

# Expander Flows, Geometric Embeddings and Graph Partitioning

(Longer version of an ACM STOC 2004 paper)

Sanjeev Arora\*      Satish Rao†      Umesh Vazirani‡

April 2, 2007

## Abstract

We give a  $O(\sqrt{\log n})$ -approximation algorithm for SPARSEST CUT, EDGE EXPANSION, BALANCED SEPARATOR, and GRAPH CONDUCTANCE problems. This improves the  $O(\log n)$ -approximation of Leighton and Rao (1988). We use a well-known semidefinite relaxation with triangle inequality constraints. Central to our analysis is a geometric theorem about projections of point sets in  $\Re^d$ , whose proof makes essential use of a phenomenon called measure concentration.

We also describe an interesting and natural “approximate certificate” for a graph’s expansion, which involves embedding an  $n$ -node expander in it with appropriate dilation and congestion. We call this an expander flow.

## 1 Introduction

Partitioning a graph into two (or more) large pieces while minimizing the size of the “interface” between them is a fundamental combinatorial problem. Graph partitions or separators are central objects of study in the theory of Markov chains, geometric embeddings and are a natural algorithmic primitive in numerous settings, including clustering, divide and conquer approaches, PRAM emulation, VLSI layout, and packet routing in distributed networks. Since finding optimal separators is NP-hard, one is forced to settle for approximation algorithms (see [31]). Here we give new approximation algorithms for some of the important problems in this class.

Graph partitioning involves finding a cut with few crossing edges conditioned on or normalized by the size of the smaller side of the cut. This gives rise to several related measures of the quality of the cut, depending on precisely how size is measured, such as conductance, expansion, normalized or sparsest cut. These measures are irreducible within a constant factor, are all NP-hard, and arise naturally in different contexts.

A weak approximation for GRAPH CONDUCTANCE follows from the connection—first discovered in the context of Riemannian manifolds [8]—between conductance and the eigenvalue gap of the Laplacian:  $2\Phi(G) \geq \lambda \geq \Phi(G)^2/2$  [3, 2, 21]. The approximation factor is  $1/\Phi(G)$ , and hence  $\Omega(n)$  in the worst case, and  $O(1)$  only if  $\Phi(G)$  is a constant. This connection between eigenvalues and expansion has had enormous influence in a variety of fields (see e.g. [9]).

---

\*Computer Science Department, Princeton University. [www.cs.princeton.edu/~arora](http://www.cs.princeton.edu/~arora). Work done partly while visiting UC Berkeley (2001-2002) and the Institute for Advanced Study (2002-2003). Supported by David and Lucille Packard Fellowship, and NSF Grants CCR-0098180 and CCR-0205594 MSPA-MCS 0528414, CCF 0514993, ITR 0205594.

†Computer Science Division, UC Berkeley. [www.cs.berkeley.edu/~satishr](http://www.cs.berkeley.edu/~satishr). Partially supported by NSF award CCR-0105533

‡Computer Science Division, UC Berkeley. [www.cs.berkeley.edu/~vazirani](http://www.cs.berkeley.edu/~vazirani) Partially supported by NSF ITR Grant CCR-0121555

Leighton and Rao [22] designed the first true approximation by giving  $O(\log n)$ -approximations for SPARSEST CUT and GRAPH CONDUCTANCE and  $O(\log n)$ -pseudo-approximations<sup>1</sup> for  $c$ -BALANCED SEPARATOR. They used a linear programming relaxation of the problem based on multicommodity flow proposed in [30]. These ideas led to approximation algorithms for numerous other NP-hard problems, see [31]. We note that the integrality gap of the LP is  $\Omega(\log n)$ , and therefore improving the approximation factor necessitates new techniques.

In this paper, we give  $O(\sqrt{\log n})$ -approximations for SPARSEST CUT, EDGE EXPANSION, and GRAPH CONDUCTANCE and  $O(\sqrt{\log n})$ -pseudo-approximation to  $c$ -BALANCED SEPARATOR. As we describe below, our techniques have also led to new approximation algorithms for several other problems, as well as a breakthrough in geometric embeddings.

The key idea underlying algorithms for graph partitioning is to spread out the vertices in some abstract space while not stretching the edges too much. Finding a good graph partition is then accomplished by partitioning this abstract space. In the eigenvalue approach, the vertices are mapped to points on the real line such that the average squared distance is constant, while the average squared distance between the endpoints of edges is minimized. Intuitively, snipping this line at a random point should cut few edges, thus yielding a good cut. In the linear programming approach, lengths are assigned to edges such that the average distance between pairs of vertices is fixed, while the average edge length is small. In this case, it can be shown that a ball whose radius is drawn from a particular distribution defines a relatively balanced cut with few expected crossing edges, thus yielding a sparse cut.

Note that by definition, the distances in the linear programming approach form a metric (i.e., satisfy the triangle inequality) while in the eigenvalue approach they don't. On the other hand, the latter approach works with a geometric embedding of the graph, whereas there isn't such an underlying geometry in the former. In this paper, we work with an approach that combines both geometry and the metric condition. In particular, this approach maps the vertices to points in an  $n$  dimensional space such that the average square distance between vertices is a fixed constant, but the average squared distance between the endpoints of edges is minimized. Furthermore, we insist that these squared distances form a metric. This embedding, which we refer to as an  $\ell_2^2$  representation, can be computed in polynomial time using semi-definite programming. The crux of this paper is the extraction of a partitioning of the graph from this embedding.

The key to achieving this is a new result (Theorem 1) about the geometric structure of  $\ell_2^2$ -representations: they contain  $\Omega(n)$  sized sets  $S$  and  $T$  that are well-separated, in the sense that every pair of points  $v_i \in S$  and  $v_j \in T$  must be at least  $\Delta = \Omega(1/\sqrt{\log n})$  apart in  $\ell_2^2$  (squared Euclidean) distance. The set  $S$  can be used to find a good cut as follows: consider all points within some uniformly chosen distance  $\delta < \Delta$  from  $S$ . The typical number of edges crossing such a cut must be small, since the length of a typical edge is short relative to  $\Delta$ .

Furthermore, the sets  $S$  and  $T$  (with worse bounds) can be constructed algorithmically by projecting the points on a random line as follows (Section 3): for suitable constant  $c$ , the leftmost  $cn$  and rightmost  $cn$  points on the line are our first candidates for  $S$  and  $T$ . However, they can contain pairs of points  $v_i \in S, v_j \in T$  whose squared distance is less than  $\Delta$ , which we discard. The technically hard part in the analysis is to prove that not too many points get discarded. This delicate argument makes essential use of a phenomenon called *measure concentration*, a cornerstone of modern convex geometry [5].

The result on well-separated sets is tight for an  $n$ -vertex hypercube—whose natural embedding into  $\mathbb{R}^{\log n}$  defines an  $\ell_2^2$  metric—where any two large sets are within  $O(1/\sqrt{\log n})$  distance. What does this indicate about the possibility of improving the  $O(\sqrt{\log n})$  approximation ratio for sparsest cuts? The hypercube example would appear to be a serious obstacle, since the dimension cuts elude all our techniques. In addition, hypercube like examples show that this particular SDP approach has an integrality gap of ... In section ... we formulate a sequence of conditions that would imply improved approximations.

The graph partitioning problem is closely related to metric embeddings. An  $\alpha(n)$  approximation for conductance and other measures for graph partitioning would follow from any general upper bound  $\alpha(n)$  for the minimum distortion for embedding  $\ell_2^2$  metrics into  $\ell_1$ . Our result on well separated sets has recently been

---

<sup>1</sup>For any fixed  $c' < c$  the pseudo-approximation algorithm finds a  $c'$ -balanced cut whose expansion is at most  $O(\log n)$  times expansion of the best  $c$ -balanced cut.

used [ALN05] to bound this distortion by  $O(\sqrt{\log n} \log \log n)$ , improving on Bourgain’s classical  $O(\log n)$  bound that applies to any metric. In fact, they give a stronger result: an embedding of  $\ell_2^2$  metrics into  $\ell_2$ . Since any  $\ell_2$  metric is an  $\ell_1$  metric, which in turn is an  $\ell_2^2$  metric, their embedding also gives an  $\ell_1$  to  $\ell_2$  embedding with the same distortion. This almost resolves a central problem in the area of metric embeddings, since any  $\ell_2$  embedding of the hypercube, regarded as an  $\ell_1$  metric, must suffer an  $\Omega(\sqrt{\log n})$  distortion [Enf69].

The challenge in producing an embedding is to ensure that the images of any two points from the original  $\ell_2^2$  metric, are far enough apart in the resulting  $\ell_1$  metric. This is accomplished in [ALN05] by generating coordinates at each distance scale  $\Delta$ . The main idea to generate such a coordinate is to find a well-separated pair of sets,  $S$  and  $T$ , and then to use “distance to  $S$ ” as the value of this coordinate. Since each node in  $T$  is at least  $\Delta/\sqrt{\log n}$  away from every node in  $S$ , this ensures that a large number of  $\Delta$ -separated pairs of nodes are far enough apart. The  $O(\sqrt{\log n} \log \log n)$  bound on the distortion requires a clever accounting scheme across distance scales.

