigenvalues of $L(G)$, we say that G *has δ-eigenvalue separation* ("G has Eig(δ)") if $1 - \delta \leq \lambda_2 [L(G)] \leq \lambda_{|V|} [L(G)] \leq 1 + \delta$.

As the eigenvalues of $L(G)$ can be computed in polynomial time (within arbitrary numerical precision), we can essentially check efficiently whether G has Eig(δ) or not.

It is not difficult to see that Eig(δ) provides a *sufficient* condition for Disc(ε). That is, for any $\varepsilon > 0$ there is a $\delta > 0$ such that any graph G that has Eig(δ) also has Disc(ε). However, while the converse implication is true if G is dense (i.e., $\mathrm{vol}(V) = \Omega(|V|^2)$), it is false for sparse graphs. In fact, providing a *necessary* condition for Disc(ε) in terms of eigenvalues has been an open problem in the area of sparse quasi-random graphs since the work of Chung and Graham [11]. Concerning this problem, we basically observe that the reason why Disc(ε) does in general not imply Eig(δ) is the existence of a small set of "exceptional" vertices.

ess-Eig(δ): We say that G *has essential δ-eigenvalue separation* ("G has ess-Eig(δ)") if there is a set $W \subset V$ of volume $\mathrm{vol}(W) \geq (1-\delta)\mathrm{vol}(V)$ such that the following is true. Let $L(G)_W = (\ell_{vw})_{v,w \in W}$ denote the minor of $L(G)$ induced on $W \times W$, and let $\lambda_1 [L(G)_W] \leq \cdots \leq \lambda_{|W|} [L(G)_W]$ signify its eigenvalues. Then we require that $1 - \delta < \lambda_2 [L(G)_W] \leq \lambda_{|W|} [L(G)_W] < 1 + \delta$.

THEOREM 1.1. *There is a constant $\gamma > 0$ such that the following is true for all graphs $G = (V, E)$ and all $\varepsilon > 0$.*

1. *If G has* ess-Eig(ε)*, then G satisfies* Disc($10\sqrt{\varepsilon}$)*.*
2. *If G has* Disc($\gamma\varepsilon^2$)*, then G satisfies* ess-Eig(ε)*.*

The main contribution is the second implication. Its proof is based on Grothendieck's inequality and the duality theorem for semidefinite programs. In effect, the proof actually provides us with an efficient algorithm that computes a set W as in the definition of ess-Eig(ε). The second part of Theorem 1.1 is best possible, up to the precise value of the constant γ (see Section 6).

1.2. The Algorithmic Regularity Lemma. Loosely speaking, a regular partition of a graph $G = (V, E)$ is a partition of $(V_1, \ldots, V_t)$ of V such that for "most" index pairs $1 \leq i < j \leq t$ the bipartite subgraph spanned by V_i and V_j is quasi-random. Thus, a regular partition approximates G by quasi-random graphs. Furthermore, the number t of classes may depend on a parameter ε that rules the accuracy of the approximation, but it does *not* depend on the order of the graph G itself. Therefore, if for some class of graphs we can compute regular partitions in polynomial time, then this graph class will admit polynomial time algorithms for various problems that are NP-hard in general.

In the sequel we introduce a new concept of regular partitions that takes into account a given "ambient" weight distribution $\boldsymbol{D} = (D_v)_{v \in V}$, which is an arbitrary sequence of rationals between 1 and $n = |V|$. We will see at the end of this section how this relates to the notion of quasi-randomness discussed in the previous section. Let $G = (V, E)$ be a graph. For subsets $X, Y \subset V$ we set

$$\varrho(X,Y) = \frac{e(X,Y)}{D(X)D(Y)}, \quad \text{where } D(U) = \textstyle\sum_{u \in U} D_u \text{ for any } U \subset V.$$

Further, we say that for disjoint sets $X, Y \subset V$ the pair (X, Y) is $(\varepsilon, \boldsymbol{D})$*-regular* if for all $X' \subset X$, $Y' \subset Y$ satisfying $D(X') \geq \varepsilon D(X)$, $D(Y') \geq \varepsilon D(Y)$ we have

$$|e(X', Y') - \varrho(X,Y)D(X')D(Y')| \leq \varepsilon \cdot \frac{D(X)D(Y)}{D(V)}\,. \tag{1.2}$$

Roughly speaking, (1.2) states that the bipartite graph spanned by X and Y is "quasi-random" with respect to the vertex weights $\boldsymbol{D}$.

In the present notation, Szemerédi's original regularity lemma [23] states that *every* graph G admits a regular partition with respect to the weight distribution $D(v) = n$ for all $v \in V$. However, if G is sparse (i.e., $|E| \ll |V|^2$), then such a regular partition is not helpful because the bound on the r.h.s. of (1.2) exceeds $|E|$. To obtain an appropriate bound, we would have to consider a weight distribution such that $D(v) \ll n$ for (at least) some $v \in V$. But with respect to such weight distributions regular partitions do not necessarily exist. The basic obstacle is the presence of large "dense spots" (X, Y), where $e(X, Y)$ is far bigger than the term $D(X)D(Y)$ suggests. To rule these out, we consider the following notion.

$(C, \eta, \boldsymbol{D})$*-boundedness.* Let $C \geq 1$ and $\eta > 0$. A graph G is $(C, \eta, \boldsymbol{D})$*-bounded* if for all $X, Y \subset V$ with $D(X), D(Y) \geq \eta D(V)$ we have $\varrho(X,Y)D(V) \leq C$.

To illustrate the boundedness condition, consider a random graph $G(\boldsymbol{D})$ with expected degree sequence $\boldsymbol{D}$ such that $D(V) \gg n$. Then for any two disjoint sets $X, Y \subset V$ we have $\mathbb{E}[e(X,Y)] = D(X)D(Y)/D(V) + o(D(V))$. Hence, Chernoff bounds imply that for all X, Y simultaneously we have $e(X,Y) = D(X)D(Y)/D(V) + o(D(V))$ with probability $1 - o(1)$ as $n \to \infty$. Therefore, for any fixed $\eta > 0$ the random graph $G(\boldsymbol{D})$ is $(1 + o(1), \eta, \boldsymbol{D})$ bounded with probability $1 - o(1)$.

Now, we can state the following *algorithmic regularity lemma* for graphs with general degree distributions, which does not only ensure the *existence* of regular partitions, but also that such a partition can be computed efficiently. We let $\langle \boldsymbol{D} \rangle$ signify the *encoding length* of a weight distribution $\boldsymbol{D} = (D_v)_{v\in V}$, i.e., the number of bits that are needed to write down the rationals $(D_v)_{v\in V}$. Observe that $\langle \boldsymbol{D} \rangle \geq n$.

THEOREM 1.2. *For any two numbers $C \geq 1$ and $\varepsilon > 0$ there exist $\eta > 0$ and $n_0 > 0$ such that for all $n \geq n_0$ and every sequence of rationals $\boldsymbol{D} = (D_v)_{v\in V}$ with $|V| = n$ and $1 \leq D_v \leq n$ for all $v \in V$ the following holds. If $G = (V, E)$ is a $(C, \eta, \boldsymbol{D})$-bounded graph and $D(V) \geq \eta^{-1}n$, then there is a partition $\mathcal{P} = \{V_i : 0 \leq i \leq t\}$ of V that satisfies the following two properties:*

REG1. *(a) $\eta D(V) \leq D(V_i) \leq \varepsilon D(V)$ for all $i = 1, \dots, t$,*
(b) $D(V_0) \leq \varepsilon D(V)$, and
(c) $|D(V_i) - D(V_j)| < \max_{v\in V} D_v$ for all $1 \leq i < j \leq t$.

REG2. *Let $\mathcal{L}$ be the set of all pairs (i, j) of indices $1 \leq i < j \leq t$ such that (V_i, V_j) is not $(\varepsilon, \boldsymbol{D})$-regular. Then*

$$\sum_{(i,j)\in\mathcal{L}} D(V_i)D(V_j) \leq \varepsilon D(V)^2.$$

Furthermore, for fixed C and ε the partition $\mathcal{P}$ can be computed in polynomial time. More precisely, there exist a function f and a polynomial Π such that the partition $\mathcal{P}$ can be computed in time $f(C, \varepsilon) \cdot \Pi(\langle \boldsymbol{D} \rangle)$.

Condition **REG1** states that all of the classes $V_1, \dots, V_t$ have approximately the same, non-negligible weight, while the "exceptional" class V_0 has a "small" weight. Also note that due to **REG1(a)** the number of classes t of the partition $\mathcal{P}$ is bounded by $1/\eta$, which only depends on C and ε, but not on G, $\boldsymbol{D}$, or n. Moreover, **REG2** requires that the total weight of the irregular pairs (V_i, V_j) is small relative to the total weight. Thus, a partition $\mathcal{P}$ that satisfies **REG1** and **REG2** approximates G by a bounded number of bipartite quasi-random graphs.

We illustrate the use of Theorem 1.2 with the example of the MAX CUT problem. While approximating MAX CUT within a ratio better than $\frac{16}{17}$ is NP-hard on general graphs [19, 24], the following theorem provides a polynomial time approximation scheme for $(C, \eta, \boldsymbol{D})$-bounded graphs.

THEOREM 1.3. *For any $\delta > 0$ and $C \geq 1$ there exist two numbers $\eta > 0$, $n_0 > 0$ and a polynomial time algorithm* **ApxMaxCut** *such that for all $n \geq n_0$ and every sequence of rationals $\boldsymbol{D} = (D_v)_{v\in V}$ with $|V| = n$ and $1 \leq D_v \leq n$ for all $v \in V$ the following is true. If $G = (V, E)$ is a $(C, \eta, \boldsymbol{D})$-bounded graph and $D(V) > \eta^{-1}n$, then* **ApxMaxCut** *outputs a cut of G that approximates the maximum cut up to an additive error of $\delta|D(V)|$.*

Finally, let us discuss a few examples and applications of the above results.

1. If we let $D(v) = n$ for all $v \in V$, then Theorem 1.2 is just an algorithmic version of Szemerédi's regularity lemma. Such a result was established previously in [3].

2. Suppose that $D(v) = \bar{d}$ for some number $\bar{d} = \bar{d}(n) = o(n)$. Then the above notions of regularity and boundedness coincide with those of the classical "sparse regularity lemma" of Kohayakawa [21] and Rödl (unpublished). Hence, Theorem 1.2 provides an algorithmic version of this regularity concept. This result has not been published previously (although it may have been known to experts in the area that this can be derived from Alon and Naor [4]). Actually devising an algorithm for computing a sparse regular partition is mentioned as an open problem in [21].
3. For a given graph $G = (V, E)$ we could just use the degree sequence as a weight distribution, i.e., $D(v) = d_v$ for all $v \in V$. Then $D(U) = \mathrm{vol}(U)$ for all $U \subset V$. Hence, the notion of regularity (1.2) is closely related to the notion of quasi-randomness from Section 1.1. The resulting regularity concept is a generalization of the "classical" sparse regularity lemma. The new concept allows for graphs with highly irregular degree distributions.

1.3. Further Related Work. Quasi-random graphs with general degree distributions were first studied by Chung and Graham [10]. They considered the properties $\mathrm{Disc}(\varepsilon)$ and $\mathrm{Eig}(\delta)$, and a number of further related ones (e.g., concerning weighted cycles). Chung and Graham observed that $\mathrm{Eig}(\delta)$ implies $\mathrm{Disc}(\varepsilon)$, and that the converse is true in the case of *dense* graphs (i.e., $\mathrm{vol}(V) = \Omega(|V|^2)$).

Regarding the step from discrepancy to eigenvalue separation, Butler [8] proved that any graph G such that for all sets $X, Y \subset V$ the bound

$$|e(X,Y) - \mathrm{vol}(X)\mathrm{vol}(Y)/\mathrm{vol}(V)| \leq \varepsilon\sqrt{\mathrm{vol}(X)\mathrm{vol}(Y)} \tag{1.3}$$

holds, satisfies $\mathrm{Eig}(O(\varepsilon(1 - \ln \varepsilon)))$. His proof builds upon the work of Bilu and Linial [6], who derived a similar result for regular graphs, and on the earlier related work of Bollobás and Nikiforov [7].

Butler's result relates to the second part of Theorem 1.1 as follows. The r.h.s. of (1.3) refers to the volumes of the sets X, Y, and may thus be significantly smaller than $\varepsilon\mathrm{vol}(V)$. By comparison, the second part of Theorem 1.1 just requires that the "original" discrepancy condition $\mathrm{Disc}(\delta)$ is true, i.e., we just need to bound $|e(S) - \mathrm{vol}(S)^2/\mathrm{vol}(V)|$ in terms of the *total* volume $\mathrm{vol}(V)$. Hence, Butler shows that the "original" eigenvalue separation condition Eig follows from a stronger version of the discrepancy property. By contrast, Theorem 1.1 shows that the "original" discrepancy condition Disc implies a weak form of eigenvalue separation ess-Eig, thereby answering a question posed by Chung and Graham [10, 11]. Furthermore, relying on Grothendieck's inequality and SDP duality, the proof of Theorem 1.1 employs quite different techniques than [6, 7, 8].

In the present work we consider a concept of quasi-randomness that takes into account vertex degrees. Other concepts that do not refer to the degree sequence (and are therefore restricted to approximately regular graphs) were studied by Chung, Graham and Wilson [12] (dense graphs) and by Chung and Graham [11] (sparse graphs). Also in this setting it has been an open problem to derive eigenvalue separation from low discrepancy. Concerning this simpler concept of quasi-randomness, our techniques yield a similar result as Theorem 1.1 as well. The proof is similar and we omit the details.

Szemerédi's original regularity lemma [23] has become an important tool in various areas, including extremal graph theory and property testing. Alon, Duke, Lefmann, Rödl, and Yuster [3] presented an algorithmic version, and showed how this lemma can be used to provide polynomial time approximation schemes for dense instances of NP-hard problems (see also [22] for a faster algorithm). Moreover, Frieze and Kannan [13] introduced a different algorithmic regularity concept, which yields better efficiency in terms of the desired approximation guarantee. Both [3, 13] encompass Theorem 1.3 in the case that $D(v) = n$ for all $v \in V$. The sparse regularity lemma from Kohayakawa [21] and Rödl (unpublished) is related to the notion of quasi-randomness from [11]. This concept of regularity has proved very useful in the theory of random graphs, see Gerke and Steger [15].

