Quasi-Randomness and Algorithmic Regularity for Graphs with General Degree Distributions

Noga Alon[1⋆], Amin Coja-Oghlan[2⋆⋆], Hiệp Hàn[2⋆⋆⋆], Mihyun Kang[2†], Vojtěch Rödl[3‡], and Mathias Schacht[2§]

[1] 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

[2] Humboldt-Universität zu Berlin, Institut für Informatik, Unter den Linden 6, 10099 Berlin, Germany
{coja,hhan,kang,schacht}@informatik.hu-berlin.de

[3] Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA
rodl@mathcs.emory.edu

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 present a novel concept of regularity that takes into account the graph's degree distribution, 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 [4], we provide an algorithm that computes a regular partition of a given (possibly sparse) graph G in polynomial time.

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

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. References to related work can be found in Section 2, and the remaining sections contain proof sketches and detailed descriptions of the algorithms.

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

⋆ Supported by an ISF grant, and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University.
⋆⋆ Supported by DFG grant FOR 413/1-2.
⋆⋆⋆ Supported by DFG within the research training group "Methods for Discrete Structures".
† Supported by DFG grant PR 296/7-3.
‡ Supported by NSF Grant DMS 0300529.
§ Supported by DFG grant SCHA 1263/1-1 and GIF Grant 889/05.

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, $e(S)$ denotes the number of edges spanned by S.

$\mathrm{Disc}(\varepsilon)$: We say that G *has discrepancy at most* ε ("G has $\mathrm{Disc}(\varepsilon)$" for short) if

$$\forall S \subset V : \left| e(S) - \frac{\mathrm{vol}(S)^2}{2\mathrm{vol}(V)} \right| < \varepsilon \cdot \mathrm{vol}(V). \tag{1}$$

To explain (1), let $\boldsymbol{d} = (d_v)_{v \in V}$, and let $G(\boldsymbol{d})$ signify a uniformly distributed random graph with degree distribution $\boldsymbol{d}$. Then the probability p_{vw} that two vertices $v, w \in V$ are adjacent in $G(\boldsymbol{d})$ is proportional to the degrees of both v and w, and hence to their product. Further, as the total number of edges is determined by the sum of the degrees, we have $\sum_{(v,w) \in V^2} p_{vw} = \mathrm{vol}(V)$, whence $p_{vw} \sim d_v d_w / \mathrm{vol}(V)$. Therefore, in $G(\boldsymbol{d})$ the *expected* number of edges inside of $S \subset V$ equals $\frac{1}{2} \sum_{(v,w) \in S^2} p_{vw} \sim \frac{1}{2} \mathrm{vol}(S)^2 / \mathrm{vol}(V)$. Consequently, (1) just says that for *any* set S the actual number $e(S)$ 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) is that it is quite difficult to check whether $G = (V, E)$ satisfies this condition. This is because 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

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

$L(G)$ turns out to be appropriate for representing graphs with general degree distributions.

$\mathrm{Eig}(\delta)$: 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 $\mathrm{Eig}(\delta)$") 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 $\mathrm{Eig}(\delta)$ or not.

It is not difficult to see that $\mathrm{Eig}(\delta)$ provides a *sufficient* condition for $\mathrm{Disc}(\varepsilon)$. That is, for any $\varepsilon > 0$ there is a $\delta > 0$ such that any graph G that has $\mathrm{Eig}(\delta)$ also has $\mathrm{Disc}(\varepsilon)$. 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 $\mathrm{Disc}(\varepsilon)$ in terms of eigenvalues has been an open problem in the area of sparse quasi-random graphs since the work of Chung and Graham [9]. Concerning this problem, we basically observe that the reason why $\mathrm{Disc}(\varepsilon)$ does in general not imply $\mathrm{Eig}(\delta)$ is the existence of a small set of "exceptional" vertices. With this in mind we refine the definition of Eig as follows.

ess-$\mathrm{Eig}(\delta)$: We say that G *has essential δ-eigenvalue separation* ("G has ess-$\mathrm{Eig}(\delta)$") 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 $1 - \delta < \lambda_2(L(G)_W) < \lambda_{|W|}(L(G)_W) < 1 + \delta$.

Theorem 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 $\mathrm{ess\text{-}Eig}(\varepsilon)$, then G satisfies $\mathrm{Disc}(10\sqrt{\varepsilon})$.*
2. *If G has $\mathrm{Disc}(\gamma\varepsilon^3)$, then G satisfies $\mathrm{ess\text{-}Eig}(\varepsilon)$.*

The proof of Theorem 1 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 $\mathrm{ess\text{-}Eig}(\varepsilon)$, provided that the input graph has $\mathrm{Disc}(\delta)$. In Appendix E we show that the second part of Theorem 1 is best possible, up to the precise value of the constant γ.

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 i, j 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 quite a few problems that are NP-hard in general.

In the sequel we introduce a new concept of regular partitions that takes into account the degree distribution of the graph. If $G = (V, E)$ is a graph and $A, B \subset V$ are disjoint, then the *relative density* of (A, B) in G is $\varrho(A, B) = \frac{e(A,B)}{\mathrm{vol}(A)\mathrm{vol}(B)}$. Further, we say that the pair (A, B) is *ε-volume regular* if for all $X \subset A$, $Y \subset B$ satisfying $\mathrm{vol}(X) \geq \varepsilon\mathrm{vol}(A)$, $\mathrm{vol}(Y) \geq \varepsilon\mathrm{vol}(B)$ we have

$$|e(X, Y) - \varrho(A, B)\mathrm{vol}(X)\mathrm{vol}(Y)| \leq \varepsilon \cdot \mathrm{vol}(A)\mathrm{vol}(B)/\mathrm{vol}(V), \tag{2}$$

where $e(X, Y)$ denotes the number of X-Y-edges in G. This condition essentially means that the bipartite graph spanned by A and B is quasi-random, given the degree distribution of G. Indeed, in a random graph the proportion of edges between X and Y should be proportional to both $\mathrm{vol}(X)$ and $\mathrm{vol}(Y)$, and hence to $\mathrm{vol}(X)\mathrm{vol}(Y)$. Moreover, $\varrho(A, B)$ measures the overall density of (A, B).

Finally, we state a condition that ensures the existence of regular partitions. While *every* dense graph G (of volume $\mathrm{vol}(V) = \Omega(|V|^2)$) admits a regular partition, such partitions do not necessarily exist for sparse graphs, the basic obstacle being extremely "dense spots". To rule out such dense spots, we say that a graph G is *(C, η)-bounded* if for all $X, Y \subset V$ with $\mathrm{vol}(X \cup Y) \geq \eta\mathrm{vol}(V)$ we have $\varrho(X, Y)\mathrm{vol}(V) \leq C$.

Theorem 2. *For any two numbers $C > 0$ and $\varepsilon > 0$ there exist $\eta > 0$ and $n_0 > 0$ such that for all $n > n_0$ the following holds. If $G = (V, E)$ is a (C, η)-bounded graph on n vertices such that $\mathrm{vol}(V) \geq \eta^{-1}n$, then there is a partition $\mathcal{P} = \{V_i \colon 0 \leq i \leq t\}$ of V that enjoys the following two properties.*

REG1. *For all $1 \leq i \leq t$ we have $\eta\mathrm{vol}(V) \leq \mathrm{vol}(V_i) \leq \varepsilon\mathrm{vol}(V)$, and $\mathrm{vol}(V_0) \leq \varepsilon\mathrm{vol}(V)$.*
REG2. *Let $\mathcal{L}$ be the set of all pairs $(i, j) \in \{1, \ldots, t\}^2$ such that (V_i, V_j) is not ε-volume-regular. Then $\sum_{(i,j)\in\mathcal{L}} \mathrm{vol}(V_i)\mathrm{vol}(V_j) \leq \varepsilon\mathrm{vol}^2(G)$.*

Furthermore, for fixed $C > 0$ and $\varepsilon > 0$ such a partition $\mathcal{P}$ of V can be computed in time polynomial in n.

Condition **REG1** states that each of the classes $V_1, \ldots, V_t$ has some non-negligible volume, and that the "exceptional" class V_0 is not too big. Moreover, **REG2** requires that the share of edges of G

that belongs to irregular pairs (V_i, V_j) is small. Thus, a partition $\mathcal{P}$ that satisfies **REG1** and **REG2** approximates G by a bounded number of bipartite quasi-random graphs, i.e., the number t of classes can be bounded solely in terms of ε and the boundedness parameter C.