In this paper, we also explore a combinatorial approach to graph partitioning that leads to a closely related algorithm with the same performance guarantee, but one which has led to fast implementations. The resulting algorithm does not use SDP’s but relies on the notion of *expander flows*, which constitute an interesting and natural “certificate” of a graph’s expansion. Note that any algorithm that approximates edge expansion  $\alpha = \alpha(G)$  must implicitly certify that every cut has large expansion. One way to do this is certification is to embed<sup>2</sup> a complete graph into the given graph with minimum congestion, say  $\mu$ . (Determining  $\mu$  is a polynomial-time computation.) Then it follows that every cut must have expansion at least  $n/\mu$ . (See Section 7.) This is exactly the certificate used in the Leighton-Rao paper, where it is shown that congestion  $O(n \log n / \alpha(G))$  suffices (and this amount of congestion is required on some worst-case graphs). Thus embedding the complete graph suffices to certify that the expansion is at least  $\alpha(G)/\log n$ .

Our certificate can be seen as a generalization of this approach, whereby we embed not the complete graph but some flow that is an “expander” (a graph whose conductance is  $\Omega(1)$ ). We show that for every graph there is an expander flow that certifies an expansion of  $\Omega(\alpha(G)/\sqrt{\log n})$  (see Section 7). This near-optimal embedding of an expander inside an arbitrary graph may be viewed as a purely structural result in graph theory. It is therefore interesting that this graph-theoretic result is proved using geometric arguments similar to the ones used to analyse our SDP-based algorithm. (There is a natural connection between the two algorithms since expander flows can be viewed as a natural family of dual solutions to the above-mentioned SDP, see Section 7.4.) In fact, the expander flows approach was the original starting point of our work.

How does one efficiently compute an expander flow? The condition that the embedded graph is an expander can be imposed by exponentially many linear constraints, one for each cut. Since the eigenvalue approach gives a constant factor approximation when the conductance is large, a violated constraint (to within constant factors) can be efficiently found, and thus the linear program can be approximately solved by the Ellipsoid method. This approach gives a polynomial time algorithm for computing expander flows. Arora, Hazan, and Kale [AHK04] used this framework to design an efficient  $\tilde{O}(n^2)$  time approximation algorithm, thus matching the running time of the best implementations of the Leighton-Rao algorithm. Their algorithm may be thought of as a primal dual algorithm and proceeds by alternately computing eigenvalues and solving min-cost multicommodity flows. Khandekar, Rao, and Vazirani [KRV06] uses the expander flow approach to break the  $\tilde{O}(n^2)$  multicommodity flow barrier and rigorously analyse an algorithm that resemble some practically effective heuristics [LR04]. The analysis gives a worse  $O(\log^2 n)$  approximation for sparsest cut but requires only  $O(\log^2 n)$  invocations of a maximum flow algorithm.

## Related prior and subsequent work.

**Semidefinite programming and approximation algorithms:** Semidefinite programs (SDPs) have numerous applications in optimization. They are solvable in polynomial time via the ellipsoid method [18], and

---

<sup>2</sup>Note that this notion of graph embedding has no connection in general to the area of geometric embeddings. It is somewhat confusing that this paper features both notions and shows some connection between them.

more efficient interior point methods are now known [1, 27]. In a seminal paper, Goemans and Williamson [17] used SDPs to design good approximation algorithms for MAX-CUT and MAX- $k$ -SAT. Researchers soon extended their techniques to other problems [20, 19, 15], but lately progress in this direction had stalled. Especially in the context of minimization problems, the GW approach of analysing “random hyperplane” rounding in an edge-by-edge fashion runs into well-known problems. By contrast, our theorem about well separated sets in  $\ell_2^2$  spaces (and the “rounding” technique that follows from it) takes a more global view of the metric space. It is the mathematical heart of our technique, just as the *region-growing* argument was the heart of the Leighton-Rao technique for analysing LPs.

Several papers have pushed this technique further, using our main theorem to design new  $\sqrt{\log n}$ -approximation algorithms for graph deletion problems such as 2CNF-DELETION and MIN-UNCUT [ACMM05], for MIN-LINEAR ARRANGEMENT [CHKR06, FL], and VERTEX SEPARATORS [FHL05]. Similarly, MIN-VERTEX COVER can now be approximated upto a factor  $2 - \Omega(1/\sqrt{\log n})$  [Kar05], an improvement over  $2 - \Omega(\log \log n / \log n)$  achieved by prior algorithms.

However, we note that the Structure Theorem does not provide a blanket improvement for all the problems for which  $O(\log n)$  algorithms were earlier designed using the Leighton-Rao technique. In particular, the integrality gap of the SDP relaxation for MIN-MULTICUT was recently shown to be  $\Omega(\log n)$  [ACMM05], the same (upto a constant factor) as the integrality gap for the LP relaxation.

**Analysis of random walks:** The mixing time of a random walk on a graph is related to the first nonzero eigenvalue of the Laplacian, and hence to the conductance. Of various techniques known for upperbounding the mixing time, most rely on lowerbounding the conductance. Diaconis and Saloff-Coste [11] describe a very general idea called the *comparison technique*, whereby the conductance of a graph is lowerbounded by embedding a *known* graph with *known* conductance into it. (The embedding need not be constructive; existence suffices.) Sinclair [32] suggested a similar technique and also noted that the Leighton-Rao multicommodity flow can be viewed as a generalization of the Jerrum-Sinclair [21] canonical path argument. Our results on expander flows imply that the comparison technique can be used to always get to within  $O(\sqrt{\log n})$  of the proper bound for conductance.

**Metric spaces and relaxations of the cut cone:** The *cut cone* is the cone of all cut semi-metrics, and is equivalent to the cone of all  $\ell_1$  semi-metrics. Graph separation problems can often be viewed as the optimization of a linear function over the cut cone (possibly with some additional constraints imposed). Thus optimization over the cut cone is NP-hard. However, one could relax the problem and optimize over some other metric space, embed this metric space in  $\ell_1$  (hopefully, with low distortion), and then derive an approximation algorithm. This approach was pioneered in [23] and [4]; see Shmoys’s survey [31]. These papers showed that the integrality gap of the obvious generalization of our SDP for NONUNIFORM SPARSEST CUT is *exactly* the same as as the worst-case distortion incurred in embedding  $n$ -point  $\ell_2^2$  metrics into  $\ell_1$ . A famous result of Bourgain shows that the distortion is at most  $O(\log n)$ , which yields an  $O(\log n)$ -approximation for NONUNIFORM SPARSEST CUT. Improving this been a major open problem.

Our above-mentioned main theorem can be viewed as a weak “average case” embedding from  $\ell_2^2$  into  $\ell_1$  whose distortion is  $O(\sqrt{\log n})$ . This is pointed out by Chawla, Gupta, and Racke [CGR05], who used it together with the measured descent technique of Krauthgamer, Lee, Mendel, Naor [KLMN04] to show that  $n$ -point  $\ell_2^2$  metrics embed into  $\ell_2$  (and hence into  $\ell_1$ ) with distortion  $O(\log^{3/4} n)$ . As mentioned previously, Arora, Lee, and Naor [ALN05] improved this distortion to  $O(\sqrt{\log n} \log \log n)$ , which is optimal up to  $\log \log n$  factor.

The above developments showed that the integrality gap for NONUNIFORM SPARSEST CUT is  $O(\sqrt{\log n} \log \log n)$ . There was speculation that the integrality gap may even be  $O(1)$ , but very recently Khot and Vishnoi [KV05] showed it is at least  $(\log \log n)^\epsilon$  for some  $\epsilon > 0$ .

## 2 Definitions and Results

We first define the problems considered in this paper; all are easily proved to be NP-hard [22]. Given a graph  $G = (V, E)$ , the goal in the UNIFORM SPARSEST CUT problem is to determine the cut  $(S, \bar{S})$  (where

$|S| \leq |\overline{S}|$  without loss of generality) that minimizes

$$\frac{|E(S, \overline{S})|}{|S| |\overline{S}|}. \quad (1)$$

Since  $|V|/2 \leq |\overline{S}| \leq |V|$ , up to a factor 2 computing the sparsest cut is the same as computing the *edge expansion* of the graph, namely,

$$\alpha(G) = \min_{S \subseteq V, |S| \leq |V|/2} \frac{|E(S, \overline{S})|}{|S|}.$$

Since factors of 2 are inconsequential in this paper, we use the two problems interchangeably. Furthermore, we often shorten UNIFORM SPARSEST CUT to just SPARSEST CUT. In another related problem, *c*-BALANCED-SEPARATOR, the goal is to determine  $\alpha_c(G)$ , the minimum edge expansion among all *c*-balanced cuts. (A cut  $(S, \overline{S})$  is *c*-balanced if both  $S, \overline{S}$  have at least  $c|V|$  vertices.) In the GRAPH CONDUCTANCE problem we wish to determine

$$\Phi(G) = \min_{S \subseteq V, |E(S)| \leq |E|/2} \frac{|E(S, \overline{S})|}{|E(S)|}, \quad (2)$$

where  $E(S)$  denotes the set of edges incident to nodes in  $S$ . There are well-known interreducibilities among these problems.

As mentioned, all our algorithms depend upon a geometric representation of the graph.

