Towards Tight Bounds for Spectral Sparsification of Hypergraphs*

Michael Kapralov†
École Polytechnique Fédérale de Lausanne
Switzerland
michael.kapralov@epfl.ch

Robert Krauthgamer‡
Weizmann Institute of Science
Israel
robert.krauthgamer@weizmann.ac.il

Jakab Tardos§
École Polytechnique Fédérale de Lausanne
Switzerland
jakab.tardos@epfl.ch

Yuichi Yoshida¶
National Institute of Informatics
Japan
yyoshida@nii.ac.jp

ABSTRACT

Cut and spectral sparsification of graphs have numerous applications, including e.g. speeding up algorithms for cuts and Laplacian solvers. These powerful notions have recently been extended to hypergraphs, which are much richer and may offer new applications. However, the current bounds on the size of hypergraph sparsifiers are not as tight as the corresponding bounds for graphs.

Our first result is a polynomial-time algorithm that, given a hypergraph on n vertices with maximum hyperedge size r, outputs an ϵ-spectral sparsifier with $O^*(nr)$ hyperedges, where O^* suppresses $(\epsilon^{-1}\log n)^{O(1)}$ factors. This size bound improves the two previous bounds: $O^*(n^3)$ [Soma and Yoshida, SODA'19] and $O^*(nr^3)$ [Bansal, Svensson and Trevisan, FOCS'19]. Our main technical tool is a new method for proving concentration of the nonlinear analogue of the quadratic form of the Laplacians for hypergraph expanders.

We complement this with lower bounds on the bit complexity of any compression scheme that $(1+\epsilon)$-approximates all the cuts in a given hypergraph, and hence also on the bit complexity of every ϵ-cut/spectral sparsifier. These lower bounds are based on Ruzsa-Szemerédi graphs, and a particular instantiation yields an $\Omega(nr)$ lower bound on the bit complexity even for fixed constant ϵ. In the case of hypergraph cut sparsifiers, this is tight up to polylogarithmic factors in n, due to recent result of [Chen, Khanna and Nagda, FOCS'20]. For spectral sparsifiers it narrows the gap to $O^*(r)$.

Finally, for directed hypergraphs, we present an algorithm that computes an ϵ-spectral sparsifier with $O^*(n^2r^3)$ hyperarcs, where r is the maximum size of a hyperarc. For small r, this improves over $O^*(n^3)$ known from [Soma and Yoshida, SODA'19], and is getting close to the trivial lower bound of $\Omega(n^2)$ hyperarcs.

CCS CONCEPTS

• **Theory of computation → Sparsification and spanners**; **Design and analysis of algorithms**; **Sketching and sampling**.

KEYWORDS

hypergraphs, edge sparsification, spectral sparsification

ACM Reference Format:
Michael Kapralov, Robert Krauthgamer, Jakab Tardos, and Yuichi Yoshida. 2021. Towards Tight Bounds for Spectral Sparsification of Hypergraphs. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC '21), June 21–25, 2021, Virtual, Italy*. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3406325.3451061

1 INTRODUCTION

Sparsification is an algorithmic paradigm where a dense object is replaced by a sparse one with similar features, which often leads to significant improvements in efficiency of algorithms, including running time, space complexity, and communication. We study edge-sparsification of hypergraphs, which replaces a hypergraph $G = (V, E, w)$ with a sparse hypergraph $\widetilde{G}$ that has the same vertex set V and only a few hyperedges, often a reweighted subset of E. This is a natural extension of edge-sparsification of ordinary graphs, which includes key concepts such as cut sparsifiers, spectral sparsifiers, and flow sparsifiers. These were studied extensively from numerous angles, including various constructions, tight size bounds, related variants, and practical applications. As this literature is too vast to cover here, we quickly recap the basics for graphs before discussing hypergraphs, which are our focus here.

Graphs. Let $G = (V, E, w)$ be an edge-weighted graph, where $w \in \mathbb{R}_+^E$. The *energy* of a vector $x \in \mathbb{R}^V$ in G is defined as

$$Q_G(x) = \sum_{uv \in E} w_{uv}(x_u - x_v)^2,$$

and can also be written as $x^\top L_G x$, where L_G is the Laplacian matrix of G. Spielman and Teng [32] introduced the notion of an *ϵ-spectral*

*Full version available at arXiv:2011.06530.
†Supported in part by ERC Starting Grant 759471.
‡Work partially supported by ONR Award N00014-18-1-2364, the Israel Science Foundation grant #1086/18, and a Minerva Foundation grant.
§Supported by ERC Starting Grant 759471.
¶Supported in part by JSPS KAKENHI Grant Number 18H05291 and 20H05965.

STOC '21, June 21–25, 2021, Virtual, Italy

ACM ISBN 978-1-4503-8053-9/21/06…$15.00
https://doi.org/10.1145/3406325.3451061

sparsifier of G, which is a graph $\widetilde{G} = (V, \widetilde{E}, \widetilde{w})$ that satisfies (for $0 \le \epsilon \le 1/2$)

$$\forall x \in \mathbb{R}^V, \qquad Q_{\widetilde{G}}(x) \in (1 \pm \epsilon) Q_G(x). \tag{1}$$

The *size* of a spectral sparsifier $\widetilde{G}$ is $|\widetilde{E}|$.

We say that an edge $e \in E$ is *cut* by $S \subseteq V$ if one endpoint of e belongs to S and another one belongs to $V \setminus S$. The total weight of edges cut by S is clearly $Q_G(1_S)$, where $1_S \in \mathbb{R}^V$ denotes the characteristic vector of a set $S \subseteq V$.

A spectral sparsifier $\widetilde{G}$ of a graph G preserves many important properties of G: its cuts have approximately the same weight as those in G; its Laplacian $L_{\widetilde{G}}$ approximates every eigenvalue of L_G; electrical flows in $\widetilde{G}$ approximate those in G. It is extremely useful to have a spectral sparsifier with a small number of edges because algorithms that involve these quantities can be applied on the sparsifier $\widetilde{G}$ instead of on G, with only a small loss in accuracy.

A spectral sparsifier of size $O(n/\epsilon^2)$ can be computed in almost linear time [21], where n is the number of vertices in G.

Hypergraphs. A hypergraph is a natural extension of a graph, which can represent relations between three or more entities, and has proved useful to solve problems in practical areas such as computer vision [16, 25], bioinformatics [19], and information retrieval [14]. Many of those problems, such as semi-supervised learning [15, 35, 39] and link prediction [36], involve the notion of energy for hypergraphs, where the *energy* of a vector $x \in \mathbb{R}^V$ in an edge-weighted hypergraph $G = (V, E, w)$ is defined as

$$Q_G(x) = \sum_{e \in E} w_e \max_{u,v \in e} (x_u - x_v)^2. \tag{2}$$

This definition matches the one for graphs when every hyperedge in G is of size two. As before, $Q_G(1_S)$ gives the total weight of hyperedges cut by S, where we regard a hyperedge $e \in E$ as *cut* if $e \cap S \neq \emptyset$ and $e \cap (V \setminus S) \neq \emptyset$.

Spectral sparsification of hypergraphs was first defined by Soma and Yoshida [29], as follows. Similarly to graphs, an *ϵ-spectral sparsifier* of G is a hypergraph $G = (V, \widetilde{E}, \widetilde{w})$ that satisfies (1). This is a strictly stronger notion than that of the hypergraph cut sparsifier which has been previously studied in [24] and [20].

Besides the applications mentioned above, spectral sparsifiers for hypergraphs were used to show agnostic learnability of a certain subclass of submodular functions [29].

Soma and Yoshida [29] showed that every hypergraph G admits an ϵ-spectral sparsifier with $\widetilde{O}(n^3/\epsilon^2)$ hyperedges,[1] which is non-trivial because a general hypergraph can have $2^n - 1$ (non-empty) hyperedges. Moreover, they provide an algorithm recovering this sparsifier, that runs in close to linear time (in the input size). Later, Bansal, Svensson and Trevisan [6] showed that every hypergraph G admits a spectral sparsifier with $\widetilde{O}(nr^3/\epsilon^2)$ hyperedges, where r is the maximum size of a hyperedge in G. Note that this bound is incomparable to [29] because r could be as large as n.

1.1 Results

Spectral sparsification of undirected hypergraphs. Our first contribution is an algorithm that constructs an ϵ-spectral sparsifier of a hypergraph with only $\widetilde{O}(nr/\epsilon^{O(1)})$ hyperedges, which improves upon the previous constructions mentioned above. (See Table 1 for known bounds for hypergraph sparsification.)

Table 1: Bounds on the size of hypergraph sparsifiers

cut sparsification	spectral sparsification	reference
$\widetilde{O}(n^2/\epsilon^2)$		[24] implicitly
$\widetilde{O}(nr/\epsilon^2)$		[20]
	$\widetilde{O}(n^3/\epsilon^2)$	[29]
	$\widetilde{O}(nr^3/\epsilon^2)$	[6]
$\widetilde{O}(n/\epsilon^2)$		[10]
	$\widetilde{O}(nr/\epsilon^{O(1)})$	Theorem 1.1

Theorem 1.1. *Given an r-uniform hypergraph $G = (V, E, w)$ and $1/n \le \epsilon \le 1/2$, one can compute in polynomial time with probability $1 - o(1)$ an ϵ-spectral sparsifier of G with $nr(\epsilon^{-1} \log n)^{O(1)}$ hyperedges. The running time is $O(mr^2) + n^{O(1)}$, where $m = |E|$.*

To simplify notation, our entire technical analysis considers a hypergraph $G = (V, E)$ that is unweighted (i.e., unit weight hyperedges), reserving the letter w for the edge weights in the sparsifier. This is actually without loss of generality, see the full version of the paper.

We stress that Theorem 1.1 in fact applies to hypergraphs with maximum size of a hyperedge at most r. Indeed, in our analysis every hyperedge is a multiset of vertices, and therefore a hyperedge with less than r vertices can be trivially extended to a multiset of exactly r vertices by copying an arbitrary vertex, without changing the energy (but it might affect vertex degrees).

Bit-complexity lower bound. To complement Theorem 1.1, we consider lower bounds on the bit complexity of sparsifiers. Here, we consider *ϵ-cut sparsifiers*, which require that (1) holds only for vectors of the form $x = 1_S$. This notion actually predates spectral sparsification and was first defined by Benczúr and Karger [7] for graphs, and by Kogan and Krauthgamer [20] for hypergraphs. Obviously, lower bounds for cut sparsifiers directly imply the same lower bounds also for spectral sparsifiers.

The second contribution of this work is a surprising connection between *a Ruzsa-Szemerédi (RS) graph* [28], which is a well-studied notion in extremal graph theory, and a lower bound on the bit complexity of a hypergraph cut sparsifier. Here, an (ordinary) graph is called a *(t, a)-RS graph* if its edge set is the union of t induced matchings of size a. Then, we show the following.

Theorem 1.2. *Suppose that there exists a (t, a)-Ruzsa-Szemerédi graph on n vertices with $a \ge 6000\sqrt{n \log n}$. Assume also one can compress unweighted $(t+1)$-uniform hypergraphs $G = (V, E)$ on $2n$ vertices into k bits, from which $Q_G(1_S)$ can be approximated for every $S \subseteq V$ within factor $1 \pm \epsilon$, where $\epsilon = O(a/n)$. Then, $k = \Omega(at)$.*

For example, by instantiating Theorem 1.2 with the $(n^{\Omega(1/\log\log n)}, n/3 - o(n))$-Ruzsa-Szemerédi graphs known due to Fischer et al. [13], we deduce that $\Omega(nr)$ bits are necessary to encode all the cut values of an arbitrary r-uniform hypergraph with $r = n^{O(1/\log\log n)}$, even within a fixed constant ratio $1 + \epsilon$.

[1] Throughout, we write $\widetilde{O}(\cdot)$ to suppress a factor of $\log^{O(1)} n$.

