

A unified view of graph regularity via matrix decompositions

Greg Bodwin¹ | Santosh Vempala²

¹EECS, University of Michigan, Ann Arbor, Michigan, USA

²College of Computing, Georgia Tech, Atlanta, Georgia, USA

Correspondence

Greg Bodwin, EECS, University of Michigan, Ann Arbor, MI, USA.

Email: bodwin@umich.edu

Funding information

NSF, Grant/Award Numbers: CCF-1717349, DMS-183932, CCF-1909756

Abstract

We give a unified proof of algorithmic weak and Szemerédi regularity lemmas for several well-studied classes of sparse graphs, for which only weak regularity lemmas were previously known. These include core-dense graphs, low threshold rank graphs, and (a version of) L^p upper regular graphs. More precisely, we define *cut pseudorandom graphs*, we prove our regularity lemmas for these graphs, and then we show that cut pseudorandomness captures all of the above graph classes as special cases. The core of our approach is an abstracted matrix decomposition, which can be computed by a simple algorithm by Charikar. Using work of Oveis Gharan and Trevisan, it also implies new PTASes for MAX-CUT, MAX-BISECTION, MIN-BISECTION for a significantly expanded class of input graphs. (It is NP Hard to get PTASes for these graphs in general.)

KEYWORDS

approximation algorithms, graph algorithms, regularity lemmas

1 | INTRODUCTION

In *graph regularity*, the goal is roughly to partition the nodes of an input graph into a few parts so that the graph looks like a random bipartite graph between most pairs of parts. The area was pioneered by a classic lemma of Szemerédi [31], which guarantees the existence of a good partition for any sufficiently dense input graph. This Szemerédi regularity lemma has many deep algorithmic and combinatorial applications [22]. It has also spawned a line of work on regularity lemmas themselves; notable examples include the weak regularity lemma of Frieze and Kannan [12], the strong regularity lemma of Alon, Fischer, Krivelevich, and Szegedy [2], and so forth. These lie at the center of a deep mathematical theory of graph limits [3,24].

A limitation of these regularity lemmas is that they only comment on sufficiently *dense* input graphs; for example, if a graph has n nodes and $\leq n^{1.99}$ edges, then none of these regularity lemmas promise a partition with nontrivial properties. There are constructions demonstrating that this is unavoidable, and there can be no perfectly analogous regularity lemmas that extend to all sparse graphs [11,15]. However, it may still be possible to prove regularity lemmas for some graph classes of interest. This is the basis for *sparse graph regularity*, a research area that investigates which useful classes of sparse graphs admit nontrivial regularity lemmas. Some notable successes in the area include various sparse regularity lemmas and applications for *jumbled graphs* and (n, d, λ) *graphs* (see survey [23]), *upper regular graphs* [21], L_p *upper regular graphs* [6,7], graphs of *low threshold rank* [14], and *core-dense graphs* [19].

The objective of the current article is to unify and strengthen the linear algebraic foundations of sparse regularity. We abstract a theorem of Lovász and Szegedy [25] into a *matrix decomposition*, and we use this decomposition to define a new class of graphs that we call *cut pseudorandom*. We then prove:

- Essentially all classes of graphs mentioned above are cut pseudorandom, and
- Cut pseudorandom graphs have sparse weak and Szemerédi-type regularity lemmas. For many of the above graph classes, the sparse Szemerédi regularity lemmas were not previously known.
- Additionally, there is a simple polynomial-time algorithm that *computes* decompositions satisfying these sparse regularity lemmas for cut pseudorandom graphs.

In the following, we overview the matrix decomposition and its consequences more formally.

1.1 | Cut decompositions

In the following, a *cut vector* is a vector with entries in $\{0, 1\}$, a *cut matrix* is the scaled outer product of two cut vectors (i.e., it is a matrix that is constant on some $S \times T$ block and zero elsewhere), and the *cut norm* $\|A\|_{\square}$ of a matrix A is $\max_{v,w} |v^T A w|$, where the max is taken over cut vectors v, w . We write $\mathbb{1}$ for the cut vector where all entries are ones. In their seminal work on weak algorithmic graph regularity, Frieze and Kannan proved:

Theorem 1 (Cut approximation [12]). *For any binary matrix $A \in \mathbb{R}^{n \times n}$ and $\epsilon > 0$, there is \hat{A} that is a sum of $O(\epsilon^{-2})$ cut matrices such that*

$$\|A - \hat{A}\|_{\square} \leq \epsilon n^2.$$

The proof of Theorem 1 uses a greedy method: in each round we “select” a new pair of cut vectors (v, w) , and we define \hat{A} as the projection of A onto the span of the matrices wv^T from pairs chosen so far. Ideally, one would select cut vectors v, w in each round that achieve the cut norm on $A - \hat{A}$, but from a computational perspective this task is NP-Hard. So instead, the selection process typically uses *approximation* algorithms to find v, w ; when A is dense an extremely fast one was given by Frieze and Kannan [12], and for sparse graphs it is common in the current literature to use a nice approximation algorithm for the cut norm by Alon and Naor [4], based on semidefinite programming.

Later, Lovász and Szegedy [25] proved a highly generalized version of Theorem 1. They used a slightly different selection rule from [12]: in each round they select the cut vectors (v, w) maximizing the mass (measured in Frobenius norm) removed from $A - \hat{A}$ as the result of their choice. This method is extremely flexible, and it implies a generalization of Theorem 1 to essentially any inner product space (cf., [25, Lemma 4.1]). But the drawback of this method is that it is not algorithmic in the context

relevant to graph regularity: it is not clear how to compute (or even approximate) in polynomial time the cut vector pair that removes the most mass from $A - \hat{A}$.

In this article, we propose yet another selection rule: in each round, we select the nonzero cut vectors v, w maximizing

$$\left| \frac{v^T (A - \hat{A}) w}{\|v\|_2 \|w\|_2} \right|. \quad (1)$$

We prove that this approach also leads to an approximation as in Theorem 1. Additionally, an advantage over the previous two methods is that the maximizing cut vectors v, w can be computed exactly in polynomial time using a simple LP-rounding algorithm by Charikar [8]. Our method also generalizes to other inner product spaces in the same way as [25], and in the Appendix, we show that our method remains algorithmic when a diagonal inner product is used.

An important conceptual point is that, since we have a deterministic selection rule, we may treat this process as an *explicit matrix decomposition*. We define the i th *projection value* ϕ_i of the decomposition to be the mass removed from A in the i th round (as measured by Frobenius norm in the “main” decomposition, or a different norm when a different inner product is used—see Section 2.2 for details). There is an analogy here to the singular values, which can be viewed as the mass removed in each round of a similar greedy process where v, w are not constrained to be cut vectors. The analogy is not perfect, linear algebraically speaking: for example, there can be n^2 nonzero projection values but there only n singular values, and unlike singular values, the projection values are not necessarily decreasing. But, as we discuss next, the analogy is useful in the context of regularity lemmas. Using this, we revisit a recent line of work that proves regularity lemmas by analysis of their singular values, and we show that simpler and stronger proofs can be obtained when projection values are used instead.

1.2 | Regularity for cut pseudorandom graphs

First, we give a formal background on graph regularity. In this exposition, let G be an n -node undirected unweighted graph with adjacency matrix A_G . The *weak regularity lemma* of Frieze and Kannan states:

Theorem 2 (Weak regularity lemma [12]). *For any $\varepsilon > 0$, there is a vertex partition Π of G into $|\Pi| \leq 2^{\varepsilon^{-2}}$ parts and a weighted graph H satisfying:*

- The adjacency matrix A_H of H is constant on its $V_1 \times V_2$ block for all $V_1, V_2 \in \Pi$, and
- $\|A_G - A_H\|_{\square} = O(\varepsilon n^2)$.

It is common to merge the blocks of Π into supernodes, and treat H as a node- and edge-weighted graph on a small number of nodes that approximates the cuts of G . In parallel, the Szemerédi regularity lemma controls a more stringent norm, at the price of a worse dependence on ε in partition size. For a matrix $A \in \mathbb{R}^{n \times n}$ and a partition Π of $[n]$, let us define the *partition norm* as the quantity

$$\|A\|_{\Pi} := \sum_{V_1, V_2 \in \Pi} \|A[V_1 \times V_2]\|_{\square},$$

where $A[V_1 \times V_2]$ denotes the matrix that agrees with A on its $V_1 \times V_2$ block and is 0 elsewhere.

Theorem 3 (Szemerédi regularity lemma [25,31]). *For any $\epsilon > 0$, there is a vertex partition Π into $|\Pi| \leq \text{tower}(\epsilon^{-2})$ parts¹ and a weighted graph H satisfying:*

- *The adjacency matrix A_H of H is constant on its $V_1 \times V_2$ block for all $V_1, V_2 \in \Pi$, and*
- *$\|A_G - A_H\|_{\Pi} = O(\epsilon n^2)$.*

We have that $\|A\|_{\square} \leq \|A\|_{\Pi}$, and so it is generally stronger to control the partition norm, and this is sometimes required in applications [1,22]. For an unweighted adjacency matrix, both norms are bounded by n^2 , so these regularity lemmas improve on the trivial control by a factor of ϵ . The norm controls of ϵn^2 in Theorems 2 and 3 apply only when $G = (V, E)$ has unweighted nodes and edges, and they are nontrivial only when $|E| \gg \epsilon n^2$. It is natural to ask whether analogous theorems can be proved for sparser graphs. Unfortunately, there are barriers to achieving analogs that work for *all* sparse input graphs [11,15,26].² However, sparse regularity lemmas can be proved for certain classes of input graphs that satisfy various “pseudorandomness” criteria. We propose a new such graph class here, with the goal of simultaneously generalizing pseudorandomness concepts that have been studied previously in the literature.