DEFINITION 1 ( $\ell_2^2$  REPRESENTATION) *An  $\ell_2^2$ -representation of a graph is an assignment of a point (vector) to each node, say  $v_i$  assigned to node  $i$ , such that for all  $i, j, k$ :*

$$|v_i - v_j|^2 + |v_j - v_k|^2 \geq |v_i - v_k|^2 \quad (\text{triangle inequality}) \quad (3)$$

*An  $\ell_2^2$ -representation is called a unit- $\ell_2^2$  representation if all points lie on the unit sphere (or equivalently, all vectors have unit length.)*

REMARK 1 We mention two alternative descriptions of unit  $\ell_2^2$  representations which however will not play an explicit role in this paper. (i) Geometrically speaking, the above triangle inequality says that every  $v_i$  and  $v_k$  subtend a nonobtuse angle at  $v_j$ . (ii) Every positive semidefinite  $n \times n$  matrix has a *Cholesky factorization*, namely, a set of  $n$  vectors  $v_1, v_2, \dots, v_n$  such that  $M_{ij} = \langle v_i, v_j \rangle$ . Thus a unit- $\ell_2^2$ -representation for an  $n$  node graph can be alternatively viewed as a positive semidefinite  $n \times n$  matrix  $M$  whose diagonal entries are 1 and  $\forall i, j, k, M_{ij} + M_{jk} - M_{ik} \leq 1$ .

Every cut  $(S, \overline{S})$  gives rise to a natural unit- $\ell_2^2$  representation, namely, one that assigns some unit vector  $v_0$  to every vertex in  $S$  and  $-v_0$  to every vertex in  $\overline{S}$ . Thus the following SDP is a relaxation for  $\alpha_c(G)$  (scaled by  $cn$ ).

$$\min \quad \frac{1}{4} \sum_{\{i,j\} \in E} |v_i - v_j|^2 \quad (4)$$

$$\forall i \quad |v_i|^2 = 1 \quad (5)$$

$$\forall i, j, k \quad |v_i - v_j|^2 + |v_j - v_k|^2 \geq |v_i - v_k|^2 \quad (6)$$

$$\sum_{i < j} |v_i - v_j|^2 \geq 4c(1-c)n^2 \quad (7)$$

This SDP motivates the following definition.

DEFINITION 2 *An  $\ell_2^2$ -representation is  $c$ -spread if equation (7) holds.*

Similarly the following is a relaxation for sparsest cut (up to scaling by  $n$ ; see Section 6).

$$\min \sum_{\{i,j\} \in E} |v_i - v_j|^2 \tag{8}$$

$$\forall i, j, k \quad |v_i - v_j|^2 + |v_j - v_k|^2 \geq |v_i - v_k|^2 \tag{9}$$

$$\sum_{i < j} |v_i - v_j|^2 = 1 \tag{10}$$

As we mentioned before the SDPs subsume both the eigenvalue approach and the Leighton-Rao approach [15]. We show that the optimum value of the SPARSEST CUT SDP is  $\Omega(\alpha(G)n/\sqrt{\log n})$ , which shows that the integrality gap is  $O(\sqrt{\log n})$ .

## 2.1 Main theorem about $\ell_2^2$ representations

In general,  $\ell_2^2$ -representations are not well-understood<sup>3</sup> This is not surprising since in  $\mathfrak{R}^d$  the representation can have at most  $2^d$  distinct vectors [10], so our three-dimensional intuition is of limited use for graphs with more than  $2^3$  vertices. The technical core of our paper is a new theorem about unit  $\ell_2^2$  representations.

**DEFINITION 3 ( $\Delta$ -SEPARATED)** *If  $v_1, v_2, \dots, v_n \in \mathfrak{R}^d$ , and  $\Delta \geq 0$ , two disjoint sets of vectors  $S, T$  are  $\Delta$ -separated if for every  $v_i \in S, v_j \in T$ ,  $|v_i - v_j|^2 \geq \Delta$ .*

**THEOREM 1 (MAIN)**

*For every  $c > 0$ , there are  $c', b > 0$  such that every  $c$ -spread unit- $\ell_2^2$  representation with  $n$  points contains  $\Delta$ -separated subsets  $S, T$  of size  $c'n$ , where  $\Delta = b/\sqrt{\log n}$ . Furthermore, there is a randomized polynomial-time algorithm for finding these subsets  $S, T$ .*

**REMARK 2** The value of  $\Delta$  in this theorem is best possible up to a constant factor, as demonstrated by the natural embedding (scaled to give unit vectors) of the boolean hypercube  $\{-1, 1\}^d$ . These vectors form a unit  $\ell_2^2$  representation, and the isoperimetric inequality for hypercubes shows that every two subsets of  $\Omega(n)$  vectors contain a vector from each whose squared distance is  $O(1/\sqrt{\log n})$  (here  $n = 2^d$ ). Theorem 1 may thus be viewed as saying that every  $\Omega(1)$ -spread unit  $\ell_2^2$ -representation is “hypercube-like” with respect to the property mentioned in the theorem. Some may view this as evidence for the conjecture that  $\ell_2^2$ -representations are indeed like hypercubes (in particular, embed in  $\ell_1$  with low distortion). The recent results in [CGR05, ALN05], which were inspired by our paper, provide partial evidence for this belief.

### 2.1.1 Corollary to Main Theorem: $\sqrt{\log n}$ -approximation

Let  $W = \sum_{\{i,j\} \in E} \frac{1}{4} |v_i - v_j|^2$  be the optimum value for the SDP defined by equations (4)–(7). Since the vectors  $v_i$ 's obtained from solving the SDP satisfy the hypothesis of Theorem 1, as an immediate corollary to the theorem we show how to produce a  $c'$ -balanced cut of size  $O(W\sqrt{\log n})$ .

**COROLLARY 2**

*There is a randomized polynomial-time algorithm that finds with high probability a cut that is  $c'$ -balanced, and has size  $O(W\sqrt{\log n})$ .*

**PROOF:** We use the algorithm of Theorem 1 to produce  $\Delta$ -separated subsets  $S, T$  for  $\Delta = b/\sqrt{\log n}$ . Let  $V_0$  denote the vertices whose vectors are in  $S$ . Associate with each edge  $e = \{i, j\}$  a length  $w_e = \frac{1}{4} |v_i - v_j|^2$ . (Thus  $W = \sum_{e \in E} w_e$ .)  $S$  and  $T$  are at least  $\Delta$  apart with respect to this distance. Let  $V_s$  denote all vertices within distance  $s$  of  $S$ . Now we produce a cut as follows: pick a random number  $r$  between 0 and  $\Delta$ , and output the cut  $(V_r, V - V_r)$ . Since  $S \subseteq V_r$  and  $T \subseteq V - V_r$ , this is a  $c'$ -balanced cut.

<sup>3</sup>As mentioned in the introduction, it has been conjectured that they are closely related to the better-understood  $\ell_1$  metrics.

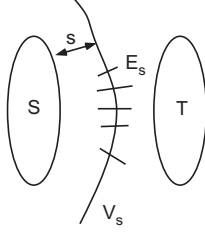


Figure 1:  $V_s$  is the set of nodes whose distance on the weighted graph to  $S$  is at most  $s$ . Set of edges leaving this set is  $E_s$ .

To bound the size of the cut, we denote by  $E_s$  the set of edges leaving  $V_s$ . Each edge  $e = \{i, j\}$  only contributes to  $E_s$  for  $s$  in the open interval  $(s_1, s_2)$ , where  $s_1 = d(i, V_0)$  and  $s_2 = d(j, V_0)$ . Triangle inequality implies that  $|s_2 - s_1| \leq w_e$ . Hence

$$W = \sum_e w_e \geq \int_{s=0}^{\Delta} |E_s|.$$

Thus, the expected value of  $|E_s|$  over the interval  $\{0, \Delta\}$  is at most  $W/\Delta$ . The algorithm thus produces a cut of size at most  $2W/\Delta = O(W\sqrt{\log n})$  with probability at least  $1/2$ .

□

### 3 $\Delta = \Omega(\log^{-2/3} n)$ -separated sets

We now describe an algorithm that given a  $c$ -spread  $\ell_2^2$  representation finds  $\Delta$ -separated sets of size  $\Omega(n)$  for  $\Delta = \Theta(1/\log^{2/3} n)$ . Our correctness proof assumes a key lemma (Lemma 7) whose proof appears in Section 4. The algorithm will be improved in Section 5 to allow  $\Delta = \Theta(1/\sqrt{\log n})$ .

The algorithm, SET-FIND, is given a  $c$ -spread  $\ell_2^2$ -representation. Constants  $c', \sigma > 0$  are chosen appropriately depending on  $c$ .

SET-FIND:

Input: A  $c$ -spread unit- $\ell_2^2$  representation  $v_1, v_2, \dots, v_n \in \mathbb{R}^d$ .

Parameters: Desired separation  $\Delta$ , desired balance  $c'$ , and projection gap,  $\sigma$ .

**Step 1:** Project the points on a random line  $u$ , and compute a value  $m$  where half the points  $v_i$  have  $\langle v_i, u \rangle > m$ .

Then, we specify that

$$S_u = \{v_i : \langle v_i, u \rangle \geq m + \frac{\sigma}{\sqrt{d}}\},$$

$$T_u = \{v_i : \langle v_i, u \rangle \leq m - \frac{\sigma}{\sqrt{d}}\}.$$

If  $|S_u| < 2c'n$  or  $|T_u| < 2c'n$ , HALT.<sup>a</sup>

**Step 2:** Pick any  $v_i \in S_u, v_j \in T_u$  such that  $|v_i - v_j|^2 \leq \Delta$ , and delete  $v_i$  from  $S_u$  and  $v_j$  from  $T_u$ . Repeat until no such  $v_i, v_j$  can be found and output the remaining sets  $S, T$ .