We illustrate the use of Theorem 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 [16, 21], the following theorem provides a polynomial time approximation scheme for (C, η)-bounded graphs.

Theorem 3. *For any $\delta > 0$ and $C > 0$ there exist two numbers $\eta > 0$, n_0 and a polynomial time algorithm* `ApxMaxCut` *such that for all $n > n_0$ the following is true. If $G = (V, E)$ is a (C, η)-bounded graph on n vertices and $\mathrm{vol}(V) > \eta^{-1}|V|$, then* `ApxMaxCut`$(G)$ *outputs a cut $(S, \bar{S})$ of G that approximates the maximum cut within a factor of $1 - \delta$.*

The corresponding result for dense graphs was obtained by Frieze and Kannan [11].

2 Related Work

Quasi-random graphs. Quasi-random graphs with general degree distributions were first studied by Chung and Graham [8]. 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 $\mathrm{Disc}(\varepsilon)$ to $\mathrm{Eig}(\delta)$, Butler [7] proved that any graph G such that for all sets $X, Y \subset V$ the bound

$$|e(X, Y) - \mathrm{vol}(X)\mathrm{vol}(Y)/\mathrm{vol}(V)| \leq \varepsilon\sqrt{\mathrm{vol}(X)\mathrm{vol}(Y)} \tag{3}$$

holds, satisfies $\mathrm{Eig}(O(\varepsilon(1 - \ln \varepsilon)))$. His proof builds heavily on the work of Bilu and Linial [5], who derived a similar result for regular graphs, and on the earlier related work of Bollobás and Nikiforov [6].

Butler's result relates to the second part of Theorem 1 as follows. The r.h.s. of (3) refers to the volumes of the sets X, Y, and may thus be significantly smaller than $\varepsilon\mathrm{vol}(V)$. By contrast, the second part of Theorem 1 just requires that the "original" discrepancy condition $\mathrm{Disc}(\delta)$ is true, i.e., we just need to bound $|e(S) - \frac{1}{2}\mathrm{vol}(S)^2/\mathrm{vol}(V)|$ in terms of the *total* volume $\mathrm{vol}(V)$. Thus, Theorem 1 requires a considerably weaker assumption. Indeed, providing a characterization of $\mathrm{Disc}(\delta)$ in terms of eigenvalues, Theorem 1 answers a question posed by Chung and Graham [8, 9]. Furthermore, relying on Grothendieck's inequality and SDP duality, the proof of Theorem 1 employs quite different techniques than those used in [5–7].

In the present work we consider a concept of quasi-randomness that takes into account the graph's degree sequence. 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 [10] (dense graphs) and by Chung and Graham [9] (sparse graphs). Also in this setting it has been an open problem to derive eigenvalue separation from low discrepancy, and concerning this simpler concept of quasi-randomness, our techniques yield a similar result as Theorem 1 as well (details omitted).

Regular partitions. Szemerédi's original regularity lemma [20] shows that any *dense* graph $G = (V, E)$ (with $|E| = \Omega(|V|^2)$) can be partitioned into a bounded number of sets $V_1, \ldots, V_t$ such that almost all pairs (V_i, V_j) are quasi-random. This statement has become an important tool in various areas, including extremal graph theory and property testing. Furthermore, 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 [18] for a faster algorithm). Moreover, Frieze and Kannan [11] introduced a different algorithmic regularity concept, which yields better efficiency in terms of the desired approximation guarantee.

A version of the regularity lemma that applies to sparse graphs was established independently by Kohayakawa [17] and Rödl (unpublished). This result is of significance, e.g., in the theory of random graphs, cf. Gerke and Steger [12]. The regularity concept of Kohayakawa and Rödl is related to the notion of quasi-randomness from [9] and shows that any graph that satisfies a certain boundedness condition has a regular partition.

In comparison to the Kohayakawa-Rödl regularity lemma, the new aspect of Theorem 2 is that it takes into account the graph's degree distribution. Therefore, Theorem 2 applies to graphs with very irregular degree distributions, which were not covered by prior versions of the sparse regularity lemma. Further, Theorem 2 yields an efficient algorithm for computing a regular partition (see e.g. [13] for a non-polynomial time algorithm in the sparse setting). To achieve this algorithmic result, we build upon the algorithmic version of Grothendieck's inequality due to Alon and Naor [4]. Besides, our approach can easily be modified to obtain a polynomial time algorithm for computing a regular partition in the sense of Kohayakawa and Rödl, which was not known previously.

3 Preliminaries

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 entries corresponding to elements of S, and 0 otherwise. 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 $S \times S$. In addition, 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. Further, for a vector $\xi \in \mathbf{R}^V$ we let $\|\xi\|$ signify the ℓ_2-norm, and for a matrix we let $\|M\| = \sup_{0 \neq \xi \in \mathbf{R}^V} \frac{\|M\xi\|}{\|\xi\|}$ denote the spectral norm. If M is symmetric, then $\lambda_{\max}(M)$ denotes the largest eigenvalue of M.

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 $\|M\|_{\mathrm{cut}} = \max_{I,J \subset \mathcal{I}} \left| \sum_{i \in I, j \in J} m_{ij} \right|$. In addition, consider the following optimization problem:

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

Then $\mathrm{SDP}(M)$ 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 (cf. Alizadeh [2]). Hence, an optimal solution to $\mathrm{SDP}(M)$ can be approximated within any numerical precision, e.g., via the ellipsoid method [15]. Grothendieck [14] established the following relation between $\mathrm{SDP}(M)$ and $\|M\|_{\mathrm{cut}}$.

Theorem 4. *There is a constant $\theta > 1$ such that for all matrices M we have $\|M\|_{\mathrm{cut}} \leq \mathrm{SDP}(M) \leq \theta \cdot \|M\|_{\mathrm{cut}}$.*

The best current bounds on the above constant are $\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2\ln(1+\sqrt{2})}$ [14, 19]. Furthermore, 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 5. *There are a constant $\theta' > 0$ and a polynomial time algorithm* `ApxCutNorm` *that computes on input M 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 $\theta' \geq 0.03$.

4 Quasi-Randomness: Proof of Theorem 1

The proof of the first part of Theorem 1 is similar to the proof given in [8, Section 4] (cf. Appendix D). Thus, we focus on the second implication, and hence assume that $G = (V, E)$ is a graph that has $\mathrm{Disc}(\gamma\varepsilon^3)$, where $\gamma > 0$ signifies some small enough constant (e.g., $\gamma = (6400\theta)^{-1}$ suffices for the proof below). Moreover, we let d_v denote the degree of $v \in V$, $n = |V|$, and $\bar{d} = n^{-1}\sum_{v\in V} d_v$. In addition, we introduce a further property.

$\mathrm{Cut}(\varepsilon)$: We say G has $\mathrm{Cut}(\varepsilon)$, 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 $\|M\|_{\mathrm{cut}} < \varepsilon \cdot \mathrm{vol}(V)$, where $e(\{v\}, \{w\}) = 1$ if $\{v, w\} \in E$ and 0 otherwise.

Since for any $S \subset V$ we have $\langle M\mathbf{1}_S, \mathbf{1}_S\rangle = \frac{\mathrm{vol}(S)^2}{\mathrm{vol}(V)} - 2e(S)$, one can easily derive the following (cf. Appendix A).

Proposition 6. *Each graph that has* $\mathrm{Disc}(0.01\delta)$ *enjoys* $\mathrm{Cut}(\delta)$.

To show that $\mathrm{Disc}(\gamma\varepsilon^3)$ implies ess-Eig(ε), we proceed as follows. By Proposition 6, $\mathrm{Disc}(\gamma\varepsilon^3)$ implies $\mathrm{Cut}(100\gamma\varepsilon^3)$. Moreover, if G satisfies $\mathrm{Cut}(100\gamma\varepsilon^3)$, then Theorem 4 entails that not only the cut norm of M is small, but even the semidefinite relaxation $\mathrm{SDP}(M)$ satisfies $\mathrm{SDP}(M) < \beta\varepsilon^3\mathrm{vol}(V)$, for some β with $0 < \beta \leq 100\theta\gamma$. This bound on $\mathrm{SDP}(M)$ can be rephrased in terms of an eigenvalue minimization problem for a matrix closely related to M. More precisely, using the duality theorem for semidefinite programs, we can infer the following (cf. Appendix B).