In the following, let $G = (V, E, w, \text{vol})$ be a graph with edge weights (w) and node weights (vol). Let $w^*(u, v) := w(u, v)\text{vol}(u)\text{vol}(v)$, let $w^*(E)$ be the sum of $w^*(e)$ over all edges $e \in E$, and let A_G^* be the *fully weighted adjacency matrix*: that is, the (i, j) th entry of $A_G^* := w^*(v_i, v_j)$, or 0 if (v_i, v_j) is a non-edge. The *cut decomposition* is the greedy matrix decomposition outlined in the previous section, and $\phi^{(r)}$ is the vector that holds the projection values from the first r rounds of this process. We define:

Definition 1 (Cut pseudorandomness). *The graph G is r cut pseudorandom with respect to a diagonal positive definite matrix D if, when we take the cut decomposition of A_G^* with respect to inner product $\langle x, y \rangle := x^T D y$, the projection values satisfy*

$$\|\phi^{(r)}\|_2 = O\left(\frac{w^*(E)}{\langle \mathbf{1}, \mathbf{1} \rangle}\right).$$

Note that the definition of cut pseudorandomness depends not just on the graph G , but also on our selection rule and hence the inner product in play—a graph that is r cut pseudorandom with respect to one matrix D might not be so under another matrix D' . We then prove the following sparse regularity lemmas for cut pseudorandom graphs:

Theorem 4 (Regularity lemmas for cut pseudorandom graphs). *If G is $r = \Theta(\epsilon^{-2})$ cut pseudorandom, then it has a sparse weak regularity lemma, that is, Theorem 2 holds with $w^*(E)$ in place of n^2 . If G is $r = \text{tower}(\epsilon^{-2})$ cut pseudorandom, then it has a sparse Szemerédi regularity lemma, that is, Theorem 3 holds with $w^*(E)$ in place of n^2 . (See Theorems 6 and 7 for self-contained statements.)*

Moreover, we show that sparse weak and Szemerédi decompositions for cut pseudorandom graphs can be computed in polynomial time. The point of cut pseudorandomness is not so much that the definition is clearly interpretable, but rather, that it simultaneously captures several previously studied graph classes that do have clear interpretations. We discuss these next.

¹Recall that $\text{tower}(x) = 2^{2^{\dots}}$ with total height x .

²A general sparse regularity lemma is possible if one allows a large “exceptional set” in which many edges can be hidden [27,29]. Szemerédi’s original regularity lemma allowed an exceptional set; for dense unweighted input graphs, this is equivalent to the phrasing by partition norm. For sparser input graphs, the phrasing by partition norm is stronger and often preferable [25], and our exposition in this article refers to the natural sparse generalization by partition norm.

1.3 | Graph classes captured by cut pseudorandomness

First, we discuss the following version of L^p upper regularity:

Definition 2 (L^p upper regularity [6,7]). The graph $G = (V, E)$ is L^p upper η regular if, for any vertex partition Π of size $|\Pi| \leq \eta^{-1}$, we have

$$\left(\sum_{V_1, V_2 \in \Pi} \left(\frac{\text{vol}(V_1)(V_2)}{\text{vol}(V)^2} \right) \left(\frac{w^*(V_1, V_2)}{\text{vol}(V_1)\text{vol}(V_2)} \right)^p \right)^{1/p} \leq O \left(\frac{w^*(E)}{\text{vol}(V)^2} \right),$$

where $w^*(V_1, V_2)$ is the sum of $w^*(e)$ over all edges going between V_1 and V_2 .

The original definition of L^p upper regular considers a somewhat narrower class of partitions; namely, those that satisfy $\text{vol}(V_i) \geq \eta \text{vol}(V)$ for all $V_i \in \Pi$. We discuss the significance of this change in Section 4.1. L^p upper regular graphs are important mainly because, in [6,7], the authors extend a theory of graph limits from the case L^∞ to general p ; their interpretation is further discussed in these papers. The case $p = \infty$ was first explored by Kohayakawa [21] and Rödl [28], and the version with node weights was considered in [1]. These papers develop sparse Szemerédi regularity lemmas for L^∞ upper regular graphs, although their lemmas do not readily extend to L^p upper regular graphs when $p < \infty$. L^∞ upper regular graphs include the notable special case of *jumbled graphs* [9,23,33,34], which in turn include (n, d, λ) graphs with appropriate parameters; these cases are studied in particular because they enjoy some expanded applications over the general case [10,13].

Alongside this theory, there are also sparse regularity lemmas for different spectrally motivated classes of input graphs. These include the following.

Definition 3 (Low threshold rank [5,14,16–18]). For a graph G , let A be its adjacency matrix, D its diagonal matrix of node degrees, and $\bar{A} = D^{-1/2}AD^{-1/2}$ its normalized adjacency matrix. For $\varepsilon > 0$, the ε threshold rank of G is

$$t_\varepsilon(G) := \sum_{\lambda \text{ an eigenvalue of } \bar{A}, \lambda > \varepsilon} \lambda^2.$$

We say that G has *low ε -threshold rank* if $t_\varepsilon(G) = O(1)$.

We also discuss the “weighted” version of this definition in Section 4.2. Graphs of low threshold rank gained prominence due to a sequence of papers achieving approximation algorithms for these graphs, where comparable approximation algorithms are known to be hard in the general case (e.g., [5,16–18]). Oveis Gharan and Trevisan [14] gave a nice explanation for these by supplying a weak regularity lemma for low threshold rank graphs, and using it for further improved approximation algorithms. A related property was considered by Kannan and Vempala [19]:

Definition 4 (Core density [19]). The core strength of a graph G with adjacency matrix A is the quantity

$$\sum_{ij} \frac{A_{ij}}{(\text{deg}(i) + \bar{d})(\text{deg}(j) + \bar{d})},$$

where \bar{d} is the average node degree. We say that G is *core dense* if its core strength is $O(1)$.

TABLE 1 Sparse regularity lemmas for various graph classes in prior work are indicated with single checkmarks

Sparse graph class	Sparse Szemerédi regularity?	Sparse weak regularity?
L_∞ upper regular	✓ [21,28]	✓ [21,28]
L_p upper regular	✓✓	✓ [6,7]
Low threshold rank	✓✓	✓ [14]
Core dense	✓✓	✓ [19]

Note: We give a unified proof of all entries in this table, thus recovering all checkmarks and proving the double checkmarks as new results.

This also has a natural weighted extension, discussed in Section 4.2. The main advantage of core density is that it can additionally be extended to tensors; we refer to [19] for details. They developed a weak regularity lemma for these graphs, and similarly used it toward approximation algorithms for core dense graphs.

Currently, all of the above graph classes have related but somewhat different proofs of weak regularity lemmas. Our point is that all of these graph classes can be viewed as special cases of cut pseudorandomness:

Theorem 5 (Cut pseudorandomness of graph classes).

- If a graph is L^p upper η -regular for any $p \geq 2$ (with the change in definition mentioned above), then it is $O(\log(1/\eta))$ -cut-pseudorandom.
- If a graph has low ϵ threshold rank, then it is $O(1/\epsilon^2)$ -cut-pseudorandom.
- If a graph is core dense, then it is r -cut-pseudorandom for every r .

Hence, we get sparse weak and Szemerédi regularity lemmas for these graphs in one shot; except in the case of L^∞ upper regularity, the sparse Szemerédi regularity lemmas are new. We remark that [6,7] establish weak regularity lemmas for L^p upper regular graphs for all $p \geq 1$, so we leave a small gap here, since our theorem begins at $p \geq 2$. References are given in Table 1.

1.4 | Algorithmic applications

Regularity lemmas are commonly used in approximation algorithms of NP-hard problems that can be related in some way to graph cuts. For example, it is NP-Hard to find a MAX-CUT PTAS for general graphs [20], but the weak regularity lemma of Frieze and Kannan [12] was initially used for an extremely efficient MAX-CUT PTAS for dense unweighted input graphs. A natural direction is then to determine what other natural classes of graphs admit a PTAS for MAX-CUT or similar problems. Oveis Gharan and Trevisan [14] proved PTASes for MAX-CUT, MAX-BISECTION, and MIN-BISECTION for graphs of low threshold rank, following prior work in [5,16–18]. In fact, they essentially prove more generally that any graph class admitting suitable algorithmic regularity lemmas have PTASes for these problems. Using their work, we get the following algorithmic corollaries of our decomposition:

Corollary 1. For any $\epsilon > 0$ and any n -node unweighted graph $G = (V, E)$ that is $O(\epsilon^{-2})$ cut pseudorandom (with respect to any matrix D), one can solve any of the following problems in $2^{\tilde{O}(1/\epsilon^3)} + \text{poly}(n)$ time:

- MAX-CUT within $\pm\epsilon|E|$ error,

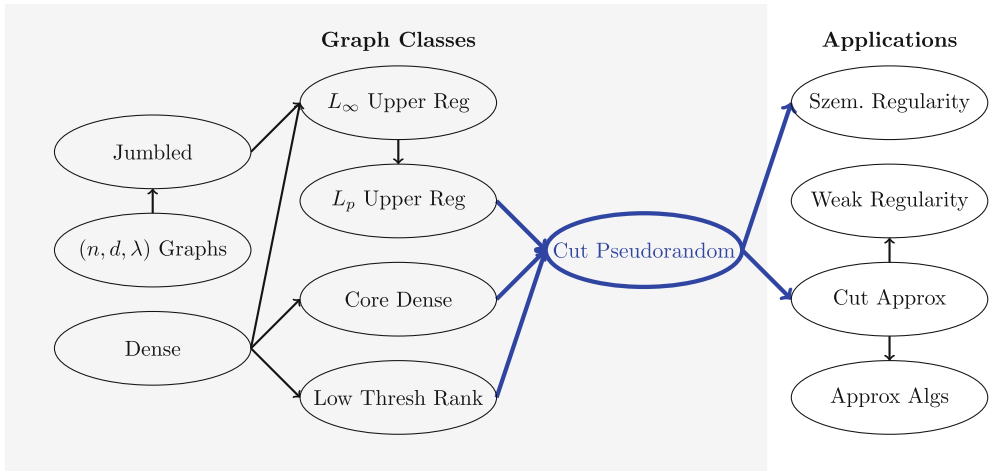
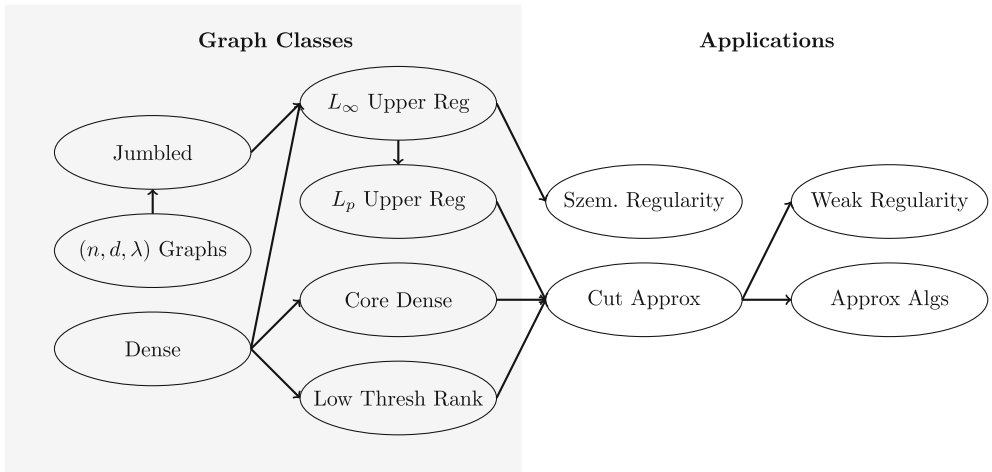


FIGURE 1 Some existing implications in the area of graph regularity (top), and how the results in this article change the picture (bottom). For clarity we have omitted arrows in these diagrams that are implied transitively by a chain of other arrows, even when these direct implications may have advantages (e.g., better quantitative bounds or faster computation). Citations and explanations are given in the text below.

- *MAX-BISECTION* within $\pm\epsilon|E|$ error, and
- *MIN-BISECTION* within $\pm\epsilon|E|$ error.

For example, this implies that L_p upper regular graphs (and others) have PTASes for all these problems, whereas none was previously known (Figure 1). These results are all discussed in more detail in Section 5.

2 | CUT DECOMPOSITIONS

The *Frobenius inner product* of two matrices $\langle \cdot, \cdot \rangle_F$ is defined by

$$\langle A, B \rangle_F := \text{Tr}(AB^T),$$

which may be equivalently viewed as the Euclidean inner product of A, B as vectors. The *Frobenius norm* $\|A\|_F$ is the corresponding matrix norm, which may also be viewed either as the L_2 -norm of the entries or of the singular values of A . The definitions of cut vectors, cut matrices, cut norm, partition norm, and the $\mathbb{1}$ vector will continue to be used in the following technical content; see Section 1.1 for a reminder of these definitions.

2.1 | Basic cut decompositions

As a warmup, we will first present the “basic” cut decomposition. The next part presents a generalization that is needed for a few applications later in the article.

Let $A \in \mathbb{R}^{n \times n}$, and let $S = \emptyset$ be an initially empty subset of cut matrices (recall that a cut matrix is a scaled outer product of two cut vectors). At all times, the matrix \hat{A} is defined as the projection, by $\langle \cdot, \cdot \rangle_F$, of A onto $\text{span}(S)$. Since S is initially empty, \hat{A} is initially 0.

2.1.1 | Decomposition algorithm

Repeat the following process until $\hat{A} = A$. Let

$$v^*, w^* := \arg \max_{v, w \neq 0 \in X} \left| \frac{v^T}{\|v\|_2} \left(A - \hat{A} \right) \frac{w}{\|w\|_2} \right|,$$

breaking ties arbitrarily, and add the cut matrix w^*v^{*T} to S . Note that v^*, w^* can be computed by a simple polynomial time algorithm of Charikar [8]; see the Appendix for more details. We repeat the process at most n^2 times, since $\mathbb{R}^{n \times n}$ is an n^2 dimensional space.

2.1.2 | Definitions and properties

Let \hat{A}_i be the value of \hat{A} after the i th round of this process, and for $i \geq 1$ let $A_i := \hat{A}_i - \hat{A}_{i-1}$. The i th *cut projection value* is the quantity $\phi_i := \|A_i\|_F$. Throughout, we use the notation $\phi^{(j)}$ to denote the vector containing the first j projection values. Finally, we note that by construction the matrices $\{A_i\}$ are pairwise orthogonal (under $\langle \cdot, \cdot \rangle_F$), and that each \hat{A}_i is a linear combination of at most i cut matrices.

2.2 | Cut decompositions under diagonal inner products

Let $D \in \mathbb{R}^{n \times n}$ be a diagonal positive definite matrix. The cut decomposition of a matrix A with respect to D is defined as follows. The basic cut decomposition presented above is the same as the cut decomposition of A with respect to the identity matrix I . First, the inner product $\langle \cdot, \cdot \rangle_D$ is defined by

$$\langle A, B \rangle_D := \langle D^{1/2} A D^{1/2}, D^{1/2} B D^{1/2} \rangle_F,$$

and $\|\cdot\|_D$ is the corresponding matrix norm. Again, S is an initially empty set of cut matrices. At all times, \hat{A} is the projection of the matrix $D^{-1} A D^{-1}$ onto the span of the matrices in S , and now the projection is by the inner product $\langle \cdot, \cdot \rangle_D$. In each round, we select the cut matrix M satisfying

$$M := \arg \max_{M^* \text{ cut matrix}} \left| \left\langle \frac{M^*}{\|M^*\|_D}, D^{-1} A D^{-1} - \hat{A} \right\rangle_D \right|, \quad (2)$$

and we add the cut matrix M to S . Roughly as before, we define:

- \hat{A}_i is the value of \hat{A} after i rounds (so $\hat{A}_0 = 0$),
- $A_i := \hat{A}_i - \hat{A}_{i-1}$, and
- the i th projection value ϕ_i is $\|A_i\|_D$.

The first thing to point out is that the algorithm of Charikar [8] no longer directly finds the cut matrix M maximizing (2) in polynomial time, essentially due to the renormalization of M^* by $\|\cdot\|_D$. Thus, in the Appendix, we show that Charikar's algorithm can be straightforwardly extended to solve (2), so long as the matrix D holds integer entries on its diagonal between 1 and $\text{poly}(n)$ (which it does for all cases of interest in this article).

2.3 | Projection values versus matrix norms

Here, we prove some inequalities that relate the projection values of the decomposition to various matrix norms of interest. In the following lemmas, let A, D be matrices where D is diagonal positive definite, and suppose we decompose A with respect to D giving projection values $\{\phi_i\}$.

Lemma 1. *For all indices $0 \leq i < n^2$, we have $(\mathbb{1}^T D \mathbb{1}) \phi_{i+1} \geq \|A - D\hat{A}_i D\|_{\square}$.*

Proof. We have

$$\begin{aligned} \phi_{i+1} &= \|A_{i+1}\|_D = \left\| D^{1/2} (\hat{A}_{i+1} - \hat{A}_i) D^{1/2} \right\|_F \\ &\geq \left\| D^{1/2} (\hat{A}_{i+1} - \hat{A}_i) D^{1/2} \right\|_2 \\ &\geq \max_{v, w \in X} \left| \frac{v^T D^{1/2}}{\|D^{1/2} v\|_2} D^{1/2} (\hat{A}_{i+1} - \hat{A}_i) D^{1/2} \frac{D^{1/2} w}{\|D^{1/2} w\|_2} \right| \\ &= \max_{v, w \in X} \left| \left\langle \frac{w w^T}{\|w v^T\|_D}, \hat{A}_{i+1} - \hat{A}_i \right\rangle_D \right|. \end{aligned}$$