This lower bound is in fact near-tight. Indeed, Chen, Khanna, and Nagda [10] showed very recently that every hypergraph G admits an ϵ-cut sparsifier with $O(n \log n/\epsilon^2)$ hyperedges, which are actually sampled from G. Applying this construction with fixed ϵ and $r = n^{O(1/\log\log n)}$ yields a sparsifier of G with $O(n \log n)$ hyperedges; encoding a hyperedge (including its weight, which is bounded by n^r) takes at most $O(r \log n)$ bits, and thus one can encode all the cuts of G using $O(nr \log^2 n)$ bits. It follows that our lower bound is optimal up to a lower order factor $O(\log^2 n)$. Instantiating our lower bound with the original construction of Ruzsa and Szemerédi [28], we can rule out the possibility of compressing the cut structure of a hypergraph with n vertices and maximum hyperedge size r with significantly less than nr space, and a polynomial scaling in the error (that is with $nr^{1-\Omega(1)}\epsilon^{-O(1)}$ space), *for any* r. See the full version of the paper for more details.

In fact, our space lower bound for hypergraphs far exceeds the $O(n \log n/\epsilon^2)$ bits that suffices to approximately represent all the cuts of an (ordinary) graph by simply storing a cut sparsifier. We thus obtain the first provable separation between the bit complexity of approximating all the cuts of a graph vs. of a hypergraph.

Spectral sparsification of directed hypergraphs. We also consider spectral sparsification of directed hypergraphs. Here, a hyperarc e consists of two disjoint sets, called the *head* $h(e) \subseteq V$ and the *tail* $t(e) \subseteq V$, and the *size* of the hyperarc is $|t(e)| + |h(e)|$. A directed hypergraph $G = (V, E)$ then consists of a vertex set V and a hyperarc set E. For an edge-weighted directed hypergraph $G = (V, E, w)$ and a vector $x \in \mathbb{R}^V$, the *energy* of x in G is defined as

$$Q_G(x) = \sum_{e \in E} w_e \max_{u \in t(e), v \in h(e)} (x_u - x_v)_+^2, \tag{3}$$

where $(a)_+ = \max\{a, 0\}$. Again, it is defined so that $Q_G(1_S)$ is the total weight of hyperarcs that are cut by S, where a hyperarc e is *cut* if $t(e) \cap S \neq \emptyset$ and $h(e) \cap (V \setminus S) \neq \emptyset$.

It is not difficult to see that a spectral sparsifier might require (in the worst-case) at least $\Omega(n^2)$ hyperarcs, even for an ordinary directed graph. Indeed, consider a balanced bipartite clique directed from one side of the bipartition towards the other. Here, every arc is the unique arc crossing some particular directed cut, and hence a sparsifier must keep all the $\Omega(n^2)$ arcs (see also [11, 18]). However, Soma and Yoshida [29] showed that every directed hypergraph admits an ϵ-spectral sparsifier with $\widetilde{O}(n^3/\epsilon^2)$ hyperarcs. We tighten this gap by showing that $\widetilde{O}(n^2/\epsilon^2)$ hyperarcs are sufficient when every hyperarc is of constant size.

Theorem 1.3. *Given a directed hypergraph $G = (V, E)$ with maximum hyperarc size at most r such that $11r \leq \sqrt{\epsilon n}$, and a value $\epsilon \leq 1/2$, one can compute in polynomial time with probability $1 - o(1)$ an ϵ-spectral sparsifier of G with $O(n^2 r^3 \log^2 n/\epsilon^2)$ hyperarcs.*

We note that Theorem 1.3 is stated under the assumption $11r \leq \sqrt{\epsilon n}$, which is useful for our analysis for technical reasons. For larger values of r the result of [29] gives a better bound on the number of hyperedges in the sparsifier, and therefore this assumption is not restrictive.

1.2 Related Work

The first construction of cut sparsifiers for hypergraphs was given by Kogan and Krauthgamer [20] and uses $O(n(r + \log n)/\epsilon^2)$ hyperedges. They also mention that an upper bound of $O(n^2 \log n/\epsilon^2)$ hyperedges follows implicitly from the results of Newman and Rabinovich [24]. Very recently (and independent of our work), Chen, Khanna, and Nagda [10] improved this bound to $O(n \log n/\epsilon^2)$ hyperedges, which is near-optimal because the current lower bound is $\Omega(n/\epsilon^2)$ edges, and actually holds for (ordinary) graphs [4, 8].

Louis [23] (later merged with Chan et al. [9]) initiated the spectral theory for hypergraphs, in which the Laplacian operator $L : \mathbb{R}^V \to \mathbb{R}^V$ of a hypergraph is defined so that its "quadratic form" $xL(x)$ coincides with the energy (2). As opposed to the graph case, here the Laplacian operator is merely piecewise linear, and hence computing its eigenvalues/vectors is hard. He showed that $O(\log r)$-approximation is possible, and that obtaining a better approximation ratio is NP-hard assuming the Small-Set Expansion (SSE) hypothesis [26]. He further showed a Cheeger inequality for hypergraphs, which implies that, given a vector $x \in \mathbb{R}^V$ with a small energy, one can efficiently find a set $S \subseteq V$ of small expansion, which roughly means that the number of hyperedges cut by S is small relative to the number of hyperedges incident to vertices in S (see Section 2 for details). Since then, several other algorithms for finding sets of small expansion have been proposed [17, 33].

Yoshida [37] proposed another piecewise linear Laplacian for directed graphs and used it to study structures of real-world networks. Generalizing the Laplacians for hypergraphs and directed graphs, Laplacian L for directed hypergraphs was proposed [22, 38], whose quadratic form $x^\top L(x)$ coincides with (3).

1.3 Discussion

An obvious open question is the existence of a spectral sparsifier with $\widetilde{O}(n)$ hyperedges. As we will see in Section 3, our overall strategy to construct a spectral sparsifier is decomposing the input hypergraph into good expanders (in a non-trivial way) and then sparsifying each expander. Here a *good expander* is a hypergraph with the maximum possible expansion up to a constant factor (see Section 2.1 for the details). However, we do not even know whether we can spectrally sparsify hypergraph expanders with $\widetilde{O}(n)$ hyperedges. To see the difficulty, note that a graph expander has expansion $\Theta(1)$ whereas an r-uniform hypergraph expander has expansion $\Theta(1/r)$. Let $x \in \mathbb{R}^V$ be a vector with $\sum_{v \in V} x_v^2 d(v) = 1$, where $d(v)$ is the degree of a vertex $v \in V$. Then by the Cheeger inequality for hypergraphs (Theorem 2.3), the energy of x in a graph expander is $\Omega(1)$ whereas that in an r-uniform hypergraph expander is merely $\Omega(1/r)$. Hence preserving the latter energy is seemingly a harder problem.

2 PRELIMINARIES

In the paper, we will often need to deal with additive or multiplicative errors of various approximations. For simplicity of notation we use $\widetilde{A} = A \pm \delta$ to denote $A - \delta \leq \widetilde{A} \leq A + \delta$, and we use $\widetilde{A} = (1 \pm \epsilon)A$ to denote $(1 - \epsilon)A \leq \widetilde{A} \leq (1 + \epsilon)A$.

2.1 Hypergraph and Expansion

A hypergraph $G = (V, E)$ on a vertex set V is usually defined so that E is a set of hyperedges, each of which is an arbitrary (non-empty) subset of V (as opposed to ordinary graphs, where it is a subset of size two). In a slight departure from the norm, we allow the hyperedges in E to be multisets instead. That is, a hyperedge may contain certain vertices multiple times. This may be thought of as a generalization of the use of self-loops in ordinary graphs, which can be considered as multisets containing a single vertex with multiplicity two — and thus having size two. This slight change in the definition allows us to consider r-uniform hypergraphs throughout the paper without loss of generality, which makes the analysis much simpler. We call a hypergraph *r-uniform* if all of its hyperedges have size r.

Let us denote the multiplicity of a vertex $v \in V$ in $e \in E$ by $\mu_e(v)$. Then the *size* of e is $\sum_{v \in V} \mu_e(v)$ (as is normal for multisets). The *degree* of a vertex v is $d(v) = \sum_{e \in E} \mu_e(v)$.

Furthermore, we also allow hyperedges in E to appear with multiplicity, i.e., parallel edges. This means that E itself is a multiset. We call a hypergraph that has neither multiset edges nor multiple instances of the same edge a *simple hypergraph*.

For a hypergraph $G = (V, E)$ and a set $S \subseteq V$, let $E(S) \subseteq E$ be the multiset of hyperedges $e \in E$ such that every vertex in e belongs to S. Then, let $G[S] = (S, E(S))$ denote the subgraph of G induced by S.

Let $G = (V, E)$ be a hypergraph and $S \subseteq V$ be a vertex set. The *volume* of S, denoted by $\mathrm{vol}(S)$, is $\sum_{v \in S} d(v)$. We say that a hyperedge $e \in E$ is *cut* by S if $e \cap S \neq \emptyset$ and $e \cap (V \setminus S) \neq \emptyset$. In this context, we often call a pair $(S, V \setminus S)$ a *cut*. Let $E(S, V \setminus S)$ denote the set of hyperedges cut by S. Then, the *expansion* of S (or a cut $(S, V \setminus S)$) is

$$\Phi(S) = \frac{|E(S, V \setminus S)|}{\min\{\mathrm{vol}(S), \mathrm{vol}(V \setminus S)\}}.$$

The expansion of a hypergraph $G = (V, E)$ is defined to be $\Phi(G) := \min_{S \subseteq V} \Phi(S)$. For $\Phi \geq 0$, we say that G is a *Φ-expander* if $\Phi(G) \geq \Phi$.

2.2 Spectral Hypergraph Theory

We briefly review spectral theory for hypergraphs. See, e.g., [9, 38] for more details.

Definition 2.1. *Let $G = (V, E)$ be a hypergraph and $x \in \mathbb{R}^V$ be a vector. The* energy *of a hyperedge $e \in E$ with respect to x is defined as $Q_x(e) = \max_{a,b \in e} (x_a - x_b)^2$, and the energy of a subset of hyperedges $E' \subseteq E$ is $Q_x(E') = \sum_{e \in E'} Q_x(e)$, respectively. Finally, the* entire energy *of x is defined as the energy of all hyperedges combined, that is, $Q(x) = Q_x(E)$. If the underlying hypergraph G is unclear from context, we specify by writing $Q_G(x)$.*

Definition 2.2. *Let $G = (V, E)$ be a hypergraph and $\epsilon > 0$. $\widetilde{G} = (V, \widetilde{E}, w)$ is a weighted subgraph of G if w is a vector in $\mathbb{R}_+^E$, mapping each hyperedge $e \in E$ to a non-negative value, and $\widetilde{E}$ denotes $\{e \in E \mid w_e > 0\}$. Such a weighted subgraph is called an ϵ-spectral sparsifier if for any vector $x \in \mathbb{R}^V$, $\widetilde{Q}(x) = (1 \pm \epsilon) \cdot Q(x)$, where $\widetilde{Q}$ denotes energy with respect to the graph $\widetilde{G}$, that is*

$$\widetilde{Q}(x) = \sum_{e \in \widetilde{E}} w_e \cdot Q_x(e).$$

The size of such a sparsifier is $|\widetilde{E}|$.

Given a hypergraph $G = (V, E)$ and a vector $x \in \mathbb{R}^V$, we can define an ordinary graph $G_x = (V, E_x)$ so that the energy of x on G and that on G_x are equal. Specifically, we define E_x as the multiset

$$E_x = \left\{ \left(\operatorname*{argmax}_{a \in e} x_a, \operatorname*{argmin}_{b \in e} x_b \right) \;\middle|\; e \in E \right\},$$

where ties are broken arbitrarily.

The following Cheeger's inequality is a cornerstone of spectral hypergraph theory. Although a similar theorem has been proven in [9, Theorem 6.1], we include the proof in Appendix A for completeness because we do not know whether their proof goes through when we allow for multiset hyperedges.