<sup>a</sup>We note that we could have simply chosen  $m$  to be 0 rather than to be the median, but this version of the procedure also applies for finding sparsest cuts.

REMARK 3 The procedure SET-FIND can be seen as a rounding procedure of sorts. It starts with a “fat” random hyperplane cut (cf. Goemans-Williamson [17]) to identify the sets  $S_u, T_u$  of vertices that project far apart. It then prunes these sets to find sets  $S, T$ .

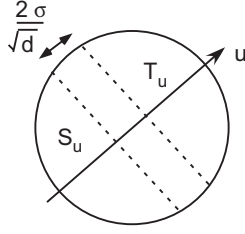


Figure 2: The sets  $S_u$  and  $T_u$  found during SET-FIND.

Notice that if SET-FIND does not HALT prematurely, it returns a  $\Delta$ -separated pair of sets. It is easy to show that the likelihood of premature halting is small, and we defer it to the next subsection. The main challenge is to show that the  $\Delta$ -separated sets are large, i.e., that no more than  $c'n$  pairs of points are deleted from  $S_u$  and  $T_u$ . This occupies the bulk of the paper and we start by giving a very high level picture.

### 3.1 Broad Outline of Proof

First observe that the Theorem is almost trivial if we desire  $O(\log n)$ -separated sets  $S$  and  $T$  (in other words,  $\Delta = \Omega(1/\log n)$ ). The reason is that the expected projection of a vector of length  $\ell$  is  $\ell/\sqrt{d}$  and the chance that it is  $k$  times larger is  $\exp(-k^2)$  (see Lemma 5). Since a pair is deleted only if the projection of  $v_i - v_j$  on  $u$  exceeds  $\Omega(1/\sqrt{d})$ , the chance that a pair satisfying  $|v_i - v_j| = O(1/\sqrt{\log n})$  or  $|v_i - v_j|^2 = O(1/\log n)$  is deleted is  $\exp(-O(\log n))$ , which is polynomially small. So with high probability *no* pair is deleted. Unfortunately, this reasoning breaks for the case of interest, namely  $\Delta = \Omega(\log^{-2/3} n)$ , when many pairs may get deleted.

So assume the algorithm fails with high probability when  $\Delta = \Omega(\log^{-2/3} n)$ . We show that for most directions  $u$  there is a sequence of  $k = \log^{1/3} n$  points  $v_1, v_2, \dots, v_k$  such that every successive pair is close, i.e.  $|v_i - v_{i+1}|^2 = \Delta$ , and their difference has a large projection on  $u$ , i.e.  $\langle v_i - v_{i+1}, u \rangle \geq 2\sigma/\sqrt{d}$ . The main point is that these projections all have the same *sign*, and thus adding them we obtain  $\langle v_1 - v_k, u \rangle \geq 2\sigma k/\sqrt{d}$ . Thus the projection scales linearly in  $k$  whereas the Euclidean distance between the first and last point in the sequence scales as  $\sqrt{k}$ , since the points come from an  $\ell_2^2$  metric. For the chosen values of  $k, \Delta$ , this means that Euclidean distance between the first and last point is  $O(1)$  whereas the projection is  $\Omega(\sqrt{\log n}/\sqrt{d})$ —large enough that the probability that such a large projection exists for *any* of the  $\binom{n}{2}$  pairs is  $o(1)$ . This is a contradiction, allowing us to conclude that the algorithm did not fail with high probability (in other words, did manage to output large  $\Delta$ -separated sets).

The idea in finding this sequence of points is to find them among deleted pairs. This uses the observation that the deleted pairs for a direction  $u$  form a matching, and that the path for a direction  $u$  can in fact use deleted pairs from a “nearby” direction  $u'$ ; something that is made formal using measure concentration. The argument uses induction, and actually shows the existence of many sequences of the type described above, not just one.

### 3.2 Covers and Matching Covers

Let us start by formalizing the condition that must hold if the algorithm fails with high probability, i.e., for most choices of directions  $\vec{u}$ ,  $\Omega(n)$  pairs must be deleted.

DEFINITION 4 *A  $(\sigma, \delta, c')$ -matching cover of a set of points is a set  $\mathcal{M}$  of matchings such that for at least a fraction  $\delta$  of directions  $u$ , there exists a matching  $M_u \in \mathcal{M}$  of at least  $c'n$  pairs of points, such that each pair  $(v_i, v_j) \in M_u$ , satisfies*

$$\langle v_i - v_j, u \rangle \geq 2\sigma/\sqrt{d}.$$

The associated matching graph  $M$  is defined as the multigraph consisting of the union of all partial matchings  $M_u$ .

In the next section, we will show that step 1 succeeds with probability  $\delta$ . This implies that if with probability at least  $\delta/2$  the algorithm did *not* produce sufficiently large  $\Delta$ -separated pair of sets, then with probability at least  $\delta/2$  many pairs were deleted. Thus the following lemma is straightforward.

LEMMA 3

If SET-FIND fails with probability greater than  $1 - \delta/2$ , on a unit  $c$ -spread  $\ell_2^2$  representation, then the set of points has a  $(\sigma, \delta/2, c')$ -matching cover.

The definition of  $(\sigma, \delta, c')$ -matching cover  $\mathcal{M}$  of pointsets suggests that for many directions there are many disjoint pairs of points that have long projections. We will work with a related notion.

DEFINITION 5 ( $(\sigma, \delta)$ -UNIFORM-MATCHING-COVER) A set of matchings  $\mathcal{M}$   $(\sigma, \delta)$ -uniform-matching-covers a set of points  $V \subseteq \mathbb{R}^d$  if for every unit vector  $u \in \mathbb{R}^d$ , there is a (partial) matching  $M_u$  of  $V$  such that every  $(v_i, v_j) \in M_u$  satisfies  $|\langle u, v_i - v_j \rangle| \geq \frac{2\sigma}{\sqrt{d}}$ , and for every  $i$ ,  $\mu(u : v_i \text{ matched in } M_u) \geq \delta$ . We refer to the set of matchings  $M_u$  to be the matching cover of  $V$ .

REMARK 4 The main difference from Definition 4 is that every point participates in  $M_u$  with constant probability for a random direction  $u$ , whereas in Definition 4 this probability could be even 0.

LEMMA 4

If a set of  $n$  vectors is  $(\sigma, \gamma, \beta)$ -matching covered by  $\mathcal{M}$ , then they contain a subset  $X$  of  $\Omega(n\delta c)$  vectors that are  $(\sigma, \delta)$ -uniformly matching covered by  $\mathcal{M}$ , where  $\delta = \Omega(\beta\gamma)$ .

PROOF: Consider the multigraph consisting of the union of all partial matchings  $M_u$ 's as described in Definition 4. The average node is in  $M_u$  for  $\gamma\beta$  measure of directions. Remove all nodes that are matched on fewer than  $\delta c/2$  measure of directions (and remove the corresponding matched edges from the  $M_u$ 's). Repeat. The aggregate measure of directions removed is at most  $\gamma\beta n/2$ . Thus at least  $\gamma\beta c n/2$  aggregate measure on directions remains. This implies that there are at least  $\gamma\beta n/4$  nodes left, each matched in at least  $\gamma\beta/4$  measure of directions. This is the desired subset  $X$ .

□

To carry out our induction we need the following weaker notion of covering a set:

DEFINITION 6 ( $(\epsilon, \delta)$ -COVER) A set  $\{w_1, w_2, \dots\}$  of vectors in  $\mathbb{R}^d$  is an  $(\epsilon, \delta)$ -cover if every  $|w_j| \leq 1$  and for at least  $\delta$  fraction of unit vectors  $u \in \mathbb{R}^d$ , there exists a  $j$  such that  $\langle u, w_j \rangle \geq \epsilon$ . We refer to  $\delta$  as the covering probability and  $\epsilon$  as the projection length of the cover.

A point  $v$  is  $(\epsilon, \delta)$ -covered by a set of points  $X$  if the set of vectors  $\{x - v | x \in X\}$  is an  $(\epsilon, \delta)$ -cover.

REMARK 5 It is important to note that matching covers are substantially stronger notion than covers. For example the  $d$ -dimensional hypercube  $\frac{1}{\sqrt{d}}\{-1, 1\}^d$  is easily checked to be an  $(\Omega(1), \Omega(1))$ -cover (equivalently, the origin is  $(\Omega(1), \Omega(1))$  covered), even though it is only  $(\Theta(1), \Theta(1/\sqrt{d}))$  uniformly matching covered. This example shows how a single outlier in a given direction can cover all other points, but this is not sufficient for a matching cover.

### 3.3 Gaussian Behavior of Projections

Before we get to the main lemma and theorem, we introduce a basic fact about projections and use it to show step 1 of SET-FIND usually succeeds.