Recall that we have defined \hat{A}_{i+1} as the projection of $D^{-1} A D^{-1}$, by $\langle \cdot, \cdot \rangle_D$, onto the span of a set of cut matrices, including $w v^T$ where v, w maximize the above expression. Thus we may continue

$$\begin{aligned} &= \max_{v, w \in X} \left| \left\langle \frac{w v^T}{\|w v^T\|_D}, D^{-1} A D^{-1} - \hat{A}_i \right\rangle_D \right| \\ &= \max_{v, w \in X} \left| \frac{v^T D^{1/2}}{\|D^{1/2} v\|_2} D^{1/2} (D^{-1} A D^{-1} - \hat{A}_i) D^{1/2} \frac{D^{1/2} w}{\|D^{1/2} w\|_2} \right| \\ &= \max_{v, w \in X} \left| \frac{v^T (A - D\hat{A}_i D) w}{\sqrt{v^T D v} \sqrt{w^T D w}} \right| \\ &\geq \frac{\|A - D\hat{A}_i D\|_{\square}}{\mathbb{1}^T D \mathbb{1}}. \end{aligned}$$

Lemma 2. *For any indices $i \leq j$ and any partition Π of $[n]$, we have*

$$\left\| D (\hat{A}_j - \hat{A}_i) D \right\|_{\Pi} \leq (\mathbb{1}^T D \mathbb{1}) \|\phi(i :: j)\|_2.$$

Proof. First, we notice that $\left\| D \left(\hat{A}_j - \hat{A}_i \right) D \right\|_{\Pi}$ is maximized when Π is the completely refined partition that puts each index into its own singleton subset. So without loss of generality we may prove the inequality in this setting. In the following, absolute value symbols on matrices $|M|$ are applied entrywise, and we will use the fact that they commute with diagonal matrices; for example, $D|M| = |DM|$.

$$\begin{aligned} \left\| D \left(\hat{A}_j - \hat{A}_i \right) D \right\|_{\Pi} &= \left(\sum_{a,b=1}^n \left| \left(D \left(\hat{A}_j - \hat{A}_i \right) D \right)_{a,b} \right| \right) \\ &= \left\langle \mathbb{1} \mathbb{1}^T, \left| D \left(\hat{A}_j - \hat{A}_i \right) D \right| \right\rangle_F \\ &= \left\langle D^{1/2} \mathbb{1} \mathbb{1}^T D^{1/2}, \left| D^{1/2} \left(\hat{A}_j - \hat{A}_i \right) D^{1/2} \right| \right\rangle_F && D \text{ diagonal} \\ &\leq \left\| D^{1/2} \mathbb{1} \mathbb{1}^T D^{1/2} \right\|_F \left\| D^{1/2} \left(\hat{A}_j - \hat{A}_i \right) D^{1/2} \right\|_F && \text{Cauchy-Schwarz} \\ &= \left(\mathbb{1}^T D \mathbb{1} \right) \left\| \hat{A}_j - \hat{A}_i \right\|_D \\ &= \left(\mathbb{1}^T D \mathbb{1} \right) \left\| \sum_{k=i+1}^j A_k \right\|_D. \end{aligned}$$

Next, since the matrices $\{A_k\}$ are pairwise orthogonal under $\langle \cdot, \cdot \rangle_D$, we have

$$\left\| \sum_{k=i+1}^j A_k \right\|_D^2 = \sum_{k=i+1}^j \|A_k\|_D^2 = \|\phi(i : j)\|_2^2$$

which combined with the above, gives

$$\left\| D \left(\hat{A}_j - \hat{A}_i \right) D \right\|_{\Pi} \leq \left(\mathbb{1}^T D \mathbb{1} \right) \|\phi(i : j)\|_2.$$

Lemma 3. For any index r , we have

$$\left\| \phi^{(r)} \right\|_2 \leq \left\| \Lambda^{(r)} \right\|_2,$$

where Λ is the vector containing the singular values of the matrix $D^{-1/2} A D^{-1/2}$ in non-increasing order.

Proof. First, we have

$$\begin{aligned} \left\| \phi^{(r)} \right\|_2^2 &= \sum_{i=1}^r \|A_i\|_D^2 \\ &= \left\| \sum_{i=1}^r A_i \right\|_D^2 && \{A_i\} \text{ orthogonal} \\ &= \left\| \hat{A}_r \right\|_D^2 \\ &= \left\| D^{1/2} \hat{A}_r D^{1/2} \right\|_F^2. \end{aligned}$$

Recall that \widehat{A}_r is the projection of $D^{-1}AD^{-1}$, by $\langle \cdot, \cdot \rangle_D$, onto the span of r cut matrices, which all have rank one. Equivalently, we can say that $D^{1/2}\widehat{A}_rD^{1/2}$ is the projection of the matrix $D^{-1/2}AD^{-1/2}$, by standard Frobenius inner product $\langle \cdot, \cdot \rangle_F$, onto the span of r matrices of rank one. Let M^* be the matrix maximizing $\|M^*\|_F$ over the matrices that are projections of $D^{-1/2}AD^{-1/2}$ onto the span of r rank one matrices; thus, we have

$$\|M^*\|_F \geq \left\| D^{-1/2}\widehat{A}_rD^{-1/2} \right\|_F.$$

Since M^* is found by taking the leading r terms of the singular value decomposition of $D^{-1/2}AD^{-1/2}$, it follows that $\|M^*\|_F$ is exactly the 2-norm of the leading r singular values. Thus

$$\left\| \phi^{(r)} \right\|_2^2 = \left\| D^{1/2}\widehat{A}_rD^{1/2} \right\|_F^2 \leq \|M^*\|_F^2 = \left\| \Lambda^{(r)} \right\|_2^2. \quad \blacksquare$$

3 | SPARSE REGULARITY LEMMAS FOR CUT PSEUDORANDOM GRAPHS

We will prove various regularity lemmas for *cut pseudorandom graphs* in this section. Let us first introduce some notation. The graphs $G = (V, E)$ discussed in the rest of this article are undirected and have both edge weights and node weights, all of which are positive. In principle, we can allow self-loops if desired.

- We write $\text{vol}(v)$ for the weight of a node v , or $\text{vol}(V)$ for the summed weight of the node set V .
- We write $w(e)$ for the weight of an edge e , or $w(E)$ for the sum of $w(e)$ over all $e \in E$, or $w(V_1, V_2)$ for the sum of $w(e)$ over all $e \in E \cap (V_1 \times V_2)$.
- We write $w^*(u, v) := w(u, v)\text{vol}(u)\text{vol}(v)$, and similarly $w^*(E)$ is the sum of $w^*(e)$ over all edges $e \in E$, and $w^*(V_1, V_2)$ is the sum of $w^*(e)$ over all edges $e \in E \cap (V_1 \times V_2)$.
- The *edge-weighted adjacency matrix* A has entries $A_{ij} = w(v_i, v_j)$ for each edge (v_i, v_j) , or $A_{ij} = 0$ if (v_i, v_j) is a non-edge. The *fully weighted adjacency matrix* A^* instead has entries $A_{ij}^* = w^*(v_i, v_j)$, or $A_{ij}^* = 0$ if (v_i, v_j) is a non-edge.

Definition 5 (Cut pseudorandomness). We say that a graph G is r *cut pseudorandom* with respect to a diagonal positive definite matrix D if, when we take the cut decomposition of its fully weighted adjacency matrix A^* with respect to D , the projection values $\{\phi_i\}$ satisfy

$$\left\| \phi^{(r)} \right\|_2 = O\left(\frac{w^*(E)}{\mathbb{1}^T D \mathbb{1}}\right).$$

We will sometimes just say “cut pseudorandom,” making D implicit, when we do not need to reference it.

We notice that cut pseudorandomness implies $r' < r$ cut pseudorandomness, so it is typically strongest to state our results with r as small as possible.

3.1 | Sparse cut approximation and sparse weak regularity

Here, we prove two closely related results on approximating the cut norm of a graph:

Lemma 4 (Sparse cut approximation). *Let $\epsilon > 0$ and let $G = (V, E)$ be $r = \Theta(\epsilon^{-2})$ cut pseudorandom. Then there is a graph $H = (V, E_H)$, possibly with different node and edge weights from G , such that:*

- *The edge-weighted adjacency matrix of H is the sum of $O(\epsilon^{-2})$ cut matrices, and*
- *$\|A_G^* - A_H^*\|_{\square} = O(\epsilon w^*(E))$, where A_G^*, A_H^* are the fully weighted adjacency matrices of G, H , respectively.*

Proof. Compute a cut decomposition of the fully weighted adjacency matrix A_G^* of G with respect to the matrix D implying cut pseudorandomness. Let $\{\phi_i\}$ be the projection values. The smallest entry ϕ_i in $\phi^{(r)}$ satisfies

$$\phi_i \leq \frac{\|\phi^{(r)}\|_2}{\sqrt{r}} = O\left(\epsilon \|\phi^{(r)}\|_2\right) = O\left(\epsilon \cdot \frac{w^*(E)}{\mathbb{1}^T D \mathbb{1}}\right),$$

where the last equality follows by cut pseudorandomness of G . By Lemma 1, we have

$$\phi_i \geq \frac{\|A_G^* - D\hat{A}_{i-1}^*D\|_{\square}}{\mathbb{1}^T D \mathbb{1}},$$

where \hat{A}_{i-1}^* is from the decomposition of A_G^* . Combining these inequalities, we have

$$\|A_G^* - D\hat{A}_{i-1}^*D\|_{\square} = O(\epsilon w^*(E)).$$

Then, we recall by construction that \hat{A}_{i-1}^* is a linear combination of $i - 1 \leq r$ cut matrices. We interpret this as the edge-weighted adjacency matrix of the graph H , and we interpret the diagonal entries of the matrix D as the node weights of H . Thus $D\hat{A}_{i-1}^*D = A_H^*$, and the lemma follows. ■