Theorem 2.3 (Hypergraph Cheeger's inequality). *Let $G = (V, E)$ be an r-uniform hypergraph with expansion at least $\Phi \leq 2/r$. Then for any vector $x \in \mathbb{R}^V$ with $\sum_{v \in V} x_v d(v) = 0$, we have*

$$Q(x) \geq \frac{r\Phi^2}{32} \sum_{v \in V} x_v^2 d(v).$$

Remark 2.4. *In fact, for simple hypergraphs the requirement $\Phi \leq 2/r$ is unnecessary and the statements holds in full generality. In our setting, this requirement is crucial, as non-simple r-uniform hypergraphs may have expansion $\omega(1/r)$, in which case the statement clearly does not hold.*

3 TECHNICAL OVERVIEW

In this section we briefly outline the techniques used in the proofs of our main results.

3.1 Spectral Sparsification of Expanders

We begin by constructing spectral sparsifiers for "good" hypergraph expanders, where we call a hypergraph a good expander if it has expansion at least $\widetilde{\Omega}(1/r)$. Even in this restricted case, no result better than $\widetilde{O}(nr^3/\epsilon^2)$ [6] was known previously. Our plan will then be to partition general input hypergraphs into a series of good expanders. The expansion $\widetilde{\Omega}(1/r)$ is in some sense the best we can hope for. In fact, r-unifrom simple hypergraphs cannot have an expansion better than $\Theta(1/r)$ and consequently no expander decomposition algorithm can guarantee expansion more than that.

To construct our spectral sparsifier for a good expander, we apply importance sampling to the input hypergraph. We sample each hyperedge e independently with some probability p_e and scale it up with weight $1/p_e$ if sampled. This guarantees that $\mathbb{E}(\widetilde{G}) = G$ and so for any vector $x \in \mathbb{R}^V$ we have $\mathbb{E}\widetilde{Q}(x) = Q(x)$, where $\widetilde{Q}$ denotes the energy with respect to the sparsifier. In our case, p_e is inversely proportional to $\min_{v \in e} d(v)$, and then the expected number of sampled hyperedges is proportional to n — simply charge each hyperedge e to a vertex $v \in e$ of minimum degree, then each vertex is in charge of $O(1)$ sampled hyperedges in expectation. It remains to prove that the random quantity $Q(x)$ concentrates well around its expectation *for all vectors x simultaneously.*

So far this is a known technique: similar approaches to constructing spectral sparsifiers in ordinary graphs have appeared in many works, starting from [30, 31]. However, all of these rely on concentration inequalities for linear functions of independent random

variables related to the matrix Bernstein inequality – see, e.g., [34][2]. Unfortunately, the energy of a hypergraph is not a linear transformation and such tools cannot be applied to it. Two recent works on spectral sparsification of hypergraphs developed methods for circumventing this problem, namely [29] and [6]. The former uses a rather crude union bound plus Chernoff bound argument, and loses a factor of n in the size of the sparsifier, both for undirected and directed hypergraphs. The latter, namely the recent work of [6] uses Talagrand's comparison inequality and generic chaining to compare the hypergraph sampling process to effective resistance sampling of [30], and loses a factor of r^3 in the size of the sparsifier. In this work we derive a simultaneous concentration inequality for $\widetilde{Q}(x)$ for all $x \in \mathbb{R}^V$ from more basic principles, and obtain a sparsifier with $\approx nr$ hyperedges as a result – a bound that is seemingly best that can be obtained through the expander decomposition route.

Note that for a single, fixed vector $x \in \mathbb{R}^V$, the concentration inequality $\widetilde{Q}(x) = (1 \pm \epsilon)Q(x)$ holds with high probability by the Chernoff bound (Theorem A.1). Our broad strategy will be to prove concentration over individual choices of x, and combine these results through a union bound. An obvious issue is that x is a continuous variable, making a direct union bound infeasible. We therefore have to discretize it, rounding each x to some $\widetilde{x}$ from a finite net. Our plan then becomes to prove the chain of approximations

$$Q(x) \cong Q(\widetilde{x}) \cong \widetilde{Q}(\widetilde{x}) \cong \widetilde{Q}(x),$$

where the second approximation ($Q(\widetilde{x}) \cong \widetilde{Q}(\widetilde{x})$) utilizes the idea above of a Chernoff bound for each $\widetilde{x}$ plus a union bound over the net.

This turns out to be too simplistic, and the analysis requires a more technical discretization of x. Recall that the energy of the whole hypergraph can be written as a sum of the energies of the individual hyperedges:

$$Q(x) = \sum_{e \in E} Q_x(e).$$

We categorize hyperedges based on a carefully chosen metric $\max_{v \in e} x_v^2 \cdot \min_{v \in e} d(v)$, which we will call the hyperedge's *power*. If a hyperedge's power is approximately 2^{-i}, then it resides in the i^{th} category E_i (see Section 4.2). We have in total a logarithmic number of categories. This categorization is important, because the power of a hyperedge turns out to be closely related to the strength of the Chernoff bound applicable to it, as well as to the required accuracy of the approximation $\widetilde{x}$. That is, some cruder approximation $\widetilde{x}$ may be sufficient to guarantee $Q_x(E_1) \cong Q_{\widetilde{x}}(E_1)$, but it might not be able to guarantee the same for a later category. Conversely, the Chernoff bound is stronger (i.e., the failure probability is smaller) at larger values of i. Thus, for each i we discretize x into a different vector $x^{(i)}$ (rather than the same $\tilde{x}$) and we prove individually for each i that

$$Q_x(E_i) \cong Q_{x^{(i)}}(E_i) \cong \widetilde{Q}_{x^{(i)}}(E_i) \cong \widetilde{Q}_x(E_i).$$

[2]More precisely, the proof of the necessary concentration properties in [31] heavily relies on linearity of the graph Laplacian (specifically, the proof proceeded by bounding the trace of a high power of a corresponding matrix using combinatorial methods), and the analysis of [30] relies on a concentration inequality for linear functions of independent random variables due to Rudelson and Vershynin [27]. Both of these proofs can also be reproduced using the matrix Bernstein inequality.

Here, "$\cong$" necessarily covers both multiplicative *and additive* errors. Indeed, we have no guarantee on the sizes of these categories. Some E_i could contain only a single hyperedge, in which case a simple Chernoff bound would yield no concentration whatsoever. This is where we utilize the additive-multiplicative version (Theorem A.2). Since we have $\Theta(\log n)$ categories to sum over, we naturally allow additive error $\Theta(\epsilon Q(x)/\log n)$.

Note that $x^{(i)}$ is a discretization of x specialized to preserve the energies of hyperedges in E_i. Intuitively, the energy of such a hyperedge e is dictated by the largest value of x_v^2 within it. This value necessarily belongs to a vertex satisfying $x_v^2 d(v) \gtrapprox 2^{-i}$. Thus, it should be enough for our rounding to preserve the x-values of vertices that satisfy this. To this end, we round the x-values of vertices with $x_v^2 d(v) \gtrapprox 2^{-i}$ *carefully* — by an inverse polynomial amount in n. However, we round the x-values of all other vertices to 0 — which is obviously a crude (non-careful) rounding. Thus, if there are only k_i vertices we have to be careful about, the number of possible settings of $x^{(i)}$ becomes $\approx \exp(k_i)$.

Recall the formula of the additive-multiplicative Chernoff bound from Theorem A.2. In our case, the allowable multiplicative error is always $\approx 1+\epsilon$, while the allowable additive error is always $\approx \epsilon Q(x)$. The only quantity that varies from level to level is the range of the random variables involved. If a specific hyperedge is sampled, it is scaled up by $1/p_e \approx \min_{v \in e} d(v)$, and the energy of this weighted hyperedge can be upper bounded by $\approx \max_{v \in e} x_v^2 \cdot \min_{v \in e} d(v)$ — exactly the power of the hyperedge. Thus, at level i, the additive-multiplicative Chernoff bound guarantees a failure probability of $\approx \exp(-2^i Q(x))$. (Here we omit the ϵ terms, along with others, for simplicity.)

Finally, we want to equate the terms in the exponents of the Chernoff bound with the enumeration of $x^{(i)}$'s, so as to bound the total failure probability. We use hypergraph Cheeger (Theorem 2.3) to relate k_i to $Q(x)$. Suppose that x is normalized in the sense that $\sum_{v \in V} x_v^2 d(v) = 1$. This immediately gives that $k_i \le 2^i$ by definition. On the other hand, we can finally use our assumption that the input hypergraph G was a good expander, since hypergraph Cheeger gives us that $Q(x) \gtrapprox 1/r$. This makes the error probability for individual $x^{(i)}$'s $\approx \exp(-2^i/r)$ (from Chernoff bounds), while the enumeration of all $x^{(i)}$ becomes $\approx \exp(2^i)$. To bridge this gap, we must sacrifice a factor r in the sampling ratio p_e, and correspondingly in the size of the output sparsifier (see proof of Claim 4.7).

The formal proof is far more involved, and can be found in Section 4.

3.2 General Spectral Sparsification of Hypergraphs

Having constructed spectral sparsifiers for good expanders, we move our attention to arbitrary input hypergraphs. We decompose the vertex set of the input hypergraph $G = (V, E)$ into clusters of good expansion, while being careful not to cut too many hyperedges between the clusters. We adapt well-known techniques to the setting of hypergraphs, and is detailed for completeness in the full version of the paper. As is common for expander decompositions, we partition V into clusters $C_1, \ldots, C_k$ such that the internal expansion of each cluster (along with its induced hyperedges) is at least

$\widetilde{\Omega}(1/r)$ while cutting only a constant fraction of the hyperedges between the clusters.

In ordinary graphs, this would immediately yield the desired result: We could simply decompose G into expanders and sparsify these, then repeat this process on the discarded hyperedges. Since the number of hyperedges decreases by a constant factor at each level, this process terminates after $O(\log n)$ levels of expander decomposition; each vertex only participates in $O(\log n)$ expanders, and thus the size bound of the overall sparsifier only suffers a logarithmic factor compared to the sparsifiers of expanders. For hypergraphs, this is not the case. Even simple, r-uniform hypergraphs may have up to nearly n^r hyperedges. This means that such a decomposition process could require $r \log n$ levels to terminate, introducing another factor r in the size of the sparsifier.

To combat this problem, we contract clusters into individual supernodes after sparsifying them (see the full version of the paper). This allows us to simply bound the number of clusters a single vertex can participate in, and consequently the size of the output sparsifier. However, proving the correctness of this more complicated algorithm introduces new challenges.

We denote the contracted version of the input hypergraph G by $G/\approx$, where $u \approx v$ if the two vertices u and v have been contracted into the same supernode. We can equate between the hyperedges of G and those of $G/\approx$ using the natural bijection between them (this means that a hyperedge e in G refers also to the corresponding hyperedge in $G/\approx$, and vice versa). Note that this operation can produce multiple parallel hyperdges, as well as *vertices appearing within the same hyperedge with multiplicity*, even if these phenomena were not allowed in the input hypergraph. It is important to note that our expander sparsification algorithm from Section 4 works equally well in this setting. Furthermore, by allowing hyperedges to contain vetices with multiplicity higher than 1, we may continue to work with r-uniform hypergraphs throughout this process of repeatedly contracting vertices. This technicality is crucial, since our expander decomposition algorithm is designed for this setting, and does not work when hyperedges have different sizes (by more than a constant factor).

The main technical contribution of this section is to show that a sparsifier computed *after contraction* still sufficiently approximates the energy of the input hypergraph *before contraction*. Here we take a simplified example: Suppose we wish sparsify a cluster $C \subset V$ and subsequently contract it into a supernode v_C. At a later level we might wish to sparsify some other cluster C' that contains v_C as one of its vertices (see Figure 1). The result is a (weighted) subset of hyperedges that well-approximates the spectral structure of C', but will this still be the case when we un-contract v_C?

Denote the hyperedges of C' by E', and let their sparsifier be $\widetilde{E}'$ (which is a weighted subset of E'). Being a sparsifier with respect to the contracted hypergraph can be viewed as being a sparsifier on the original hypergraph *only when $x \in \mathbb{R}^V$ is uniform, i.e., takes the same value, on all verices of C*, as in this case we can simply assign that same value to v_C, and the energy of the original and contracted hypergraphs will be the same. Unfortunately, we have to deal with general vectors $x \in \mathbb{R}^V$, so we quantify how far x is from satisfing that uniformity requirement. We consider the maximum discrepancy between the x-values of C, defined as $\delta = \max_{u,v \in C} |x_u - x_v|$. We show that the additive error introduced by taking $\widetilde{E}'$ as a sparsifier to E' in the original hypergraph – as opposed to the contracted hypergraph where it is guaranteed to be a good sparsifier – is proportional to δ^2 per hyperedge (see the full version of the paper).