Lemma 7. *For any symmetric* $n \times n$ *matrix* Q *we have*

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

Let $D = \mathrm{diag}(d_v)_{v\in V}$. Then Lemma 7 entails the following.

Lemma 8. *Suppose that* $\mathrm{SDP}(M) < \beta\varepsilon^3\mathrm{vol}(V)$ *for some* β, $0 < \beta < 1/64$. *Then there exists a subset* $W \subset V$ *of volume* $\mathrm{vol}(W) \geq (1-\varepsilon)\cdot\mathrm{vol}(V)$ *such that the matrix* $\mathcal{M} = D^{-\frac{1}{2}} M D^{-\frac{1}{2}}$ *satisfies* $\|\mathcal{M}_W\| < \varepsilon$.

Proof. Let $U = \{v \in V : d_v > \beta^{\frac{1}{3}}\varepsilon\bar{d}\}$. Then

$$\mathrm{vol}(V \setminus U) \leq \beta^{\frac{1}{3}}\varepsilon\bar{d}|V \setminus U| \leq \varepsilon\mathrm{vol}(V)/2. \tag{4}$$

Since $\mathrm{SDP}(M_U) \leq \mathrm{SDP}(M)$, Lemma 7 entails that there is a vector $\mathbf{1} \perp z \in \mathbf{R}^U$ such that $\lambda_{\max}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0\end{pmatrix} \otimes M_U - \mathrm{diag}\binom{z}{z}\right] < \beta\varepsilon^3\bar{d}$. Hence, setting $y = D_U^{-1}z$, we obtain

$$\lambda_{\max}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0\end{pmatrix} \otimes \mathcal{M}_U - \mathrm{diag}\begin{pmatrix} y \\ y\end{pmatrix}\right] < \beta^{\frac{2}{3}}\varepsilon^2, \tag{5}$$

because all entries of the diagonal matrix D_U exceed $\beta^{\frac{1}{3}}\varepsilon\bar{d}$. Moreover, as $z \perp \mathbf{1}$, we have

$$y \perp D_U\mathbf{1}. \tag{6}$$

Now, let $W = \{v \in U : |y_v| < \beta^{\frac{1}{3}}\varepsilon\}$ consist of all vertices v on which the "correcting vector" y is small. Since on W all entries of the diagonal matrix $\mathrm{diag}\binom{y}{y}$ are smaller than $\beta^{\frac{1}{3}}\varepsilon$ in absolute value, (5) yields

$$\lambda_{\max}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \mathcal{M}_W\right] < \beta^{\frac{1}{3}}\varepsilon + \beta^{\frac{2}{3}}\varepsilon^2 \leq 2\beta^{\frac{1}{3}}\varepsilon; \tag{7}$$

in other words, on W the effect of y is negligible. Further, (7) entails that $\|\mathcal{M}_W\| \leq 2\beta^{\frac{1}{3}}\varepsilon < \varepsilon$.

Finally, we need to show that $\mathrm{vol}(W)$ is large. To this end, we consider the set $S = \{v \in U : y_v < 0\}$ and let $\zeta = D_U^{\frac{1}{2}}\mathbf{1}_S$. Thus, for each $v \in U$ the entry ζ_v equals $d_v^{\frac{1}{2}}$ if $y_v < 0$, while $\zeta_v = 0$ if $y_v \geq 0$, so that $\|\zeta\|^2 = \mathrm{vol}(S)$. Hence, (5) yields that

$$\begin{aligned} 2\beta^{\frac{2}{3}}\varepsilon^2\mathrm{vol}(S) = 2\beta^{\frac{2}{3}}\varepsilon^2\|\zeta\|^2 &\geq \left\langle \left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \mathcal{M}_U - \mathrm{diag}\begin{pmatrix} y \\ y \end{pmatrix}\right] \cdot \begin{pmatrix} \zeta \\ \zeta \end{pmatrix}, \begin{pmatrix} \zeta \\ \zeta \end{pmatrix} \right\rangle \\ &= 2\left\langle \mathcal{M}_U\zeta, \zeta \right\rangle - 2\sum_{v \in S} d_v y_v = 2\left\langle M_U\mathbf{1}_S, \mathbf{1}_S \right\rangle - 2\sum_{v \in S} d_v y_v. \end{aligned} \tag{8}$$

Further, as $\mathrm{SDP}(M_U) \leq \mathrm{SDP}(M) \leq \beta\varepsilon^3\mathrm{vol}(V)$, Theorem 4 entails that $\langle M_U\mathbf{1}_S, \mathbf{1}_S\rangle \leq \|M_U\|_{\mathrm{cut}} \leq \beta\varepsilon^3\mathrm{vol}(V)$. Plugging this bound into (8) and recalling that $y_v < 0$ for all $v \in S$, we conclude that

$$\sum_{v \in S} d_v|y_v| \leq \beta^{\frac{2}{3}}\varepsilon^2\mathrm{vol}(S) + \beta\varepsilon^3\mathrm{vol}(V) \leq 2\beta^{\frac{2}{3}}\varepsilon^2\mathrm{vol}(V). \tag{9}$$

Hence, (6) entails that actually $\sum_{v \in U} d_v|y_v| \leq 4\beta^{\frac{2}{3}}\varepsilon^2\mathrm{vol}(V)$. As $|y_v| \geq \beta^{\frac{1}{3}}\varepsilon$ for all $v \in U \setminus W$, we obtain $\mathrm{vol}(U \setminus W) \leq 4\beta^{\frac{1}{3}}\varepsilon\mathrm{vol}(V) < \frac{1}{2}\varepsilon\mathrm{vol}(V)$. Thus, (4) yields $\mathrm{vol}(V \setminus W) < \varepsilon\mathrm{vol}(V)$, as desired. □

Finally, setting $\gamma = (6400\theta)^{-1}$ and combining Theorem 4, Proposition 6, and Lemma 8, we conclude if G has $\mathrm{Disc}(\gamma\varepsilon^3)$, then there is a set W such that $\mathrm{vol}(W) \geq (1-\varepsilon)\mathrm{vol}(V)$ and $\|\mathcal{M}_W\| < \varepsilon$. As $\mathcal{M}$ is closely related to the normalized Laplacian $L(G)$, one can infer via elementary linear algebra that the minor $L(G)_W$ corresponding to W satisfies $1 - \varepsilon \leq \lambda_2(L(G)_W) \leq \lambda_{|W|}(L(G)_W) \leq 1 + \varepsilon$, whence G has $\mathrm{ess\text{-}Eig}(\varepsilon)$ (cf. Appendix C for details).

5 The Algorithmic Regularity Lemma

In this section we present a polynomial time algorithm `Regularize` that computes for a given graph $G = (V, E)$ a partition satisfying **REG1** and **REG2**, provided that G satisfies the assumptions of Theorem 2. In particular, this will show that such a partition exists. We will outline `Regularize` in Section 5.1. The crucial ingredient is a subroutine `Witness` for checking whether a given pair (A, B) of subsets of V is ε-volume regular. This subroutine is the content of Section 5.2.

Throughout this section, we let $\varepsilon > 0$ be an arbitrarily small but fixed and $C > 0$ an arbitrarily large but fixed number. In addition, we define a sequence $(t_k)_{k\geq 1}$ by letting $t_1 = \lceil 2/\varepsilon \rceil$ and $t_{k+1} = t_k 2^{t_k}$. Let $k^* = \lceil C\varepsilon^{-3} \rceil$, $\eta = t_{k^*}^{-6}\varepsilon^{-8k^*}$, and choose $n_0 > 0$ big enough.

We always assume that $G = (V, E)$ is a graph on $n = |V| > n_0$ vertices that is (C, η)-bounded, and that $\mathrm{vol}(V) \geq \eta^{-1}n$.

5.1 The Algorithm `Regularize`

In order to compute the desired regular partition of its input graph G, the algorithm `Regularize` proceeds as follows. In its first step, `Regularize` computes any initial partition $\mathcal{P}^1 = \{V_i^1 : 0 \leq i \leq s_1\}$ such that each class V_i $(1 \leq i \leq s_1)$ has a decent volume.

Algorithm 9. `Regularize`(G)
Input: A graph $G = (V, E)$. *Output:* A partition of V.

1. Compute an initial partition $\mathcal{P}^1 = \{V_0^1 : 0 \leq i \leq s_1\}$ such that $\frac{1}{4}\varepsilon\mathrm{vol}(V) \leq \mathrm{vol}(V_i^1) \leq \frac{3}{4}\varepsilon\mathrm{vol}(V)$ for all $1 \leq i \leq s_1$; thus, $s_1 \leq 4\varepsilon^{-1}$. Set $V_0^1 = \emptyset$.