Theorem 6 (Sparse weak regularity lemma). *Let $\epsilon > 0$ and let $G = (V, E)$ be $r = \Theta(\epsilon^{-2})$ cut pseudorandom. Then there is a graph $H = (V, E_H)$, possibly with different node and edge weights from G , such that:*

- *There is a vertex partition Π into $|\Pi| = 2^{O(\epsilon^{-2})}$ parts such that for all $V_i, V_j \in \Pi$ the edge-weighted adjacency matrix of H is constant on its $V_i \times V_j$ block, and*
- *$\|A_G^* - A_H^*\|_{\square} = O(\epsilon w^*(E))$, where A_G^*, A_H^* are the fully weighted adjacency matrices of G, H , respectively.*

Proof. This follows almost immediately from Lemma 4. We simply recall that the edge-weighted adjacency matrix of H in Lemma 4 is the sum of $\leq r$ cut matrices, and we notice that each cut matrix is the outer product of two cut vectors, which each define a bipartition of the vertex set. Thus, the common refinement Π of all these bipartitions has $|\Pi| \leq 2^{2r} = 2^{O(r)}$ parts, and the edge-weighted adjacency matrix of H is constant on its $V_i \times V_j$ block for all $V_i, V_j \in \Pi$. ■

We remark here that it is common to merge together the parts in Π into supernodes, setting the weight of each supernode to be the sum of the node weights that comprise it. In this sense, one can view H as a graph on very few nodes that approximates G in cut norm. We will stay away from this phrasing, however, as it requires one to formalize the embedding of the nodes of H into the nodes of G which can get cumbersome.

3.2 | Sparse Szemerédi regularity

We now prove our sparse Szemerédi regularity lemma. Our approach is a thematic expansion on ideas from [25,30], which give proofs of regularity lemmas based on the SVD. We show that our projection values can be effectively substituted for singular values in a way that makes the approach work for sparse regularity as well. This particular exposition is influenced by [32].

Theorem 7 (Sparse Szemerédi regularity lemma). *Let $\epsilon > 0$ and let $G = (V, E)$ be $r = \text{tower}(\Theta(\epsilon^{-2}))$ cut pseudorandom. Then there is a graph $H = (V, E_H)$, possibly with different node and edge weights from G , and a vertex partition Π into $|\Pi| = \text{tower}(O(\epsilon^{-2}))$ parts such that:*

- For all $V_i, V_j \in \Pi$ the edge-weighted adjacency matrix of H is constant on its $V_i \times V_j$ block, and
- $\|A_G^* - A_H^*\|_{\Pi} = O(\epsilon w^*(E))$, where A_G^*, A_H^* are the fully weighted adjacency matrices of G, H , respectively.

Proof. Let G be r cut pseudorandom with respect to the matrix D . We will treat the parameter r as an integer that can be selected later in the proof, verifying at the end that we only need to use a bound of the form $r = \text{tower}(\Omega(\epsilon^{-2}))$. Let q be an integer and let $f : \mathbb{N} \rightarrow \mathbb{N}$ be a function satisfying $f(n) > n$ for all n , both of which we will also choose later in the proof. Compute the cut decomposition of A_G^* with respect to D . By Lemma 4, so long as $r \geq f(q)$, there exists an index $j \leq f(q)$ such that

$$\|A_G^* - D\hat{A}_j^*D\|_{\square} = O\left(\frac{w^*(E)}{f(q)^{1/2}}\right), \quad (3)$$

where \hat{A}_j^* is from the cut decomposition. Additionally, let $i := \min\{q, j\}$. As before, \hat{A}_i^* (also from the cut decomposition) is a linear combination of i cut matrices, each of which is formed using two cut vectors, which each define two bipartitions of the vertices. Let Π be the common refinement of these bipartitions, which thus has $|\Pi| \leq 2^{O(i)}$ parts, and we note that \hat{A}_i^* is constant on each $V_1 \times V_2$ block for $V_1, V_2 \in \Pi$. We define H to be the graph whose node weights are given by D and whose edge-weighted adjacency matrix is given by \hat{A}_i^* ; thus H satisfies the first point in the theorem. Our goal is now to bound $\|A_G^* - A_H^*\|_{\Pi}$. Using the triangle inequality we can write

$$\|A_G^* - A_H^*\|_{\Pi} = \|A_G^* - D\hat{A}_i^*D\|_{\Pi} \leq \|A_G^* - D\hat{A}_j^*D\|_{\Pi} + \|D\hat{A}_j^*D - D\hat{A}_i^*D\|_{\Pi}.$$

So it suffices to bound each of the two norms on the right-hand side as $O(\epsilon w^*(E))$.

The first norm. To bound $\|A_G^* - D\hat{A}_j^*D\|_{\Pi}$, we leverage our freedom to choose f and (partially) our freedom to choose q . We have:

$$\begin{aligned} \|A_G^* - D\hat{A}_j^*D\|_{\Pi} &= \sum_{V_1, V_2 \in \Pi} \max_{S \subseteq V_1, T \subseteq V_2} \left| (A_G^* - D\hat{A}_j^*D)(S, T) \right| \\ &\leq 2^{O(q)} \|A_G^* - D\hat{A}_j^*D\|_{\square} \\ &\leq \frac{2^{O(q)}}{f(q)^{1/2}} w^*(E), \end{aligned}$$

where the second-to-last inequality is by unioning over the $\leq 2^{O(q)}$ pairs of parts in the partition, and the last inequality is from (3). We now choose f to be an exponential function of the form $f(q) = C^q$

with a large enough base C , and we enforce a lower bound of $q \geq \log(1/\epsilon)$, giving

$$\begin{aligned} \left\| A_G^* - D\hat{A}_j^*D \right\|_{\Pi} &\leq \frac{2^{O(q)}}{f(q)^{1/2}} w^*(E) \\ &\leq w^*(E) \cdot 2^{-O(q)} \\ &\leq O(\epsilon w^*(E)) \quad (\text{first norm control}). \end{aligned}$$

The second norm. We next bound $\left\| D\hat{A}_j^*D - D\hat{A}_i^*D \right\|_{\Pi}$. In the case where $j \leq q$, we have $i = j$ and so $\hat{A}_i^* = \hat{A}_j^*$ and so this term is identically zero. So we will assume in the following that $i = q < j \leq f(q)$. The idea here is to consider the projection values $\{\phi_i\}$ from the decomposition, and choose the value of q to be sure that not too much of the total mass of the projection values lies between ϕ_{i+1} and ϕ_j , and then we can apply Lemma 2 to convert this to a bound on the partition norm. Let $f^{(i)}$ denote the function f iterated i times, and suppose $r \geq f^{\Omega(\epsilon^{-2})}(0)$. Letting $\{\phi_i\}$ be the projection values of the decomposition, and denote by $\phi(x :: y)$ the vector holding the subsequence of projection values (ϕ_x, \dots, ϕ_y) . By the pigeonhole principle there is an integer $\log(1/\epsilon) \leq k = O(\epsilon^{-2})$ satisfying³

$$\left\| \phi(f^{(k)}(0) + 1 :: f^{(k+1)}(0)) \right\|_2^2 = O\left(\epsilon^2 \left\| \phi^{(r)} \right\|_2^2\right). \tag{4}$$

We set $q := f^{(k)}(0)$ according to this index k satisfying (4). Now we bound:

$$\begin{aligned} \left\| D\hat{A}_j^*D - D\hat{A}_i^*D \right\|_{\Pi} &\leq (1^T D \mathbb{1}) \left\| \phi(i :: j) \right\|_2 \quad \text{Lemma 8} \\ &\leq (1^T D \mathbb{1}) \cdot \left\| \phi(i :: f^{(i)}) \right\|_2 \quad j \leq f(i) \\ &= (1^T D \mathbb{1}) \cdot O\left(\epsilon \left\| \phi^{(r)} \right\|_2\right) \quad \text{using (4)} \\ &= O(\epsilon w^*(E)) \quad (\text{second norm control}), \end{aligned}$$

where the last step follows from the assumption that G is cut pseudorandom.

Parameter checks. Our control on the first term requires us to choose f to be an exponential function. In the second norm control, we set $q = f^{O(\epsilon^{-2})}(0)$, and thus $q = \text{tower}(O(\epsilon^{-2}))$. The partition size is $|\Pi| = 2^{O(q)} = \text{tower}(O(\epsilon^{-2}))$ as well. We pick up a lower bound of $r \geq f(q)$ in the setup, and $r \geq f^{\Omega(\epsilon^{-2})}(0)$ in the second norm control; thus both are satisfied so long as $r = \text{tower}(\Omega(\epsilon^{-2}))$. ■

4 | CUT PSEUDORANDOMNESS OF SOME GRAPH CLASSES

In the following, we will consider some graph classes studied in prior work on sparse regularity and show that they are cut pseudorandom.

4.1 | L^p upper regularity

In this part, we will discuss a slightly modified version of L^p upper regularity:

³The lower bound $k \geq \log(1/\epsilon)$ is only used to enforce the lower bound $q \geq \log(1/\epsilon)$ picked up in the analysis of the first norm; it does not play a role in bounding the second norm.