We handle this additive error by arguing that it is dwarfed by energy of x with respect to C. On the one hand, we introduce δ^2 error per hyperedge of C' for a total of at most $\approx \delta^2 d' n$, where d' is the typical degree in C'. On the other hand, we know that the range of x within C is δ, so by hypergraph Cheeger (Theorem 2.3) the energy of C is at least $\approx \delta^2 d/r$, where d is the typical degree in C. (Here we assume that there are no outlier vertices with extremely low degree, which can be guaranteed by a slight adaptation of the expander decomposition subroutine.) Recall that the number of hyperedges — and therefore the typical degree — decreases by a constant factor per level. If we can simply guarantee that the sparsifiaction of C precedes the sparsification of C' by at least $\Omega(\log n)$ levels, then d will dwarf d' by an arbitrarily large $n^{\Theta(1)}$ factor. We accomplish this by simply waiting $\Omega(\log n)$ levels to contract a cluster after sparsifying it.

The formal proof is far more involved, but relies on the same concept of charging additive errors to previous clusters, until we ultimately achieve the desired overall error of $\epsilon Q(x)$. The details appear in the full version of the paper.

3.3 Lower Bounds

The most common method for approximating the Laplacian of a (hyper)graph is to take a weighted subset of the original (hyper)edges. While asympotically optimal for graphs [4, 8], this method has obvious limitations as a data structure: it is not hard to come up with an example where $\Omega(n)$ hyperedges are required even for the sparsifier to be connected, and if the input hypergraph is r-uniform, this translates into $\Omega(nr \log n)$ bit complexity, a linear loss in the arity r of the hypergraph. It is therefore natural to ask whether there are more efficient ways of storing a spectral approximation to a hypergraph. As concrete example, we could permit the inclusion of hyperedges not in the original hypergraph – could this or another scheme lead to a data structure that can approximate the spectral structure of a hypergraph using $\widetilde{O}(n)$ space, avoiding a dependence on r?

We study this question in full generality:

> Is it possible to compress a hypergraph into a $o(n \cdot r)$ size data structure that can approximate the energy $Q_G(x)$ (defined in (2)) simultaneously for all $x \in \mathbb{R}^V$?

We show a space lower bound of $\Omega(nr)$ for sparsifying a hypergraph on n vertices with maximum hyperedge-size r[3]. In fact, our lower bound applies even to the weaker notion of cut sparsification (where one only wants to approximate $Q_G(x)$ for all $x \in \{0,1\}^V$), and is tight by the recent result of [10], who gave a sampling-based cut sparsification algorithm that produces hypergraph sparsifiers with $O(n \log^{O(1)} n)$ hyperedge. In what follows we give an outline of our lower bound.

[3]With some limits on the range of r. For more formal statements of our results see the full version of the paper.

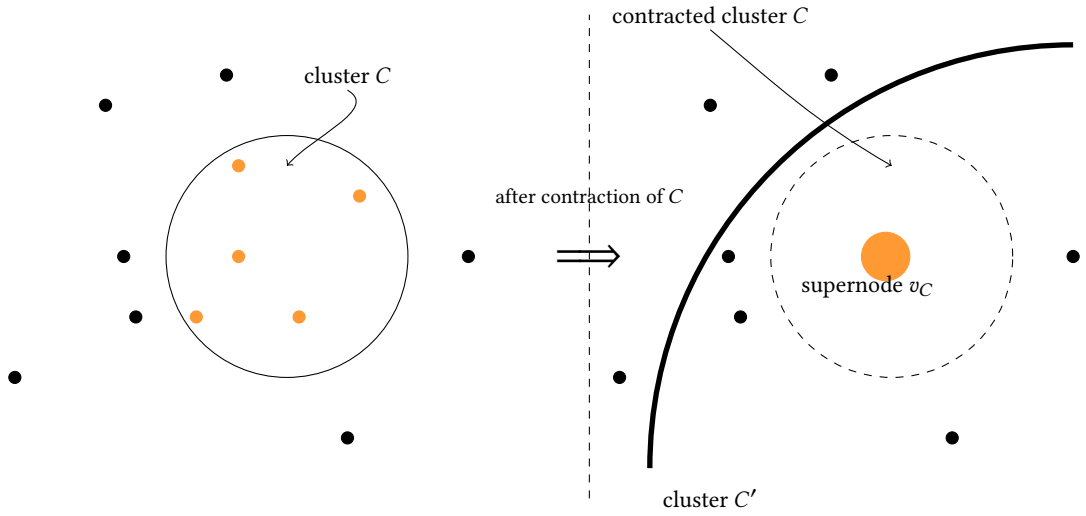

Figure 1: Illustration of the contraction process. Vertices inside C are contracted into a single supernode v_C. This is then contained in a later cluster C'.

We start by formally defining the data structure for approximating the cut structure of a hypergraph that we prove a lower bound for. A *hypergraph cut sparsification scheme (HCSS)* is an algorithm for compressing the cut structure of a hypergraph such that queries on the size of cuts can be answered within a small multiplicative error:

Definition 3.1. *Let $\mathfrak{H}(n,r)$ be the set of hypergraphs on a vertex set $[n]$ with each hyperedge having size at most r. A pair of functions $\textsc{Sparsify} : \mathfrak{H}(n,r) \to \{0,1\}^k$ and $\textsc{Cut} : \{0,1\}^k \times 2^{[n]} \to \mathbb{N}$ is said to be an (n,r,k,ε)-HCSS if for all inputs $G = (V,E) \in \mathfrak{H}(n,r)$ the following holds.*

- *For every query $S \in 2^{[n]}$, $\left|\textsc{Cut}(\textsc{Sparsify}(G), S) - |E(S,\overline{S})|\right| \leq \varepsilon \cdot |E(S,\overline{S})|$.*

To argue a lower bound on the space requirement (parameter k above), we use a reduction to string compression. It is known that $\{0,1\}$-strings of length ℓ cannot be significantly compressed to a small space data structure that allows even extremely crude additive approximations to subset sum queries — see, e.g., the LP decoding paper of [12] (here we only need a lower bound for computationally unbounded adversaries), or the full version of this paper. We manage to encode a $\{0,1\}$-string of length ℓ into the cut structure of a hypergraph H with *fewer hyperedges than ℓ* — a testament to the higher complexity of hypergraph cut structures, as opposed to the cut structures of ordinary graphs.

Our string encoding construction utilizes Ruzsa-Szemerédi graphs. These are (ordinary) graphs whose edge-sets are the union of *induced* matchings. Our construction works generally on any Ruzsa-Szemerédi graphs and as a result we get several lower bounds in various parameter regimes (values of the hyperedge arity r and the precision parameter ϵ) based on the specific Ruzsa-Szemerédi graph constructions we choose to utilize. In particular, for the setting where $r = n^{O(1/\log\log n)}$ we are able to conclude that any hypergraph cut sparsification scheme requires $\Omega(rn)$ bits of space even for constant ϵ, matching the upper bound of [10] to within logarithmic factors. For larger r we get a lower bound of $n^{1-o(1)}r$ bits of space for $\epsilon = n^{-o(1)}$. The latter in particular rules out the possibility of an ϵ-sparsifier that can be described with asymptotically fewer than $(\epsilon^{-1})^{O(1)}nr$ bits of space.

Here we briefly describe how we encode strings into hypergraphs generated from Ruzsa-Szemerédi graphs. Let G be a bipartite Ruzsa-Szemerédi-graph (with bipartition $P \cup Q$) composed of t induced matchings of size a each. We can then use the $a \cdot t$ edges of the graph to encode a string s of length $\ell = at$: simply order the edges of G and remove any edges corresponding to 0 coordinates in s, while keeping edges corresponding to 1's. This graph — which we call G_s — already encodes s when taken as a whole. However, its cut structure is not sufficient for decoding it. For that we need to turn G_s into a hypergraph H_s as follows: For each vertex u on one side of the bipartition, say P, we combine all edges adjacent on u into one hyperedge containing $\{u\} \cup \Gamma(u)$. This means that each hyperedge will have only a single vertex in P, but many vertices in Q (see Figure 2).

To decode the original string s from the cut structure of H, we must be able to answer subset sum queries $q \subseteq [at]$, that is return how many 1-coordinates s has, restricted to q. (For more details see the full version of the paper.) To do this, consider each induced matching one at a time and decode s restricted to the corresponding coordinates. We measure the size of a carefully chosen cut in H_s. Consider Figure 2: We restrict our view to a single matching M_j supported on P_j and Q_j in the two sides of the bipartition. Suppose for simplicity that q is entirely contained in this matching, and we are interested in the Hamming-weight of s restricted to a subset of coordinates q. To create our cut, in the top half of the hypergraph (P), we take the endpoints of edges corresponding to q – we call this set A. In the bottom half (Q), we take everything except for Q_j. The cut, which we call S, is depicted in red in Figure 2.

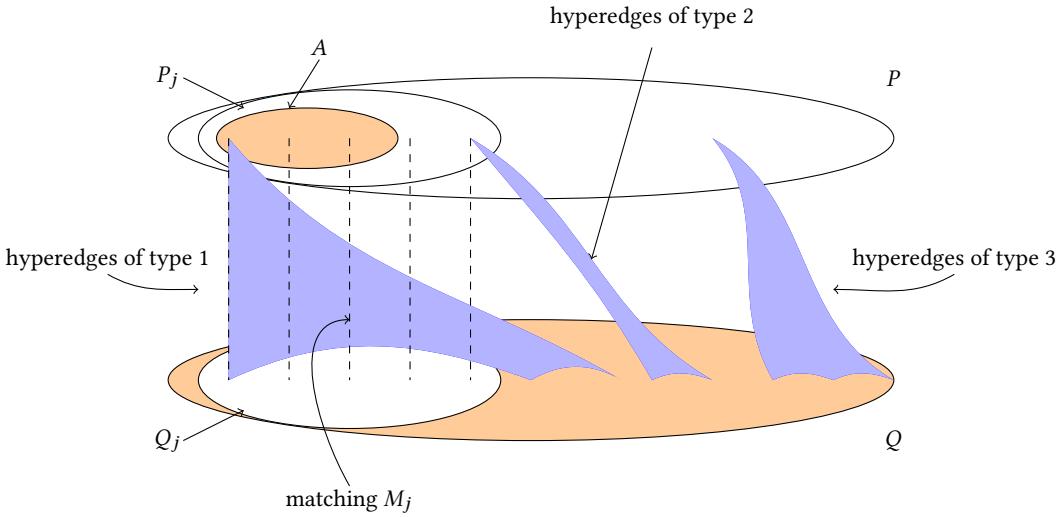

Figure 2: Illustration of the decoding process. One side of the cut S is depicted in orange.

Informally, the crux of the decoding is the observation that the number of hyperedges crossing from A to Q_j is exactly the quantity we want to approximate. Indeed, consider a coordinate in q. If it has value 1 in s, the corresponding hyperedge crosses from A to Q_j, thus crossing the cut S. If however this coordinate is 0 in s, the corresponding hyperedge does not cross to Q_j, thus not crossing the cut. These types of hyperedges are denoted by 1 in Figure 2.

Unfortunately, there are more hyperedges crossing S, adding noise to our measurement of s. One might hope to prove that the noise is small, i.e., can be attributed to measurement error, but this is not the case. Instead, we show that while this noise is not small, it is predictable enough to subtract accurately without knowing s. Hyperedges denoted 2 in Figure 2 cross from $P_j \setminus A$ to $Q \setminus Q_j$. Here we observe that nearly all hyperedges from $P_j \setminus A$ do in fact cross the cut, for almost all choices of s. Hyperedges denoted 3 in Figure 2 cross from $P \setminus P_j$ to $Q \setminus Q_j$. Here we cannot say much about the quantity of such hyperedges crossing the cut. However, we observe that this quantity does not depend on q, and therefore we can use Chernoff bounds (Theorem A.1) to prove that it concentrates around its expectation with high probability *over* s. This allows us to predict and subtract the noise caused by type 3 hyperedges, for whatever instance of Ruzsa-Szemerédi-graph we use .