2. Preliminaries.

2.1. Notation. We let $\mathbf{1}$ denote the vector with all entries equal to one (in any dimension). If $S \subset V$ is a subset of some set V, then we let $\mathbf{1}_S \in \mathbf{R}^V$ denote the vector whose entries are 1 on the components corresponding to elements of S, and 0 otherwise. More generally, if $\xi \in \mathbf{R}^V$ is a vector, then $\xi_S \in \mathbf{R}^V$ signifies the vector obtained from ξ by replacing all components with indices in $V \setminus S$ by 0. Moreover, if $A = (a_{vw})_{v,w\in V}$ is a matrix, then $A_S = (a_{vw})_{v,w\in S}$ denotes the minor of A induced on $S \times S$. Further, for a vector $\xi \in \mathbf{R}^V$ we let $\|\xi\|$ signify the ℓ_2-norm, and for a matrix $M \in \mathbf{R}^{V\times V}$ we let

$$\|M|| = \max_{0\neq\xi\in\mathbf{R}^V} \frac{\|M\xi\|}{\|\xi\|} = \max_{\xi,\eta\in\mathbf{R}^V\setminus\{0\}} \frac{\langle M\xi, \eta\rangle}{\|\xi\|\cdot\|\eta\|}$$

denote the spectral norm.

If $\xi = (\xi_v)_{v\in V}$ is a vector, then $\mathrm{diag}(\xi)$ signifies the $V \times V$ matrix with diagonal ξ and off-diagonal entries equal to 0. In particular, $\boldsymbol{E} = \mathrm{diag}(\mathbf{1})$ denotes the identity matrix (of any size). Moreover, if M is a $\nu\times\nu$ matrix, then $\mathrm{diag}(M) \in \mathbf{R}^\nu$ signifies the vector comprising the diagonal entries of M. If both $A = (a_{ij})_{1\leq i,j\leq\nu}$, $B = (b_{ij})_{1\leq i,j\leq\nu}$ are $\nu \times \nu$ matrices, then we let $\langle A, B\rangle = \sum_{i,j=1}^{\nu} a_{ij}b_{ij}$.

If M is a symmetric $\nu \times \nu$ matrix, then

$$\lambda_1 [M] \leq \cdots \leq \lambda_\nu [M] = \lambda_{\max} [M]$$

denote the eigenvalues of M. We will occasionally need the Courant-Fischer characterizations of λ_2 and $\lambda_{\max}$, which read (see [5, Chapter 7])

$$\lambda_2 [M] = \max_{0\neq\zeta\in\mathbf{R}^\nu} \min_{\xi\perp\zeta,\|\xi\|=1} \langle M\xi, \xi\rangle , \lambda_{\max} [M] = \max_{\zeta\in\mathbf{R}^\nu,\|\zeta\|=1} \langle M\zeta, \zeta\rangle . \qquad (2.1)$$

Recall that a symmetric matrix M is positive semidefinite if $\lambda_1 [M] \geq 0$. In this case we write $M \geq 0$. Furthermore, M is positive definite if $\lambda_1 [M] > 0$, denoted as $M > 0$. If M, M' are symmetric, then $M \geq M'$ (resp. $M > M'$) denotes the fact that $M - M' \geq 0$ (resp. $M - M' > 0$).

2.2. Grothendieck's inequality. An important ingredient to our proofs and algorithms is Grothendieck's inequality. Let $M = (m_{ij})_{i,j\in\mathcal{I}}$ be a matrix. Then the *cut-norm* of M is

$$\|M\|_{\mathrm{cut}} = \max_{I,J\subset\mathcal{I}} \left| \sum_{(i,j)\in I\times J} m_{ij} \right| .$$

In addition, consider the following optimization problem:

$$\mathrm{SDP}(M) = \max \sum_{i,j\in\mathcal{I}} m_{ij} \langle x_i, y_j\rangle \ \text{ s.t. } \forall i \in \mathcal{I} : \|x_i\| = \|y_i\| = 1, \ x_i, y_i \in \mathbf{R}^{2|\mathcal{I}|}. \quad (2.2)$$

This can be reformulated as a *linear* optimization problem over the cone of positive semidefinite $2|\mathcal{I}| \times 2|\mathcal{I}|$ matrices, i.e., as a semidefinite program (see Alizadeh [2]).

LEMMA 2.1. *For any $\nu \times \nu$ matrix M we have*

$$\mathrm{SDP}(M) = \frac{1}{2} \max \left\langle \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes M, X \right\rangle \ \textit{s.t.}\ \mathrm{diag}(X) = \mathbf{1},\ X \geq 0,\ X \in \mathbf{R}^{2\nu \times 2\nu}. \tag{2.3}$$

Proof. Let $x_1, \ldots, x_{2\nu} \in \mathbf{R}^{2\nu}$ be a family of unit vectors such that $\mathrm{SDP}(M) = \sum_{i,j=1}^{\nu} m_{ij} \langle x_i, x_{j+\nu} \rangle$. We obtain a positive semidefinite matrix $X = (x_{i,j})_{1 \leq i,j \leq 2\nu}$ by setting $x_{i,j} = \langle x_i, x_j \rangle$. Since $x_{i,i} = \|x_i\|^2 = 1$ for all i, this matrix satisfies $\mathrm{diag}(X) = \mathbf{1}$. Moreover,

$$\left\langle \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes M, X \right\rangle = 2 \sum_{i,j=1}^{\nu} m_{ij} x_{i,j+\nu} = 2 \sum_{i,j=1}^{\nu} m_{ij} \langle x_i, x_{j+\nu} \rangle . \tag{2.4}$$

Hence, the optimization problem on the r.h.s. of (2.3) yields an upper bound on $\mathrm{SDP}(M)$.

Conversely, if $X = (x_{i,j})$ is a feasible solution to (2.3), then there exist vectors $x_1, \ldots, x_{2\nu} \in \mathbf{R}^{2\nu}$ such that $x_{i,j} = \langle x_i, x_j \rangle$, because X is positive semidefinite. Moreover, since $\mathrm{diag}(X) = \mathbf{1}$, we have $1 = x_{i,i} = \|x_i\|^2$. Thus, $x_1, \ldots, x_{2\nu}$ is a feasible solution to (2.2), and (2.4) shows that the resulting objective function values coincide. ∎

Grothendieck [17] established the following relation between $\mathrm{SDP}(M)$ and the cut norm $\|M\|_{\mathrm{cut}}$.

THEOREM 2.2. *There is a constant $\theta > 1$ such that for all matrices M we have* $\|M\|_{\mathrm{cut}} \leq \mathrm{SDP}(M) \leq \theta \cdot \|M\|_{\mathrm{cut}}$.

Since by Lemma 2.1 $\mathrm{SDP}(M)$ can be stated as a semidefinite program, an optimal solution to $\mathrm{SDP}(M)$ can be approximated in polynomial time within any numerical precision, e.g., via the ellipsoid method [18]. By applying an appropriate rounding procedure to a near-optimal solution to $\mathrm{SDP}(M)$, Alon and Naor [4] obtained the following algorithmic result.

THEOREM 2.3. *There are a constant $\theta' > 0$ and a polynomial time algorithm* ***ApxCutNorm*** *that on input M computes two sets $I, J \subset \mathcal{I}$ such that*

$$\theta' \cdot \|M\|_{\mathrm{cut}} \leq \left| \sum_{i \in I, j \in J} m_{ij} \right| .$$

Alon and Naor presented a randomized algorithm that guarantees an approximation ration $\theta' > 0.56$, and a deterministic one with $\theta' \geq 0.03$. Finally, we need the following dual characterization of SDP. The proof can be found in the next section, Section 2.3.

LEMMA 2.4. *For any symmetric $n \times n$ matrix Q we have*

$$\mathrm{SDP}(Q) = n \cdot \min_{z \in \mathbf{R}^n,\, z \perp \mathbf{1}} \lambda_{\max} \left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes Q - \mathrm{diag}\begin{pmatrix} z \\ z \end{pmatrix} \right].$$

2.3. Proof of Lemma 2.4. The proof of Lemma 2.4 relies on the duality theorem for semidefinite programs. For a symmetric $n \times n$ matrix Q set $\mathcal{Q} = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes Q$. Furthermore, let

$$\mathrm{DSDP}(Q) = \min \langle \mathbf{1}, y\rangle \ \text{ s.t. } \mathcal{Q} \leq \mathrm{diag}(y), \quad y \in \mathbf{R}^{2n}.$$

LEMMA 2.5. *We have* $\mathrm{SDP}(Q) = \mathrm{DSDP}(Q)$.

Proof. By Lemma 2.1 we can rewrite the vector program $\mathrm{SDP}(Q)$ in the standard form of a semidefinite program:

$$\mathrm{SDP}(Q) = \max \langle \mathcal{Q}, X\rangle \ \text{ s.t. } \mathrm{diag}(X) = \mathbf{1},\ X \geq 0,\ X \in \mathbf{R}^{(2n)\times(2n)}.$$

Since $\mathrm{DSDP}(Q)$ is the dual of $\mathrm{SDP}(Q)$, the lemma follows directly from the SDP duality theorem as stated in [20, Corollary 2.2.6]. ▯

To infer Lemma 2.4, we shall simplify DSDP and reformulate this semidefinite program as an eigenvalue minimization problem. First, we show that it suffices to optimize over $y' \in \mathbf{R}^n$ rather than $y \in \mathbf{R}^{2n}$.

LEMMA 2.6. *Let* $\mathrm{DSDP}'(Q) = \min 2\langle \mathbf{1}, y'\rangle$ *s.t.* $\mathcal{Q} \leq \mathrm{diag}(\binom{1}{1} \otimes y')$, $y' \in \mathbf{R}^n$. *Then* $\mathrm{DSDP}(Q) = \mathrm{DSDP}'(Q)$.

Proof. Since for any feasible solution y' to $\mathrm{DSDP}'(Q)$ the vector $y = \binom{1}{1} \otimes y'$ is a feasible solution to $\mathrm{DSDP}(Q)$, we conclude that $\mathrm{DSDP}(Q) \leq \mathrm{DSDP}'(Q)$. Thus, we just need to establish the converse inequality $\mathrm{DSDP}'(Q) \leq \mathrm{DSDP}(Q)$.

To this end, let $\mathcal{F}(Q) \subset \mathbf{R}^{2n}$ signify the set of all feasible solutions y to $\mathrm{DSDP}(Q)$. We shall prove that $\mathcal{F}(Q)$ is closed under the linear operator

$$\mathcal{I} : \mathbf{R}^{2n} \to \mathbf{R}^{2n}, \quad (y_1, \ldots, y_n, y_{n+1}, \ldots, y_{2n}) \mapsto (y_{n+1}, \ldots, y_{2n}, y_1, \ldots, y_n),$$

i.e., $\mathcal{I}(\mathcal{F}(Q)) \subset \mathcal{F}(Q)$; note that $\mathcal{I}$ just swaps the first and the last n entries of y. To see that this implies the assertion, consider an optimal solution $y = (y_i)_{1\leq i\leq 2n} \in \mathcal{F}(Q)$. Then $\frac{1}{2}(y + \mathcal{I}y) \in \mathcal{F}(Q)$, because $\mathcal{F}(Q)$ is convex. Now, let $y' = (y'_i)_{1\leq i\leq n}$ be the projection of $\frac{1}{2}(y + \mathcal{I}y)$ onto the first n coordinates. Since $\frac{1}{2}(y + \mathcal{I}y)$ is a fixed point of $\mathcal{I}$, we have $\frac{1}{2}(y + \mathcal{I}y) = \binom{1}{1} \otimes y'$. Hence, the fact that $\frac{1}{2}(y + \mathcal{I}y)$ is feasible for $\mathrm{DSDP}(Q)$ implies that y' is feasible for $\mathrm{DSDP}'(Q)$. Thus, we conclude that $\mathrm{DSDP}'(Q) \leq 2\langle \mathbf{1}, y'\rangle = \langle \mathbf{1}, y\rangle = \mathrm{DSDP}(Q)$.

To show that $\mathcal{F}(Q)$ is closed under $\mathcal{I}$, consider a vector $y \in \mathcal{F}(Q)$. Since $\mathrm{diag}(y) - \mathcal{Q}$ is positive semidefinite, we have

$$\forall \eta \in \mathbf{R}^{2n} : \langle (\mathrm{diag}(y) - \mathcal{Q})\eta, \eta\rangle \geq 0. \tag{2.5}$$

The objective is to show that $\mathrm{diag}(\mathcal{I}y) - \mathcal{Q}$ is positive semidefinite, i.e.,

$$\forall \xi \in \mathbf{R}^{2n} : \langle (\mathrm{diag}(\mathcal{I}y) - \mathcal{Q})\xi, \xi\rangle \geq 0. \tag{2.6}$$

To derive (2.6) from (2.5), we decompose y into its two halfs $y = \binom{u}{v}$ $(u, v \in \mathbf{R}^n)$. Then $\mathcal{I}y = \binom{v}{u}$. Moreover, let $\xi = \binom{\alpha}{\beta} \in \mathbf{R}^{2n}$ be any vector, and set $\eta = \mathcal{I}\xi = \binom{\beta}{\alpha}$. We obtain

$$\begin{aligned} \langle (\mathrm{diag}(\mathcal{I}y) - \mathcal{Q})\xi, \xi\rangle &= \langle \mathrm{diag}(v)\alpha, \alpha\rangle + \langle \mathrm{diag}(u)\beta, \beta\rangle - \frac{\langle Q\alpha, \beta\rangle + \langle Q\beta, \alpha\rangle}{2} \\ &= \langle (\mathrm{diag}(y) - \mathcal{Q})\eta, \eta\rangle \overset{(2.5)}{\geq} 0, \end{aligned}$$

thereby proving (2.6). □

Proof of Lemma 2.4. Let

$$\mathrm{DSDP}''(Q) = n \cdot \min_{z\in\mathbf{R}^n,\, z\perp\mathbf{1}} \lambda_{\max}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes Q + \mathrm{diag}\begin{pmatrix}1\\1\end{pmatrix} \otimes z\right].$$

By Lemmas 2.5 and 2.6, it suffices to prove that $\mathrm{DSDP}'(Q) = \mathrm{DSDP}''(Q)$.