Definition 6 (L^p upper regularity [6,7]). A graph $G = (V, E)$ is L^p upper regular with parameters (C, η) if, for any vertex partition Π of size $|\Pi| \leq \eta^{-1}$, we have

$$\left(\sum_{V_1, V_2 \in \Pi} \frac{\text{vol}(V_1)\text{vol}(V_2)}{\text{vol}(V)^2} \cdot \left(\frac{w^*(V_1, V_2)}{\text{vol}(V_1)\text{vol}(V_2)} \right)^p \right)^{1/p} \leq C \frac{w^*(E)}{\text{vol}(V)^2}.$$

We also allow $p = \infty$, in which case we amend the above sum to a max in the usual way, giving the condition

$$\max_{\substack{V_i, V_j \subseteq V, \\ \text{vol}(V_i), \text{vol}(V_j) \geq \eta \text{vol}(V)}} \frac{w^*(V_1, V_2)}{\text{vol}(V_1)\text{vol}(V_2)} \leq C \frac{w^*(E)}{\text{vol}(V)^2}.$$

The original definition considers partitions where $\text{vol}(V_i) \geq \eta \text{vol}(V)$ for all parts $V_i \in \Pi$. Our change to possibly unbalanced partitions is in line with the take on regularity throughout this article inspired by [11,25], which de-emphasizes the need for balance between different parts in the partition. It is very likely possible to enforce balance by adding some massaging steps to the decomposition itself; for examples of this in prior work, see [1] (cf., Algorithm 4.3, Step 6.2, and its proof of correctness), or [6] (cf., Lemmas 3.1–3.3). However, we have not included similar massaging here.

This change in definition is rather restrictive for $p = \infty$: for example, supposing G has unit node and edge weights, the left-hand side is maximized at 1 by choosing V_1, V_2 to be singleton sets containing nodes connected by an edge, and thus it forces $|E| \geq Cn^2$, hence requiring the graph to be dense in the classical sense. At the other extreme, the change does not matter at all when $p = 1$: all graphs are L^1 upper regular in either definition. As we will see shortly, the changed definition only needs to be applied at the case $p = 2$, which we believe is more toward the unrestrictive end.

Theorem 8. *For any $\eta > 0$, $p \geq 2$, every G that is L^p upper regular with parameters $(\eta, O(1))$ is also $r = \Theta(\log 1/\eta)$ cut pseudorandom.*

Proof. First, as shown in [6,7], Hölder's inequality implies that L^p upper regularity implies $L^{p' < p}$ upper regularity, so it suffices to prove the claim for $p = 2$ only. To do so, let A^*, D be the fully weighted adjacency matrix and node weight matrix of G , respectively. Take the cut decomposition of A^* with respect to D , and let Π be the common refinement of all the $\leq 2r$ bipartitions used in the leading r cut matrices of the decomposition, which have the matrix \hat{A}_r^* from the decomposition in their span. Thus we have $|\Pi| < 1/\eta$, assuming $r = O(\log 1/\eta)$ with a small enough implicit constant.

By construction, the matrix \hat{A}_r^* is the projection of the matrix $D^{-1}A^*D^{-1}$, by $\langle \cdot, \cdot \rangle_D$, onto the span of the first r selected cut matrices. Let \hat{B}_r^* be the projection of $D^{-1}A^*D^{-1}$, by $\langle \cdot, \cdot \rangle_D$, onto the span of all cut matrices corresponding to parts in Π (i.e., all cut matrices of the form $v_1 v_2^T$, where v_1, v_2 are the indicator vectors of some $V_1, V_2 \in \Pi$). Notice that the span of these cut matrices from Π contains the span of the leading r cut matrices in the decomposition, and thus we have

$$\|\hat{A}_r^*\|_D \leq \|\hat{B}_r^*\|_D.$$

For all $V_1, V_2 \in \Pi$, we have that \hat{B}_r^* is constant on its $V_1 \times V_2$ block, and each entry on the block is given by

$$\left\langle \frac{v_1 v_2^T}{\|v_1 v_2^T\|_D^2}, D^{-1}A^*D^{-1} \right\rangle_D = \frac{v_2^T A^* v_1}{(v_1^T D v_1)(v_2^T D v_2)} = O\left(\frac{w^*(V_1, V_2)}{\text{vol}(V_1)\text{vol}(V_2)}\right). \quad (5)$$

Using this, we bound

$$\begin{aligned}
\|\phi^{(r)}\|_2^2 &= \sum_{i=1}^r \|A_i^*\|_D^2 = \left\| \sum_{i=1}^r A_i^* \right\|_D^2 = \|\widehat{A}_r^*\|_D^2 \quad \{A_i^*\} \text{ orthogonal} \\
&\leq \|\widehat{B}_r^*\|_D^2 \\
&= \left\| D^{1/2} \widehat{B}_r^* D^{1/2} \right\|_F^2 \\
&= \sum_{V_1, V_2 \in \Pi} \sum_{a \in V_1, b \in V_2} \left(\widehat{B}_r^* \right)_{ab}^2 D_{aa} D_{bb} \\
&= \sum_{V_1, V_2 \in \Pi} \text{vol}(V_1) \text{vol}(V_2) \cdot O\left(\frac{w^*(V_1, V_2)}{\text{vol}(V_1) \text{vol}(V_2)} \right)^2 \quad \text{using (5)} \\
&= \text{vol}(V)^2 \cdot \sum_{V_1, V_2 \in \Pi} \frac{\text{vol}(V_1) \text{vol}(V_2)}{\text{vol}(V)^2} \cdot O\left(\frac{w^*(V_1, V_2)}{\text{vol}(V_1) \text{vol}(V_2)} \right)^2 \\
&= \text{vol}(V)^2 \cdot O\left(\frac{w^*(E)}{\text{vol}(V)^2} \right)^2 \quad G \text{ is } L^2 \text{ dense, } |\Pi| \leq \eta^{-1} \\
&= O\left(\frac{w^*(E)}{\text{vol}(V)} \right)^2 \\
&= O\left(\frac{w^*(E)}{\|1^T D 1\|} \right)^2. \quad \blacksquare
\end{aligned}$$

4.2 | Core density and low threshold rank

In this part, let $\deg(v)$ denote the *weighted* degree of the node v ; that is, $\deg(v) := \sum_{v \in e} w(e)$.

Definition 7 (Core density [19]). For a graph $G = (V, E)$ with unit node weights, the core density is the quantity

$$\sum_{(u,v) \in E} \frac{1}{(\deg(u) + \bar{d})(\deg(v) + \bar{d})}.$$

We say that G is core dense if its core density is $O(1)$.

Theorem 9. *Let G be a core dense graph with average weighted node degree \bar{d} . Then G is r cut pseudorandom, for any parameter r .*

Proof. Let A be the (edge- or fully) weighted adjacency matrix of G , and let D be the diagonal matrix where $D_{ii} := \deg(v_i) + \bar{d}$. The entry (u, v) of the matrix $D^{-1/2} A D^{-1/2}$ is exactly

$$\frac{1}{\sqrt{(\deg(u) + \bar{d})(\deg(v) + \bar{d})}},$$

so the sum square of entries of this matrix is exactly the core density of A . Thus, since G is core dense, we have

$$\|A\|_{D^{-1}}^2 = \left\| D^{-1/2} A D^{-1/2} \right\|_F^2 = O(1).$$

Using this together with Lemma 3, letting $\Phi = \phi^{(n^2)}$ be the vector holding *all* projection values of the decomposition of G with respect to D , we have

$$\|\Phi\|_2^2 \leq \|D^{-1/2}AD^{-1/2}\|_F^2 = O(1).$$

Additionally, in G we have

$$\mathbb{1}^T D \mathbb{1} = \sum_{v \in V} \deg(v) + \bar{d} = 2w(E) + 2w(E) = O(w(E)),$$

and so

$$\frac{w(E)}{\mathbb{1}^T D \mathbb{1}} = O(1),$$

thus implying cut pseudorandomness. \blacksquare

Finally, we consider graphs of *low threshold rank*:

Definition 8 (Low threshold rank graphs [5,14,16–18]). Let G be a graph with unit node weights and (edge- or fully) weighted adjacency matrix A , and let D be the diagonal matrix where $D_{ii} = \deg(v_i)$. We say that G has *low ϵ threshold rank* if

$$\sum_{\lambda \text{ is an eigenvalue of } D^{-1/2}AD^{-1/2}, \lambda > \epsilon} \lambda^2 = O(1).$$

Theorem 10. For any $\epsilon > 0$, any graph G with low ϵ threshold rank is $r = \Theta(\epsilon^{-2})$ cut pseudorandom.

Proof. Let $\Lambda^{(r)}$ be the vector containing the leading r eigenvalues of the matrix $D^{-1/2}AD^{-1/2}$. Using Lemma 3, when we decompose A with respect to D , the projection values $\{\phi_i\}$ satisfy

$$\begin{aligned} \|\phi^{(r)}\|_2^2 &\leq \|\Lambda^{(r)}\|_2^2 \\ &= \left(\sum_{\lambda \in \Lambda^{(r)}, \lambda > \epsilon} \lambda^2 \right) + \left(\sum_{\lambda \in \Lambda^{(r)}, \lambda \leq \epsilon} \lambda^2 \right) \\ &\leq O(1) + r\epsilon^2 \\ &= O(1), \end{aligned}$$

where the second-to-last inequality is obtained using the definition of cut pseudorandomness and a union over the r eigenvalues in $\Lambda^{(r)}$, and the last one is obtained using the parameter restriction $r = O(\epsilon^{-2})$. Finally, as before we notice that $\mathbb{1}^T D \mathbb{1} = 2w(E)$, so $w(E)/(\mathbb{1}^T D \mathbb{1}) = O(1)$, thus implying cut pseudorandomness. \blacksquare

5 | ALGORITHMIC APPLICATIONS

Here, we will survey some of the algorithmic applications that arise from our new regularity lemmas.