Ultimately, we show that efficient cut sparsification for such hypergraphs would result in an equally efficient compression of $\{0,1\}$-strings, which implies our lower bounds. For more details see the full version of the paper.

3.4 Directed Spectral Sparsification of Hypergraphs

In the full version of the paper, we also apply our discretization technique from Section 4 to the spectral sparsification of directed hypergraphs. As a testiment to the versitility of this technique, we are able to produce an $O(n^2 r^3 \log^2 n/\epsilon^2)$-sized ϵ-spectral sparsifier. This is a factor n better than the previous state of the art by [29], and nearly optimal in the setting where r is constant.

The broad arc of the proof is very similar to that of Section 4: We construct our sparsifier using importance sampling. We then divide the set of hyperarcs into a logarithmic number of categories, E_i. For each category separately, we show using discretization that the energy of the proposed sparsifier approximates the energy of the input hypergraph with respect to *all* $x \in \mathbb{R}^V$ simultaneously with high probability.

However, the details of each of these steps differ from their corresponding step in Section 4. Here we mention only a few key differences. Instead of looking at degrees or expansion, we define a novel quantity characterizing each hyperarc we call its *overlap*. Intuitively, this denotes the highest density of an induced subgraph in which the paericular hyperarc resides. We then sample each hyperarc with probability inverse proportional to its overlap. We show that this produces a sufficiently small sparsifier with high probability .

Perhaps the most crucial departure from Section 4 occurs during the discretization step when proving $Q_x(E_i) = \widetilde{Q}_x(E_i)$. Instead of discretizing the vector $x \in \mathbb{R}^V$, we discretize the derived vector of energies on the hyperarcs, that is $Q_x \in \mathbb{R}^E$. So for each x and i we define a vector $Q_x^{(i)}$ — from a finite set of possibilities — such that, informally

$$Q_x(E_i) \cong Q_x^{(i)}(E_i) \cong \widetilde{Q}_x^{(i)}(E_i) \cong \widetilde{Q}_x(E_i).$$

This additional trick is necessary; we do not know of a way to make the discretization argument work by rounding x itself.

For more details on the construction of directed hypergraph sparsifiers and their analysis see the full version of the paper.

4 SPECTRAL SPARSIFICATION OF EXPANDERS

In this section, we prove the following.

Theorem 4.1. *There is an algorithm that, given a parameter n, given $100/n \le \epsilon \le 1/2$ and an r-uniform hypergraph $G = (V, E)$ with $|V| \le n$ and expansion at least $350\sqrt{(\log n)/(\epsilon r n)} \le \Phi \le 2/r$, outputs an ϵ-spectral sparsifier of G with $O(|V| \cdot (\frac{1}{\epsilon} \log n)^{O(1)}/(\Phi^2 r))$ hyperedges with probability $1 - O((\log n)/n^2)$ in $O(r|E|)$ time.*

Remark 4.2. *Note that n here does not denote the size of V but an arbitrary parameter larger than that. n serves only as an indirect error parameter, as the failure probability of the algorithm is allowed to be $1 - O((\log n)/n^2)$. The reason for this notation is that later on, we apply Theorem 4.1 to subgraphs of the input hypergraph. In this context, n will denote the the size of the input hypergraph, whereas $|V|$ will denote the (potentially much smaller) size of the cluster within it, to be sparsified. Note that the size of the sparsifier scales linearly in the size of the cluster, but only logarithmically in the size of the input hypergraph. The latter is because the desired failure probability is always defined in terms of n.*

For more details on this, see the full version of the paper.

Remark 4.3. *The guarantee of Theorem 4.1 translates to $|\widetilde{E}| = \widetilde{O}(|V|r)$ when $\Phi(G) = \Omega(1/r)$, i.e., when G is a nearly-optimal expander.*

We show our construction of the sparsifier in Section 4.1 and discuss its correctness in Section 4.2, where some proofs are deferred to Section 4.3.

The following lemma is useful throughout this section.

Lemma 4.4. *For any hypergraph $G = (V, E)$, we have*

$$\sum_{e \in E} \frac{1}{\min_{v \in e} d(v)} \le |V|.$$

Proof. Consider each hyperedge $e \in E$ to be directed towards its vertex with the lowest degree, i.e., $\operatorname{argmin}_{v \in e} d(v)$, breaking ties arbitrarily. Each vertex $v \in V$ has at most $d(v)$ incoming hyperedges, and each such hyperedge contributes to the above sum by $1/d(v)$. Hence the total contribution of all the incoming hyperedges to v is at most 1. It follows that the overall summation is at most $|V|$. □

4.1 Construction

The construction of $\widetilde{G}$ is quite simple. Sample each hyperedge $e \in E$ with probability $p_e = \min\left(\frac{\lambda}{\min_{v \in e} d(v)}, 1\right)$ for

$$\lambda = (\epsilon^{-1} \log n)^{O(1)}/(\Phi^2 r). \tag{4}$$

Each sampled hyperedge e is given weight $w_e = 1/p_e$, and for every non-sampled hyperedge e define $w_e = 0$. Let $\widetilde{G}$ contain the sampled hyperedges, i.e., $\widetilde{E} = \{e \in E \mid w_e > 0\}$. Notice that each random variable w_e has expectation $\mathbb{E}[w_e] = 1$, and thus informally $\mathbb{E}[\widetilde{G}] = G$.

Clearly we can compute the output in time $O(r|E|)$. Also, we can bound the size of the sparsifier with high probability as follows.

Lemma 4.5. *We have*

$$\mathbb{P}[|\widetilde{E}| \ge 2\lambda|V|] \le O(1/n^2),$$

when $|\widetilde{E}| = \Omega(\log n)$.

Proof. First, we have

$$\mathbb{E}[|\widetilde{E}|] \le \sum_{e \in E} p_e \le \lambda \sum_{e \in E} \frac{1}{\min_{v \in e} d(v)} \le \lambda|V|,$$

where the last inequality is due to Lemma 4.4. Noting that $|\widetilde{E}|$ is a sum of independent indicator random variables, the claimed inequality is a direct consequence of the Chernoff bound (Theorem A.1). □

4.2 Correctness

Let us now consider the spectral properties of $\widetilde{G}$. We must prove that with high probability

$$\forall x \in \mathbb{R}^V, \quad \widetilde{Q}(x) = (1 \pm \epsilon) \cdot Q(x). \tag{5}$$

We stress that this gives an error bound that holds for all x *simultaneously*. We may assume without loss of generality that $\sum_{v \in V} x_v d(v) = 0$ and $\sum_{v \in V} x_v^2 d(v) = 1$, because Equation (5) is invariant under translation and scaling of x. Let the set of such centered and normalized vectors be $\overline{\mathbb{R}^V}$. This guarantees that every non-isolated vertex v has $x_v^2 \le 1/d(v) \le 1$, and by Theorem 2.3 we get $Q(x) \ge \frac{r\Phi^2}{32}$.

Now fix one such vector $x \in \overline{\mathbb{R}^V}$, and use it to partition the hyperedge multiset E into $O(\log n)$ subsets as follows. For each $i = 1, \ldots, i^*$, where $i^* = \lceil 2 \log n \rceil$, let

$$E_i = \left\{ e \in E \mid \max_{v \in e} x_v^2 \cdot \min_{v \in e} d(v) \in (2^{-i}, 2^{-i+1}] \right\},$$

and let

$$E_* = E \setminus \bigcup_{i=1}^{i^*} E_i = \left\{ e \in E \mid \max_{v \in e} x_v^2 \cdot \min_{v \in e} d(v) \le 2^{-i^*} \right\}.$$

To justify the second equality in the equation above, note that $\sum_{v \in V} x_v^2 d(v) = 1$ implies $x(v)^2 \le 1/d(v)$, and therefore for every $e \in E$

$$\max_{v \in e} x_v^2 \cdot \min_{v \in e} d(v) \le \max_{v \in e} 1/d(v) \cdot \min_{v \in e} d(v) = 1.$$

Informally, we would like to show that with high probability, for all x and all i we have $\widetilde{Q}_x(E_i) \cong Q_x(E_i)$. Note that the multisets E_i and E_* are dependent on x, but we omit this from the notation for better readability. Our plan is to define another vector $x^{(i)} \in \mathbb{R}^V$ by rounding the coordinates of x, that preserves $Q(E_i)$ up to small multiplicative and additive error. Using this rounded vector, we will then show

$$\widetilde{Q}_x(E_i) \cong \widetilde{Q}_{x^{(i)}}(E_i) \cong Q_{x^{(i)}}(E_i) \cong Q_x(E_i),$$

and similarly also $\widetilde{Q}_x(E_*) \cong Q_x(E_*)$.

Formally, for each $v \in V$ define $x_v^{(i)}$ as follows:

- If $x_v^2 d(v) \ge \epsilon^2 2^{-i}/2500$, then round x_v to the nearest integer multiple of $1/(n^2\sqrt{d(v)})$.
- If $x_v^2 d(v) < \epsilon^2 2^{-i}/2500$, then round x_v to 0.

We implement the above plan using the following four claims.

First, we show that for every scale i the energy of E_i with respect to the rounded vector $x^{(i)}$ is quite close to the energy of E_i with respect to the original vector x:

Claim 4.6. *For all $x \in \overline{\mathbb{R}^V}$ and all $i = 1, \dots, i^*$,*

$$Q_{x^{(i)}}(E_i) = \left(1 \pm \frac{\epsilon}{10}\right) Q_x(E_i) \pm \frac{20}{n}.$$

Next, we show that for every scale i our sampling process preserves energy of E_i on rounded version of all x simultaneously:

Claim 4.7. *For all $i = 1, \dots, i^*$,*

$$\mathbb{P}\left[\forall x \in \overline{\mathbb{R}^V},\ \widetilde{Q}_{x^{(i)}}(E_i) = \left(1 \pm \frac{\epsilon}{10}\right) Q_{x^{(i)}}(E_i) \pm \frac{\epsilon Q(x)}{10 \log n}\right] \geq 1 - \frac{1}{n^2}.$$

We then relate the energy of the sampled E_i on rounded versions of x to the corresponding energy on original x:

Claim 4.8. *For all $i = 1, \dots, i^*$,*

$$\mathbb{P}\left[\forall x \in \overline{\mathbb{R}^V},\ \widetilde{Q}_{x^{(i)}}(E_i) = \left(1 \pm \frac{\epsilon}{10}\right) \widetilde{Q}_x(E_i) \pm \frac{60}{n}\right] \geq 1 - \frac{1}{n^2}.$$

Finally we bound the error introduced on the hyperedges of E_*.

Claim 4.9.

$$\mathbb{P}\left[\forall x \in \overline{\mathbb{R}^V},\ \widetilde{Q}_x(E_*) = Q_x(E_*) \pm \frac{12}{n}\right] \geq 1 - \frac{1}{n^2}.$$

Before proving these claims, which we do in the next section, let us show how to use them to show the correctness of the sparsifier.

Lemma 4.10. *The hypergraph $\widetilde{G}$ is an ϵ-spectral sparsifier of G with probability $1 - O((\log n)/n^2)$.*

Proof. Assume henceforth that the events in Claims 4.7, 4.8, and 4.9 all hold simultaneously for every i — we know that this happens with probability $1 - O(\log n/n^2)$ — and let us compare $Q_x(E_i)$ with $\widetilde{Q}_x(E_i)$ for each i. If the above claims had no additive error, we could conclude that $\widetilde{Q}_x(E_i) = (1 \pm 4\epsilon/10) Q_x(E_i)$. Similarly, if they had no multiplicative error, we could conclude that $|\widetilde{Q}_x(E_i) - Q_x(E_i)| \leq \frac{80}{n} + \frac{\epsilon Q(x)}{10 \log n}$; we could then use the assumed lower bound on Φ to bound $\frac{80}{n} \leq \frac{\epsilon Q(x)}{10 \log n}$, and sum up these additive errors over all $i = 1, \dots, i^*$ to a total that is bounded by $\frac{4}{10} \epsilon Q(x)$. These arguments extend easily also to E_*.