LEMMA 5 (GAUSSIAN BEHAVIOR OF PROJECTIONS)

If  $v$  is a vector of length  $\ell$  in  $\mathbb{R}^d$  and  $u$  is a randomly chosen unit vector then

1. for  $x \leq 1$ ,  $\Pr[|\langle v, u \rangle| \leq \frac{x\ell}{\sqrt{d}}] \leq 3x$ .

2. for  $x \leq \sqrt{d}/4$ ,  $\Pr[|\langle v, u \rangle| \geq \frac{x\ell}{\sqrt{d}}] \leq e^{-x^2/4}$ .

Formally, to show that step 1 of SET-FIND succeeds with good probability, we prove that the pointsets satisfy conditions in the following definition.

**DEFINITION 7** *A set of  $n$  points  $\{v_1, \dots, v_n\} \in \mathbb{R}^d$  is  $(\sigma, \gamma, c')$ -projection separated if for at least  $\gamma$  fraction of the directions  $u$ , there exist two sets  $S_u, T_u$  each of size at least  $2c'n$ , where*

$$S_u = \{i : \langle v_i, u \rangle \geq m + \sigma/\sqrt{d}\},$$

$$T_u = \{i : \langle v_i, u \rangle \leq m - \sigma/\sqrt{d}\},$$

with  $m$  being the median value of  $\langle v_i, u \rangle$ .

Using part (i) of Lemma 5, we show that the points are sufficiently projection separated. Here, it is convenient for the points to be on the surface of the unit sphere. Later, for approximating sparsest cut we prove a similar statement without this condition.

**LEMMA 6**

*For every positive  $c < 1/3$ , there are  $c', \sigma, \delta > 0$ , such that every  $c$ -spread unit  $\ell_2^2$  representation is  $(\sigma, \delta, c')$ -projection separated.*

**PROOF:**

The  $c$ -spread condition states that  $\sum_{k,j} |v_k - v_j|^2 \geq 4c(1-c)n^2$ . Since the maximum value of  $|v_k - v_j|$  is at most 2, we can conclude that  $\sum_{k,j} |v_k - v_j| \geq 2c(1-c)n^2$ .

We claim for every  $i$  that  $\sum_j |v_i - v_j| \geq c(1-c)n$ . This follows since by the triangle inequality on  $\ell_2$  we have

$$\sum_{k,j} |v_k - v_j| \leq \sum_{k,j} |v_k - v_i| + |v_i - v_j| \leq 2n \sum_j |v_i - v_j|.$$

Furthermore, since the maximum distance is 2, we can conclude  $|v_i - v_j| \geq c(1-c)/2$  for at least  $c(1-c)n/4$  values of  $j$ . For these pairs  $i, j$ , condition (i) of Lemma 5 implies that for a random  $u$ , that  $|\langle v_i, u \rangle - \langle v_j, u \rangle|$  is at least  $c(1-c)/(6\sqrt{d})$  with probability at least  $1/3$ .

This implies that for  $\sigma = c(1-c)/3$ , the expected number of pairs where  $|\langle v_i, u \rangle - \langle v_j, u \rangle|$  is at least  $\sigma/\sqrt{d}$  is at least  $c(1-c)n^2/12$ . Since, the number of such pairs can never be larger than  $n^2$ , we can conclude that the number of separated pairs is larger than  $c(1-c)n^2/24$  for at least  $\delta = c(1-c)/24$  fraction of the directions  $u$ .

For such a direction  $u$ , we can conclude  $|\{j : |\langle u, v_j \rangle - m| \geq c(1-c)/(12\sqrt{d})\}| \geq c(1-c)n^2/24$ . Since, half the nodes are on either side of  $m$ , we have the condition required by the lemma with  $c' = c(1-c)/24$ .

□

### 3.4 Main Lemma and Theorem

We are now ready to formally state a lemma about the existence, for most directions, of many pairs of points with large projections. For each  $k$ , we inductively ensure that there is a large set of points which participate as one endpoint of such a (large projection) pair for at least half the directions. This lemma (and its variants) is a central technical contribution of this paper.

**DEFINITION 8** *Given a set of  $n$  points that is  $(\sigma, \delta, c')$ -matching covered by  $\mathcal{M}$  with associated matching graph  $M$ , we define  $v$  to be in the  $k$ -core,  $S_k$ , if  $v$  is  $(k\frac{\sigma}{2\sqrt{d}}, 1/2)$ -covered by points which are within  $k$  hops of  $v$  in the matching graph  $M$ .*

**LEMMA 7**

*For every set of  $n$  points that is  $(\sigma, \delta, c')$ -matching covered by  $\mathcal{M}$  with associated matching graph  $M$ , there are positive constants  $a = a(\delta, c')$  and  $b = b(\delta, c')$  such that for every  $k \geq 1$ :*

1. Either  $|S_k| \geq a^k n$ .
2. Or there is a pair  $(v_i, v_j)$  with distance at most  $k$  in the matching graph  $M$ , such that

$$|v_i - v_j| \geq b\sigma/\sqrt{k}.$$

We use this lemma below to show that SET-FIND finds  $\Omega(1/\log^{1/3} n)$ -separated sets. Later, we develop a modified algorithm for finding  $\Omega(1/\sqrt{\log n})$ -separated sets.

Subsequently, the aforementioned paper by James Lee [] strengthened case 1 of the lemma to show the existence of  $S_k$  of size at least  $|M|/2$ , and thus that case 2 meets the stronger condition that  $|v_i - v_j| \geq g\sigma$ . This implies that SET-FIND without modification finds large  $\Omega(1/\sqrt{\log n})$ -separated sets.

We can now prove the following theorem.

**THEOREM 8**

*Set-find finds an  $\Delta$ -separated set for a  $c$ -spread unit  $\ell_2^2$  representation with constant probability for some  $\Delta = \Omega(1/\log^{2/3} n)$ .*

**PROOF:**

Recall, that, by Lemma 3, if SET-FIND fails, then the set of points is  $(\sigma, \delta, c')$ -matching covered. Moreover, each pair  $(v_i, v_j)$  in any matching  $M_u$  for a direction  $u$  satisfies  $|v_i - v_j| \leq \sqrt{\Delta}$  and  $\langle v_i - v_j, u \rangle \geq 2\sigma/\sqrt{d}$ .

We will now apply Lemma 7 to show that such a matching covered set does not exist for  $\Delta = \Theta(1/\log^{2/3} n)$ , which means that SET-FIND cannot fail for this  $\Delta$ . Concretely, we show that neither case of the Lemma holds for  $k = g\sigma/\sqrt{\Delta}$  where  $g$  is the constant defined in the Lemma.

Let us start by dispensing with Case 2. Since the metric is  $\ell_2^2$ , a pair of points  $(v_i, v_j)$  within distance  $k$  in the matching graph satisfies  $|v_i - v_j|^2 \leq k\Delta$ , and thus  $|v_i - v_j| \leq \sqrt{k\Delta}$ . For our choice of  $k$  this is less than  $g\sigma/\sqrt{k}$ , so Case 2 does not happen.

Case 1 requires that for at least  $\frac{1}{2}$  of the directions  $u$  there is a pair of vectors  $(v_i, v_j)$  such that  $|v_i - v_j| \leq \sqrt{k\Delta}$  and where  $\langle v_i - v_j, u \rangle \geq k\sigma/2\sqrt{d}$ .

For any fixed  $i, j$ , the probability that a randomly chosen direction satisfies this condition is at most  $\exp(-k/\sigma\Delta)$ , which can be made smaller than  $1/2n^2$  by choosing some  $\Delta = \Omega(1/\log^{2/3} n)$  since  $k = \Theta(\log^{1/3} n)$ . Therefore, by the union bound, the fraction of directions for which any pair  $i, j$  satisfies this condition is less than  $n^2/2n^2 \leq 1/2$ . Thus Case 1 also cannot happen for this value of  $\Delta$ .

This proves the theorem.

□

## 4 Proof of Lemma 7

### 4.1 Measure concentration

We now introduce an important geometric property of covers. If a point is covered in a non-negligible fraction of directions, then it is also covered (with slightly smaller projection length) in almost all directions. This property follows from measure concentration which can be stated as follows.

Let  $S^{d-1}$  denote the surface of the unit ball in  $\mathbb{R}^d$  and let  $\mu(\cdot)$  denote the standard measure on it. For any set of points  $A$ , we denote by  $A_\gamma$  the  $\gamma$ -neighborhood of  $A$ , namely, the set of all points that have distance at most  $\gamma$  to some point in  $A$ .

**LEMMA 9 (CONCENTRATION OF MEASURE)**

*If  $A \subseteq S^{d-1}$  is measurable and  $\gamma > \frac{2\sqrt{\log(1/\mu(A))+t}}{\sqrt{d}}$ , where  $t > 0$ , then  $\mu(A_\gamma) \geq 1 - \exp(-t^2/2)$ .*

PROOF: P. Levy's isoperimetric inequality (see [5]) states that  $\mu(A_\gamma)/\mu(A)$  is minimized for spherical caps<sup>4</sup>. The lemma now follows by a simple calculation using the standard formula for (d-1)-dimensional volume of spherical caps, which says that the cap of points whose distance is at least  $s/\sqrt{d}$  from an equatorial plane is  $\exp(-s^2/2)$ .  $\square$

The following lemma applies measure concentration to boost covering probability.