5.1 | Graph algorithms

Oveis Gharan and Trevisan [14] developed the following algorithmic applications of cut decompositions:

Theorem 11 ([14]). *Let $\epsilon > 0$, let $G = (V, E)$ be a node- and edge-unweighted graph, and let H be a graph on the same vertex set V which may have node and edge weights. Suppose H satisfies the premises of Lemma 4—that is,*

- *The edge-weighted adjacency matrix of H is the sum of $O(\epsilon^{-2})$ cut matrices (given as input), and for each cut matrix C we have $|w^*(C)| = O(w^*(E))$, and*
- *$\|A_G - A_H^*\|_{\square} = O(\epsilon|E|)$, where A_G is the adjacency matrix of G and A_H^* is the fully weighted adjacency matrix of H .*

Then in $2^{\tilde{O}(1/\epsilon^3)} + \text{poly}(n)$ time, one can find any of:

- *A MAX-CUT on G within $\pm\epsilon|E|$ error,*
- *A MAX-BISECTION on G within $\pm\epsilon|E|$ error, and*
- *A MIN-BISECTION on G within $\pm\epsilon|E|$ error.*

Note that MAX-CUT, MAX-BISECTION, and MIN-BISECTION all have $\text{OPT} = \Theta(|E|)$, so this algorithm is an efficient FPTAS. Theorem 11 is highly nontrivial; the proof includes a number of detailed and clever technical ideas that we will not recap here. Their proof is for a specific node-weight function for H (specifically, each node v is weighted by $\text{deg}(v)^{-1/2}$), but it extends easily to arbitrary positive node weights.

In [14], the authors observe that graphs of low threshold rank satisfy the premises, and so they enjoy these approximation algorithms. By Lemma 4, the premises of Theorem 11 also hold for the appropriate class of cut pseudorandom graphs. We thus have:

Corollary 2. *For any $\epsilon > 0$ and any unweighted graph that is $r = O(\epsilon^{-2})$ cut pseudorandom (with respect to any matrix D), one can compute MAX-CUT, MAX-BISECTION, and MIN-BISECTION with an efficient FPTAS.*

(The hypothesis that $|w^*(C)| = O(w^*(E))$ was not stated explicitly in the proof of Lemma 4, but it follows instantly from the proof.)

5.2 | MAX CSP algorithms and tensor decompositions

MAX-CUT can be viewed as MAX-2-CSP; this naturally leads to the question of whether the regularity method also gives PTASes for MAX- k -CSP. Such a PTAS was proved by Kannan and Vempala in [19], for core-dense graphs. The core of their approach was the following tensor decomposition theorem. Recall that a tensor $T \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_s}$ is an s -dimensional array. While there is no direct analog of SVD for tensors, we observe next that the cut decomposition extends naturally. For a tensor T as above, define its s -form

$$T(x^1, \dots, x^s) = \sum_{i_1, i_2, \dots, i_s=1}^{n_1, \dots, n_s} T_{i_1, \dots, i_s} x_{i_1}^1 \dots x_{i_s}^s,$$

and define the cut norm of T as

$$\|T\|_{\square} := \max_{(u^1, \dots, u^s) \in X} |T(u^1, \dots, u^s)|.$$

Theorem 12. *Let $T \in \mathbb{R}^{n_1 \times \dots \times n_s}$ and let $\varepsilon > 0$. Then there is \hat{T} that is a linear combination of $r = O(\varepsilon^{-2})$ “cut tensors,” that is, each is formed by an outer product of a tuple of cut vectors, satisfying*

$$\|T - \hat{T}\|_{\square} = O\left(\varepsilon \|T\|_F \sqrt{n_1 \times \dots \times n_s}\right).$$

Here $\|T\|_F$ denotes the L^2 vector norm of the entries of T . We will point out that our cut decomposition approach can be extended to prove this tensor decomposition theorem *existentially*. This is only half the work needed for a PTAS: one also must compute the decomposition quickly, and unfortunately, our algorithms do not readily extend to tensors. However, Kannan and Vempala [19] gave a different algorithm that computed the decomposition in the case of core-dense tensors (we refer to [19] for details on how the definition of core-dense matrices extends to tensors).

Define the inner product $\langle \cdot, \cdot \rangle_F$ over tensors that works as the standard Euclidean inner product over their entries. Let S be an initially empty set of tensors, and at all times we define the tensor \hat{T} as the projection of the input tensor T onto the span of the tensors in S , under $\langle \cdot, \cdot \rangle_F$. In each round, we find the cut vectors maximizing the quantity

$$\arg \max_{(u^1, \dots, u^s) \in X} \left| \frac{(T - \hat{T})(u^1, \dots, u^s)}{\|u^1\|_2, \dots, \|u^s\|_2} \right|.$$

Then, we add the tensor $u^1 \otimes \dots \otimes u^s$ to S (where \otimes denotes outer product). Letting \hat{T}_i be the value of \hat{T} in the i th round of the decomposition, and $T_i := \hat{T}_i - \hat{T}_{i-1}$, we define the i th projection value as $\phi_i := \|T_i\|_F$. Similar to the case of matrices, the tensors $\{T_i\}$ are orthogonal under $\langle \cdot, \cdot \rangle_F$, and it follows that $\phi_i \leq \varepsilon \|T - \hat{T}_i\|_F$ for some $i = O(\varepsilon^{-2})$. This matrix \hat{T}_i satisfies the theorem, and the proof is completed by the chain of inequalities

$$\frac{\|T - \hat{T}_i\|_{\square}}{\sqrt{n_1 \times \dots \times n_s}} \leq \max_{(u^1, \dots, u^s) \in X} \left| \frac{(T - \hat{T}_i)(u^1, \dots, u^s)}{\|u^1\|_2, \dots, \|u^s\|_2} \right| \leq \phi_i \leq \varepsilon \|T - \hat{T}_i\|_F.$$

ACKNOWLEDGMENTS

We are grateful to Keaton Hamm and to several anonymous reviewers for corrections, useful technical discussions, and references to prior work.

REFERENCES

1. N. Alon, A. Coja-Oghlan, H. Hàn, M. Kang, V. Rödl, and M. Schacht, *Quasi-randomness and algorithmic regularity for graphs with general degree distributions*, SIAM J. Comput. **39** (2010), no. 6, 2336–2362.
2. N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy, *Efficient testing of large graphs*, Combinatorica **20** (2001), no. 4, 451–476.
3. N. Alon, E. Fischer, I. Newman, and A. Shapira, *A combinatorial characterization of the testable graph properties: It's all about regularity*, SIAM J. Comput. **39** (2009), 143–167.
4. N. Alon and A. Naor, *Approximating the cut-norm via Grothendieck's inequality*, Proc. 36th Annu. ACM Symp. Theory Comput., ACM, 2004, pp. 72–80.