To see that $\mathrm{DSDP}''(Q) \le \mathrm{DSDP}'(Q)$, let y' be an optimal solution to $\mathrm{DSDP}'(Q)$. Let $\lambda = n^{-1}\langle \mathbf{1}, y'\rangle$ and $z = 2(\lambda\mathbf{1} - y')$. Then $\langle z, \mathbf{1}\rangle = 2(n\lambda - \langle \mathbf{1}, y'\rangle) = 0$, whence z is a feasible solution to $\mathrm{DSDP}''(Q)$. Furthermore, as y' is a feasible solution to $\mathrm{DSDP}'(Q)$, we have

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes Q = 2\mathcal{Q} \le 2\mathrm{diag}\begin{pmatrix}1\\1\end{pmatrix} \otimes y' = 2\lambda\boldsymbol{E} - \mathrm{diag}\begin{pmatrix}1\\1\end{pmatrix} \otimes z,$$

where $\boldsymbol{E}$ is the identity matrix. Hence, the matrix $2\lambda\boldsymbol{E} - \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes Q - \mathrm{diag}\binom{1}{1} \otimes z$ is positive semidefinite. This implies that all eigenvalues of $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes Q + \mathrm{diag}\binom{1}{1} \otimes z$ are bounded by 2λ, i.e., $\lambda_{\max}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes Q + \mathrm{diag}\binom{1}{1} \otimes z\right] \le 2\lambda$. As a consequence,

$$\begin{aligned}\mathrm{DSDP}''(Q) &\le n\lambda_{\max}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes Q + \mathrm{diag}\begin{pmatrix}1\\1\end{pmatrix} \otimes z\right] \\ &\le 2n\lambda = 2\langle \mathbf{1}, y'\rangle = \mathrm{DSDP}'(Q).\end{aligned}$$

Conversely, consider an optimal solution z to $\mathrm{DSDP}''(Q)$. Set

$$\mu = \lambda_{\max}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes Q + \mathrm{diag}\begin{pmatrix}1\\1\end{pmatrix} \otimes z\right] = n^{-1}\mathrm{DSDP}''(Q), \quad y' = \frac{1}{2}(\mu\mathbf{1} - z).$$

Since all eigenvalues of $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes Q + \mathrm{diag}\binom{1}{1} \otimes z$ are bounded by μ, the matrix $\mu\boldsymbol{E} - \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes Q - \mathrm{diag}\binom{1}{1} \otimes z$ is positive semidefinite, i.e., $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes Q \le \mu\boldsymbol{E} - \mathrm{diag}\binom{1}{1} \otimes z$. Therefore,

$$\mathcal{Q} = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes Q \le \frac{1}{2}\left(\mu\boldsymbol{E} - \mathrm{diag}\begin{pmatrix}1\\1\end{pmatrix} \otimes z\right) = \mathrm{diag}\begin{pmatrix}1\\1\end{pmatrix} \otimes y'.$$

Hence, y' is a feasible solution to $\mathrm{DSDP}'(Q)$. Furthermore, since $z \perp \mathbf{1}$ we obtain

$$\mathrm{DSDP}'(Q) \le 2\langle \mathbf{1}, y'\rangle = \mu n = \mathrm{DSDP}''(Q),$$

as desired. □

3. Quasi-Randomness: Proof of Theorem 1.1.

3.1. From Essential Eigenvalue Separation to Low Discrepancy. We prove the first part of Theorem 1.1. Suppose that $G = (V, E)$ is a graph that admits a set $W \subset V$ of volume $\mathrm{vol}(W) \ge (1-\varepsilon)\mathrm{vol}(V)$ such that the eigenvalues of the minor L_W of the normalized Laplacian satisfy

$$1 - \varepsilon \le \lambda_2[L_W] \le \lambda_{\max}[L_W] \le 1 + \varepsilon. \tag{3.1}$$

We may assume without loss of generality that $\varepsilon < 0.01$. Our goal is to show that G has $\mathrm{Disc}(10\sqrt{\varepsilon})$.

Let $\Delta = (\sqrt{d_v})_{v \in W} \in \mathbf{R}^W$. Hence, Δ is a real vector indexed by the elements of W. Moreover, let $\mathcal{L}_W$ denote the matrix whose vw'th entry is $(d_v d_w)^{-\frac{1}{2}}$ if v, w are adjacent, and 0 otherwise ($v, w \in W$), so that $L_W = \boldsymbol{E} - \mathcal{L}_W$. Further, let $\mathcal{M}_W = \mathrm{vol}(V)^{-1}\Delta\Delta^T - \mathcal{L}_W$. Then for all unit vectors $\xi \perp \Delta$ we have

$$L_W\xi - \xi = -\mathcal{L}_W\xi = \mathcal{M}_W\xi. \tag{3.2}$$

Moreover, for all $S \subset W$

$$|\langle \mathcal{M}_W \Delta_S, \Delta_S\rangle| = \left| \frac{\mathrm{vol}(S)^2}{\mathrm{vol}(V)} - e(S) \right|. \tag{3.3}$$

The key step of the proof is to derive the following bound on the operator norm of $\mathcal{M}_W$.

Lemma 3.1. *We have $\|\mathcal{M}_W\| \le 10\sqrt{\varepsilon}$.*

If it were the case that $W = V$, then Lemma 3.1 would be immediate. For if $W = V$, then Δ is an eigenvector of $L = L_W$ with eigenvalue 0. Hence, the definition $\mathcal{M}_W = \|\Delta\|^{-2}\Delta\Delta^T - \boldsymbol{E} + L_W$ ensures that $\mathcal{M}_W\Delta = 0$. Moreover, for all $\xi \perp \Delta$ we have $\mathcal{M}_W\xi = (L_W - \boldsymbol{E})\xi$, whence (3.1) implies that $\|\mathcal{M}_W\| \le \max\{|\lambda_2[L_W] - 1|, |\lambda_{\max}[L_W] - 1|\} \le \varepsilon$.

But of course generally W is a proper subset of V. In this case Δ is not necessarily an eigenvector of L_W. In fact, the smallest eigenvalue of L_W may be strictly positive. In order to prove Lemma 3.1 we will investigate the eigenvector ζ of L_W with the smallest eigenvalue $\lambda_1[L_W]$ and show that it is "close" to Δ. Then, we will use (3.1) to derive the desired bound on $\|\mathcal{M}_W\|$.

Proof of Lemma 3.1. Let ζ be a unit length eigenvector of L_W with eigenvalue $\lambda_1[L_W]$. There is a decomposition $\Delta = \|\Delta\| \cdot (s\zeta + t\chi)$, where $s^2 + t^2 = 1$ and $\chi \perp \zeta$ is a unit vector. Since $\langle L_W\Delta, \Delta\rangle = e(W, V \setminus W) \le \mathrm{vol}(V \setminus W) \le \varepsilon\mathrm{vol}(V)$ and $\|\Delta\|^2 = \mathrm{vol}(W) \ge (1-\varepsilon)\mathrm{vol}(V) \ge 0.99\mathrm{vol}(V)$, we have

$$2\varepsilon \ge \|\Delta\|^{-2}\langle L_W\Delta, \Delta\rangle = s^2\langle L_W\zeta, \zeta\rangle + t^2\langle L_W\chi, \chi\rangle. \tag{3.4}$$

Because χ is perpendicular to the eigenvector ζ with eigenvalue $\lambda_1[L_W]$, Courant-Fischer (2.1) and (3.1) yield $\langle L_W\chi, \chi\rangle \ge \lambda_2[L_W] \ge \frac{1}{2}$. Hence, (3.4) implies $2\varepsilon \ge t^2/2$. Consequently,

$$t^2 \le 4\varepsilon, \quad \text{and thus} \quad s^2 \ge 1 - 4\varepsilon. \tag{3.5}$$

Now, let $\xi \perp \Delta$ be a unit vector, and decompose $\xi = x\zeta + y\eta$, where $\eta \perp \zeta$ is a unit vector. Because $\zeta = s^{-1}\left(\frac{\Delta}{\|\Delta\|} - t\chi\right)$, we have $x = \langle\zeta, \xi\rangle = s^{-1}\left\langle\frac{\Delta}{\|\Delta\|}, \xi\right\rangle - \frac{t}{s}\langle\chi, \xi\rangle = -\frac{t}{s}\langle\chi, \xi\rangle$. Hence, (3.5) implies $x^2 \le 5\varepsilon$ and $y^2 \ge 1 - 5\varepsilon$. Combining these two estimates with (3.1) and (3.2), we conclude that $\|\mathcal{M}_W\xi\| = \|L_W\xi - \xi\| \le x(1 - \lambda_1[L_W]) + y\|L_W\eta - \eta\| \le 3\sqrt{\varepsilon}$. Hence, we have established that

$$\sup_{0 \ne \xi \perp \Delta} \frac{\|\mathcal{M}_W\xi\|}{\|\xi\|} \le 3\sqrt{\varepsilon}. \tag{3.6}$$

Furthermore, since $\|\Delta\|^2 = \mathrm{vol}(W)$, (3.3) implies

$$\begin{aligned}\frac{|\langle \mathcal{M}_W \Delta, \Delta\rangle|}{\|\Delta\|^2} &= \left|\frac{\mathrm{vol}(W)}{\mathrm{vol}(V)} - \frac{e(W)}{\mathrm{vol}(W)}\right| \\ &\le \left|\frac{\mathrm{vol}(W)}{\mathrm{vol}(V)} - \frac{e(W)}{\mathrm{vol}(V)}\right| + \left|\frac{e(W)}{\mathrm{vol}(W)} - \frac{e(W)}{\mathrm{vol}(V)}\right| \\ &= \frac{e(W, V\setminus W)}{\mathrm{vol}(V)} + \frac{e(W)(\mathrm{vol}(V) - \mathrm{vol}(W))}{\mathrm{vol}(V)\mathrm{vol}(W)} \\ &\le \frac{e(W, V\setminus W)}{\mathrm{vol}(V)} + \frac{\mathrm{vol}(V\setminus W)}{\mathrm{vol}(V)} \le \frac{2\mathrm{vol}(V\setminus W)}{\mathrm{vol}(V)}. \end{aligned} \tag{3.7}$$

As we are assuming that $\mathrm{vol}(W) \ge (1-\varepsilon)\mathrm{vol}(V)$, we obtain $\|\Delta\|^{-2}\,|\langle \mathcal{M}_W\Delta, \Delta\rangle| \le 2\varepsilon$. Finally, combining this last estimate with (3.6), we conclude that $\|\mathcal{M}_W\| \le 10\sqrt{\varepsilon}$. □

Lemma 3.1 easily implies that G has $\mathrm{Disc}(10\sqrt{\varepsilon})$. For let $R \subset V$ be arbitrary, set $S = R \cap W$, and let $T = R \setminus W$. Since $\|\Delta_S\|^2 = \mathrm{vol}(S) \le \mathrm{vol}(V)$, Lemma 3.1 and (3.3) imply that

$$\left|\frac{\mathrm{vol}(S)^2}{\mathrm{vol}(V)} - e(S)\right| \le \|\mathcal{M}_W\| \cdot \|\Delta_S\|^2 \le 10\sqrt{\varepsilon}\mathrm{vol}(V). \tag{3.8}$$

Furthermore, as $\mathrm{vol}(W) \ge (1-\varepsilon)\mathrm{vol}(V)$,

$$e(R) - e(S) \le e(T) + 2e(S,T) \le 2\mathrm{vol}(T) \le 2\mathrm{vol}(V\setminus W) \le 2\varepsilon\mathrm{vol}(V), \text{ and}$$

$$\begin{aligned}\frac{\mathrm{vol}(R)^2 - \mathrm{vol}(S)^2}{2\mathrm{vol}(V)} &\le \frac{\mathrm{vol}(T)^2}{2\mathrm{vol}(V)} + \frac{\mathrm{vol}(S)\mathrm{vol}(T)}{\mathrm{vol}(V)} \\ &\le \frac{\mathrm{vol}(V\setminus W)^2}{2\mathrm{vol}(V)} + \mathrm{vol}(V\setminus W) \le 2\varepsilon\mathrm{vol}(V).\end{aligned}$$

Combining these two estimates with (3.8), we see that $\left|\frac{\mathrm{vol}(R)^2}{\mathrm{vol}(V)} - e(R)\right| < 20\sqrt{\varepsilon}\mathrm{vol}(V)$, i.e., G satisfies $\mathrm{Disc}(10\sqrt{\varepsilon})$.

3.2. From Low Discrepancy to Essential Eigenvalue Separation. In this section we establish the second part of Theorem 1.1. Let θ denote the constant from Theorem 2.2 and set $\gamma = 10^{-6}/\theta$. Assume that $G = (V, E)$ is a graph that has $\mathrm{Disc}(\gamma\varepsilon^2)$ for some $\varepsilon < 0.001$. In addition, we may assume without loss of generality that G has no isolated vertices. Let d_v denote the degree of $v \in V$, let $n = |V|$, and set $\bar{d} = \mathrm{vol}(V)/n = \sum_{v\in V} d_v/n$. Our goal is to show that G has ess-Eig(ε). To this end, we introduce an additional property.

Cut(δ): We say G has Cut(δ) if the matrix $M = (m_{vw})_{v,w\in V}$ with entries

$$m_{vw} = \frac{d_v d_w}{\mathrm{vol}(V)} - e(v,w)$$

has cut norm $\|M\|_{\mathrm{cut}} < \delta \cdot \mathrm{vol}(V)$; here $e(v,w) = 1$ if $\{v,w\} \in E$ and $e(v,w) = 0$ otherwise.

PROPOSITION 3.2. *For any $\delta > 0$ the following is true: if G satisfies* $\mathrm{Disc}(0.01\delta)$*, then G satisfies* Cut(δ).