For the formal calculation, consider first one direction,

$$\begin{aligned}
&\widetilde{Q}_x(E_i) \\
&\leq \left(1 - \frac{\epsilon}{10}\right)^{-1} \left[\frac{60}{n} + \widetilde{Q}_{x^{(i)}}(E_i)\right] \\
&\leq \left(1 - \frac{\epsilon}{10}\right)^{-1} \left[\frac{60}{n} + \frac{\epsilon Q(x)}{10 \log n} + \left(1 + \frac{\epsilon}{10}\right) Q_{x^{(i)}}(E_i)\right] \\
&\leq \left(1 - \frac{\epsilon}{10}\right)^{-1} \left[\frac{60}{n} + \frac{\epsilon Q(x)}{10 \log n} + \left(1 + \frac{\epsilon}{10}\right) \left[\frac{20}{n} + \left(1 + \frac{\epsilon}{10}\right) Q_x(E_i)\right]\right] \\
&\leq \frac{120}{n} + \frac{2\epsilon Q(x)}{10 \log n} + \left(1 + \frac{4\epsilon}{10}\right) Q_x(E_i).
\end{aligned}$$

Here the second line follows by Claim 4.8, the third line follows by Claim 4.7, the fourth line follows by Claim 4.6, and the last line follows since $\epsilon \leq 1/2$. Now sum this over all i and combine it with the bound from Claim 4.9 on the error introduced by E_*, to get

$$\begin{aligned}
\widetilde{Q}(x) &= \widetilde{Q}_x(E_*) + \sum_{i=1}^{i^*} \widetilde{Q}_x(E_i) \\
&\leq \left[Q_x(E_*) + \frac{12}{n}\right] + i^* \left[\frac{120}{n} + \frac{2\epsilon Q(x)}{10 \log n}\right] + \left(1 + \frac{4\epsilon}{10}\right) \sum_{i=1}^{i^*} Q_x(E_i) \\
&\leq \frac{250 \log n}{n} + \frac{5\epsilon Q(x)}{10} + \left(1 + \frac{4\epsilon}{10}\right) Q(x) \\
&\leq (1 + \epsilon) Q(x).
\end{aligned}$$

The last inequality follows from $Q(x) \geq \Phi^2 r/32$ (by Theorem 2.3), and the theorem's assumption that $\Phi \geq 350\sqrt{(\log n)/(\epsilon r n)}$.

The other direction, $\widetilde{Q}(x) \geq (1 - \epsilon) Q(x)$, follows similarly. □

Theorem 4.1 then follows by Lemmas 4.5 and 4.10 and a union bound.

4.3 Proofs of Claims 4.6, 4.7, 4.8, and 4.9

We begin by presenting a preliminary lemma about the effects of approximating x on a general quadratic form, which will be useful in proving the four claims.

Lemma 4.11. *Let $G = (V, E)$ be a hypergraph and let $x, \widetilde{x}$ be two vectors in $\mathbb{R}^V$ such that $|x_v - \widetilde{x}_v| \leq \delta$ on every coordinate $v \in V$ for some $\delta \geq 0$. Then for any $e \in E$,*

$$\left|Q_x(e) - Q_{\widetilde{x}}(e)\right| \leq 4\delta \left(\sqrt{Q_x(e)} + \delta\right).$$

Proof. Given $e \in E$, we begin by finding two vertices $u^*, v^* \in e$ such that

$$\left|Q_x(e) - Q_{\widetilde{x}}(e)\right| \leq \left|(x_{u^*} - x_{v^*})^2 - (\widetilde{x}_{u^*} - \widetilde{x}_{v^*})^2\right|.$$

It is indeed possible to find such vertices. If $Q_x(e) \geq Q_{\widetilde{x}}(e)$, set u^*, v^* such that $Q_x(e) = (x_{u^*} - x_{v^*})^2$, and we get

$$\left|Q_x(e) - Q_{\widetilde{x}}(e)\right| = Q_x(e) - Q_{\widetilde{x}}(e) \leq (x_{u^*} - x_{v^*})^2 - (\widetilde{x}_{u^*} - \widetilde{x}_{v^*})^2,$$

since $Q_{\widetilde{x}}(e) \geq (\widetilde{x}_u - \widetilde{x}_v)^2$ for every $u, v \in e$, and in particular $Q_{\widetilde{x}}(e) \geq (\widetilde{x}_{u^*} - \widetilde{x}_{v^*})^2$. Otherwise, i.e., $Q_{\widetilde{x}} > Q_x(e)$, similarly set u^*, v^* such that $Q_{\widetilde{x}}(e) = (\widetilde{x}_{u^*} - \widetilde{x}_{v^*})^2$.

Using these u^*, v^*, we have

$$\begin{aligned}
\left|Q_x(e) - Q_{\widetilde{x}}(e)\right| &\leq \left|(x_{u^*} - x_{v^*})^2 - (\widetilde{x}_{u^*} - \widetilde{x}_{v^*})^2\right| \\
&= |x_{u^*} + \widetilde{x}_{u^*} - x_{v^*} - \widetilde{x}_{v^*}| \cdot |x_{u^*} - \widetilde{x}_{u^*} - x_{v^*} + \widetilde{x}_{v^*}|
\end{aligned}$$

Let us now bound each of these two factors. The second one is clearly bounded by 2δ by the lemma's assumption. To bound the first term, we use that $Q_x(e) = \max_{u,v \in e} (x_u - x_v)^2 \geq (x_{u^*} - x_{v^*})^2$, and therefore

$$\begin{aligned}
|x_{u^*} + \widetilde{x}_{u^*} - x_{v^*} - \widetilde{x}_{v^*}| &\leq 2|x_{u^*} - x_{v^*}| + |x_{u^*} - \widetilde{x}_{u^*}| + |x_{v^*} - \widetilde{x}_{v^*}| \\
&\leq 2\sqrt{Q_x(e)} + 2\delta.
\end{aligned}$$

Putting these two bounds together, we obtain the result of the lemma. □

The following claim examines the effects of rounding from x to $x^{(i)}$ (from the previous section) on a single hyperedge of E_i. This is the main technical claim that allows as to then easily prove both Claims 4.6 and 4.8.

Claim 4.12. *For all $x \in \mathbb{R}^V$, all $i = 1, \dots, i^*$, and every hyperedge $e \in E_i$,*

$$Q_{x^{(i)}}(e) = \left(1 \pm \frac{\epsilon}{10}\right) Q_x(e) \pm \frac{20}{n^2 \min_{v \in e} d(v)}.$$

Proof. We examine the difference $\left|Q_x(e) - Q_{x^{(i)}}(e)\right|$. Recall that $x_v^{(i)}$ is either rounded to the nearest multiple of $1/(n^2\sqrt{d(v)})$ or rounded to zero. We consider two cases:

(1) No vertex in e is rounded to zero.
(2) At least one vertex in e is rounded to zero.

For simplicity, denote $x_+ = \max_{v\in e}|x_v|$ and $d_- = \min_{v\in e} d(v)$. Recall that by definition of E_i,

$$x_+^2 d_- \in \left(2^{-i}, 2^{-i+1}\right]. \tag{6}$$

In the first case, the value of x on every vertex $v \in e$ changes by at most $1/(n^2\sqrt{d(v)}) \le 1/(n^2\sqrt{d_-})$. Thus, we can apply Lemma 4.11 with $\delta = 1/(n^2\sqrt{d_-})$ to get

$$\left|Q_x(e) - Q_{x^{(i)}}(e)\right| \le \frac{4}{n^2\sqrt{d_-}} \cdot \left(\sqrt{Q_x(e)} + \frac{1}{n^2\sqrt{d_-}}\right).$$

We can use (6) to bound $Q_x(e) \le 4x_+^2 \le 4 \cdot 2^{1-i}/d_- \le 4/d_-$. Substituting this in, we get

$$\left|Q_x(e) - Q_{x^{(i)}}(e)\right| \le \frac{4}{n^2\sqrt{d_-}} \cdot \left(\frac{2}{\sqrt{d_-}} + \frac{1}{n^2\sqrt{d_-}}\right) \le \frac{20}{n^2 d_-}.$$

In the second case, the value of x on a vertex in e can still change by at most $1/(n^2\sqrt{d_-})$ by rounding to a non-zero value. It can additionally be rounded to a zero, as long as $x_v^2 d(v) < \epsilon^2 2^{-i}/2500$, which amounts to additive error per coordinate of at most $|x_v| < \epsilon/\sqrt{2500 \cdot 2^i d(v)} \le \epsilon/\sqrt{2500 \cdot 2^i d_-}$. Therefore we can apply Lemma 4.11 with $\delta = \epsilon/\sqrt{2500 \cdot 2^i d_-} \ge 1/(n^2\sqrt{d_-})$, which gives us that

$$\left|Q_x(e) - Q_{x^{(i)}}(e)\right| \le \frac{4\epsilon}{\sqrt{2500 \cdot 2^i d_-}} \cdot \left(\sqrt{Q_x(e)} + \frac{\epsilon}{\sqrt{2500 \cdot 2^i d_-}}\right).$$

This time we use a lower bound on $Q_x(e)$. Recall that we assumed that at least one vertex in e is rounded to zero. Let one such vertex be v_0. This means that $x_{v_0}^2 d(v_0) \le \epsilon^2 2^{-i}/2500$, but at the same time $x_+^2 d(v_0) \ge x_+^2 d_- \ge 2^{-i}$. Using these two facts, we get our lower bound

$$\sqrt{Q_x(e)} \ge x_+ - |x_{v_0}| \ge x_+ - \epsilon x_+/50 \ge \frac{49}{50\sqrt{2^i d_-}}.$$

Substituting this in, we get

$$\begin{aligned}\left|Q_x(e) - Q_{x^{(i)}}(e)\right| &\le \frac{4\epsilon\sqrt{Q_x(e)}}{49} \cdot \left(\sqrt{Q_x(e)} + \frac{\epsilon\sqrt{Q_x(e)}}{49}\right) \\ &\le \frac{4\epsilon Q_x(e)}{49} \cdot \left(1 + \frac{\epsilon}{49}\right) \\ &\le \frac{\epsilon}{10} Q_x(e).\end{aligned}$$

In conclusion, in the first case we get the claimed additive error, while in the second case we get the claimed multiplicative error. □

We are now ready to proceed to proving Claims 4.6 and 4.8.