LEMMA 10 (BOOSTING LEMMA)

Let  $\{v_1, v_2, \dots\}$  be a finite set of vectors that is an  $(\epsilon, \delta)$ -cover, and  $|v_i| \leq \ell$ . Then, for any  $\gamma > \frac{\sqrt{2 \log(2/\delta)} + t}{\sqrt{d}}$ , the vectors are also a  $(\epsilon - 2\ell\gamma, \delta')$ -cover, where  $\delta' = 1 - \exp(-t^2/2)$ .

PROOF: Let  $A$  denote the set of directions  $u$  for which there is an  $i$  such that  $\langle u, v_i \rangle \geq \epsilon$ . Since  $|v_i - u|^2 = 1 + |v_i|^2 - 2\langle u, v_i \rangle$  we also have:

$$A = S^{d-1} \cap \bigcup_i \text{Ball} \left( v_i, \sqrt{1 + |v_i|^2 - 2\epsilon} \right),$$

which also shows that  $A$  is measurable. Also,  $\mu(A) \geq \delta$  by hypothesis. Thus by Lemma 9,  $\mu(A_\gamma) \geq 1 - \exp(-t^2/2)$ .

We argue that for each direction  $u$  in  $A_\gamma$ , there is a vector  $v_i$  in the  $(\epsilon, \delta)$  cover with  $\langle v_i, u \rangle \geq \epsilon - 2\ell\gamma$  as follows. Let  $u \in A, u' \in A_\gamma \cap S^{d-1}$  be such that  $|u - u'| \leq \gamma$ .

The projection length of  $v_i$  on  $u$  is  $|v_i| \cos \theta$  where  $\theta$  is the angle between  $v_i$  and  $u$ . The projection length of  $v_i$  on  $u'$  is  $|v_i| \cos \theta'$  where  $\theta'$  is the angle between  $v_i$  and  $u'$ . The angle formed by  $u$  and  $u'$  is at most  $2\gamma$  since  $\alpha \leq 2 \sin \alpha$ , so  $\theta - 2\gamma \leq \theta' \leq \theta + 2\gamma$ . Since the absolute value of the slope of the cosine function is at most 1, we can conclude that the differences in projection is at most  $2\gamma|v_i| \leq 2\gamma\ell$ . That is,  $\langle v_i, u' \rangle \geq \epsilon - 2\gamma\ell$ .

Combined with the lower bound on the  $\mu(A_\gamma)$ , we conclude that the set of directions  $u'$  such that there is an  $i$  such that  $\langle u', v_i \rangle \geq \epsilon - 2\ell\gamma$  has measure at least  $1 - \exp(-t^2/2)$ .  $\square$

## 4.2 Cover Composition

In this section, we give the construction that lies at the heart of the inductive step in the proof of Lemma 7.

Let  $X \subseteq \mathbb{R}^d$  be a pointset that is uniform  $(\epsilon, \delta)$ -matching-covered, where  $M_u$  denotes as usual the matching associated with direction  $u$ . Suppose  $Z \subseteq X$  consists of points in  $X$  that are  $(\epsilon_1, 1 - \delta/2)$ -covered, and for  $v_i \in Z$  let  $Z_i$  denote points that cover  $v_i$ .

The following lemma shows how to combine the cover and matching cover to produce a cover for a set  $Z'$  with projection length the sum of the projection lengths of the original covers.

LEMMA 11 (COVER COMPOSITION)

If  $|Z| \geq \tau |X|$ , then there is a set  $Z'$  consisting of at least  $\delta\tau/4$  fraction of points in  $X$  that are  $(\epsilon_1 + \epsilon, \delta\tau/4)$ -covered. Furthermore, the cover for each such point  $v_k$  can even be found with only vectors  $v_j - v_k$  that can be expressed as

$$v_j - v_k = v_j - v_i + v_i - v_k,$$

where (i)  $v_j - v_i \in Z_i$  (ii)  $\{v_i, v_j\}$  is an edge in  $M_u$  for some direction  $u$ .

REMARK 6 This is called ‘‘cover composition’’ because each vector  $v_j - v_k$  in the new cover can be written as a sum of two vectors, one of which is in the matching cover and the other in the cover of  $v_i$ .

PROOF:

Let  $v_i \in Z$ . For  $1 - \delta/2$  fraction of directions  $u$ , there is a point  $v_j \in T_i$  such that  $\langle v_j - v_i, u \rangle \geq \epsilon_1$ . Also for  $\delta$  fraction of directions  $u$ , there is a point in  $v_k \in X_i$  such that  $\langle v_k - v_i, u \rangle \leq -\epsilon$  and  $v_k$  is matched to  $v_i$

<sup>4</sup>Levy's isoperimetric inequality is not trivial; see [29] for a sketch. However, results qualitatively the same—but with worse constants—as Lemma 9 can be derived from the more elementary Brunn-Minkowski inequality; this ‘‘approximate isoperimetric inequality’’ of Ball, de Arias and Villa also appears in [29].

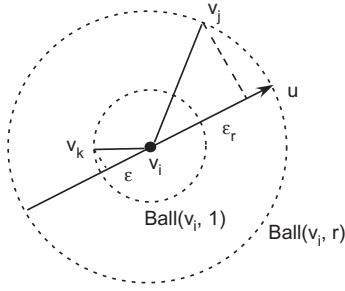


Figure 3: Direction  $u$  is chosen randomly. The vector  $v_j - v_k$  has projection  $\epsilon_1 + \epsilon$  on  $u$  and it gets “assigned” to  $k$ .

in the matching  $M_u$ . Thus for a  $\delta/2$  fraction of directions  $u$ , both events happen and thus the pair  $(v_j, v_k)$  satisfies  $\langle v_j - v_k, u \rangle \geq \epsilon_r + \epsilon$ . Since  $v_j \in Z_i$  and  $v_k \in X_i$ , we “assign” this vector  $v_j - v_k$  to point  $v_k$  for direction  $u$ , as a step towards building a cover centered at  $v_k$ . Now we argue that for many  $v_k$ ’s, the vectors assigned to it in this way form a  $(\epsilon_1 + \epsilon, \delta|Z|/2|X|)$ -cover.

For each point  $v_i \in Z$ , for  $\delta/2$  fraction of the directions  $u$  the process above assigns a vector to a point in  $X$  for direction  $u$  according to the matching  $M_u$ . Thus on average for each direction  $u$ , at least  $\delta|Z|/2$  vectors get assigned by the process. Thus, for a random point in  $X$ , the expected measure of directions for which the point is assigned a vector is at least  $\delta|Z|/2|X|$ .

Furthermore, at most one vector is assigned to any point for a given direction  $u$  (since the assignment is governed by the matching  $M_u$ ). Therefore at least  $\delta|Z|/4|X|$  fraction of the points in  $X$  must be assigned a vector for  $\delta|Z|/4|X|$  fraction of the directions.

We will define all such points of  $X$  to be the set  $Z'$  and note that the size is at least  $\delta|Z|/4$  as required.  $\square$

For the inductive proof of Lemma 7, it is useful to boost the covering probability of the cover obtained in Lemma 11 back up to  $1 - \delta/2$ . This is achieved by invoking the boosting lemma as follows.

**COROLLARY 12**

*If the vectors in the covers for  $Z'$  have length at most*

$$\frac{\epsilon\sqrt{d}}{8\sqrt{\log \frac{8}{\tau\delta}} + 4\sqrt{\log \frac{2}{\delta}}}$$

*then the  $(\epsilon_1 + \epsilon, \tau\delta/4)$ -covers for  $Z'$  are also  $(\epsilon_1 + \epsilon/2, 1 - \delta/2)$ -covers.*

**PROOF:**

This follows from Lemma 10, since the loss in projection in this corollary is  $\epsilon/2$  which should be less than  $2\gamma\ell$  where  $\ell$  is the length of the vectors in the cover and  $\gamma$  is as defined in Lemma 10. Solving for  $\ell$  yields the upper bound in the corollary.  $\square$

Though the upper bound on the length in the hypothesis appears complicated, we only use the lemma with  $\epsilon = \Theta(1/\sqrt{d})$  and constant  $\delta$ . Thus, the upper bound on the length is  $\theta(1/\sqrt{\log(1/\tau)})$ . Furthermore,  $1/\tau$  will often be  $\theta(1)$ .

**4.3 Proof of Lemma 7**

The proof of the lemma is carried out using Lemmas 10, and Corollary 12 in a simple induction. We assume that the pointset is  $(\gamma, \beta, \sigma)$ -matching covered. Thus, by Lemma 4 there exists an  $(\sigma, \delta)$ -uniform matching cover  $\mathcal{M}$  of a subset  $X$  of the pointset.

Before proving the lemma, let us restate a slight strengthening to make our induction go through.

**DEFINITION 9** *Given a set of  $n$  points that is  $(\sigma, \delta, c')$ -matching covered by  $\mathcal{M}$  with associated matching graph  $M$ , we define  $v$  to be in the  $(k, \rho)$ -core,  $S_{k, \rho}$ , if  $v$  is  $(k \frac{\sigma}{2\sqrt{d}}, 1 - \rho)$ -covered by points which are within  $k$  hops of  $v$  in the matching graph  $M$ .*

**CLAIM 13**

For constant  $\delta$ , and every  $k \geq 1$ ,

1.  $|S_{k, \delta/2}| \geq (\frac{\delta}{4})^k |X|$
2. or there is a pair  $(v_i, v_j)$  such that  $|v_i - v_j| = \Omega(\sigma/\sqrt{k})$  and  $v_i$  and  $v_j$  are within  $k$  hops in the matching graph  $M$ .