Proof. Suppose that $G = (V, E)$ has $\mathrm{Disc}(0.01\delta)$. We shall prove below that for any two $S, T \subset V$

$$|\langle M\mathbf{1}_S, \mathbf{1}_T\rangle| \le 0.03\delta\mathrm{vol}(V) \text{ if } S \cap T = \emptyset, \tag{3.9}$$

$$|\langle M\mathbf{1}_S, \mathbf{1}_T\rangle| \le 0.02\delta\mathrm{vol}(V) \text{ if } S = T. \tag{3.10}$$

To see that (3.9) and (3.10) imply the assertion, consider two arbitrary subsets $X, Y \subset V$. Letting $Z = X \cap Y$ and combining (3.9) and (3.10), we obtain

$$\begin{aligned}|\langle M\mathbf{1}_X, \mathbf{1}_Y\rangle| &\leq \left|\langle M\mathbf{1}_{X\setminus Z}, \mathbf{1}_{Y\setminus Z}\rangle\right| + \left|\langle M\mathbf{1}_Z, \mathbf{1}_{Y\setminus Z}\rangle\right| + \left|\langle M\mathbf{1}_Z, \mathbf{1}_{X\setminus Z}\rangle\right| \\ &\quad +2\left|\langle M\mathbf{1}_Z, \mathbf{1}_Z\rangle\right| \\ &\leq \delta \mathrm{vol}(V).\end{aligned}$$

Since this bound holds for any X, Y, we conclude that $\|M\|_{\text{cut}} \leq \delta\mathrm{vol}(V)$.

To prove (3.9), we note that $\mathrm{Disc}(0.01\delta)$ implies for disjoint sets S and T

$$\left|e(S) - \frac{\mathrm{vol}(S)^2}{\mathrm{vol}(V)}\right| \leq 0.02\delta\mathrm{vol}(V), \qquad \left|e(T) - \frac{\mathrm{vol}(T)^2}{\mathrm{vol}(V)}\right| \leq 0.02\delta\mathrm{vol}(V), \tag{3.11}$$

$$\left|e(S\cup T) - \frac{(\mathrm{vol}(S)+\mathrm{vol}(T))^2}{\mathrm{vol}(V)}\right| \leq 0.02\delta\mathrm{vol}(V). \tag{3.12}$$

If S and T are disjoint, (3.11)–(3.12) yield

$$\begin{aligned}|\langle M\mathbf{1}_S, \mathbf{1}_T\rangle| &= \left|e(S,T) - \frac{\mathrm{vol}(S)\mathrm{vol}(T)}{\mathrm{vol}(V)}\right| \\ &= \frac{1}{2}\left|e(S\cup T) - e(S) - e(T) - \frac{(\mathrm{vol}(S)+\mathrm{vol}(T))^2 - \mathrm{vol}(S)^2 - \mathrm{vol}(T)^2}{\mathrm{vol}(V)}\right| \\ &\leq \frac{1}{2}\left|e(S) - \frac{\mathrm{vol}(S)^2}{\mathrm{vol}(V)}\right| \\ &\qquad + \frac{1}{2}\left|e(T) - \frac{\mathrm{vol}(T)^2}{\mathrm{vol}(V)}\right| + \frac{1}{2}\left|e(S\cup T) - \frac{(\mathrm{vol}(S)+\mathrm{vol}(T))^2}{\mathrm{vol}(V)}\right| \\ &\leq 0.03\delta\mathrm{vol}(V),\end{aligned}$$

whence (3.9) follows. Finally, as

$$|\langle M\mathbf{1}_S, \mathbf{1}_S\rangle| = \left|e(S) - \frac{\mathrm{vol}(S)^2}{\mathrm{vol}(V)}\right|,$$

(3.10) follows from (3.11). □

Let $D = \mathrm{diag}(d_v)_{v\in V}$ be the matrix with the vertex degrees on the diagonal. Let $\mathcal{M} = D^{-\frac{1}{2}}MD^{-\frac{1}{2}}$. Then the vw'th the entry of $\mathcal{M}$ is $\frac{\sqrt{d_v d_w}}{\mathrm{vol}(V)} - (d_v d_w)^{-1/2}$ if v, w are adjacent, and $\frac{\sqrt{d_v d_w}}{\mathrm{vol}(V)}$ otherwise. Establishing the following lemma is the key step.

Lemma 3.3. *Suppose that* $\mathrm{SDP}(M) < \varepsilon^2\mathrm{vol}(V)/64$. *Then there exists a subset* $W \subset V$ *of volume* $\mathrm{vol}(W) \geq (1-\varepsilon)\cdot\mathrm{vol}(V)$ *such that* $\|\mathcal{M}_W\| < \varepsilon$.

Proof. Recall that $\bar{d} = \mathrm{vol}(V)/n$. Lemma 2.4 implies that there is a vector $\mathbf{1} \perp z \in \mathbf{R}^V$ such that

$$\lambda_{\max}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0\end{pmatrix} \otimes M - \mathrm{diag}\begin{pmatrix} z \\ z\end{pmatrix}\right] = \mathrm{SDP}(M)/n < \varepsilon^2\bar{d}/64. \tag{3.13}$$

Basically W is going to be the set of all v such that $|z_v|$ is small (and such that d_v is not too small). On the minor induced on $W \times W$ the diagonal matrix $\mathrm{diag}\binom{z}{z}$ has little effect, and thus (3.13) will imply the desired bound on $\|\mathcal{M}_W\|$. To carry out the details we need to define W precisely, bound $\|\mathcal{M}_W\|$, and prove that $\mathrm{vol}(W) \geq (1-\varepsilon)\mathrm{vol}(V)$.

Let $y = D^{-1}z$ and $U = \{v \in V : d_v > \varepsilon\bar{d}/8\}$. Let $y' = (y_v)_{v\in U}$ and $z' = (z_v)_{v\in U}$. Since all entries of the restricted diagonal matrix D_U exceed $\varepsilon\bar{d}/8$, we have

$$\begin{aligned}
&\lambda_{\max}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \mathcal{M}_U - \mathrm{diag}\begin{pmatrix} y' \\ y' \end{pmatrix}\right] \\
&\quad = \lambda_{\max}\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes D_U^{-\frac{1}{2}} \cdot \left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes M_U - \mathrm{diag}\begin{pmatrix} z' \\ z' \end{pmatrix}\right] \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes D_U^{-\frac{1}{2}}\right] \\
&\quad \le 8\left(\varepsilon\bar{d}\right)^{-1}\lambda_{\max}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes M_U - \mathrm{diag}\begin{pmatrix} z' \\ z' \end{pmatrix}\right] \\
&\quad \le 8(\varepsilon\bar{d})^{-1}\lambda_{\max}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes M - \mathrm{diag}\begin{pmatrix} z \\ z \end{pmatrix}\right] \overset{(3.13)}{<} \varepsilon/8. \qquad (3.14)
\end{aligned}$$

Let $W = \{v \in U : |y_v| < \varepsilon/8\}$ and let $y'' = (y_v)_{v\in W}$. Then $\|\mathrm{diag}\binom{y''}{y''}\| < \varepsilon/8$, because the norm of a diagonal matrix equals the largest absolute value of an entry on the diagonal. Therefore, (3.14) yields

$$\begin{aligned}
\lambda_{\max}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \mathcal{M}_W\right] &\le \lambda_{\max}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \mathcal{M}_W - \mathrm{diag}\begin{pmatrix} y'' \\ y'' \end{pmatrix}\right] + \left\|\mathrm{diag}\begin{pmatrix} y'' \\ y'' \end{pmatrix}\right\| \\
&\le \lambda_{\max}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \mathcal{M}_U - \mathrm{diag}\begin{pmatrix} y' \\ y' \end{pmatrix}\right] + \left\|\mathrm{diag}\begin{pmatrix} y'' \\ y'' \end{pmatrix}\right\| \\
&\le \varepsilon/4. \qquad (3.15)
\end{aligned}$$

Further, (3.15) implies that $\|\mathcal{M}_W\| < \varepsilon$. To see this, consider a pair $\xi, \eta \in \mathbf{R}^W$ of unit vectors. Then (3.15) and Courant-Fischer (2.1) yield

$$\begin{aligned}
\varepsilon/2 \ge 2\lambda_{\max}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \mathcal{M}_W\right] &\ge \left\langle \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \mathcal{M}_W \cdot \begin{pmatrix} \xi \\ \eta \end{pmatrix}, \begin{pmatrix} \xi \\ \eta \end{pmatrix} \right\rangle \\
&= \left\langle \begin{pmatrix} \mathcal{M}_W\eta \\ \mathcal{M}_W\xi \end{pmatrix}, \begin{pmatrix} \xi \\ \eta \end{pmatrix} \right\rangle = \langle \mathcal{M}_W\eta, \xi\rangle + \langle \mathcal{M}_W\xi, \eta\rangle \\
&= 2\langle \mathcal{M}_W\xi, \eta\rangle \qquad \text{[because } \mathcal{M}_W \text{ is symmetric]}.
\end{aligned}$$

Since this holds for any pair ξ, η, we conclude that $\|\mathcal{M}_W\| \le \varepsilon/4 < \varepsilon$.

Finally, we need to show that $\mathrm{vol}(W)$ is large. To this end, we consider the set $S = \{v \in V : z_v < 0\}$. Since $\mathrm{vol}(V) = \bar{d}n \ge \bar{d}|S|$, we have

$$\begin{aligned}
\frac{\varepsilon^2\mathrm{vol}(V)}{32} \ge \frac{\varepsilon^2\bar{d}|S|}{32} &= \frac{\varepsilon^2\bar{d}}{64} \cdot \left\|\begin{pmatrix} \mathbf{1}_S \\ \mathbf{1}_S \end{pmatrix}\right\|^2 \\
&\ge \lambda_{\max}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes M - \mathrm{diag}\begin{pmatrix} z \\ z \end{pmatrix}\right] \cdot \left\|\begin{pmatrix} \mathbf{1}_S \\ \mathbf{1}_S \end{pmatrix}\right\|^2 \qquad \text{[due to (3.13)]} \\
&\ge \left\langle \left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes M - \mathrm{diag}\begin{pmatrix} z \\ z \end{pmatrix}\right] \cdot \begin{pmatrix} \mathbf{1}_S \\ \mathbf{1}_S \end{pmatrix}, \begin{pmatrix} \mathbf{1}_S \\ \mathbf{1}_S \end{pmatrix} \right\rangle \qquad \text{[by (2.1)]} \\
&= 2\langle M\mathbf{1}_S, \mathbf{1}_S\rangle - 2\sum_{v\in S} z_v. \qquad (3.16)
\end{aligned}$$

Further, Theorem 2.2 implies $|\langle M\mathbf{1}_S, \mathbf{1}_S\rangle| \le \|M\|_{\mathrm{cut}} \le \mathrm{SDP}(M) \le \varepsilon^2\mathrm{vol}(V)/64$. Inserting this into (3.16) and recalling that $z_v < 0$ for all $v \in S$, we conclude that $\sum_{v\in S}|z_v| \le \varepsilon^2\mathrm{vol}(V)/32$. Since $z \perp \mathbf{1}$, this actually implies $\sum_{v\in V}|z_v| \le \varepsilon^2\mathrm{vol}(V)/16$.

As $z = Dy$ and $|y_v| > \varepsilon/8$ for all $v \in U \setminus W$, we obtain

$$\varepsilon\mathrm{vol}(U \setminus W)/8 \le \sum_{v\in U\setminus W} d_v|y_v| = \sum_{v\in U\setminus W} |z_v| \le \varepsilon^2\mathrm{vol}(V)/8. \tag{3.17}$$

Furthermore, by the definition of U we have

$$\mathrm{vol}(V \setminus U) \le \varepsilon\bar{d}n/8 \le \varepsilon\mathrm{vol}(V)/8. \tag{3.18}$$

Combining (3.17) and (3.18), we see that $\mathrm{vol}(V \setminus W) \le \varepsilon\mathrm{vol}(V)$, which implies $\mathrm{vol}(W) \ge (1-\varepsilon)\mathrm{vol}(V)$. □

Finally, we show how Lemma 3.3 implies that G has ess-Eig(ε). Assume that G has Disc($\gamma\varepsilon^2$). By Proposition 3.2 this implies that G satisfies Cut($100\gamma\varepsilon^2$). Hence, Theorem 2.2 shows $\mathrm{SDP}(M) \le \beta\varepsilon^2\mathrm{vol}(V)$ for some $0 < \beta \le 100\theta\gamma$. Thus, by Lemma 3.3 and our choice of γ there is a set W such that $\mathrm{vol}(W) \ge (1-\varepsilon/10)\mathrm{vol}(V)$ and $\|\mathcal{M}_W\| < \varepsilon/10$. Furthermore, $\mathcal{M}_W$ relates to the minor L_W of the Laplacian as follows. Let $\mathcal{L}_W = \boldsymbol{E} - L_W$ be the matrix whose vw'th entry is $(d_v d_w)^{-1/2}$ if $v, w \in W$ are adjacent, and 0 otherwise. Moreover, let $\Delta = (\sqrt{d_v})_{v\in W} \in \mathbf{R}^W$. Then $\mathcal{M}_W = \mathrm{vol}(V)^{-1}\Delta\Delta^T - \mathcal{L}_W$. Therefore, for all unit vectors $\xi \perp \Delta$ we have

$$|\langle L_W\xi, \xi\rangle - 1| = |\langle \mathcal{L}_W\xi, \xi\rangle| = |\langle \mathcal{M}_W\xi, \xi\rangle| \le \|\mathcal{M}_W\| < \varepsilon/10. \tag{3.19}$$

Combining (3.19) with Courant-Fischer (2.1), we obtain

$$\lambda_2[L_W] = \max_{0\ne\zeta\in\mathbf{R}^W} \min_{\xi\perp\zeta,\,\|\xi\|=1} \langle L_W\xi,\xi\rangle \ge \min_{\xi\perp\Delta,\,\|\xi\|=1} \langle L_W\xi,\xi\rangle \ge 1-\varepsilon. \tag{3.20}$$

To bound $\lambda_{\max}[L_W]$ as well, we need to compute $\|L_W\Delta\|^2$. To this end, recall that the row of L_W corresponding to a vertex $v \in V$ contains a one at position v. For $w \ne v$ the entry is $-(d_v d_w)^{-\frac{1}{2}}$ if v and w are adjacent, and 0 otherwise. Hence, the v-entry of the vector $L_W\Delta$ equals

$$\Delta_v - \sum_{w\in W:\{v,w\}\in E} \frac{\Delta_w}{\sqrt{d_v d_w}} = \sqrt{d_v} - \frac{e(v,W)}{\sqrt{d_v}} = \frac{d_v - e(v,W)}{\sqrt{d_v}}.$$