Then, 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 $\mathcal{P}^k$ $(k \geq 1)$. As soon as `Regularize` can verify that $\mathcal{P}^k$ satisfies both **REG1** and **REG2**, the algorithm stops.

To check whether the current partition $\mathcal{P}^k = \{V_i^k : 1 \leq i \leq s_1\}$ satisfies **REG2**, `Regularize` employs a subroutine `Witness`. Given a pair (V_i^k, V_j^k), `Witness` tries to check whether (V_i^k, V_j^k) is ε-volume-regular.

Proposition 10. *There is a polynomial time algorithm* `Witness` *that satisfies the following. Let* $A, B \subset V$ *be disjoint.*

1. *If* `Witness`(G, A, B) *answers "yes", then the pair* (A, B) *is* ε*-volume regular.*
2. *On the other hand, if the answer is "no", then* (A, B) *is not* $\varepsilon/200$*-volume regular. In this case* `Witness` *outputs a pair* (X^*, Y^*) *of subsets* $X^* \subset A$, $Y^* \subset B$ *such that* $\mathrm{vol}(X^*) \geq \frac{\varepsilon}{200}\mathrm{vol}(A)$, $\mathrm{vol}(Y^*) \geq \frac{\varepsilon}{200}\mathrm{vol}(B)$, *and* $|e(X^*, Y^*) - \varrho(A, B)\mathrm{vol}(X^*)\mathrm{vol}(Y^*)| > \frac{\varepsilon\mathrm{vol}(A)\mathrm{vol}(B)}{200\mathrm{vol}(V)}$.

We call a pair (X^*, Y^*) as in 2. an $\frac{\varepsilon}{200}$-*witness* for (A, B).

By applying `Witness` to each pair (V_i^k, V_j^k) of the partition $\mathcal{P}^k$, `Regularize` can single out a set $\mathcal{L}^k$ such that all pairs V_i, V_j with $(i, j) \notin \mathcal{L}^k$ are ε-volume regular. Hence, if

$$\sum_{(i,j)\in\mathcal{L}^k} \mathrm{vol}(V_i^k)\mathrm{vol}(V_j^k) < \varepsilon\mathrm{vol}(V)^2,$$

then $\mathcal{P}^k$ satisfies **REG2**. As we will see below that by construction $\mathcal{P}^k$ satisfies **REG1** for all k, in this case $\mathcal{P}^k$ is a feasible regular partition, whence `Regularize` stops.

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 the previously partition $\mathcal{P}^k$
4. call the procedure `Witness`$(G, V_i^k, V_j^k, \varepsilon)$.
   If it answers "no" and hence outputs an $\frac{\varepsilon}{200}$-witness (X_{ij}^k, X_{ji}^k) for (V_i^k, V_j^k), then add (i, j) to $\mathcal{L}^k$.
5. If $\sum_{(i,j)\in\mathcal{L}^k} \mathrm{vol}(V_i^k)\mathrm{vol}(V_j^k) < \varepsilon\mathrm{vol}(V)^2$, then output the partition $\mathcal{P}^k$ and halt.

If Step 5 does not halt, `Regularize` constructs a refinement $\mathcal{P}^{k+1}$ of $\mathcal{P}^k$. To this end, the algorithm decomposes each class V_i^k of $\mathcal{P}^k$ into up to 2^{s_k} pieces. Consider the sets X_{ij} with $(i, j) \in \mathcal{L}^k$ and define an 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 $u \in X_{ij} \leftrightarrow v \in X_{ij}$. Thus, the equivalence classes of $\equiv_i^k$ are the regions of the Venn diagram of the sets V_i and X_{ij} with $(i, j) \in \mathcal{L}^k$. Then `Regularize` obtains $\mathcal{P}^{k+1}$ as follows.

6. Let $\mathcal{C}^k$ be the set of all equivalence classes of the relations $\equiv_i^k$ $(1 \leq i \leq s_k)$. Moreover, let $\mathcal{C}_*^k = \{V_1^{k+1}, \dots, V_{s_{k+1}}^{k+1}\}$ be the set of all classes $W \in \mathcal{C}$ such that $\mathrm{vol}(W) > \varepsilon^{4(k+1)}\mathrm{vol}(V)/(15t_{k+1}^3)$. Finally, let $V_0^{k+1} = V_0^k \cup \bigcup_{W \in \mathcal{C}^k \setminus \mathcal{C}_*^k} W$, and set $\mathcal{P}^{k+1} = \{V_i^{k+1} : 0 \leq i \leq s_{k+1}\}$.

Since for each i there are at most s_k indices j such that $(i,j) \in \mathcal{L}^k$, in $\mathcal{P}^{k+1}$ every class V_i^k gets split into at most 2^{s_k} pieces. Hence, $s_{k+1} \leq s_k 2^{s_k}$. Thus, as $s_1 \leq t_1$, we conclude that $s_k \leq t_k$ for all k. Therefore, our choice of η ensures that $\mathrm{vol}(V_i^{k+1}) \geq \eta \mathrm{vol}(V)$ for all $1 \leq i \leq s_{k+1}$ (because Step 6 puts all equivalence classes $W \in \mathcal{C}^k$ of "extremely small" volume into the exceptional class). Moreover, it is easily seen that $\mathrm{vol}(V_0^{k+1}) \leq \varepsilon(1 - 2^{k+2})\mathrm{vol}(V)$. In effect, $\mathcal{P}^{k+1}$ satisfies **REG1**.

Thus, to complete the proof of Theorem 2 it just remains to show that `Regularize` will actually succeed and output a partition $\mathcal{P}^k$ for some $k \leq k^*$. To show this, we define the *index* of a partition $\mathcal{P} = \{V_i : 0 \leq i \leq s\}$ as

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

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

Proposition 11. *If $G = (V, E)$ is a (C, η)-bounded graph and $\mathcal{P} = \{V_i\colon\ 0 \leq 1 \leq t\}$ is a partition of V with $\mathrm{vol}(V_i) \geq \eta \mathrm{vol}(V)$ for all $i \in \{1, \dots, t\}$, then $\mathrm{ind}(\mathcal{P}) \leq C$.*

Lemma 11 entails that $\mathrm{ind}(\mathcal{P}^k) \leq C$ 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 12. *If $\sum_{(i,j) \in \mathcal{L}^k} \mathrm{vol}(V_i^k)\mathrm{vol}(V_j^k) \geq \varepsilon \mathrm{vol}(V)^2$, then $\mathrm{ind}(\mathcal{P}^{k+1}) \geq \mathrm{ind}(\mathcal{P}^k) + \varepsilon^3/8$.*

Since the index of the initial partition $\mathcal{P}^1$ is non-negative, Lemmas 11 and 12 readily imply that `Regularize` will terminate and output a feasible partition $\mathcal{P}^k$ for some $k < k^*$.

Finally, we point out that the overall running time of `Regularize` is polynomial. For the running time of Steps 1–3 and 5–6 is $O(\mathrm{vol}(V))$, and the running time of Step 4 is polynomial due to Proposition 10.

5.2 The Procedure `Witness`

The subroutine `Witness` for Proposition 10 employs the algorithm `ApxCutNorm` from Theorem 5 for approximating the cut norm as follows.

Algorithm 13. `Witness`(G, A, B)
Input: A graph $G = (V, E)$, disjoint sets $A, B \subset V$, and a number $\varepsilon > 0$.
Output: A partition of V.

1. Set up a matrix $M = (m_{vw})_{(v,w) \in A \times B}$ with entries $m_{vw} = 1 - \varrho(A, B) d_v d_w$ if v, w are adjacent in G, and $m_{vw} = -\varrho(A, B) d_v d_w$ otherwise.
   Call `ApxCutNorm`(M) to compute sets $X \subset A$, $Y \subset B$ such that $|\langle M\mathbf{1}_X, \mathbf{1}_Y\rangle| \geq \frac{3}{100}\|M\|_{\mathrm{cut}}$.
2. If $|\langle M\mathbf{1}_X, \mathbf{1}_Y\rangle| < 3\varepsilon/100$, then return "yes".
3. Otherwise, pick an arbitrary set $X' \subset A \setminus X$ of volume $\frac{3\varepsilon}{100}\mathrm{vol}(A) \leq \mathrm{vol}(X') \leq \frac{4\varepsilon}{100}\mathrm{vol}(A)$.
   - If $\mathrm{vol}(X) \geq \frac{3\varepsilon}{100}\mathrm{vol}(A)$, then let $X^* = X$.
   - If $\mathrm{vol}(X) < \frac{3\varepsilon}{100}\mathrm{vol}(A)$ and $|e(X', Y) - \varrho(A, B)\mathrm{vol}(X')\mathrm{vol}(Y)| > \frac{\varepsilon \mathrm{vol}(A)\mathrm{vol}(B)}{100\mathrm{vol}(V)}$, set $X^* = X'$.
   - Otherwise, set $X^* = X \cup X'$.