**PROOF:**

For  $k = 1$ , the claim follows by setting  $S_{1, \delta/2} = X$ . Now we proceed by induction on  $k$ .

Clearly if case 2 holds for  $k$ , then it also holds for  $k + 1$ . So assume that case 1 holds for  $k$ , i.e.,  $S_{k, \delta/2}$  satisfies the conditions of case 1. Composing the covers of  $S_{k, \delta/2}$  with the matching cover  $\mathcal{M}$  using Lemma 11 yields a cover for a set  $T$  of size at least  $\frac{\delta}{4}|S_k|$  with covering probability  $\Theta(|S_k|/|X|)$ , but with projection length that is larger by  $\epsilon = \Omega(1/\sqrt{d})$ . To finish the induction step the covering probability must be boosted to  $1 - \delta/2$  which by Lemma 10 decreases the projection length by

$$\ell \sqrt{\log \Theta(|X|/|S_k|) / \sqrt{d}} = O(\ell \sqrt{k} / \sqrt{d}) \quad (11)$$

where  $\ell$  upper-bounds the length of the vectors in the cover. If this decrease is less than  $\epsilon/2$ , then the points in  $T$  are  $(k\epsilon/2 + \epsilon/2, 1 - \delta/2)$ -covered and  $S_{k+1, \delta/2}$  is large as required in case 1. Otherwise we have  $\epsilon/2 = \Theta(\sigma/\sqrt{d}) = O(\ell \sqrt{k} / \sqrt{d})$ , which simplifies to  $\ell \geq g\sigma/\sqrt{k}$ , for some constant  $g$  and thus case 2 holds for  $k + 1$ .

□

**REMARK 7** An essential aspect of this proof is that the  $k$ -length path (from  $M^k$ ) pieces together pairs from many different matchings  $M_u$ . Consider, for example, given a direction  $u$  and a matching  $M_u$  with projection  $\epsilon$  on  $u$ , how we produce any pair of points with larger projection on  $u$ . Clearly, edges from the matching  $M_u$  do not help with this. Instead, given any pair  $(x, y) \in M_u$ , we extend it with a pair  $(y, z) \in M_{u'}$  for some different direction  $u'$  where  $(y, z)$  also has a large projection on  $u$ . The existence of such a pair  $(y, z)$  follows from the boosting lemma.

## 5 Achieving $\Delta = \Omega(1/\sqrt{\log n})$

Theorem 1 requires  $\Delta = \Omega(1/\sqrt{\log n})$  whereas we proved above that SET-FIND finds large  $\Delta$ -separated sets for  $\Delta = \Omega(\log^{-2/3} n)$ . Now we would like to run SET-FIND for  $\Delta = \Omega(1/\sqrt{\log n})$ . The main problem in proving that SET-FIND succeeds for larger  $\Delta$  is that our version of Lemma 7 is too weak. Before seeing how to remedy this it is useful to review how the proof of Theorem 1 constrains the different parameters and results in a  $\Theta(1/\log^{2/3} n)$  upperbound on  $\Delta$ .

1. In the proof of Lemma 7 we can continue to inductively assert case 1 (producing covered set  $S_k$ ) of Claim 13 as long as the right hand side of Equation (11) is less than  $\sigma/2\sqrt{d}$ , i.e., as long as  $\ell^2 \ll 1/k$ , where  $\ell$  bounds the lengths of the vectors.
2. By the triangle inequality,  $\ell = O(\sqrt{k\Delta})$ .
3. Deriving the final contradiction in the proof of Theorem 8 requires that  $\exp(-k/\Delta) \ll 1/n^2$ .

The limit on the vector lengths in item 1, combined with item 2, requires that  $k \leq 1/\Delta^{1/3}$ , and thus item 3 constrains  $\Delta$  to be  $O(1/\log^{2/3} n)$ .

The improvement derives from addressing the limit on vector lengths in item 1, which arises from the need to boost the covering probability for  $S_k$  (whose size decreases exponentially in  $k$ ) from  $|S_k|/|X|$  to  $1 - \delta/2$  as detailed in Equation (11). Thus, we would like to prevent the size of  $S_k$  from decreasing below  $\Omega(|X|)$ . This will allow the induction to continue until the vectors have length  $\Omega(1)$ . To do so, we will use the fact that any point close to a point which is  $(\epsilon, \delta)$ -covered is also  $(\epsilon', \delta')$ -covered where  $\epsilon' \approx \epsilon$  and  $\delta' \approx \delta$ . This is quantified in the Lemma 14.

LEMMA 14 (COVERING CLOSE POINTS)

Suppose  $v_1, v_2, \dots \in \mathbb{R}^d$  form an  $(\epsilon, \delta)$ -cover for  $v_0$ . Then, they also form an  $(\epsilon - \frac{ts}{\sqrt{d}}, \delta - e^{-t^2/4})$ -cover for every  $v'_0$  such that  $|v_0 - v'_0| \leq s$ .

PROOF: If  $u$  is a random unit vector,  $\Pr_u[\langle u, v_0 - v'_0 \rangle \geq \frac{ts}{\sqrt{d}}] \leq e^{-t^2/4}$ .  $\square$

REMARK 8 Lemma 14 shows that  $(\epsilon, \delta)$  covers of a point are remarkably robust. In our context where we can afford to lose  $\Theta(1/\sqrt{d})$  in projection length, even a  $\Theta(1)$ -neighborhood of that point is well covered.

DEFINITION 10 ( $\zeta$ -PROXIMATE GRAPH) A graph  $G$  on a pointset  $v_1, \dots, v_n \in \mathbb{R}^d$  is called  $\zeta$ -proximate if for each edge  $(v_i, v_j)$  in  $G$ , satisfies  $|v_i - v_j| \leq \zeta$ . A set of vertices  $S$  is non-magnifying if  $|S \cup \Gamma(S)| < |V|/2$ .

Notice that for any set  $S$  in the  $\zeta$ -proximate graph if  $T = V - S - \Gamma(S)$  then  $S, T$  is a  $\zeta$ -separated pair of sets. We will use  $\zeta = \Delta$ , thus if we ever find a non-magnifying set  $S$  of size  $\Omega(n)$  then we are done.

We prove a stronger version of Lemma 7 where case 2 only occurs when vectors have  $\Omega(1)$  length by ensuring in case 1 that the  $(k, \delta/2)$ -core,  $S_{k, \delta/2}$ , has size at least  $|X|/2$ .

The induction step in Lemma 11 now yields a set, which we denote by  $S'_{k+1}$ , of cardinality  $\delta|X|$ , which is covered by  $X$  with constant covering probability and with projection length increased by  $\Omega(1/\sqrt{d})$ . If  $S'_{k+1}$  is a non-magnifying set in the  $\zeta$ -proximate graph, we can halt (we have found a  $\zeta$ -separated pair of sets). Otherwise,  $S'_{k+1}$  can be augmented with  $\Gamma(S'_{k+1})$  to obtain a set  $T$  of size  $|X|/2$  which by Lemma 14 is covered with projection length that is smaller by  $O(\zeta/\sqrt{d})$  and is contained in  $S_{k+1, \delta/2}$ .

Then, we boost the covering probability at a cost of reducing the projection length by  $O(\ell/\sqrt{d})$  where  $\ell$  is the length of the covering vectors. Thus when  $\zeta \ll 1$ , we either increase the projection length by  $\Omega(1/\sqrt{d})$  or  $\ell$  is too large or we produce a non-magnifying set. We have proved the following enhanced version of Lemma 7:

LEMMA 15

For every set of points  $X$ , that is  $(\delta, \sigma)$ -uniform matching covered by  $\mathcal{M}$  with associated matching graph  $M$ , and a  $\zeta$ -proximate graph on the points for  $\zeta \leq \zeta_0(\delta, c', \sigma)$ , at least one of the following is true for every  $k \geq 1$ , where  $g = g(\delta, c')$  are positive constants.

1. The  $(k, \delta/2)$ -core,  $S_{k, \delta/2}$ , has cardinality at least  $|X|/2$
2. There is a pair  $(v_i, v_j)$  with distance at most  $k$  in the matching graph  $M$ , such that

$$|v_i - v_j| = g.$$

3. The set,  $S'_{k'}$ , is non-magnifying in the  $\zeta$ -proximate graph for some  $k' < k$ .

We can now prove Theorem 1. We use set-find with some parameter  $\Delta = \Theta(1/\sqrt{\log n})$ . If it fails to find a  $\Delta$ -separated set the pointset must have been  $(\delta, c', \sigma)$ -matching covered where  $\delta, c'$ , and  $\sigma$  are  $\Theta(1)$ . We first apply Lemma 4 to obtain a  $(\delta', \sigma)$ -uniform matching covered set  $X$  of size  $\Omega(n)$ . We then apply Lemma 15 where the edges in the  $\zeta$ -proximate graph consist of all pairs in  $X$  whose  $\ell_2^2$  distance is less than

some  $\zeta = \Theta(1/\sqrt{\log n})$ . If case 3 ever occurs, we can produce a pair of  $\zeta$ -separated of size  $\Omega(n)$ . Otherwise, we can continue the induction until  $k = \Omega(\sqrt{\log n})$  at which point we have a cover with projection length  $\sqrt{\log n}/\sqrt{d}$  consisting of vectors of  $O(1)$  length. As in our earlier proof, this contradicts the assumption that set-find failed.