Since $\|\Delta\|^2 = \sum_{v\in W} d_v = \mathrm{vol}(W) \ge (1-\varepsilon/10)\mathrm{vol}(V)$, we obtain

$$\begin{aligned}\frac{\|L_W\Delta\|^2}{\|\Delta\|^2} &= \sum_{v\in W} \frac{(e(v,W)-d_v)^2}{d_v\cdot\mathrm{vol}(W)} \\ &\le \frac{1}{1-\varepsilon/10}\sum_{v\in W}\frac{d_v - e(v,W)}{\mathrm{vol}(V)} \le \frac{2\mathrm{vol}(V\setminus W)}{\mathrm{vol}(V)} < \varepsilon/5. \end{aligned}\tag{3.21}$$

Further, decomposing any unit vector $\eta \in \mathbf{R}^W$ as $\eta = \alpha\|\Delta\|^{-1}\Delta + \beta\xi$ with a unit vector $\xi \perp \Delta$ and $\alpha^2 + \beta^2 = 1$, we get

$$\begin{aligned}\langle L_W\eta,\eta\rangle &= \left\langle L_W\left(\alpha\|\Delta\|^{-1}\Delta+\beta\xi\right), \alpha\|\Delta\|^{-1}\Delta+\beta\xi\right\rangle \\ &= \frac{\alpha^2}{\|\Delta\|^2}\cdot\langle L_W\Delta,\Delta\rangle + \frac{\alpha\beta}{\|\Delta\|}\cdot\langle L_W\Delta,\xi\rangle + \frac{\alpha\beta}{\|\Delta\|}\cdot\langle L_W\xi,\Delta\rangle + \beta^2\langle L_W\xi,\xi\rangle \\ &= \frac{\alpha^2}{\|\Delta\|^2}\cdot\langle L_W\Delta,\Delta\rangle + \frac{2\alpha\beta}{\|\Delta\|}\cdot\langle L_W\Delta,\xi\rangle + \beta^2\langle L_W\xi,\xi\rangle,\end{aligned}$$

where the last step follows from the fact that L_W is symmetric. Hence, using (3.19) and (3.21), we get

$$\begin{aligned}\langle L_W\eta,\eta\rangle &\leq \frac{\alpha^2}{\|\Delta\|^2}\cdot\|L_W\Delta\|\cdot\|\Delta\| + \frac{2\alpha\beta}{\|\Delta\|}\cdot\|L_W\Delta\|\cdot\|\xi\| + \beta^2\langle L_W\xi,\xi\rangle\\ &\leq \alpha^2\sqrt{\varepsilon/5} + 2\alpha\beta\sqrt{\varepsilon/5} + \beta^2(1+|\langle L_W\xi,\xi\rangle - 1|)\\ &\leq \sqrt{\varepsilon/5}(\alpha^2+2\alpha\beta) + \beta^2(1+\varepsilon/10)\\ &\leq 3\sqrt{\varepsilon/5}\cdot|\alpha| + (1-\alpha^2)(1+\varepsilon/10).\end{aligned}$$

Differentiating the last expression, we find that the maximum is attained at $\alpha = \frac{3}{2}\sqrt{\varepsilon/5}/(1+\varepsilon/10)$. Plugging this value in, we obtain $\langle L_W\eta,\eta\rangle \leq 1+\varepsilon$. Hence, by Courant-Fischer (2.1), $\lambda_{\max}[L_W] = \max_{\|\eta\|=1}\langle L_W\eta,\eta\rangle \leq 1+\varepsilon$. Thus, (3.20) shows that G has ess-Eig(ε).

4. The Algorithmic Regularity Lemma: Proof of Theorem 1.2. In this section we establish Theorem 1.2. The proof is conceptually similar to Szemerédi's original proof of the "dense" regularity lemma [23] and its adaptation for sparse graphs due to Kohayakawa [21] and Rödl (unpublished). A new aspect here is that we deal with a different (more general) notion of regularity; this requires various technical modifications of the previous arguments. More importantly, we present an *algorithm* for actually computing a regular partition of a sparse graph efficiently.

In order to find a regular partition efficiently, we crucially need an algorithm to check for a given weight distribution $\boldsymbol{D} = (D_v)_{v\in V}$, a given graph G, and a pair (A,B) of vertex sets whether (A,B) is $(\varepsilon,\boldsymbol{D})$-regular. While [3] features a (purely combinatorial) algorithm for this problem in dense graphs, this approach does not work in the sparse case. In Section 4.1 we present an algorithm `Witness` that does. It is based on Grothendieck's inequality and the semidefinite relaxation of the cut norm (see Theorem 2.3). Then, in Section 4.2 we will show how `Witness` can be used to compute a regular partition to establish Theorem 1.2.

Throughout this section, we let $0<\varepsilon<10^{-7}$ be an arbitrarily small but fixed number, and $C\geq 1$ signifies an arbitrarily large but fixed number. In addition, we define a sequence $(t_k)_{k\geq 1}$ by

$$t_1 = \lceil 1/\varepsilon^2\rceil \quad\text{ and }\quad t_{k+1} = \lceil 2200^2C^2t_k^6 2^{t_k}/\varepsilon^{4(k+1)}\rceil. \tag{4.1}$$

Note that due to that choice we have

$$t_{k+1} \geq 2200Ct_k^{2.5}\,. \tag{4.2}$$

Further, let

$$k^* = \lceil 10^6C^2\varepsilon^{-3}\rceil \quad\text{ and }\quad \eta = \min\left\{\frac{\varepsilon^{8k^*}}{12800^2t_{k^*}^6C^4},\frac{1}{t_{k^*}^2}\right\} \tag{4.3}$$

and choose $n_0 = n_0(C,\varepsilon) > 0$ big enough. We let $G=(V,E)$ be a graph on $n=|V|>n_0$ vertices, and let $\boldsymbol{D} = (D_v)_{v\in V}$ be a sequence of rationals with $1\leq D_v\leq n$ for all $v\in V$. We will always assume that G is $(C,\eta,\boldsymbol{D})$-bounded, and that $D(V)\geq\eta^{-1}n$.

4.1. The Procedure `Witness`. The subroutine `Witness` shown in Figure 4.1 is given a graph G, a weight distribution $\boldsymbol{D}$, vertex sets A, B, and a number $\varepsilon>0$. `Witness` either outputs "yes", in which case (A,B) is $(\varepsilon,\boldsymbol{D})$-regular in G, or "no".

ALGORITHM 4.1. $\texttt{Witness}(G, \boldsymbol{D}, A, B, \varepsilon)$

1. Set up the matrix $M = (m_{vw})_{(v,w)\in A\times B}$ with entries

$$m_{vw} = \begin{cases} 1 - \varrho(A,B)D_vD_w & \text{if } v, w \text{ are adjacent in } G, \\ -\varrho(A,B)D_vD_w & \text{otherwise.} \end{cases}$$

   Call $\texttt{ApxCutNorm}(M)$ to compute sets $X \subset A$, $Y \subset B$ such that $|\langle M\mathbf{1}_X, \mathbf{1}_Y\rangle| \geq \frac{3}{100}\|M\|_{\text{cut}}$.
2. If $|\langle M\mathbf{1}_X, \mathbf{1}_Y\rangle| < \frac{3\varepsilon}{100}\frac{D(A)D(B)}{D(V)}$, then return "yes".
3. If not, let $X' = A \setminus X$.
   - If $D(X) \geq \frac{3\varepsilon}{100}D(A)$, then let $X^* = X$.
   - If $D(X) < \frac{3\varepsilon}{100}D(A)$ and $|e(X', Y) - \varrho(A,B)D(X')D(Y)| > \frac{\varepsilon D(A)D(B)}{100D(V)}$, set $X^* = X'$.
   - Otherwise, set $X^* = X \cup X'$.
4. Let $Y' = B \setminus Y$.
   - If $D(Y) \geq \frac{\varepsilon}{200}D(B)$, then let $Y^* = Y$.
   - If $D(Y) < \frac{\varepsilon}{200}D(B)$ and

$$|e(X^*, Y') - \varrho(A,B)D(X^*)D(Y')| > \frac{\varepsilon D(A)D(B)}{200D(V)},$$

     let $Y^* = Y'$.
   - Otherwise, set $Y^* = Y \cup Y'$.
5. Answer "no" and output (X^*, Y^*) as an $(\varepsilon/200, \boldsymbol{D})$-witness.

FIG. 4.1. *The algorithm* $\texttt{Witness}$.

In the latter case the algorithm also produces a "witness of irregularity", i.e., a pair of sets $X^* \subset A$, $Y^* \subset B$ for which the regularity condition (1.2) is violated with ε replaced by $\varepsilon/200$. $\texttt{Witness}$ employs the algorithm $\texttt{ApxCutNorm}$ from Theorem 2.3.

LEMMA 4.2. *Suppose that $A, B \subset V$ are disjoint.*

1. *If $\texttt{Witness}(G, \boldsymbol{D}, A, B, \varepsilon)$ answers 'yes', then the pair (A, B) is $(\varepsilon, \boldsymbol{D})$-regular.*
2. *If the answer is 'no', then (A, B) is not $(\varepsilon/200, \boldsymbol{D})$-regular. In this case $\texttt{Witness}$ outputs an $(\varepsilon/200, \boldsymbol{D})$-witness, i.e., a pair (X^*, Y^*) of subsets $X^* \subset A$, $Y^* \subset B$ such that $D(X^*) \geq \frac{\varepsilon}{200}D(A)$, $D(Y^*) \geq \frac{\varepsilon}{200}D(B)$, and*

$$|e(X^*, Y^*) - \varrho(A,B)D(X^*)D(Y^*)| > \frac{\varepsilon}{200} \cdot \frac{D(A)D(B)}{D(V)}.$$

Moreover, there exist a function f and a polynomial Π such that the running time of $\texttt{Witness}$ is bounded by $f(C, \varepsilon) \cdot \Pi(\langle \boldsymbol{D} \rangle)$.

Proof. Note that for any two subsets $S \subset A$ and $T \subset B$ we have

$$\langle M\mathbf{1}_S, \mathbf{1}_T\rangle = e(S,T) - \varrho(A,B)D(S)D(T).$$

Therefore, if the sets $X \subset A$ and $Y \subset B$ computed by $\texttt{ApxCutNorm}$ are such that

$$|\langle M\mathbf{1}_X, \mathbf{1}_Y\rangle| < \frac{3\varepsilon}{100}\frac{D(A)D(B)}{D(V)}$$

then by Theorem 2.3 we have

$$|e(S,T) - \varrho(A,B)D(S)D(T)| \leq \|M\|_{\text{cut}} \leq \frac{100}{3}|\langle M\mathbf{1}_X, \mathbf{1}_Y\rangle| < \varepsilon\frac{D(A)D(B)}{D(V)}$$

for all $S \subset A$ and $T \subset B$. Thus, if `Witness` answers "yes" then the pair (A, B) is $(\varepsilon, \boldsymbol{D})$-regular.

On the other hand, if `ApxCutNorm` yields sets X, Y such that $\langle M\mathbf{1}_X, \mathbf{1}_Y\rangle \geq \frac{3\varepsilon}{100}\frac{D(A)D(B)}{D(V)}$ then `Witness` has to guarantee that the output pair (X^*, Y^*) is an $(\varepsilon/200, \boldsymbol{D})$-witness.

Indeed, if $D(X) \geq \frac{3\varepsilon}{100}D(A)$ and $D(Y) \geq \frac{\varepsilon}{200}D(B)$ then (X, Y) actually is an $(\varepsilon/200, \boldsymbol{D})$-witness. However, as `ApxCutNorm` does not guarantee any lower bound on $D(X)$ and $D(Y)$ let us assume first that $D(X) < \frac{3\varepsilon}{100}D(A)$ and $D(Y) \geq \frac{\varepsilon}{200}D(B)$. Then Step 3 of `Witness` sets $X' = A \setminus X$. We have $D(X') \geq \frac{3}{100}D(A)$. If X' itself satisfies $|e(X', Y) - \varrho(A, B)D(X')D(Y)| > \frac{\varepsilon D(A)D(B)}{100D(V)}$ then (X', Y) obviously is an $(\varepsilon/200, \boldsymbol{D})$-witness. Otherwise, by the triangle inequality, we deduce

$$\left| e(X \cup X', Y) - e(A, B)\frac{D(X \cup X')D(Y)}{D(A)D(B)} \right| \geq \frac{2\varepsilon}{100}\frac{D(A)D(B)}{D(V)}$$

and thus, $(X \cup X', Y)$ is an $(\varepsilon/200, \boldsymbol{D})$-witness.

In the case $D(X) < \frac{3\varepsilon}{100}D(A)$ and $D(Y) < \frac{\varepsilon}{200}D(B)$ we simply repeat the argument for Y, and hence `Witness` outputs an $(\varepsilon/200, \boldsymbol{D})$-witness for (A, B).

The running time of `Witness` is dominated by Step 1, i.e., the execution of `ApxCutNorm`. By Theorem 2.3 the running time of `ApxCutNorm` is polynomial in the encoding length of the input matrix. Moreover, the construction of M in Step 1 shows that its encoding length is of the form $f(C, \varepsilon) \cdot \Pi(\langle \boldsymbol{D}\rangle)$ for a certain function f and a polynomial Π, as claimed. □

4.2. The Algorithm `Regularize`. In order to compute the desired regular partition of the input graph G, the algorithm `Regularize` starts with an arbitrary initial partition $\mathcal{P}^1 = \{V_i^1 : 0 \leq i \leq s_1\}$ such that each class V_i^1 $(1 \leq i \leq s_1)$ has a "decent" weight $D(V_i^1)$. In the subsequent steps, `Regularize` computes a sequence $(\mathcal{P}^k)$ of partitions such that $\mathcal{P}^{k+1}$ is a "more regular" refinement of $\mathcal{P}^k$ $(k \geq 1)$. The algorithm halts as soon as it can verify that $\mathcal{P}^k$ satisfies both **REG1** and **REG2** of Theorem 1.2. To this end `Regularize` applies the subroutine `Witness` to each pair (V_i^k, V_j^k) of the current partition $\mathcal{P}^k$. By Lemma 4.2 this yields a set $\mathcal{L}^k$ of pairs (i, j) such that all (V_i^k, V_j^k) with $(i, j) \notin \mathcal{L}^k$ are $(\varepsilon, \boldsymbol{D})$-regular. Hence, $\mathcal{P}^k$ satisfies **REG2** as soon as

$$\sum_{(i,j)\in\mathcal{L}^k} D(V_i^k)D(V_j^k) < \varepsilon D(V)^2. \tag{4.4}$$

In this case the algorithm `Regularize` stops and outputs $\mathcal{P}^k$. As we will see, all partitions $\mathcal{P}^k$ satisfy **REG1** by construction. Consequently, once (4.4) holds, `Regularize` has obtained the desired regular partition. Figure 4.2 shows the pseudocode.