4. Pick a further set $Y' \subset B \setminus Y$ of volume $\frac{\varepsilon}{200}\mathrm{vol}(B) \leq \mathrm{vol}(Y') \leq \frac{2\varepsilon}{300}\mathrm{vol}(B)$.
   - If $\mathrm{vol}(Y) \geq \frac{\varepsilon}{200}\mathrm{vol}(B)$, then let $Y^* = Y$.
   - If $\mathrm{vol}(Y) < \frac{\varepsilon}{200}\mathrm{vol}(B)$ and $|e(X^*, Y') - \varrho(A,B)\mathrm{vol}(X^*)\mathrm{vol}(Y')| > \frac{\varepsilon \mathrm{vol}(A)\mathrm{vol}(B)}{200\mathrm{vol}(V)}$, let $Y^* = Y'$.
   - Otherwise, set $Y^* = Y \cup Y'$.
5. Answer “no” and output (X^*, Y^*) as an $\varepsilon/8$-witness.

Given the graph G along with two disjoint sets $A, B \subset V$, `Witness` sets up a matrix M. The crucial property of M is that for any two subsets $S \subset A$ and $T \subset B$ we have $\langle M\mathbf{1}_S, \mathbf{1}_T\rangle = e(S,T) - \varrho(A,B)\mathrm{vol}(S)\mathrm{vol}(T)$. Therefore, if $\|M\|_{\text{cut}} \leq \varepsilon \mathrm{vol}(A)\mathrm{vol}(B)/\mathrm{vol}(V)$, then the pair (A,B) is ε-volume regular. Hence, in order to find out whether (A,B) is ε-volume regular, `Witness` employs the algorithm `ApxCutNorm` to approximate $\|M\|_{\text{cut}}$. If Step 2 of `Witness` answers “yes”, then (A,B) is ε-volume regular, because `ApxCutNorm` achieves an approximation ratio $> \frac{3}{100}$ by Theorem 5.

On the other hand, if `ApxCutNorm` yields sets X, Y such that $|\langle M\mathbf{1}_X, \mathbf{1}_Y\rangle| > \frac{3\varepsilon \mathrm{vol}(A)\mathrm{vol}(B)}{100\mathrm{vol}(V)}$, then `Witness` constructs an $\varepsilon/200$-witness for (A,B). Indeed, if the volumes of X and Y are “large enough” – say, $\mathrm{vol}(X) \geq \frac{\varepsilon}{200}\mathrm{vol}(A)$ and $\mathrm{vol}(Y) \geq \frac{\varepsilon}{200}\mathrm{vol}(B)$ – then (X,Y) actually is an $\varepsilon/200$-witness. However, as `ApxCutNorm` does not guarantee any lower bound on $\mathrm{vol}(X)$, $\mathrm{vol}(Y)$, Steps 3–5 try to enlarge the sets X, Y a little if their volume is too small. Finally, it is straightforward to verify that this procedure yields an $\varepsilon/200$-witness (X^*, Y^*), which entails Proposition 10.

6 An Application: Max Cut

As an application of Theorem 2 and, in particular, the polynomial time algorithm `Regularize` for computing a regular partition, we obtain the following algorithm for approximating the max cut of a graph $G = (V,E)$ that satisfies the assumptions of Theorem 3.

Algorithm 14. `ApxMaxCut`(G)
Input: A graph $G = (V,E)$. *Output:* A cut $(S, \bar{S})$ of G.

1. Use `Regularize` to compute an $\varepsilon = \frac{\delta}{40C}$-volume regular partition $\mathcal{P} = \{V_i : 0 \leq i \leq t\}$ of G.
2. Determine an optimal solution $(c_1, \ldots, c_t)$ to the optimization problem

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

3. For each $1 \leq i \leq t$ let $S_i \subset V_i$ be a subset of volume $\mathrm{vol}(S_i) \sim \varepsilon c_i \mathrm{vol}(V_i)$. Output $S = \bigcup_{i=1}^t S_i$ and $\bar{S} = V \setminus S$.

The basic insight behind `ApxMaxCut` is the following. If (V_i, V_j) is an ε-volume regular pair of $\mathcal{P}$, then for any subsets $X, X' \subset V_i$ and $Y, Y' \subset V_j$ such that $\mathrm{vol}(X) = \mathrm{vol}(X')$ and $\mathrm{vol}(Y) = \mathrm{vol}(Y')$ the condition **REG2** ensures that $|e(X,Y) - e(X',Y')| \leq \frac{2\varepsilon \mathrm{vol}(V_i)\mathrm{vol}(V_j)}{\mathrm{vol}(V)}$; 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 volume are “interchangeable”.

Therefore, to compute a good cut $(S, \bar{S})$ of G we just have to optimize the *proportion of volume* 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 volume we choose. However, determining the optimal fraction of volume is still a somewhat involved (essential continuos) optimization problem. Hence, in order to discretize this problem, we chop each V_i into at most ε^{-1} chunks of volume $(1 + o(1))\varepsilon \mathrm{vol}(V_i)$. 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`.

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 $\mathrm{vol}(S_i) \sim \varepsilon c_i \mathrm{vol}(V_i)$.

To prove that `ApxMaxCut` does indeed guarantee an approximation ratio of $1-\delta$, we have to cope with a number of technicalities. For instance, we have to show that the number of edges belonging to irregular pairs of the partition $\mathcal{P}$ as well as the number of edges *inside* the classes V_i are negligible. We omit the details of the analysis from this extended abstract.

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 STOC (2004) 72–80
5. Bilu, Y., Linial, N.: Lifts, discrepancy and nearly optimal spectral gap. Combinatorica, to appear
6. B. Bollobás and V. Nikiforov: Graphs and Hermitian matrices: discrepancy and singular values, Discrete Math. 285 (2004), 17-32
7. Butler, S.: On eigenvalues and the discrepancy of graphs. preprint
8. Chung, F., Graham, R.: Quasi-random graphs with given degree sequences. Preprint (2005).
9. Chung, F., Graham, R.: Sparse quasi-random graphs. Combinatorica **22** (2002) 217–244
10. Chung, F., Graham, R., Wilson, R.M.: Quasi-random graphs. Combinatorica **9** (1989) 345–362
11. Frieze, A., Kannan, R.: Quick approximation to matrices and applications. Combinatorica **19** (1999) 175–200
12. 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
13. Gerke, S., Steger, A.: A characterization for sparse ε-regular pairs. The Electronic J. Combinatorics **14** (2007), R4, 12pp
14. Grothendieck, A.: Résumé de la théorie métrique des produits tensoriels topologiques. Bol. Soc. Mat. Sao Paulo **8** (1953) 1–79
15. Grötschel, M., Lovász, L., Schrijver, A.: Geometric algorithms and combinatorial optimization. Springer (1988)
16. Håstad, J.: Some optimal inapproximability results. J. of the ACM **48** (2001) 798–859
17. Kohayakawa, Y.: Szemerédi's regularity lemma for sparse graphs. In Cucker, F., Shub, M. (eds.): Foundations of computational mathematics (1997) 216–230.
18. Kohayakawa, Y., Rödl, V., Thoma, L.: An optimal algorithm for checking regularity. SIAM J. Comput., **32** (2003) 1210–1235
19. Krivine, J.L.: Sur la constante de Grothendieck. C. R. Acad. Sci. Paris Ser. A-B **284** (1977) 445–446
20. Szemerédi, E.: Regular partitions of graphs. Problémes Combinatoires et Théorie des Graphes Colloques Internationaux CNRS **260** (1978) 399-401
21. Trevisan, L., Sorkin, G., Sudan, M., Williamson, D.: Gadgets, approximation, and linear programming. SIAM J. Computing **29** (2000) 2074–2097