5. B. Barak, P. Raghavendra, and D. Steurer. *Rounding semidefinite programming hierarchies via global correlation*, Proc. 2011 IEEE 52nd Annu. Symp. Found. Comput. Sci., IEEE, 2011, pp. 472–481.
6. C. Borgs, J. Chayes, H. Cohn, and Y. Zhao, *An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions*, Trans. Amer. Math. Soc. **372** (2014), no. 5, 3019–3062.
7. C. Borgs, J. Chayes, H. Cohn, and Y. Zhao, *An L^p theory of sparse graph convergence II: LD convergence, quotients, and right convergence*, Ann. Probab. **46** (2014), 337–396.
8. M. Charikar, *Greedy approximation algorithms for finding dense components in a graph*, Proc. Int. Workshop Approx. Algorithms Comb. Optim., Springer, 2000, pp. 84–95.
9. F. R. K. Chung, R. L. Graham, and R. M. Wilson, *Quasi-random graphs*, Combinatorica **9** (1989), no. 4, 345–362.
10. D. Conlon, J. Fox, and Y. Zhao, *Extremal results in sparse pseudorandom graphs*, Adv. Math. **256** (2012), 206–290.
11. J. Fox and L. Miklós Lovász, *A tight lower bound for Szemerédi’s regularity lemma*, Combinatorica **37** (2014), no. 5, 911–951.
12. A. Frieze and R. Kannan, *Quick approximation to matrices and applications*, Combinatorica **19** (1999), no. 2, 175–220.
13. S. Gerke and A. Steger, *The sparse regularity lemma and its applications*, Surv. Comb. **327** (2005), 227–258.
14. S. O. Oveis Gharan and L. Trevisan, *A new regularity lemma and faster approximation algorithms for low threshold rank graphs*, Theory Comput. **11** (2015), 241–256.
15. W. T. Gowers, *Lower bounds of tower type for Szemerédi’s uniformity lemma*, Geom. Funct. Anal. GAFA **7** (1997), no. 2, 322–337.
16. V. Guruswami and A. K. Sinop, *Lasserre hierarchy, higher eigenvalues, and approximation schemes for graph partitioning and quadratic integer programming with PSD objectives*, Proc. 2011 IEEE 52nd Annu. Symp. Found. Comput. Sci., IEEE, 2011, pp. 482–491.
17. V. Guruswami and A. K. Sinop, *Faster SDP hierarchy solvers for local rounding algorithms*, Proc. 2012 IEEE 53rd Annu. Symp. Found. Comput. Sci., IEEE, 2012, pp. 197–206.
18. V. Guruswami and A. K. Sinop, *Approximating non-uniform sparsest cut via generalized spectra*, Proc. 24th Annu. ACM-SIAM Symp. Discrete Algorithms, SIAM, 2013, pp. 295–305.
19. R. Kannan and S. S. Vempala, *Spectral algorithms*, Found. Trends Theor. Comput. Sci. **4** (2009), no. 3-4, 157–288.
20. S. Khot, G. Kindler, E. Mossel, and R. O’Donnell, *Optimal inapproximability results for MAX-CUT and other 2-variable CSPs?* Proc. Annu. IEEE Symp. Found. Comput. Sci. FOCS, vol. 146-154, 2004, pp. 146–154.
21. Y. Kohayakawa, “*Szemerédi’s regularity lemma for sparse graphs*,” in F. Cucker, M. Shub, Eds., *Foundations of computational mathematics*, Springer, New York, NY, 1997, pp. 216–230.
22. J. Komlós, A. Shokoufandeh, M. Simonovits, and E. Szemerédi, “*The regularity lemma and its applications in graph theory*,” *Summer school on theoretical aspects of computer science*, in G. B. Khosrovshahi, A. Shokoufandeh, A. Shokrollahi, Eds., Springer, New York, NY, 2000, pp. 84–112.
23. M. Krivelevich and B. Sudakov, “*Pseudo-random graphs*,” *More sets, graphs and numbers*, in E. Györi, G. O. H. Katona, L. Lovász, T. Fleiner, Eds., Springer, New York, NY, 2006, pp. 199–262.
24. L. Lovász, *Large networks and graph limits*, American Mathematical Society Colloquium Publications, 2012.
25. L. Lovász and B. Szegedy, *Szemerédi’s lemma for the analyst*, Geom. Funct. Anal. **17** (2007), 252–270.
26. G. Moshkovitz and A. Shapira, *A short proof of Gowers’ lower bound for the regularity lemma*, Combinatorica **36** (2016), no. 2, 187–194.
27. G. Moshkovitz and A. Shapira, *A sparse regular approximation lemma*, Trans. Amer. Math. Soc. **371** (2019), no. 10, 6779–6814.
28. V. Rödl, *The dimension of a graph and generalized Ramsey theorems*, PhD thesis, Master’s thesis, Charles University, Prague, 1973.
29. A. Scott, *Szemerédi’s regularity lemma for matrices and sparse graphs*, Comb. Probab. Comput. **20** (2011), no. 3, 455–466.
30. B. Szegedy, *Limits of kernel operators and the spectral regularity lemma*, Eur. J. Comb **32** (2010), 1156–1167.
31. E. Szemerédi, *Regular partitions of graphs*. Technical report, Stanford University of California, Department of Computer Science, 1975.
32. T. Tao, *The spectral proof of the Szemerédi regularity lemma*. <https://terrytao.wordpress.com/2012/12/03/the-spectral-proof-of-the-szemeredi-regularity-lemma/>, 2012.
33. A. Thomason, *Pseudo-random graphs*, Ann Discrete Math **33** (1987), 307–331.
34. A. Thomason, *Random graphs, strongly regular graphs and pseudorandom graphs*, Lond. Math. Soc. Lect. Note Ser **123** (1987), 173–195.

How to cite this article: G. Bodwin, and S. Vempala, *A unified view of graph regularity via matrix decompositions*, Random Struct. Algorithms. **61** (2022), 62–83. <https://doi.org/10.1002/rsa.21053>

APPENDIX: CUT NORM MAXIMIZATION

Here, we prove that our cut decompositions are computable in polynomial time with respect to any diagonal matrix D with positive polynomially bounded integer entries on the diagonal. Recall that the goal is to compute cut vectors v, w in each round maximizing

$$\left| \left\langle \frac{vw^T}{\|vw^T\|_D}, D^{-1}AD^{-1} - \hat{A}_i \right\rangle_D \right| = \frac{w^T (A - D\hat{A}_iD) v}{\|w^T D w\|_2 \|v^T D v\|_2}.$$

We can reparametrize this expression by setting $A \leftarrow A - D\hat{A}_iD$, and then the goal is to solve

$$\max_{S, T \subseteq [n]} \frac{A(S, T)}{\sqrt{\sum_{i \in S} d_i \sum_{j \in T} d_j}},$$

where $d_i := D_{ii}$. We do so as follows.

Theorem 13. *There is an algorithm that, given $A \in \mathbb{R}^{n \times n}$ and nonnegative vector of integers $d \in \mathbb{Z}^n$, returns $S^*, T^* \subseteq [n]$ exactly solving*

$$\max_{S, T \subseteq [n]} \frac{A(S, T)}{\sqrt{\sum_{i \in S} d_i \sum_{j \in T} d_j}}$$

in time polynomial in n , the description length of A (number of bits) and $\sum_{i \in [n]} d_i$.

Proof. We verify that the proof of Charikar [8] works with slight extensions to edge and vertex weights. Let $c = \sum_{i \in S} d_i / \sum_{j \in T} d_j$ and enumerate all possible values of c . For each value, consider the following linear program.

$$\begin{aligned} & \max \sum_{i, j \in [n]} x_{ij}, \\ & \text{s.t.} \\ & \quad x_{ij} \leq A_{ij} s_i, A_{ij} t_j \quad \text{for all } i, j \\ & \quad \sum_{i \in [n]} d_i s_i \leq \sqrt{c} \\ & \quad \sum_{j \in [n]} d_j t_j \leq \frac{1}{\sqrt{c}}. \end{aligned}$$

First, take any two subsets S, T . Then for the LP solution we set

$$x = \frac{\sqrt{c}}{\sum_{i \in S} d_i} = \frac{1}{\sqrt{c} \sum_{j \in T} d_j}$$

and $x_{ij} = A_{ij} \cdot x$, $s_i = x$, $t_j = x$ for $i \in S, j \in T$, respectively, and all other variables to zero. Then,

$$\sum_{ij} x_{ij} = \left(\sum_{i \in S, j \in T} A_{ij} \right) \cdot x$$

$$\begin{aligned}
 &= \frac{A(S, T)}{\sqrt{c} \sum_{j \in T} d_j} \\
 &= \frac{A(S, T)}{\sqrt{\sum_{i \in S} d_i \sum_{j \in T} d_j}}.
 \end{aligned}$$

Thus the LP solution has value at least as large.

Now for the converse, take any solution to the LP. We can assume that $x_{ij} = \min\{A_{ij}s_i, A_{ij}t_j\}$. Define the following subsets for any $r \geq 0$:

$$S(r) = \{i \in [n] : s_i \geq r\}, \quad T(r) = \{j \in [n] : t_j \geq r\}$$

Also, let $E(r) = \{(i, j) : x_{ij} \geq A_{ij}r \text{ if } A_{ij} \geq 0; x_{ij} \leq A_{ij}r \text{ if } A_{ij} < 0\}$. Then, $(i, j) \in E(r)$ implies that $i \in S(r), j \in T(r)$. And since $x_{ij} = \min A_{ij}s_i, A_{ij}t_j$, it follows that $E(r)$ is precisely the set of edges (i, j) between $S(r)$ and $T(r)$ with weights A_{ij} .

Next, note that

$$\int_0^\infty \left(\sum_{i \in S(r)} d_i \right) dr = \sum_{i \in [n]} d_i s_i \leq \sqrt{c}.$$

Similarly,

$$\int_0^\infty \left(\sum_{j \in T(r)} d_j \right) dr = \sum_{j \in [n]} d_j t_j \leq \frac{1}{\sqrt{c}}.$$

Therefore, by Cauchy–Schwarz,

$$\int_0^\infty \sqrt{\sum_{i \in S(r)} d_i \sum_{j \in T(r)} d_j} dr \leq \left(\int_0^\infty \sum_{i \in S(r)} d_i dr \int_0^\infty \sum_{j \in T(r)} d_j dr \right)^{1/2} \leq 1.$$

Also,

$$\int_0^\infty \sum_{ij \in E(r)} A_{ij} dr = \sum_{ij} x_{ij}.$$

Therefore, if there is an LP solution of value v , then

$$\frac{\int_0^\infty \sum_{ij \in E(r)} A_{ij} dr}{\int_0^\infty \sqrt{\sum_{i \in S(r)} d_i \sum_{j \in T(r)} d_j} dr} \geq \sum_{ij} x_{ij} \geq v.$$

And hence there exists an r s.t.

$$\frac{\sum_{ij \in E(r)} A_{ij}}{\sqrt{\sum_{i \in S(r)} d_i \sum_{j \in T(r)} d_j}} \geq v$$

as claimed. Thus we can solve the problem using the LP, and then scanning for the best value of r . ■

Notice that the runtime of this algorithm depends polynomially on the number of possible values of $\frac{s^t D s}{t^t D t}$ where s, t range over cut vectors. Thus, when D holds integers, this number of possible values depends polynomially on $\sum_{i \in [n]} d_i$, as claimed.