Step 6 is the central step of the algorithm. In the first part of that step we construct a joint refinement of the previous partition $\mathcal{P}^k$ and all the witnesses of irregularity (X_{ij}^k, X_{ji}^k) discovered in Step 4. Similarly as in the original proof of Szemerédi's it will turn out that a bounded parameter (the so-called index defined below) of the partition $\mathcal{C}^k$ increases by $\Omega(\varepsilon^3)$ compared to $\mathcal{P}^k$. Since $\mathcal{P}_k$ consists of s_k classes and for every $i = 1, \ldots, s_k$ there are at most $s_k - 1$ witness sets X_{ij} $(j \neq i)$, the refinement $\mathcal{C}^k$ contains at most $s_k 2^{s_k - 1} < s_k 2^{s_k}$ vertex classes. In the second part of Step 6 we split the classes of $\mathcal{C}^k$ into pieces of almost equal weight. Here for each class of $\mathcal{C}^k$ we may get one class of left-over vertices $V_{0,q}^k$ of smaller weight, which together

Algorithm 4.3. $\texttt{Regularize}(G, C, \boldsymbol{D}, \varepsilon)$

1. Fix an arbitrary partition $\mathcal{P}^1 = \{V_i^1 : 0 \le i \le s_1\}$ for some $s_1 \le t_1$ with the property
   - $D(V)/t_1 - \max_{v \in V} D_v < D(V_i^1) \le D(V)/t_1$ for all $1 \le i \le s_1$ and
   - $D(V \setminus (\bigcup_{i \in [s_1]} V_i^1)) \le D(V)/t_1$.

   Set $V_0^1 = V \setminus \bigcup_{i \in [s_1]} V_i^1$ and set $k^* = \lceil 1000^2 C^2 \varepsilon^{-3} \rceil$.
2. For $k = 1, 2, 3, \ldots, k^*$ do
3. Initially, let $\mathcal{L}^k = \emptyset$.
   For each pair (V_i^k, V_j^k) $(i < j)$ of classes of partition $\mathcal{P}^k$
4. call the procedure $\texttt{Witness}(G, \boldsymbol{D}, V_i^k, V_j^k, \varepsilon)$.
   If it answers "no" and hence outputs an $(\varepsilon/200, \boldsymbol{D})$-witness (X_{ij}^k, X_{ji}^k) for (V_i^k, V_j^k), then add (i, j) to $\mathcal{L}^k$.
5. If $\sum_{(i,j) \in \mathcal{L}^k} D(V_i^k) D(V_j^k) < \varepsilon (D(V))^2$, then output the partition $\mathcal{P}^k$ and halt.
6. Else construct a refinement $\mathcal{P}^{k+1}$ of $\mathcal{P}^k$ as follows:
   - First construct the unique minimal partition $\mathcal{C}^k$ of $V \setminus V_0^k$, which refines $\{X_{ij}^k, V_i \setminus X_{ij}^k\}$ for every $i = 1, \ldots, s_k$ and every $j \ne i$. More precisely, we define the equivalence relation $\equiv_i^k$ on V_i by letting $u \equiv_i^k v$ iff for all j such that $(i, j) \in \mathcal{L}_k$ it is true that $u \in X_{ij}^k \Leftrightarrow v \in X_{ij}^k$ and we let $\mathcal{C}^k$ be the set of all equivalence classes of the relations $\equiv_i^k$ $(1 \le i \le s_k)$.
   - Set $\alpha_k = \varepsilon^{4(k+1)}/(2200^2 C^2 t_k^6 2^{t_k})$ and split each vertex class of $\mathcal{C}^k$ into blocks with weight between $\alpha_k D(V)$ and $\alpha_k D(V) + \max_{v \in V} D_v$ and possibly one exceptional block of smaller weight. More preceisely, we construct a refinement $\mathcal{C}_*^k = \{V_{0,1}^{k+1}, \ldots, V_{0,r_k}^{k+1}, V_1^{k+1}, \ldots, V_{s_{k+1}}^{k+1}\}$ of $\mathcal{C}^k$ such that:
     - $r_k \le |\mathcal{C}^k| \le s_k 2^{s_k}$,
     - $D(V_{0,q}^{k+1}) < \alpha_k D(V)$ for all $q \in [r_k]$, and
     - $\alpha_k D(V) \le D(V_i^{k+1}) < \alpha_k D(V) + \max_{v \in V} D_v$ for all $i \in [s_{k+1}]$.
   - Let $V_0^{k+1} = V_0^k \cup \bigcup_{q \in [r_k]} V_{0,q}^{k+1}$ and set $\mathcal{P}^{k+1} = \{V_i^{k+1} : 0 \le i \le s_{k+1}\}$.

Fig. 4.2. *The algorithm* $\texttt{Regularize}$.

with V_0^k form the new exceptional class V_0^{k+1}. Due to the construction in Step 6, the bound $s_1 \le t_1$, and (4.1) for any $k \ge 0$ the partition $\mathcal{P}^{k+1}$ consist of at most

$$s_{k+1} + 1 \le \lceil 2200^2 C^2 t_k^6 2^{t_k} / \varepsilon^{4(k+1)} \rceil = t_{k+1}$$

classes. Moreover, our choice (4.3) of η and the construction in Step 1 ensure that

$$\varepsilon^2 D(V) \ge D(V_i^{k+1}) \ge \sqrt{\eta} D(V) \text{ for all } 1 \le i \le s_{k+1} \tag{4.5}$$

for every $k < k^*$ (since in Step 6 we put all vertex classes of "extremely small" weight into the exceptional class). Furthermore, due to $r_i \le s_i 2^{s_i}$, $s_i \le t_i$, and $\varepsilon < 1/2$ we have

$$D(V_0^{k+1}) \leq D(V_0^1) + \sum_{i=2}^{k+1} r_i \frac{\varepsilon^{4(i+1)}}{2200^2 C^2 t_k^6 2^{t_k}} D(V)$$
$$\leq \frac{D(V)}{t_1} + D(V) \sum_{i=2}^{k+1} \varepsilon^{2i} \leq \frac{\varepsilon^2}{1-\varepsilon^2} D(V) \leq \varepsilon D(V)\,.$$

In effect, $\mathcal{P}^{k+1}$ always satisfies **REG1**, as **REG1(c)** is ensured by Step 6.

Thus, to complete the proof of Theorem 1.2 it just remains to show that Step 5 of `Regularize` will actually output a partition $\mathcal{P}^k$ for some $k \leq k^*$. More precisely, we have to show that for every input graph G there exists a $k \leq k^*$ such that $\sum_{(i,j)\in\mathcal{L}^k} \boldsymbol{D}(V_i^k)\boldsymbol{D}(V_j^k) < \varepsilon(\boldsymbol{D}(V))^2$. To show this, we use, as in the original proof of Szemerédi [23], the concept of the *index* of a partition $\mathcal{P} = \{V_i : 0 \leq i \leq s\}$ and define

$$\operatorname{ind}(\mathcal{P}) = \sum_{1\leq i<j\leq s} \varrho(V_i, V_j)^2 D(V_i) D(V_j) = \sum_{1\leq i<j\leq s} \frac{e(V_i, V_j)^2}{D(V_i)D(V_j)}.$$

Note that we do *not* take into account the (exceptional) class V_0 here. Using the boundedness condition, we derive the following.

Proposition 4.4. *If $G = (V, E)$ is a $(C, \eta, \boldsymbol{D})$-bounded graph and $\mathcal{P} = \{V_i\colon 0 \leq 1 \leq t\}$ is a partition of V with $D(V_i) \geq \eta D(V)$ for all $i \in \{1, \dots, t\}$, then $0 \leq \operatorname{ind}(\mathcal{P}) \leq C^2$.*

Proof. Since $D(V_i) \geq \eta D(V)$ for all $i \in \{1, \dots, t\}$ it follows from the $(C, \eta, \boldsymbol{D})$-boundedness of G that

$$\operatorname{ind}(\mathcal{P}) = \sum_{1\leq i<j\leq s} \frac{e(V_i, V_j)^2}{D(V_i)D(V_j)} \leq \sum_{1\leq i<j\leq s} \frac{Ce(V_i, V_j)}{D(V)} \leq C\frac{e(V, V)}{D(V)} \leq C^2,$$

as claimed. □

Proposition 4.4 and (4.5) imply that $\operatorname{ind}(\mathcal{P}^k) \leq C^2$ for all k. In addition, since `Regularize` obtains $\mathcal{P}^{k+1}$ by refining $\mathcal{P}^k$ according to the witnesses of irregularity computed by `Witness`, the index of $\mathcal{P}^{k+1}$ is actually considerably larger than the index of $\mathcal{P}^k$. More precisely, the following is true.

Lemma 4.5. *If $\sum_{(i,j)\in\mathcal{L}^k} D(V_i^k)D(V_j^k) \geq \varepsilon(D(V))^2$, then*

$$\operatorname{ind}(\mathcal{P}^{k+1}) \geq \operatorname{ind}(\mathcal{P}^k) + \varepsilon^3/8.$$

The proof of Lemma 4.5 is deferred to the next section, Section 4.3.

We close this section by pointing out that Propositions 4.4 and Lemma 4.5 readily imply that `Regularize` will terminate and output a feasible partition $\mathcal{P}^k$ for some $k \leq k^*$. Moreover, the dominant contribution to the running time of `Regularize` stems from the execution of the subroutine `Witness`, which gets called at most $O(k^* t_{k^*}^2)$ times. By Lemma 4.2 each execution takes time $f(C, \varepsilon)\cdot\Pi(\langle\boldsymbol{D}\rangle)$ for a certain function f and a polynomial Π. Hence, the total running time of `Regularize` is bounded by $f^*(C, \varepsilon)\cdot\Pi(\langle\boldsymbol{D}\rangle)$, where $f^*(C, \varepsilon) = O(k^* t_{k^*}^2)\cdot f(C, \varepsilon)$.

4.3. Proof of Lemma 4.5. As mentioned before, the proof of Lemma 4.5 follows the lines of the original proof of Szemerédi [23] with the main differences resulting from the somewhat different concept of regularity. We will use the following defect-form of the Cauchy-Schwarz inequality.

LEMMA 4.6 (Defect form of Cauchy-Schwarz inequality). *For all $i \in I$ let σ_i, d_i be positive real numbers satisfying $\sum_{i\in I} \sigma_i = 1$. Furthermore let $J \subset I$, $\varrho = \sum_{i\in I} \sigma_i \varrho_i$ and $\sigma_J = \sum_{j\in J} \sigma_j$. If $\sum_{j\in J} \sigma_j \varrho_j = \sigma_J(\varrho + \nu)$ then $\sum_{i\in I} \sigma_i \varrho_i^2 \geq \varrho^2 + \nu^2 \sigma_J$.*

Further, we will need the following technical proposition. Its proof is straightforward and we omit it here.

PROPOSITION 4.7. *Let $1/5 > \delta > 0$, $\eta > 0$, $C \geq 1$, and $\boldsymbol{D} = (D_v)_{v\in V}$ be a sequence of rationals with $1 \leq D_v \leq n$ for all $v \in V$. Let $G = (V, E)$ be a $(C, \eta, \boldsymbol{D})$-bounded graph and $A, B \subset V$ be disjoint subsets of V with $D(A), D(B) \geq \sqrt{\eta} D(V)$. If $A' \subset A$ and $B' \subset B$ satisfy $D(A \setminus A') < \delta D(A)$ and $D(B \setminus B') < \delta D(B)$, then*

$$\left| \frac{e(A,B)}{D(A)D(B)} - \frac{e(A',B')}{D(A')D(B')} \right| \leq \frac{(7\delta + 4\sqrt{\eta})C}{D(V)}$$

$$\left| \frac{e^2(A,B)}{D(A)D(B)} - \frac{e^2(A',B')}{D(A')D(B')} \right| \leq (21\delta + 9\sqrt{\eta})C^2.$$

For two partitions $\mathcal{P}' = \{V_j' \colon 0 \leq j \leq s\}$ and $\mathcal{P} = \{V_i \colon 0 \leq i \leq t\}$ we say $\mathcal{P}'$ *almost refines* $\mathcal{P}$, if for every $j \in [s]$ there exists an $i \in [t]$ such that $V_j' \subset V_i$. Note that an almost refinement may not be a refinement, since V_0' could be a proper superset of V_0.

PROPOSITION 4.8. *Let $\mathcal{P}' = \{V_j' \colon 0 \leq j \leq s\}$ and $\mathcal{P} = \{V_i \colon 0 \leq i \leq t\}$ be two partitions of V. If $\mathcal{P}'$ almost refines $\mathcal{P}$, then* $\mathrm{ind}(\mathcal{P}') \geq \mathrm{ind}(\mathcal{P})$.

Proof. For $V_i \in \mathcal{P}$, $i \in [t]$ let $I_i = \{j \colon V_j' \in \mathcal{P}', V_j' \subset V_i\}$. Then, using the Cauchy-Schwarz inequality, we conclude

$$\mathrm{ind}(\mathcal{P}') = \sum_{1\leq i<j\leq s} \frac{e^2(V_i', V_j')}{D(V_i')D(V_j')} \geq \sum_{1\leq k<l\leq t} \sum_{\substack{i\in I_k \\ j\in I_l}} \frac{e^2(V_i', V_j')}{D(V_i')D(V_j')}$$

$$\geq \sum_{1\leq k<l\leq t} \frac{\left(\sum_{i\in I_k, j\in I_l} e(V_i', V_j')\right)^2}{\sum_{i\in I_k, j\in I_l} D(V_i')D(V_j')} = \sum_{1\leq k<l\leq t} \frac{e^2(V_k, V_l)}{D(V_k)D(V_l)} = \mathrm{ind}(\mathcal{P}),$$

as desired. □

Proof of Lemma 4.5. Remember our assumption that $\varepsilon < 10^{-7}$. Let $K \subset V$ be the union of the equivalence classes with negligible weight; more precisely, in view of Step 6 we set

$$K = \bigcup_{q\in[r_k]} V_{0,q}^{k+1}.$$

Note that due to $r_k \leq s_k 2^{s_k}$ and $s_k \leq t_k$ we have

$$D(K) \leq r_k \frac{\varepsilon^{4(k+1)}}{2200^2 C^2 t_k^6 2^{t_k}} D(V) \leq \frac{\varepsilon^{4(k+1)}}{2200^2 C^2 t_k^5} D(V). \tag{4.6}$$

Now let $\mathcal{P}' = \{V_i' \colon 0 \leq i \leq s_k\}$ be the partition given by

$$V_i' = \begin{cases} V_0^k \cup K & \text{if } i = 0, \\ V_i^k \setminus K & \text{otherwise.} \end{cases}$$

To show the index increment $\mathrm{ind}(\mathcal{P}^{k+1}) \geq \mathrm{ind}(\mathcal{P}^k) + \varepsilon^3/1000^2$ we will proceed in two steps. In the first step we will compare the index of $\mathcal{P}'$ to the index of $\mathcal{P}^k$.