Claim 4.6. *For all $x \in \overline{\mathbb{R}^V}$ and all $i = 1, \dots, i^*$,*

$$Q_{x^{(i)}}(E_i) = \left(1 \pm \frac{\epsilon}{10}\right) Q_x(E_i) \pm \frac{20}{n}.$$

Proof. We can bound

$$\begin{aligned}\left|Q_x(E_i) - Q_{x^{(i)}}(E_i)\right| &\le \sum_{e\in E_i} \left|Q_x(e) - Q_{x^{(i)}}(e)\right| \\ &\le \sum_{e\in E_i} \left[\frac{\epsilon}{10} Q_x(e) + \frac{20}{n^2 \min_{v\in e} d(v)}\right] \\ &\le \frac{\epsilon}{10} Q_x(E_i) + \frac{20}{n^2} \sum_{e\in E} \frac{1}{\min_{v\in e} d(v)} \\ &\le \frac{\epsilon}{10} Q_x(E_i) + \frac{20}{n}\end{aligned}$$

by Lemma 4.4 and $n \ge |V|$, as claimed. The second line follows by Claim 4.12. □

Claim 4.8. *For all $i = 1, \dots, i^*$,*

$$\mathbb{P}\left[\forall x \in \overline{\mathbb{R}^V},\ \widetilde{Q}_{x^{(i)}}(E_i) = \left(1 \pm \frac{\epsilon}{10}\right)\widetilde{Q}_x(E_i) \pm \frac{60}{n}\right] \ge 1 - \frac{1}{n^2}.$$

Proof. Similarly to the previous proof, we first bound

$$\left|\widetilde{Q}_x(E_i) - \widetilde{Q}_{x^{(i)}}(E_i)\right| \le \sum_{e\in E_i} \left|\widetilde{Q}_x(e) - \widetilde{Q}_{x^{(i)}}(e)\right|.$$

Recall that $\widetilde{Q}_x(e) = w_e Q_x(e)$ where w_e is a random variable (independent from all others) that takes value $1/p_e$ with probability p_e, and value 0 otherwise. Similarly, $\widetilde{Q}_{x^{(i)}}(e) = w_e \widetilde{Q}_{x^{(i)}}(e)$. Applying this along with Claim 4.12, we get

$$\begin{aligned}\left|\widetilde{Q}_x(E_i) - \widetilde{Q}_{x^{(i)}}(E_i)\right| &\le \sum_{e\in E_i} w_e \left|Q_x(e) - Q_{x^{(i)}}(e)\right| \\ &\le \sum_{e\in E_i} \left[\frac{\epsilon}{10} w_e Q_x(e) + w_e \cdot \frac{20}{n^2 \min_{c\in e} d(v)}\right] \\ &= \frac{\epsilon}{10} \widetilde{Q}_x(e) + \sum_{e\in E_i} \frac{20 w_e}{n^2 \min_{v\in e} d(v)}.\end{aligned}$$

Note that in the sum here the term corresponding to e is zero unless e is sampled to $\widetilde{E}$, in which case $w_e = 1/p_e \le 1 + \min_{v\in e} d(v)/\lambda$. (Recall λ from equation 4.) Using also Lemmas 4.4 and 4.5, and the fact that $n \ge |V|$, we have that with high probability

$$\begin{aligned}\sum_{e\in E_i} \frac{20 w_e}{n^2 \min_{v\in e} d(v)} &\le \sum_{e\in \widetilde{E}_i} \frac{20}{n^2 \min_{v\in e} d(v)} + \sum_{e\in \widetilde{E}} \frac{20}{\lambda n^2} \\ &\le \sum_{e\in E_i} \frac{20}{n^2 \min_{v\in e} d(v)} + |\widetilde{E}| \cdot \frac{20}{\lambda n^2} \\ &\le \frac{60}{n}.\end{aligned}$$

□

Claim 4.9.

$$\mathbb{P}\left[\forall x \in \overline{\mathbb{R}^V},\ \widetilde{Q}_x(E_*) = Q_x(E_*) \pm \frac{12}{n}\right] \ge 1 - \frac{1}{n^2}.$$

Proof. Note that

$$\begin{aligned}|\widetilde{Q}_x(E_*) - Q_x(E_*)| &\le \widetilde{Q}_x(E_*) + Q_x(E_*) \\ &= \sum_{e\in E_*} \max_{u,v\in e} (x_u - x_v)^2 + \sum_{e\in E_*} w_e \cdot \max_{u,v\in e} (x_u - x_v)^2.\end{aligned}$$

Now, we bound each term using that $\max_{v\in e} x_v^2 \cdot \min_{v\in e} d(v) \le 1/n^2$ by definition of E_*. For the first term, we use Lemma 4.4,

$$\sum_{e\in E_*} \max_{u,v\in e} (x_u - x_v)^2 \le 4 \sum_{e\in E_*} \max_{v\in e} x_v^2 \le 4 \sum_{e\in E^*} \frac{1}{n^2 \min_{v\in e} d(v)} \le \frac{4}{n}.$$

For the second term, we have by Lemma 4.5, and the fact that $n \ge |V|$, that with high probability,

$$\begin{aligned}\sum_{e\in E_*} w_e \cdot \max_{u,v\in e} (x_u - x_v)^2 &\le 4 \sum_{e\in E_*} w_e \cdot \max_{v\in e} x_v^2 \\ &\le 4 \sum_{e\in E_*} \frac{w_e}{n^2 \min_{v\in e} d(v)} \\ &\le 4|\widetilde{E}| \cdot \frac{1}{\lambda n^2} \\ &\le \frac{8}{n}. \end{aligned}$$ □

Finally, we prove the technical crux of the theorem, Claim 4.7.

Claim 4.7. *For all $i = 1, \dots, i^*$,*

$$\mathbb{P}\left[\forall x \in \overline{\mathbb{R}^V},\ \widetilde{Q}_{x^{(i)}}(E_i) = \left(1 \pm \frac{\epsilon}{10}\right) Q_{x^{(i)}}(E_i) \pm \frac{\epsilon Q(x)}{10 \log n}\right] \ge 1 - \frac{1}{n^2}.$$

Proof. We shall prove the stronger claim

$$\mathbb{P}\left[\forall x \in \overline{\mathbb{R}^V},\ \widetilde{Q}_{x^{(i)}}(E_i) = \left(1 \pm \frac{\epsilon}{10}\right) Q_{x^{(i)}}(E_i) \pm \frac{\epsilon r \Phi^2/32}{10 \log n}\right] \ge 1 - \frac{1}{n^2}.$$

This is indeed stronger, since for all $x \in \overline{\mathbb{R}^V}$, we know that $Q(x) \ge r\Phi^2/32$ by the Hypergraph Cheeger inequality (Theorem 2.3). This allows us to argue that the probabilistic claim depends on x only through $x^{(i)}$ and E_i. These are discrete which will allow for the use of union bound later on. We will first prove a deviation bound for a single instance of $(x^{(i)}, E_i)$ using an additive-multiplicative Chernoff bound, and then extend it to hold for all instances simultaneously using a union bound.

Fix i, $x^{(i)}$, and E_i. Notice that $\widetilde{Q}_{x^{(i)}}(E_i) = \sum_{e\in E_i} w_e \cdot \max_{a,b\in e} (x_a^{(i)} - x_b^{(i)})^2$ is a sum of independent random variables whose expectation is $Q_{x^{(i)}}(E_i)$. Let us bound the maximum range of one summand, for some $e \in E_i$. If $p_e = 1$ the range is 0, and otherwise the range is bounded by

$$\begin{aligned} w_e \cdot \max_{a,b\in e} \left(x_a^{(i)} - x_b^{(i)}\right)^2 &\le \max_{a,b\in e} \frac{2(x_a^2 + x_b^2)}{p_e} \\ &\le \frac{4}{\lambda} \max_{v\in e} x_v^2 \cdot \min_{v\in e} d(v) \\ &\le \frac{2^{-i+3}}{\lambda}. \end{aligned}$$

We can thus apply Theorem A.2 and get

$$\begin{aligned}&\mathbb{P}\left[|\widetilde{Q}_{x^{(i)}}(E_i) - Q_{x^{(i)}}(E_i)| \ge \frac{\epsilon}{10} Q_{x^{(i)}}(E_i) + \frac{\epsilon r\Phi^2/32}{10\log n}\right] \\ &\le 2\exp\left(-\frac{\epsilon/10 \cdot (\epsilon r\Phi^2/32)/(10\log n)}{3 \cdot 2^{-i+3}/\lambda}\right) \\ &= 2\exp\left(-\frac{\lambda 2^i \epsilon^2 r\Phi^2}{32 \cdot 2400 \log n}\right).\end{aligned}$$

Now we wish to extend this high-probability bound to hold simultaneously for all possible $x^{(i)}$ and E_i. How many possible settings of $(x^{(i)}, E_i)$ are there? Each non-zero coordinate v of $x^{(i)}$ has $x_v^2 d(v) \ge \epsilon^2 2^{-i}/2500$, so there are at most $2500 \cdot 2^i/\epsilon^2$ such coordinates. Furthermore, each such coordinate $x_v^{(i)}$ is an integer multiple of $1/(n^2\sqrt{d(v)})$ in the range $[-1/\sqrt{d(v)}, 1/\sqrt{d(v)}]$, so there are only $2n^2$ possibilities per non-zero coordinate. Thus, the total number of vectors $x^{(i)}$ is at most

$$\binom{|V|}{2500 \cdot 2^i/\epsilon^2} \cdot (2n^2)^{2500\cdot 2^i/\epsilon^2} \le (2n^3)^{2500\cdot 2^i/\epsilon^2}$$

We still need to enumerate the number of possible hyperedge multisets E_i given $x^{(i)}$. To know whether some hyperedge $e \in E$ is in E_i, we must know whether the value of $\max_{v\in e} x_v^2 \min_{v\in e} d(v)$ is in $(2^{-i}, 2^{-i+1}]$. Unfortunately, this depends on $\max_{v\in e} x_v^2$, which is not determined by $x^{(i)}$, due to the rounding error between x and $x^{(i)}$. Let $D = \{d(v) \mid v \in V\}$ be the set of all degrees in G. It suffices to know for each v corresponding to a non-zero coordinate of $x^{(i)}$ the two values

$$\min\{d \in D \mid x_v^2 d > 2^{-i}\} \quad \text{and} \quad \max\{d \in D \mid x_v^2 d \le 2^{-i+1}\}.$$

Indeed, we need not worry about zero coordinates of $x^{(i)}$, i.e., vertices v with $x_v^2 d(v) < \epsilon^2 2^{-i}/2500$, as these cannot attain $\max_{u\in e} x_u^2$ for a hyperedge $e \in E_i$. Thus, the total number of possible multisets E_i *given* $x^{(i)}$ is at most $(|V|^2)^{2500\cdot 2^i/\epsilon^2} \le (n^2)^{2500\cdot 2^i/\epsilon^2}$.

We are now ready to apply a union bound,

$$\begin{aligned}&\mathbb{P}\left[\forall x,\ |\widetilde{Q}_{x^{(i)}}(E_i) - Q_{x^{(i)}}(E_i)| \le \frac{\epsilon}{10} Q_{x^{(i)}}(E_i) + \frac{\epsilon(r\Phi^2/32)}{10\log n}\right] \\ &\le \left(n^2 \cdot (2n^3)\right)^{2500\cdot 2^i/\epsilon^2} \cdot 2\exp\left(-\frac{\lambda 2^i \epsilon^2 r\Phi^2}{32 \cdot 2400 \log n}\right) \\ &\le 2\exp\left(\frac{15000 \cdot 2^i \log n}{\epsilon^2} - \frac{\lambda 2^i \epsilon^2 r\Phi^2}{32 \cdot 2400 \log n}\right) \le \frac{1}{n^2},\end{aligned}$$

where the last inequality holds as long as $\lambda \ge 24 \cdot 10^8 \cdot \log^2 n/(\epsilon^4 \Phi^2 r)$, which is indeed how we set λ. □

ACKNOWLEDGMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 759471).

A TECHNICAL LEMMAS

A.1 Concentration Inequalities

The following concentration bound is standard.

Theorem A.1 (Chernoff bound, see e.g. [3]). *Let $X_1, \dots, X_n$ be independent random variables in the range $[0, a]$. Let $\sum_{i=1}^n X_i = S$. Then for any $\delta \in [0, 1]$ and $\mu \ge \mathbb{E}S$,*

$$\mathbb{P}[|S - \mathbb{E}S| \ge \delta\mu] \le 2\exp\left(-\frac{\delta^2\mu}{3a}\right).$$

The following slight variation, allowing for both multiplicative and additive error, will be the most convenient for our purposes throughout the paper.