A Proof of Proposition 6

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

$$|\langle M\mathbf{1}_S, \mathbf{1}_T \rangle| \leq 0.03\delta\mathrm{vol}(V) \text{ if } S \cap T = \emptyset, \tag{10}$$

$$|\langle M\mathbf{1}_S, \mathbf{1}_T \rangle| \leq 0.02\delta\mathrm{vol}(V) \text{ if } S = T. \tag{11}$$

Letting $Z = X \cap Y$ and combining (10) and (11), we obtain

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

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

To prove (10), we note that $\mathrm{Disc}(0.01\delta)$ implies that

$$\left|e(S) - \frac{\mathrm{vol}(S)^2}{2\mathrm{vol}(V)}\right| \le 0.01\delta \mathrm{vol}(V), \tag{12}$$

$$\left|e(T) - \frac{\mathrm{vol}(T)^2}{2\mathrm{vol}(V)}\right| \le 0.01\delta \mathrm{vol}(V), \tag{13}$$

$$\left|e(S \cup T) - \frac{(\mathrm{vol}(S) + \mathrm{vol}(T))^2}{2\mathrm{vol}(V)}\right| \le 0.01\delta \mathrm{vol}(V). \tag{14}$$

Moreover, since S, T are disjoint, we have $e(S,T) = e(S \cup T) - e(S) - e(T)$. Therefore, the desired bound on $|\langle M\mathbf{1}_S, \mathbf{1}_T\rangle| = \left|e(S,T) - \frac{\mathrm{vol}(S)\mathrm{vol}(T)}{\mathrm{vol}(V)}\right|$ follows from (12)–(14). Finally, as $|\,\langle M\mathbf{1}_S, \mathbf{1}_S\rangle\,| = 2\left|e(S) - \frac{\mathrm{vol}(S)^2}{2\mathrm{vol}(V)}\right|$, we obtain (11).

B Proof of Lemma 7

Suppose that Q is an $n \times n$ matrix, and let $\mathcal{Q} = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes Q$. If $A = (a_{ij})_{1\le i,j\le n}, B = (b_{ij})_{1\le i,j\le n}$ are symmetric $n \times n$ matrices, we use the notation $A \le B$ to state that $B - A$ is positive semidefinite. Moreover, $\mathrm{diag}(A) \in \mathbf{R}^n$ signifies the diagonal of A, and $\langle A, B\rangle = \sum_{i,j=1}^n a_{ij}b_{ij}$.

Lemma 15. *Consider the semidefinite program*

$$\begin{aligned}\mathrm{DSDP}(Q) = \min\, &\langle \mathbf{1}, y\rangle \\ \textit{s.t.}\ &\mathcal{Q} \le \mathrm{diag}(y), \quad y \in \mathbf{R}^{2n}.\end{aligned}$$

Then $\mathrm{SDP}(Q) = \mathrm{DSDP}(Q)$.

Proof. We first rewrite the vector program $\mathrm{SDP}(Q)$ in the standard form of a semidefinite program:

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

Then the lemma follows directly from the SDP duality theorem as stated in [2]. □

$\mathrm{DSDP}(Q)$ is dual of $\mathrm{SDP}(Q)$. To infer Lemma 7, 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 16. *Let* $\mathrm{DSDP}'(Q) = \min 2\langle \mathbf{1}, y'\rangle$ *s.t.* $\mathcal{Q} \le \mathrm{diag}\binom{1}{1} \otimes y',\ y' \in \mathbf{R}^n$. *Then* $\mathrm{DSDP}(Q) = \mathrm{DSDP}'(Q)$.

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

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., $\mathcal{F}(Q)$ is closed under swapping 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 $\frac{1}{2}(y + \mathcal{I}y) \in \mathcal{F}(Q)$. 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 to $\mathrm{DSDP}(Q)$ implies that y' is feasible to $\mathrm{DSDP}'(Q)$. Thus, we conclude that

$$\mathrm{DSDP}'(Q) \leq 2 \left\langle \mathbf{1}, y' \right\rangle = \left\langle \mathbf{1}, y \right\rangle = \mathrm{DSDP}(Q).$$

To show that $\mathcal{F}(Q)$ is closed under $\mathcal{I}$, consider a vector $y \in \mathcal{F}(Q)$. Then we know that

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

Furthermore, our objective is to show that

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

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 $\xi = \binom{\alpha}{\beta}$, where $\alpha, \beta \in \mathbf{R}^n$, and set $\eta = \mathcal{I}\xi = \binom{\beta}{\alpha}$. Then

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

because Q is symmetric. □

Proof of Lemma 7. 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 15 and 16, it suffices to prove that $\mathrm{DSDP}'(Q) = \mathrm{DSDP}''(Q)$.

To see that $\mathrm{DSDP}''(Q) \leq \mathrm{DSDP}'(Q)$, consider an optimal solution y' to $\mathrm{DSDP}'(Q)$. Let $\lambda = n^{-1} \langle \mathbf{1}, y' \rangle$ and $z = \lambda \mathbf{1} - y'$. Then $2z \perp \mathbf{1}$, i.e., $2z$ is a feasible solution to $\mathrm{DSDP}''(Q)$. Furthermore, the feasibility of y' implies that

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

where $\boldsymbol{E}$ is the identity matrix. Hence,

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

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

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

Then the definition of μ implies that $\begin{pmatrix}0\,1\\1\,0\end{pmatrix}\otimes Q - \mathrm{diag}\binom{1}{1}\otimes z \leq \mu\boldsymbol{E}$, whence

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

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

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

as desired. □

C Proof of Theorem 1, 2nd part

Let $W \subset V$ be a set of volume $\mathrm{vol}(W) \geq (1-\varepsilon)\mathrm{vol}(V)$ such that $\|\mathcal{M}_W\| < \varepsilon$. Moreover, let $\Delta = (\sqrt{d_v})_{v\in W} \in \mathbf{R}^W$. Then we can rewrite $\mathcal{M}_W$ as a difference $\mathcal{M}_W = \mathrm{vol}(V)^{-1}\Delta\Delta^T - \mathcal{L}_W$, where for $v, w \in W$ the corresponding entry of $\mathcal{L}_W$ is $-(d_v d_w)^{-\frac{1}{2}}$ if v, w are adjacent, and 0 otherwise. Furthermore, the normalized Laplacian L of G satisfies $L_W = \boldsymbol{E} - \mathcal{L}_W$. Therefore, for all unit vectors $\xi \perp \Delta$ we have

$$|\langle L_W\xi, \xi\rangle - 1| = |\langle \mathcal{L}_W\xi, \xi\rangle| = |\langle \mathcal{M}_W\xi, \xi\rangle| \leq \|\mathcal{M}_W\| < \varepsilon. \tag{15}$$

In addition, since $\|\Delta\|^2 = \mathrm{vol}(W) \geq \frac{1}{2}\mathrm{vol}(V)$, we have

$$\frac{\|L_W\Delta\|^2}{\|\Delta\|^2} = \sum_{v\in W}\frac{2(e(v,W)-d_v)^2}{d_v\cdot\mathrm{vol}(V)} \leq 2\sum_{v\in W}\frac{d_v - e(v,W)}{\mathrm{vol}(V)} \leq \frac{\mathrm{vol}(V\setminus W)}{\mathrm{vol}(V)} < 2\varepsilon. \tag{16}$$

Finally, combining (15) and (16) with the Rayleigh characterizations of $\lambda_2(L_W)$ and $\lambda_{\max}(L_W)$, we conclude that $1-\varepsilon < \lambda_2(L_W) \leq \lambda_{\max}(L_W) < 1+\varepsilon$, thereby completing the proof.

D Proof of Theorem 1, 1st part

Let $W \subset V$ be a set of volume $\mathrm{vol}(W) \geq (1-\varepsilon)\mathrm{vol}(V)$ such that the normalized Laplacian L of G satisfies $1-\varepsilon \leq \lambda_2(L_W) \leq \lambda_{\max}(L_W) \leq 1+\varepsilon$; we may assume without loss of generality that $\varepsilon < 10^{-6}$. Our goal is to show that then G has $\mathrm{Disc}(10\sqrt{\varepsilon})$. Let $\Delta = (\sqrt{d_v})_{v\in W} \in \mathbf{R}^W$, and set $\mathcal{M}_W = \mathrm{vol}(V)^{-1}\Delta\Delta^T - \mathcal{L}_W$, where for $v, w \in W$ the corresponding entry of $\mathcal{L}_W$ is $-(d_v d_w)^{-\frac{1}{2}}$ if v, w are adjacent, and 0 otherwise. Note that the normalized Laplacian L of G satisfies $L_W = \boldsymbol{E} - \mathcal{L}_W$. Therefore, for all unit vectors $\xi \perp \Delta$ we have

$$L_W\xi - \xi = -\mathcal{L}_W\xi = \mathcal{M}_W\xi. \tag{17}$$

If $S \subset W$, then we let Δ_S denote the vector that coincides with Δ on all coordinates $v \in S$, and whose entries on $W \setminus S$ equal 0. Then

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

Therefore, it is straightforward to derive the assertion from the following lemma.