CLAIM 4.9. $|\mathrm{ind}(\mathcal{P}^k) - \mathrm{ind}(\mathcal{P}')| \leq \varepsilon^4$.

The second step will reveal the index increment of $\mathcal{P}^{k+1}$ compared to $\mathcal{P}'$.

CLAIM 4.10. $\mathrm{ind}(\mathcal{P}^{k+1}) \geq \mathrm{ind}(\mathcal{P}') + \varepsilon^3/800^2$.

As $\varepsilon < 10^{-7}$, this yields an index increment $\mathrm{ind}(\mathcal{P}^{k+1}) \geq \mathrm{ind}(\mathcal{P}^k) + \varepsilon^3/1000^2$. □

Proof of Claim 4.9. Let (V_i^k, V_j^k) be a pair of partition classes of $\mathcal{P}^k$ and let $V_i' = V_i^k \setminus K$ and $V_j' = V_j^k \setminus K$. Note that due to $D(V_i^k) \geq \varepsilon^{4k} D(V)/t_k^3$ and (4.6) we have

$$D(V_i') \geq D(V_i^k) - D(K) \geq (1 - \tfrac{\varepsilon^4}{42C^2 t_k^2}) D(V_i^k)$$

Analogously $D(V_j') \geq \left(1 - \varepsilon^4/(42C^2 t_k^2)\right) D(V_j^k)$ holds. In effect, using Proposition 4.7 we get

$$\left|\frac{e^2(V_i', V_j')}{D(V_i')D(V_j')} - \frac{e^2(V_i^k, V_j^k)}{D(V_i^k)D(V_j^k)}\right| \leq \frac{\varepsilon^4}{2t_k^2} + 9\sqrt{\eta}C^2 \overset{(4.3)}{\leq} \frac{\varepsilon^4}{t_k^2}.$$

Consequently

$$|\mathrm{ind}(\mathcal{P}^k) - \mathrm{ind}(\mathcal{P}')| \leq \sum_{1 \leq i < j \leq s_k} \left|\frac{e^2(V_i^k, V_j^k)}{D(V_i^k)D(V_j^k)} - \frac{e^2(V_i', V_j')}{D(V_i')D(V_j')}\right| \leq \varepsilon^4.$$

□

Proof of Claim 4.10. Let (V_i^k, V_j^k) be an irregular pair and $(A, B) = (V_i^k \setminus K, V_j^k \setminus K)$. Furthermore let (X_{ij}^k, X_{ji}^k) be an $(\varepsilon/200, \boldsymbol{D})$-witness. Then, for $X = X_{ij}^k \setminus K \subset A$ and $Y = X_{ji}^k \setminus K \subset B$, we have due to Proposition 4.7

$$\begin{aligned}\left|\frac{e(X,Y)}{D(X)D(Y)} - \frac{e(A,B)}{D(A)D(B)}\right| &\geq \frac{\varepsilon}{200} \frac{D(A)D(B)}{D(X_{ij}^k)D(X_{ji}^k)D(V)} - \frac{\frac{7\varepsilon^2}{2200^2} + \frac{7\cdot 200\varepsilon}{2200^2} + 8\sqrt{\eta}C}{D(V)} \\ &\geq \frac{\varepsilon}{400} \frac{D(A)D(B)}{D(X)D(Y)D(V)} - \frac{\varepsilon}{1600 D(V)} - \frac{\varepsilon}{1600 D(V)} \\ &\geq \frac{\varepsilon}{800} \frac{D(A)D(B)}{D(X)D(Y)D(V)}\,. \qquad (4.7)\end{aligned}$$

Thus, (X, Y) 'witnesses' that (A, B) is not $(\varepsilon/800, \boldsymbol{D})$-regular.

Now we will use Lemma 4.6 to prove $\mathrm{ind}(\mathcal{P}^{k+1}) \geq \mathrm{ind}(\mathcal{P}') + \varepsilon^3/4$. So let $I = A \times B$ and for all $(u, v) \in I$ let

$$\sigma_{uv} = \frac{D_u D_v}{D(A)D(B)} \quad \text{and} \quad \varrho_{uv} = \varrho(V^{k+1}(u), V^{k+1}(v))$$

where $V^{k+1}(x)$ denotes the partition class $V_i^{k+1} \in \mathcal{P}^{k+1}$ such that $x \in V_i^{k+1}$. Then

$$\sum_{(u,v)\in I} \sigma_{uv} = 1 \quad \text{and}$$

$$\sum_{(u,v)\in I} \sigma_{uv}\varrho_{uv} = \sum_{(u,v)\in I} \frac{D_u D_v}{D(A)D(B)} \frac{e(V^{k+1}(u), V^{k+1}(v))}{D(V^{k+1}(u))D(V^{k+1}(v))} = \varrho(A, B).$$

Moreover, let $J = X \times Y$ and $\sigma_J = \sum_{(u,v)\in J} \sigma_{uv} = \frac{D(X)D(Y)}{D(A)D(B)}$. Then we have

$$\frac{1}{\sigma_J} \sum_{(u,v)\in J} \sigma_{uv}\varrho_{uv} = \frac{D(A)D(B)}{D(X)D(Y)} \sum_{\substack{V_i^{k+1}\subset X\\ V_j^{k+1}\subset Y}} \sum_{\substack{u\in V_i^{k+1}\\ v\in V_j^{k+1}}} \frac{D_u D_v}{D(A)D(B)} \varrho(V_i^{k+1}, V_j^{k+1})$$
$$= \frac{e(X,Y)}{D(X)D(Y)} = \varrho(X,Y) = \varrho(A,B) + \nu$$

for some $|\nu| \geq \varepsilon D(A)D(B)/(800D(X)D(Y)D(V))$ due to (4.7).

Hence, from the Cauchy-Schwarz inequality (Lemma 4.6) we deduce

$$\frac{1}{D(A)D(B)} \sum_{\substack{V_i^{k+1}\subset A\\ V_j^{k+1}\subset B}} \varrho^2(V_i^{k+1}, V_j^{k+1}) D(V_i^{k+1}) D(V_j^{k+1})$$
$$= \sum_{u,v\in I} \frac{D_u D_v}{D(A)D(B)} \varrho^2(V^{k+1}(u), V^{k+1}(v)) = \sum_{(u,v)\in I} \sigma_{uv}\varrho_{uv}^2$$
$$\geq \varrho^2(A,B) + \left(\frac{\varepsilon D(A)D(B)}{800D(X)D(Y)D(V)}\right)^2 \frac{D(X)D(Y)}{D(A)D(B)}$$
$$\geq \frac{1}{D(A)D(B)} \left(\varrho^2(A,B)D(A)D(B) + \frac{\varepsilon^2 D(A)D(B)}{800^2 D^2(V)}\right).$$

From the last inequality we infer the amount of the index increment on the irregular pair (A,B). So, in view of Proposition 4.8, after summing over all pairs we get

$$\operatorname{ind}(\mathcal{P}^{k+1}) - \operatorname{ind}(\mathcal{P}') \geq \sum_{(i,j)\in\mathcal{L}^k} \frac{\varepsilon^2}{800^2} \frac{D(A)D(B)}{D^2(V)} \geq \frac{\varepsilon^3}{800^2}.$$

□

5. An Application: MAX CUT. As an application of Theorem 1.2 and, in particular, the polynomial time algorithm `Regularize` for computing a regular partition, we obtain the algorithm shown in Figure 5.1 for approximating the maximum cut of a graph $G = (V,E)$ that satisfies the assumptions of Theorem 1.3.

The basic insight behind `ApxMaxCut` is the following. If (V_i, V_j) is an $(\varepsilon, \boldsymbol{D})$-regular pair of $\mathcal{P}$, then for any subsets $X, X' \subset V_i$ and $Y, Y' \subset V_j$ such that $D(X) = D(X')$ and $D(Y) = D(Y')$ the condition **REG2** ensures that $|e(X,Y) - e(X',Y')| \leq \frac{2\varepsilon D(V_i)D(V_j)}{D(V)}$. That is, the difference between $e(X,Y)$ and $e(X',Y')$ is negligible. In other words, as far as the number of edges is concerned, subsets that have the same weight are "interchangeable".

Therefore, to compute a good cut $(S, \bar{S})$ of G we just have to optimize the *proportion of weight* of each V_i that is to be put into S or into $\bar{S}$, but it does not matter which subset of V_i of this weight we choose. However, determining the optimal fraction of weight is still a somewhat involved (essential continuous) optimization problem. Hence, in order to discretize this problem, we chop each V_i into at most ε^{-1} chunks of weight $\varepsilon D(V_i)$. Then, we just have to determine the number c_i of chunks of each V_i that we join to S. This is exactly the optimization problem detailed in Step 2 of `ApxMaxCut`.

ALGORITHM 5.1. `ApxMaxCut`$(G, C, \boldsymbol{D}, \delta)$
Input: A $(C, \eta, \boldsymbol{D})$-bounded graph $G = (V, E)$ and $\delta > 0$.
Output: A cut $(S, \bar{S})$ of G.

1. Use `Regularize` to compute $\varepsilon = \frac{\delta}{400C}$-regular partition $\mathcal{P} = \{V_i : 0 \le i \le t\}$ of G.
2. Determine an optimal solution $(c_1^*, \ldots, c_t^*)$ to the optimization problem

$$\max \sum_{i \neq j} \varepsilon c_i(1 - \varepsilon c_j) e(V_i, V_j) \text{ s.t. } \forall 1 \le j \le t : 0 \le c_j \le \varepsilon^{-1},\ c_j \in \mathbb{Z}.$$

3. For each $1 \le i \le t$ let $S_i \subset V_i$ be a subset such that $|D(S_i) - c_i^* \varepsilon D(V_i)| \le 2\varepsilon D(V_i)$. Output $S = \bigcup_{i=1}^t S_i$ and $\bar{S} = V \setminus S$.

FIG. 5.1. *The algorithm* ***ApxMaxCut****.*

Observe that the time required to solve this problem is *independent* of n, i.e., Step 2 has a *constant* running time. For the number t of classes of $\mathcal{P}$ is bounded by a number independent of n, and the number $\lceil \varepsilon^{-1} \rceil + 1$ of choices for each c_i does not depend on n either. In addition, Step 3 can be implemented so that it runs in linear time, because $S_i \subset V_i$ can be *any* subset that satisfies the condition stated in Step 3. Thus, the total running time of `ApxMaxCut` is polynomial.

To prove that `ApxMaxCut` does indeed guarantee an approximation within an additive $\delta D(V)$, we compare the maximum cut of G with the optimal solution μ^* of the optimization problem from Step 2, i.e.,

$$\mu^* = \max \sum_{i,j} \varepsilon c_i(1 - \varepsilon c_j) e(V_i, V_j) \text{ s.t. } \forall 1 \le j \le t : 0 \le c_j \le \varepsilon^{-1},\ c_j \in \mathbb{Z}. \tag{5.1}$$

To this end, we say that a cut $(T, \bar{T})$ of G is *compatible* with a feasible solution $(c_1, \ldots, c_t)$ to the optimization problem (5.1) if $|D(T \cap V_i) - c_i \varepsilon D(V_i)| \le 2\varepsilon D(V_i)$.

LEMMA 5.2. *Suppose that* $(T, \bar{T})$ *is compatible with the feasible solution* $(c_i)_{1 \le i \le t}$ *of (5.1). Moreover, let*

$$\mu = \sum_{i,j} \varepsilon c_i(1 - \varepsilon c_j) e(V_i, V_j)$$

be the objective function value corresponding to $(c_1, \ldots, c_t)$*. Then* $|e(T, \bar{T}) - \mu| \le \frac{\delta}{8} D(V)$*.*

Proof. Set $T_i = T \cap V_i$ and $\bar{T}_i = V_i \setminus T_i$, so that $e(T, \bar{T}) = \sum_{i \neq j} e(T_i, \bar{T}_j) + \sum_{i=0}^t e(T_i, \bar{T}_i)$, and let $\mu_{ij} = \varepsilon c_i(1 - \varepsilon c_j) e(V_i, V_j)$ $(1 \le i, j \le t)$. Moreover, let $\mathcal{L}$ be the set of all pairs (i, j) such that the pair (V_i, V_j) is not $(\varepsilon, \boldsymbol{D})$-regular. Then **REG 2** and the $(C, \eta, \boldsymbol{D})$-boundedness of G imply that

$$\begin{aligned} \sum_{(i,j) \in \mathcal{L}} \mu_{ij} \le \sum_{(i,j) \in \mathcal{L}} e(V_i, V_j) \le \sum_{(i,j) \in \mathcal{L}} \frac{C D(V_i) D(V_j)}{D(V)} \le C \varepsilon D(V) \\ = \frac{\delta}{400} D(V), \end{aligned} \tag{5.2}$$

$$\sum_{(i,j) \in \mathcal{L}} e(T_i, \bar{T}_j) \le \sum_{(i,j) \in \mathcal{L}} e(V_i, V_j) \le \frac{\delta}{400} D(V).$$

Furthermore, since $D(V_0) \le \varepsilon D(V)$ and $C \ge 1$ we have

$$e(T_0, \bar{T}) + e(\bar{T}_0, T) \le D(V_0) \le \varepsilon D(V) \le \frac{\delta}{400} D(V),$$

and as $D(V_i) \le \varepsilon D(V)$ for all i, the $(C, \eta, \boldsymbol{D})$-boundedness condition yields

$$\sum_{i=1}^{t} e(T_i, \bar{T}_i) \le \sum_{i=1}^{t} \frac{CD(V_i)^2}{D(V)} \le C\varepsilon D(V) = \frac{\delta}{400} D(V).$$

In addition, let

$$\mathcal{S} = \{(i,j) : i,j > 0,\ i \ne j \wedge (i,j) \notin \mathcal{L} \wedge (D(T_i) < \varepsilon D(V_i) \vee D(\bar{T}_j) < \varepsilon D(V_j))\}.$$

We shall prove below that for all $(i,j) \notin (\mathcal{L} \cup \mathcal{S})$, $i,j > 0$, $i \ne j$

$$\left|\mu_{ij} - e(T_i, \bar{T}_j)\right| < 5\varepsilon e(V_i, V_j) + \varepsilon \frac{D(V_i) D(V_j)}{D(V)}, \tag{5.3}$$

and that

$$\sum_{(i,j) \in \mathcal{S}} \mu_{ij} + e(T_i, \bar{T}_j) < 6\varepsilon D(V). \tag{5.4}$$

Combining (5.2)–(5.4), we thus obtain

$$\begin{aligned}
\left|e(T, \bar{T}) - \mu\right| \le & \sum_{\substack{(i,j) \notin (\mathcal{L} \cup \mathcal{S}) \\ i,j>0,\, i \ne j}} \left|\mu_{ij} - e(T_i, \bar{T}_j)\right| \\
& + \sum_{(i,j) \in (\mathcal{L} \cup \mathcal{S})} (\mu_{ij} + e(T_i, T_j)) + e(T_0, \bar{T}) + e(\bar{T}_0, T) + \sum_{i=1}^{t} e(T_i, \bar{T}_i) \\
& \le 6\varepsilon D(V) + \frac{\delta}{200} D(V) + 6\varepsilon D(V) + \frac{\delta}{400} D(V) + \frac{\delta}{400} D(V) \le \frac{\delta}{8} D(V),
\end{aligned}$$

as desired.