Theorem A.2 (Additive-multiplicative Chernoff bounds [5]). *Let $X_1, \dots X_n$ be independent random variables in the range*

$[0, a]$. *Let* $\sum_{i=1}^{n} X_i = S$. *Then for all* $\delta \in [0,1]$ *and* $\alpha > 0$,

$$\mathbb{P}[|S - \mathbb{E}S| \geq \delta \mathbb{E}S + \alpha] \leq 2\exp\left(-\frac{\delta\alpha}{3a}\right).$$

A.2 Proof of Hypergraph Cheeger's Inequality

Proof of Theorem 2.3: Recall that $\sum_{v\in V} x_v d(v) = 0$. Suppose for contradiction that there exists a vector $x \in \mathbb{R}^V$ such that $Q(x) < \frac{r\Phi^2}{32}\sum_{v\in V} x_v^2 d(v)$. Let $\widetilde{x} \in \mathbb{R}^V$ be such x shifted such that $\sum_{v\in V} x_v d_x(v) = 0$, where $d_x(v)$ denotes the degree of v in $G_x = (V, E_x)$. Then, we have

$$\begin{aligned} Q(x) &< \frac{r\Phi^2}{32}\sum_{v\in V} x_v^2 d(v) \leq \frac{r\Phi^2}{32}\sum_{v\in V} \widetilde{x}_v^2 d(v) = \frac{r\Phi^2}{32}\sum_{e\in E}\sum_{v\in e}\widetilde{x}_v^2 \\ &\leq \frac{r^2\Phi^2}{32}\sum_{e\in E}\max_{v\in e}\widetilde{x}_v^2 \leq \frac{r^2\Phi^2}{32}\sum_{(a,b)\in E_x}\left(\widetilde{x}_a^2 + \widetilde{x}_b^2\right) \\ &= \frac{r^2\Phi^2}{32}\sum_{v\in V}\widetilde{x}_v^2 d_x(v), \end{aligned}$$

The second inequality follows since x is centered, that is $\sum_{v\in V} x_v d(v) = 0$.

This means, by Cheeger's inequality for ordinary graphs [1, 2], that there exists a vertex set S of expansion $\frac{r\Phi}{4}$ in G_x. Moreover, S can be chosen to be a *sweep cut* with respect to x (regardless of the degree vector) in the sense that S is of the form $\{v \in V \mid x_v \leq \tau\}$ or $\{v \in V \mid x_v \geq \tau\}$ for some $\tau \in \mathbb{R}$. Let $S \subseteq V$ be the smaller side of the cut (in volume). Let $\eta := |E_x(S, V\setminus S)|$ and $\zeta := |E(S)|$. Then, we have

$$\eta \leq \frac{r\Phi}{4}\sum_{v\in S} d_x(v) = \frac{r\Phi}{4}(\eta + 2\zeta).$$

Since $\Phi \leq \frac{2}{r}$, $\frac{r\Phi}{4} \leq \frac{1}{2}$ and so $\zeta \geq \frac{\eta}{r\Phi}$. Since S is a sweep cut with respect to x, each edge of G_x crossing the cut $(S, V\setminus S)$ corresponds to a distinct hyperedge of G also crossing the cut, and each edge of G_x fully inside S translates to a hyperedge of G fully inside S. Therefore, the number of edges crossing the cut $(S, V\setminus S)$ in G is still η and $\sum_{v\in S} d(v) > r\zeta \geq \frac{\eta}{\Phi}$. Similarly, $\sum_{v\in V\setminus S} d(v) > \frac{\eta}{\Phi}$. Therefore, the expansion of the cut $(S, V\setminus S)$ in G is less than Φ, which is a contradiction. □

REFERENCES

[1] Noga Alon. 1986. Eigenvalues and expanders. *Combinatorica* 6, 2 (1986), 83–96. https://doi.org/10.1007/bf02579166

[2] N Alon and V.D Milman. 1985. λ_1, Isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B* 38, 1 (1985), 73–88. https://doi.org/10.1016/0095-8956(85)90092-9

[3] Noga Alon and Joel H. Spencer. 2008. *The Probabilistic Method, Third Edition.* Wiley.

[4] Alexandr Andoni, Jiecao Chen, Robert Krauthgamer, Bo Qin, David P. Woodruff, and Qin Zhang. 2016. On Sketching Quadratic Forms. In *Proceedings of the 2016 Conference on Innovations in Theoretical Computer Science (ITCS).* 311–319. https://doi.org/10.1145/2840728.2840753

[5] Ashwinkumar Badanidiyuru and Jan Vondrák. 2013. Fast algorithms for maximizing submodular functions. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA).* 1497–1514. https://doi.org/10.1137/1.9781611973402.110

[6] Nikhil Bansal, Ola Svensson, and Luca Trevisan. 2019. New Notions and Constructions of Sparsification for Graphs and Hypergraphs. In *Proceedings of the IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS).* 910–928. https://doi.org/10.1109/focs.2019.00059

[7] András A. Benczúr and David R. Karger. 2015. Randomized Approximation Schemes for Cuts and Flows in Capacitated Graphs. *SIAM J. Comput.* 44, 2 (2015), 290–319. https://doi.org/10.1137/070705970

[8] Charles Carlson, Alexandra Kolla, Nikhil Srivastava, and Luca Trevisan. 2019. Optimal Lower Bounds for Sketching Graph Cuts. In *Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA).* 2565—-2569.

[9] T.-H. Hubert Chan, Anand Louis, Zhihao Gavin Tang, and Chenzi Zhang. 2018. Spectral Properties of Hypergraph Laplacian and Approximation Algorithms. *Journal of the ACM* 65, 3 (2018), 1–48. https://doi.org/10.1145/3178123

[10] Yu Chen, Sanjeev Khanna, and Ansh Nagda. 2020. Near-linear Size Hypergraph Cut Sparsifiers. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020.* IEEE, 61–72. https://doi.org/10.1109/FOCS46700.2020.00015

[11] Yu Cheng, Debmalya Panigrahi, and Kevin Sun. 2020. Sparsification of Balanced Directed Graphs. *CoRR* abs/2006.01975 (2020). arXiv:2006.01975 https://arxiv.org/abs/2006.01975

[12] Irit Dinur and Kobbi Nissim. 2003. Revealing information while preserving privacy. In *Proceedings of the 22nd ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS).* 202–210. https://doi.org/10.1145/773153.773173

[13] Eldar Fischer, Eric Lehman, Ilan Newman, Sofya Raskhodnikova, Ronitt Rubinfeld, and Alex Samorodnitsky. 2002. Monotonicity testing over general poset domains. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC).* 474—-483. https://doi.org/10.1145/509907.509977

[14] David Gibson, Jon Kleinberg, and Prabhakar Raghavan. 2000. Clustering categorical data: an approach based on dynamical systems. *The International Journal on Very Large Data Bases* 8, 3-4 (2000), 222–236. https://doi.org/10.1007/s007780050005

[15] Matthias Hein, Simon Setzer, Leonardo Jost, and Syama Sundar Rangapuram. 2013. The Total Variation on Hypergraphs - Learning on Hypergraphs Revisited. In *Advances in Neural Information Processing Systems 26 (NIPS).* 2427–2435.

[16] Yuchi Huang, Qingshan Liu, and Dimitris Metaxas. 2009. Video object segmentation by hypergraph cut. In *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 1738–1745. https://doi.org/10.1109/cvpr.2009.5206795

[17] Masahiro Ikeda, Atsushi Miyauchi, Yuuki Takai, and Yuichi Yoshida. 2018. Finding Cheeger cuts in hypergraphs via heat equation. *arXiv preprint arXiv:1809.04396* (2018). https://arxiv.org/abs/1809.04396

[18] Motoki Ikeda and Shin-ichi Tanigawa. 2018. Cut Sparsifiers for Balanced Digraphs. In *Proceedings of the 16th International Workshop on Approximation and Online Algorithms (WAOA).* 277–294. https://doi.org/10.1007/978-3-030-04693-4_17

[19] Steffen Klamt, Utz-Uwe Haus, and Fabian Theis. 2009. Hypergraphs and Cellular Networks. *PLoS Computational Biology* 5, 5 (2009), e1000385. https://doi.org/10.1371/journal.pcbi.1000385

[20] Dmitry Kogan and Robert Krauthgamer. 2015. Sketching Cuts in Graphs and Hypergraphs. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science (ITCS).* 367–376. https://doi.org/10.1145/2688073.2688093

[21] Yin Tat Lee and He Sun. 2018. Constructing Linear-Sized Spectral Sparsification in Almost-Linear Time. *SIAM J. Comput.* 47, 6 (2018), 2315–2336. https://doi.org/10.1137/16m1061850

[22] Pan Li and Olgica Milenkovic. 2018. Submodular Hypergraphs: p-Laplacians, Cheeger Inequalities and Spectral Clustering. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, Vol. 80. 3020–3029.

[23] Anand Louis. 2015. Hypergraph Markov Operators, Eigenvalues and Approximation Algorithms. In *Proceedings of the 47th Annual ACM on Symposium on Theory of Computing (STOC).* 713–722. https://doi.org/10.1145/2746539.2746555

[24] Ilan Newman and Yuri Rabinovich. 2013. On Multiplicative λ-Approximations and Some Geometric Applications. *SIAM J. Comput.* 42, 3 (2013), 855–883. https://doi.org/10.1137/100801809

[25] P. Ochs and T. Brox. 2012. Higher order motion models and spectral clustering. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 614–621. https://doi.org/10.1109/cvpr.2012.6247728

[26] Prasad Raghavendra and David Steurer. 2010. Graph expansion and the unique games conjecture. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing (STOC).* 755–764. https://doi.org/10.1145/1806689.1806792

[27] Mark Rudelson and Roman Vershynin. 2007. Sampling from large matrices: An approach through geometric functional analysis. *J. ACM* 54, 4 (2007), 21. https://doi.org/10.1145/1255443.1255449

[28] Imre Z Ruzsa and Endre Szemerédi. 1978. Triple systems with no six points carrying three triangles. *Combinatorics (Keszthely, 1976), Coll. Math. Soc. J. Bolyai* 18 (1978), 939–945.

[29] Tasuku Soma and Yuichi Yoshida. 2019. Spectral Sparsification of Hypergraphs. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA).* 2570–2581. https://doi.org/10.1137/1.9781611975482.159

[30] Daniel A. Spielman and Nikhil Srivastava. 2011. Graph Sparsification by Effective Resistances. *SIAM J. Comput.* 40, 6 (2011), 1913–1926. https://doi.org/10.1137/080734029

[31] Daniel A. Spielman and Shang-Hua Teng. 2011. Spectral Sparsification of Graphs. *SIAM J. Comput.* 40, 4 (2011), 981–1025. https://doi.org/10.1137/08074489X

[32] Daniel A. Spielman and Shang-Hua Teng. 2011. Spectral Sparsification of Graphs. *SIAM J. Comput.* 40, 4 (2011), 981–1025. https://doi.org/10.1137/08074489x

[33] Yuuki Takai, Atsushi Miyauchi, Masahiro Ikeda, and Yuichi Yoshida. 2020. Hypergraph Clustering Based on PageRank. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*. 1970—-1978. https://doi.org/10.1145/3394486.3403248

[34] Joel A. Tropp. 2015. An Introduction to Matrix Concentration Inequalities. *Foundations and Trends® in Machine Learning* 8, 1-2 (2015), 1–230. https://doi.org/10.1561/2200000048

[35] Naganand Yadati, Madhav Nimishakavi, Prateek Yadav, Vikram Nitin, Anand Louis, and Partha P. Talukdar. 2019. HyperGCN: A New Method For Training Graph Convolutional Networks on Hypergraphs. In *Advances in Neural Information Processing Systems 32*. 1509–1520.

[36] Naganand Yadati, Vikram Nitin, Madhav Nimishakavi, Prateek Yadav, Anand Louis, and Partha Talukdar. 2020. NHP: Neural Hypergraph Link Prediction. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management (CIKM)*. 1705—-1714. https://doi.org/10.1145/3340531.3411870

[37] Yuichi Yoshida. 2016. Nonlinear Laplacian for Digraphs and its Applications to Network Analysis. In *Proceedings of the 9th ACM International Conference on Web Search and Data Mining (WSDM)*. 483–492. https://doi.org/10.1145/2835776.2835785

[38] Yuichi Yoshida. 2019. Cheeger Inequalities for Submodular Transformations. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2582–2601. https://doi.org/10.1137/1.9781611975482.160

[39] Chenzi Zhang, Shuguang Hu, Zhihao Gavin Tang, and T-H. Hubert Chan. 2020. Re-Revisiting Learning on Hypergraphs: Confidence Interval, Subgradient Method, and Extension to Multiclass. *IEEE Transactions on Knowledge and Data Engineering* 32, 3 (2020), 506–518. https://doi.org/10.1109/tkde.2018.2880448