Lemma 17. *We have* $\|\mathcal{M}_W\| \leq 10\sqrt{\varepsilon}$.

Proof. Let ζ be an eigenvector of L_W with eigenvalue $\lambda_1(L_W)$ of unit length. Since

$$\langle L_W \Delta, \Delta \rangle = e(W, V \setminus W) \leq \mathrm{vol}(V \setminus W) \leq \varepsilon \mathrm{vol}(V),$$

and $\|\Delta\|^2 = \mathrm{vol}(W) \geq 0.99\mathrm{vol}(V)$, we conclude that

$$0 \leq \lambda_1(L_W) \leq 2\varepsilon. \tag{18}$$

Thus, decomposing Δ as $\frac{\Delta}{\|\Delta\|^2} = s\zeta + t\chi$, where $s^2 + t^2 = 1$ and $\chi \perp \zeta$ is a unit vector, we obtain

$$2\varepsilon \geq \|\Delta\|^{-2} \langle L_W \Delta, \Delta \rangle = s^2 \langle L_W \zeta, \zeta \rangle + t^2 \langle L_W \chi, \chi \rangle \geq \frac{t^2}{2} \langle L_W \chi, \chi \rangle,$$

because $\lambda_2(L_W) \geq 1 - \varepsilon$. Consequently,

$$t^2 \leq 4\varepsilon, \quad s^2 \geq 1 - 4\varepsilon. \tag{19}$$

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

$$x = \langle \zeta, \xi \rangle = s^{-1} \left\langle \frac{\Delta}{\|\Delta\|}, \xi \right\rangle - \frac{t}{s} \langle \chi, \xi \rangle.$$

Hence, as $\xi \perp \Delta$, (19) entails that

$$x^2 \leq 5\varepsilon, \quad y^2 \geq 1 - 5\varepsilon. \tag{20}$$

Combining, (17), (18) and (20), we conclude that

$$\|\mathcal{M}_W \xi\| = \|L_W \xi - \xi\| \leq x(1 - \lambda_1(L_W)) + y\|L_W \eta - \eta\| \leq 3\sqrt{\varepsilon}.$$

Hence, we have established that

$$\sup_{0 \neq \xi \perp \Delta} \frac{\|\mathcal{M}_W\|}{\|\xi\|} \leq 3\sqrt{\varepsilon}. \tag{21}$$

Furthermore, a direct computation yields

$$\frac{|\langle M_W \Delta, \Delta \rangle|}{\|\Delta\|^2} \leq \frac{2\mathrm{vol}(V \setminus W)}{\mathrm{vol}(V)} < 2\varepsilon. \tag{22}$$

Finally, combining (21) and (22), we conclude that $\|M_W\| \leq 10\sqrt{\varepsilon}$. $\square$

E Construction of a Graph with a Small Spectral Gap

In this section we show that the second part of Theorem 1 is tight up to the precise value of the constant γ. To this end, we sketch a (probabilistic) construction of a graph $G = (V, E)$ on n vertices that has $\mathrm{Disc}(\varepsilon)$ but does not have ess-Eig$(C^{-2}\alpha)$, where $C > 0$ is a sufficiently large constant and $\alpha = C^{-1}\varepsilon^{\frac{1}{3}}$. We assume $\varepsilon > 0$ is a sufficiently small number, and choose $n = n(\varepsilon)$ sufficiently large.

Let $X \subset V = \{1, \ldots, n\}$ be a set of cardinality αn. Then G is a graph with the following properties chosen uniformly at random.

1. All vertices in $V \setminus X$ have degree $d = n^{\frac{1}{4}}$.
2. All vertices in X have degree αd.
3. Each vertex in X has exactly $d\mathrm{vol}(X)/\mathrm{vol}(V)$ neighbors in X.
4. All vertices in $V \setminus X$ have the same number of neighbors in X.

Thus, G features a rather large set X of vertices of relatively small degree such that X is much denser than one would expect in a random graph $G(\boldsymbol{d})$ with the same degree distribution as G. We say that G has some property *with high probability* ("w.h.p.") if the probability that X has the property is $1-o(1)$ as $n \to \infty$.

To show that w.h.p. G does not have ess-Eig$(C^{-2}\alpha)$, let $Y \subset X$ be a set of size $|Y| = \frac{1}{2}|X|$. Since G is a random graph, w.h.p. *any* such set Y satisfies $e(Y) \geq 0.1e(X)$. Furthermore, the construction of G ensures that

$$\frac{2e(Y)\mathrm{vol}(V)}{\mathrm{vol}(Y)^2} \geq 0.9\alpha^{-1}. \tag{23}$$

On the other hand, if $W \subset V$ is such that $\mathrm{vol}(V \setminus W) \leq C^{-2}\alpha\mathrm{vol}(V)$ and $1-C^{-2}\alpha \leq \lambda_2(L(G)_W) \leq \lambda_{\max}(L(G)_W) \leq 1 + C^{-2}\alpha$, then any $Y \subset X \setminus W$, $|Y| \geq \frac{1}{2}|X|$ would satisfy

$$\left|2e(Y) - \frac{\mathrm{vol}(Y)^2}{\mathrm{vol}(V)}\right| \leq C^{-1}\alpha\mathrm{vol}(Y) \tag{24}$$

w.h.p., as can be shown by similar arguments as in Appendix D. Since (23) contradicts (24), we conclude that G violates ess-Eig$(C^{-2}\alpha)$ w.h.p.

Furthermore, to prove that G has $\mathrm{Disc}(\varepsilon)$, we observe that for any $Y \subset X$, $Z \subset V \setminus X$ we have

$$\frac{2e(Y)}{\mathrm{vol}(V)} \leq \frac{\alpha dn\mathrm{vol}(X)}{\mathrm{vol}(V)^2} \leq 2\alpha^3 < 0.01\varepsilon, \tag{25}$$

$$\left|e(Y,Z) - \frac{\mathrm{vol}(Y)\mathrm{vol}(Z)}{\mathrm{vol}(G)}\right| \leq \alpha^2 dn \left|1 - \frac{dn}{\mathrm{vol}(G)}\right| \leq 3\alpha^3 dn \leq 0.01\varepsilon\mathrm{vol}(G). \tag{26}$$

Since the subgraph of G induced on $V \setminus X$ is a random regular graph, (25) and (26) entail that G satisfies $\mathrm{Disc}(\varepsilon)$ w.h.p.

F Proof of Lemma 12

To prove of the Lemma 12 we follow the lines of the original proof of Semerédi. First we need the following observation.

Proposition 18. *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 $\mathcal{P}'$ 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

$$\begin{aligned}\mathrm{ind}(\mathcal{P}') &= \sum_{1\leq i<j\leq s} \frac{e^2(V'_i, V'_j)}{\mathrm{vol}(V'_i)\mathrm{vol}(V'_j)} \leq \sum_{1\leq k<l\leq t} \sum_{\substack{i\in I_k \\ j\in I_l}} \frac{e^2(V'_i, V'_j)}{\mathrm{vol}(V'_i)\mathrm{vol}(V'_j)} \\ &\geq \sum_{1\leq k<l\leq t} \frac{\left(\sum_{\substack{i\in I_k \\ j\in I_l}} e(V'_i, V'_j)\right)^2}{\sum_{\substack{i\in I_k \\ j\in I_l}} \mathrm{vol}(V'_i)\mathrm{vol}(V'_j)} = \sum_{1\leq k<l\leq t} \frac{e^2(V_k, V_l)}{\mathrm{vol}(V_k)\mathrm{vol}(V_l)} = \mathrm{ind}(\mathcal{P}).\end{aligned}$$

$\square$

Furthermore the proof will use the following defect-form of the Cauchy-Schwarz-Lemma.