To establish (5.3), consider a pair $(i,j) \notin (\mathcal{L} \cup \mathcal{S})$, $i \ne j$. Since $D(T_i) \ge \varepsilon D(V_i)$ and $D(\bar{T}_j) \ge \varepsilon D(V_j)$ and (V_i, V_j) is $(\varepsilon, \boldsymbol{D})$-regular, we have

$$\left| e(T_i, \bar{T}_j) - \frac{D(T_i) D(\bar{T}_j)}{D(V_i) D(V_j)} e(V_i, V_j) \right| < \frac{\varepsilon D(V_i) D(V_j)}{D(V)}. \tag{5.5}$$

Moreover, as $(T, \bar{T})$ is compatible with $(c_1, \ldots, c_t)$,

$$\left| \frac{D(T_i)}{D(V_i)} - \varepsilon c_i \right| < 2\varepsilon, \qquad \left| \frac{D(\bar{T}_j)}{D(V_j)} - (1 - \varepsilon c_j) \right| < 2\varepsilon, \tag{5.6}$$

and combining (5.5) and (5.6) yields (5.3).

Finally, to prove (5.4), consider an index i such that $D(T_i) < \varepsilon D(V_i)$. Then

$$\sum_{j=1}^{t} e(T_i, \bar{T}_j) \le D(T_i) < \varepsilon D(V_i).$$

Similarly, if $D(\bar{T}_j) < \varepsilon D(V_j)$, then $\sum_{i=1}^t e(T_i, \bar{T}_j) < \varepsilon D(V_j)$. Therefore,

$$\sum_{(i,j)\in\mathcal{S}} e(T_i, \bar{T}_j) < 2\varepsilon D(V). \tag{5.7}$$

Further, if $D(T_i) < \varepsilon D(V_i)$, then $c_i \le 2$, because $(T, \bar{T})$ is compatible with $(c_1, \ldots, c_t)$. Thus $\sum_{j=1}^t \mu_{ij} \le 2\varepsilon \sum_j e(V_i, V_j) \le 2\varepsilon D(V_i)$. Analogously, if $D(\bar{T}_j) < \varepsilon D(V_j)$, then $\sum_{i=1}^t \mu_{ij} \le 2\varepsilon D(V_j)$. Consequently,

$$\sum_{(i,j)\in\mathcal{S}} \mu_{ij} < 4\varepsilon D(V). \tag{5.8}$$

Hence, (5.4) follows from (5.7) and (5.8). □

Proof of Theorem 1.3. Step 3 of `ApxMaxCut` ensures that $(S, \bar{S})$ is compatible with $(c_1^*, \ldots, c_t^*)$. Therefore, Lemma 5.2 yields

$$e(S, \bar{S}) \ge \mu^* - \frac{\delta}{8} D(V). \tag{5.9}$$

Further, let $(T, \bar{T})$ be a maximum cut of G. Then we can construct a feasible solution to (5.1) that is compatible with $(T, \bar{T})$ by letting

$$c_i = \left\lfloor \frac{D(T \cap V_i)}{\varepsilon D(V_i)} \right\rfloor \qquad (1 \le i \le t).$$

Let $\mu = \sum_{i,j} \varepsilon c_i(1-\varepsilon c_j)e(V_i, V_j)$ be the corresponding objective function value. Then Lemma 5.2 implies that

$$e(T, \bar{T}) \le \mu + \frac{\delta}{8} D(V). \tag{5.10}$$

As μ^* is the optimal value of (5.1), we have $\mu^* \ge \mu$, and thus (5.9) and (5.10) yield $e(S, \bar{S}) \ge e(T, \bar{T}) - \frac{\delta}{4} D(V)$. Consequently, `ApxMaxCut` provides the desired approximation guarantee. □

6. Conclusion.

1. Theorem 1.1 states that there is a constant $\gamma > 0$ such that $\text{Disc}(\gamma\varepsilon^2)$ implies ess-Eig(ε). This statement is best possible, up to the precise value of γ. To see this, we describe a (probabilistic) construction of a graph $G = (V, E)$ on n vertices that has $\text{Disc}(10\varepsilon)$ but does not have ess-Eig($0.01\sqrt{\varepsilon}$). Assume that $\varepsilon > 0$ is a sufficiently small number, and choose $n = n(\varepsilon)$ sufficiently large. Moreover, let $X = \{1, \ldots, \sqrt{\varepsilon} n\}$ and $\bar{X} = \{\sqrt{\varepsilon} n + 1, \ldots, n\}$. Further, let $d = n/2$ and set

   $$p_X = 1, \quad p_{X\bar{X}} = p_{\bar{X}X} = \frac{1 - 2\sqrt{\varepsilon}}{2 - 2\sqrt{\varepsilon}}, \quad p_{\bar{X}} = \frac{1 - 2\sqrt{\varepsilon} + 2\varepsilon}{2(1 - \sqrt{\varepsilon})^2}.$$

   Finally, let G be the random graph with vertex set $V = \{1, \ldots, n\}$ obtained as follows: any two vertices in X are adjacent, any two vertices in $\bar{X}$ are connected with probability $p_{\bar{X}}$ independently, and each possible X-$\bar{X}$ edge is present with probability $p_{X\bar{X}}$ independently. Thus, the vertices X form a clique. Moreover, the expected degree of each vertex is d. It easily seen that

G satisfies Disc(10ε). To see that G does not satisfy ess-Eig($\sqrt{\varepsilon}/2$), let $\mathcal{E}$ be the matrix with entries

$$\mathcal{E}_{vw} = \begin{cases} 1 & \text{if } v, w \in X, \\ p_{X\bar{X}} & \text{if } (v,w) \in X \times \bar{X} \cup \bar{X} \times X, \\ p_{\bar{X}} & \text{if } v, w \in \bar{X}. \end{cases}$$

This matrix just comprises the probabilities that v, w are adjacent. Results on the eigenvalues of random matrices [14] imply that $\|\boldsymbol{E} - L(G) - d^{-1}\mathcal{E}\| = o(1)$. Let $W \subset \{1, \ldots, n\}$ be an arbitrary set of size $|W| \geq (1 - 0.01\varepsilon)n$. Then $\|\boldsymbol{E} - L(G)_W - d^{-1}\mathcal{E}_W\| \leq \|\boldsymbol{E} - L(G) - d^{-1}\mathcal{E}\| = o(1)$. Therefore, in order to show that $\lambda_2(L(G)_W) < 1 - 0.01\sqrt{\varepsilon}$ it suffices to prove that the matrix $\boldsymbol{E} - d^{-1}\mathcal{E}_W$ satisfies

$$\lambda_2(\boldsymbol{E} - d^{-1}\mathcal{E}_W) \leq 1 - \sqrt{\varepsilon}/2. \tag{6.1}$$

Let $x = |X \cap W|$ and $\bar{x} = |\bar{X} \cap W|$. The matrix $d^{-1}\mathcal{E}_W$ has rank two, and the eigenvectors with non-zero eigenvalues lie in the space spanned by the vectors $\mathbf{1}_{X \cap W}$ and $\mathbf{1}_{\bar{X} \cap W}$. This implies that its non-zero eigenvalues coincide with those of the 2×2 matrix

$$\mathcal{E}_* = d^{-1} \cdot \begin{pmatrix} x & \bar{x} \cdot p_{X\bar{X}} \\ x \cdot p_{X\bar{X}} & \bar{x} \cdot p_{\bar{X}} \end{pmatrix},$$

which can be computed directly. The smaller eigenvalue is at least $\sqrt{\varepsilon}/(1 - \sqrt{\varepsilon}) - \varepsilon \geq \sqrt{\varepsilon}/2$. Hence, $\lambda_2(\boldsymbol{E} - d^{-1}\mathcal{E}_W) \leq 1 - \sqrt{\varepsilon}/2$.

2. In the conference version of this paper we stated erroneously that the implication "Disc($\gamma\varepsilon^3$) $\Rightarrow$ ess-Eig(ε)" is best possible.
3. The techniques presented in Section 3 can be adapted easily to obtain a similar result as Theorem 1.1 with respect to the concepts of discrepancy and eigenvalue separation from [11]. More precisely, let $G = (V, E)$ be a graph on n vertices, let $p = 2|E|n^{-2}$ be the edge density of G, and let $\gamma > 0$ denote a small enough constant. If for any subset $X \subset V$ we have $|2e(X) - |X|^2 p| < \gamma\varepsilon^2 n^2 p$, then there exists a set $W \subset V$ of size $|W| \geq (1 - \varepsilon)n$ such that the following is true. Letting $A = A(G)$ signify the adjacency matrix of G, we have $\max\{-\lambda_1[A_W], \lambda_{|W|-1}[A_W]\} \leq \varepsilon np$. That is, all eigenvalues of the minor A_W except for the largest are at most εnp in absolute value. The same example as under 1. shows that this result is best possible up to the precise value of γ.

REFERENCES

[1] Albert, R., Barabási, A.L.: Statistical mechanics of complex networks. Reviews of modern physics **74** (2002) 47–97

[2] Alizadeh, F.: Interior point methods in semidefinite programming with applications to combinatorial optimization. SIAM J. Optimization **5** (1995) 13–51

[3] Alon, N., Duke, R.A., Rödl, V., Yuster, R.: The algorithmic aspects of the regularity lemma. J. of Algorithms **16** (1994) 80–109

[4] Alon, N, Naor, A.: Approximating the cut-norm via Grothendieck's inequality. Proc. 36th ACM STOC (2004) 72–80

[5] Bellman, R: Introduction to matrix analysis, 2nd ed. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

[6] Bilu, Y., Linial, N.: Lifts, discrepancy and nearly optimal spectral gap. Combinatorica **26** (2006) 495–519

[7] Bollobás, B., Nikiforov, V.: Graphs and Hermitian matrices: discrepancy and singular values, Discrete Math. **285** (2004) 17–32
[8] Butler, S.: Using discrepancy to control singular values for nonnegative matrices. Linear Algebra and its Applications **419** (2006) 486–493
[9] Chung, F.: Spectral graph theory. American Mathematical Society (1997).
[10] Chung, F., Graham, R.: Quasi-random graphs with given degree sequences. Random Structures and Algorithms **32** (2008) 1–19
[11] Chung, F., Graham, R.: Sparse quasi-random graphs. Combinatorica **22** (2002) 217–244
[12] Chung, F., Graham, R., Wilson, R.M.: Quasi-random graphs. Combinatorica **9** (1989) 345–362
[13] Frieze, A., Kannan, R.: Quick approximation to matrices and applications. Combinatorica **19** (1999) 175–200
[14] Füredi, Z., Komloś, J.: The eigenvalues of random symmetric matrices. Combinatorica **1** (1981) 233–241
[15] Gerke, S., Steger, A.: The sparse regularity lemma and its applications. In: Surveys in Combinatorics, Proc. 20th British Combinatorial Conference, London Mathematical Society Lecture Notes Series 327, ed. Bridget S. Webb, Cambridge University Press (2005) 227–258
[16] Gerke, S., Steger, A.: A characterization for sparse ε-regular pairs. The Electronic J. Combinatorics **14** (2007), R4
[17] Grothendieck, A.: Résumé de la théorie métrique des produits tensoriels topologiques. Bol. Soc. Mat. Sao Paulo **8** (1953) 1–79
[18] Grötschel, M., Lovász, L., Schrijver, A.: Geometric algorithms and combinatorial optimization. Springer (1988)
[19] Håstad, J.: Some optimal inapproximability results. J. of the ACM **48** (2001) 798–859
[20] Helmberg, C.: Semidefinite programming for combinatorial optimization. Habilitation thesis. Report ZR-00-34, Zuse Institute Berlin (2000)
[21] Kohayakawa, Y.: Szemerédi's regularity lemma for sparse graphs. In Cucker, F., Shub, M. (eds.): Foundations of computational mathematics (1997) 216–230.
[22] Kohayakawa, Y., Rödl, V., Thoma, L.: An optimal algorithm for checking regularity. SIAM J. Comput., **32** (2003) 1210–1235
[23] Szemerédi, E.: Regular partitions of graphs. Problémes Combinatoires et Théorie des Graphes Colloques Internationaux CNRS **260** (1978) 399-401
[24] Trevisan, L., Sorkin, G., Sudan, M., Williamson, D.: Gadgets, approximation, and linear programming. SIAM J. Computing **29** (2000) 2074–2097