Lemma 19 (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

$$\sum_{i\in I} \sigma_i \varrho_i^2 \geq \varrho^2 + \nu^2 \sigma_J.$$

Lastly, for technical reasons we state the following proposition. Its proof is straightforward and we omit it here.

Proposition 20. *Let $1/5 > \delta > 0$, $G = (V, E)$ and $A, B \subset V$ be disjoint subsets of V. Furthermore let $A' \subset A$ and $B' \subset B$ with $\mathrm{vol}(A \setminus A') < \delta \mathrm{vol}(A)$ and $\mathrm{vol}(B \setminus B') < \delta \mathrm{vol}(B)$. Then the following inequalities hold*

$$\left| \frac{e(A,B)}{\mathrm{vol}(A)\mathrm{vol}(B)} - \frac{e(A',B')}{\mathrm{vol}(A')\mathrm{vol}(B')} \right| \leq \frac{5\delta}{\min\{\mathrm{vol}(A), \mathrm{vol}(B)\}} \tag{27}$$

$$\left| \frac{e^2(A,B)}{\mathrm{vol}(A)\mathrm{vol}(B)} - \frac{e^2(A',B')}{\mathrm{vol}(A')\mathrm{vol}(B')} \right| \leq 15\delta. \tag{28}$$

Proof of the Lemma 12

Without loss of generality let assume $\varepsilon \leq 1/8$ and let $K \subset V$ be the union of the equivalence classes with a negligible volume size, more precisely

$$K = \bigcup_{W \in \mathcal{C}_*^k \setminus \mathcal{C}^k} \mathrm{vol}(W) = \bigcup \left\{ W \in \mathcal{C}^k : \ \mathrm{vol}(W) \leq \frac{\varepsilon^4 (k+1) \mathrm{vol}(V)}{15 t_{k+1}^3} \right\}.$$

Now let $\mathcal{P}' = \{V_i' : \ 0 \leq i \leq s_k\}$ be an auxiliary 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 $\mathrm{ind}(\mathcal{P}^{k+1}) \geq \mathrm{ind}(\mathcal{P}^k) + \varepsilon^3/8$ we will proceed in two steps. In the first step we will compare the index of $\mathcal{P}'$ to the index of $\mathcal{P}^k$. This will yield the following.

Claim 1 $|\mathrm{ind}(\mathcal{P}^k) - ind(\mathcal{P}')| \leq \varepsilon^4$.

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

Claim 2 $\mathrm{ind}(\mathcal{P}^{k+1}) \geq \mathrm{ind}(\mathcal{P}) + \varepsilon^3/4$.

Together, with $\varepsilon \leq 1/8$, this yields an index increment

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

$\square$

Proof of Claim 1. 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 $\mathrm{vol}(V_i^k) \geq \varepsilon^{4k}\mathrm{vol}(V)/15t_k^3$. Thus we have

$$\mathrm{vol}(V_i') \geq \mathrm{vol}(V_i^k) - \mathrm{vol}(K) \geq \mathrm{vol}(V_i^k) - \varepsilon^4\left(\frac{\varepsilon^{4k}}{15}\frac{\mathrm{vol}(G)}{t_{k+1}^2}\right) \geq \left(1 - \frac{\varepsilon^4}{15t_k^2}\right)\mathrm{vol}(V_i^k).$$

Analogously $\mathrm{vol}(V_j') \geq \left(1 - \varepsilon^4/(15t_k^2)\right)\mathrm{vol}(V_j^k)$ holds. In effect, using the Proposition 20 we get

$$\left|\frac{e^2(V_i', V_j')}{\mathrm{vol}(V_i')\mathrm{vol}(V_j')} - \frac{e^2(V_i^k, V_j^k)}{\mathrm{vol}(V_i^k)\mathrm{vol}(V_j^k)}\right| \leq \frac{\varepsilon^4}{t_k^2}.$$

Consequently

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

□

Proof of Claim 2. 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 the witness of irregularity. Then, for $X = X_{ij}^k \setminus K \subset A$ and $Y = X_{ji}^k \setminus K \subset B$, we have

$$\begin{aligned}\left|\frac{e(X,Y)}{\mathrm{vol}(X)\mathrm{vol}(Y)} - \frac{e(A,B)}{\mathrm{vol}(A)\mathrm{vol}(B)}\right| &= \varepsilon\frac{\mathrm{vol}(A)\mathrm{vol}(B)}{\mathrm{vol}(X_{ij}^k)\mathrm{vol}(X_{ji}^k)\mathrm{vol}(G)} - \frac{10\varepsilon^4}{t_{k+1}\mathrm{vol}(B)}\\ &\geq \frac{\varepsilon}{2}\frac{\mathrm{vol}(A)\mathrm{vol}(B)}{\mathrm{vol}(X)\mathrm{vol}(Y)\mathrm{vol}(G)}\end{aligned}$$

due to Proposition 20. Thus, (X, Y) witnesses that (A, B) is not $\varepsilon/2$-volume-regular.

Now we'll use the Lemma 19 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

$$\sigma_{uv} = \frac{\deg(u)\deg(v)}{\mathrm{vol}(A)\mathrm{vol}(B)} \quad \text{and} \quad d_{uv} = \varrho(V^{k+1}(u), V^{k+1}(y))$$

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

$$\sum_{(u,v)\in I}\sigma_{uv} = 1 \quad \text{and} \quad d = \sum_{(u,v)\in I}\sigma_{uv}d_{uv} = \varrho(A, B).$$

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

$$\begin{aligned}\frac{1}{\sigma_J}\sum_{(u,v)\in J}\sigma_{uv}d_{uv} &= \frac{\mathrm{vol}(A)\mathrm{vol}(B)}{\mathrm{vol}(X)\mathrm{vol}(Y)}\sum_{\substack{V_i^{k+1}\subset A\\ V_j^{k+1}\subset B}}\sum_{\substack{u\in V_i^{k+1}\\ v\in V_j^{k+1}}}\frac{\deg(u)\deg(v)}{\mathrm{vol}(A)\mathrm{vol}(B)}\varrho(V_i^{k+1}, V_j^{k+1})\\ &= \frac{e(X,Y)}{\mathrm{vol}(X)\mathrm{vol}(Y)} = \varrho(X, Y) = \varrho(A, B) + \nu\end{aligned}$$

for some $|\nu| \geq \frac{\varepsilon}{2}\frac{\mathrm{vol}(A)\mathrm{vol}(B)}{\mathrm{vol}(X)\mathrm{vol}(Y)\mathrm{vol}(G)}$ due to (29).

So from Cauchy-Schwarz-inequality (Lemma 19) we deduce

$$\sum_{(u,v)\in I} \sigma_{uv} d_{uv}^2 = \sum_{u,v\in I} \frac{\deg(u)\deg(v)}{\mathrm{vol}(A)\mathrm{vol}(B)} \varrho^2(V^{k+1}(u), V^{k+1}(v)) \tag{29}$$

$$= \frac{1}{\mathrm{vol}(A)\mathrm{vol}(B)} \sum_{\substack{V_i^{k+1}\subset A \\ V_j^{k+1}\subset B}} \varrho^2(V_i^{k+1}, V_j^{k+1})\mathrm{vol}(V_i^{k+1})\mathrm{vol}(V_j^{k+1}) \tag{30}$$

$$\geq \varrho^2(A,B) + \left(\frac{\varepsilon \mathrm{vol}(A)\mathrm{vol}(B)}{2\mathrm{vol}(X)\mathrm{vol}(Y)\mathrm{vol}(G)}\right)^2 \times \frac{\mathrm{vol}(X)\mathrm{vol}(Y)}{\mathrm{vol}(A)\mathrm{vol}(B)} \tag{31}$$

$$\geq \frac{1}{\mathrm{vol}(A)\mathrm{vol}(B)} \left(\mathrm{ind}(A,B) + \frac{\varepsilon^2 \mathrm{vol}(A)\mathrm{vol}(B)}{4\mathrm{vol}^2(G)}\right). \tag{32}$$

¿From (30) and (32) we infer the amount of the index increment on the irregular pair (A,B). So, summing over all irregular pairs we get

$$\mathrm{ind}(\mathcal{P}^{k+1}) - \mathrm{ind}(\mathcal{P}') \geq \sum_{(i,j)\in L} \frac{\varepsilon^2}{4} \frac{\mathrm{vol}(A)\mathrm{vol}(B)}{\mathrm{vol}^2(G)} - \varepsilon^4 \geq \frac{\varepsilon^3}{4}.$$

□