Thus, either SET-FIND gives a  $\Delta$ -separated pair of large sets, or for some  $k$ , the  $\zeta$ -neighborhood of  $S_k$ ,  $\Gamma(S_k)$ , is small and thus,  $S_k, V - \Gamma(S_k)$  forms a  $\zeta$ -separated pair of large sets. Clearly, we can identify such an  $S_k$  by a randomized algorithm that uses random sampling of directions to check whether a vertex is well covered.

This proves Theorem 1.

REMARK 9 Lee's direct proof that SET-FIND[Lee05] produces  $\Omega(1/\sqrt{\log n})$ -separated sets relies on the following clever idea: rather than terminating the induction with a non-magnifying set (in the  $\zeta$ -proximal graph), he shows how to use the non-magnifying condition to bound the loss in covering probability. The main observation is that after cover composition, the covering probability is actually  $|S_k|/\Gamma(S_k)$  rather than our pessimistic bound of  $|S_k|/|X|$ . Thus by insisting on the invariant that  $|\Gamma(S_k)| = O(|S_k|)$ , we can ensure that the covering probability falls by only a constant factor, thus incurring only a small boosting cost and allowing the induction to continue for  $k = \Omega(1/\sqrt{\Delta})$  steps. Maintaining the invariant is easy since  $S_k$  can be replaced by  $\Gamma(S_k)$  with small cost in projection length using Lemma 14. The point being that over the course of the induction this region growing step only needs to be invoked  $O(\log n)$  times.

## 6 $O(\sqrt{\log n})$ Ratio for SPARSEST CUT

Now we describe a rounding technique for the SDP in (8)–(10) that gives a  $O(\sqrt{\log n})$ -approximation to SPARSEST CUT. Note that our results on expander flows in Section 7 give an alternative  $O(\sqrt{\log n})$ -approximation algorithm.

First we see in what sense the SDP in (8)–(10) is a relaxation for SPARSEST CUT. For any cut  $(S, \bar{S})$  consider a vector representation that places all nodes in  $S$  at one point of the sphere of squared radius  $(4|S||\bar{S}|)^{-1}$  and all nodes in  $\bar{S}$  at the diametrically opposite point. It is easy to verify that this solution is feasible and has value  $|E(S, \bar{S})|/|S||\bar{S}|$ . Since  $|\bar{S}| \in [n/2, n]$ , we can treat it as a scaling factor. We conclude that the *optimal* value of the SDP multiplied by  $n/2$  is a lower bound for SPARSEST CUT.

The next theorem implies that the integrality gap is  $O(\sqrt{\log n})$ .

### THEOREM 16

*There is a polynomial-time algorithm that, given a feasible SDP solution with value  $\beta$ , produces a cut  $(S, \bar{S})$  satisfying  $|E(S, \bar{S})| = O(\beta|S|n\sqrt{\log n})$ .*

The proof divides into two cases, one of which is similar to that of Theorem 1. The other case is dealt with the following Lemma.

### LEMMA 17

*For every constants  $c, \tau$  where  $c < 1, \tau < 1/8$  there is a polynomial-time algorithm for the following task. Given any feasible SDP solution with  $\beta = \sum_{\{i,j\} \in E} |v_i - v_j|^2$ , and a node  $k$  such that the geometric ball of squared-radius  $\tau/n^2$  around  $v_k$  contains at least  $cn$  vectors, the algorithm finds a cut  $(S, \bar{S})$  with expansion at most  $O(\beta n/c)$ .*

PROOF: Let  $d(i, j)$  denote  $|v_i - v_j|^2$ , and when  $\{i, j\}$  is an edge  $e$  we write  $w_e$  for  $d(i, j)$ . The weights  $w_e$  turn the graph into a weighted graph.

Let  $X$  be the subset of nodes that correspond to the vectors in the geometric ball of radius  $\tau/n^2$  around  $v_k$ . The algorithm consists of doing a breadth-first search on the weighted graph starting from  $X$ . For  $s \geq 0$  let  $V_s$  be the set of nodes whose distance from  $X$  is at most  $s$ , and let  $E_s$  be the set of edges leaving  $V_s$ . We identify  $s$  for which the cut  $(V_s, \bar{V}_s)$  has the lowest expansion, say  $\alpha_{obs}$ , and output this cut.

The expansion of the cut  $(V_s, \overline{V}_s)$  is

$$\frac{|E_s|}{\min\{|V_s|, |\overline{V}_s|\}}.$$

Since  $|V_s| \geq c \cdot |\overline{V}_s|$ , the expansion is at most

$$\frac{|E_s|}{c \cdot |\overline{V}_s|},$$

allowing us to conclude  $|E_s| \geq c\alpha_{obs} |\overline{V}_s|$  for all  $s$ .

To finish, we need to show that  $\alpha_{obs} = O(\beta n/c)$ .

Since  $\sum_{i < j} d(i, j) = 1$ , the triangle inequality implies that each node  $m$  also satisfies:

$$\sum_{i, j} d(i, m) + d(j, m) \geq 1,$$

which implies

$$\sum_j d(j, m) \geq 1/2n. \quad (12)$$

Separating terms corresponding to  $j \in X$  and  $j \notin X$  in (12) and using  $c\tau < 1/8$  we obtain for the special node  $k$ :

$$\sum_{j \notin X} d(j, k) \geq \frac{1}{2n} - \frac{\tau}{n^2} \cdot cn = \frac{3}{8n}. \quad (13)$$

The Lemma's hypothesis also says

$$\sum_{e \in E} w_e = \beta. \quad (14)$$

As we noticed in the proof of Corollary 2, each edge  $e = \{i, j\}$  only contributes to  $E_s$  for  $s$  in the open interval  $(s_1, s_2)$ , where  $s_1 = d(i, X)$  and  $s_2 = d(j, X)$ . Triangle inequality implies that  $|s_2 - s_1| \leq w_e$ .

Thus

$$\beta = \sum_{e \in E} w_e \geq \int_{s > 0} |E_s| ds \geq \int_{s > 0} c\alpha_{obs} |\overline{V}_s| ds.$$

Furthermore, we note that

$$\int_{s > 0} |\overline{V}_s| ds = \sum_{i \notin X} d(i, X) \geq \sum_{i \notin X} (d(i, k) - \frac{\tau}{n^2}).$$

Thus, from equation (13) we have that

$$\int_{s > 0} |\overline{V}_s| ds \geq \frac{3}{8n} - n \cdot \frac{\tau}{n^2} > \frac{1}{4n}.$$

Combining the above inequalities, we obtain  $\beta \geq c\alpha_{obs}/4n$ , or, in other words  $\alpha_{obs} = O(\beta n/c)$ .

□

Now, in case the conditions of Lemma 17 does not hold, we run SET-FIND.

The following lemma ensures that for any pointset that is  $c$ -spread and does not meet the conditions of Lemma 17 that the procedure proceeds to Step 2 of SET-FIND with a reasonable probability for a value of  $\sigma$  that is  $\Omega(1)$ .

LEMMA 18

Given a  $c$ -spread set of points where no more than  $n/10$  of them are contained in a ball of diameter  $1/10$ , there is a  $\sigma = O(c)$  such that Step 1 of SET-FIND does not HALT with probability at least  $\Omega(1)$ .

PROOF:

We begin with the following claim.

CLAIM 19

For at least  $1/10$  of the directions, at least  $1/10$  of the pairs are separated by at least  $1/20\sqrt{d}$  in the projection.

PROOF:

Consider a vertex  $v_i$ . The number of pairs  $v_i, v_j$  where  $|v_i - v_j| \geq 1/10$  is at least  $9(n-1)/10$ . Over all pairs, the number is  $9n(n-1)/10$ . For unordered pairs, this is  $9n(n-1)/20$ .

For each such pair, the probability that  $v_i$  and  $v_j$  fall at least  $1/20\sqrt{d}$  apart is at least  $1/3$  from Lemma 5. Thus, the expected number of pairs that are separated by an interval of length  $1/20\sqrt{d}$  is at least  $9n(n-1)/60$ . Using the fact that the expected number can never be larger than  $n(n-1)/2$ , we see that with probability at least  $9/120$  that at least  $9n(n-1)/120$  pairs are separated by at least  $1/20\sqrt{d}$ .  $\square$

Notice that SET-FIND with  $\sigma = 1/40$  only HALTS in step 1 if more than 90% of the nodes are within  $1/40\sqrt{d}$  of the median point. This does not occur with probability more than  $1/10$ . Thus, SET-FIND-SPARSE succeeds with  $\Omega(1)$  probability.

$\square$

To finish, we observe that if the procedure fails in step 2, then the point set must be  $(\sigma, \Theta(1), \Theta(1))$  matching covered. We can then use Lemma 15 to produce a cut or to derive a contradiction as we did before.

## 7 Expander Flows: Approximate certificates of expansion

Deciding whether a given graph  $G$  has expansion at least  $\alpha$  is coNP-complete [6] and thus has no short certificate unless the polynomial hierarchy collapses. On the other hand, the value of the SDP used in the previous sections gives an “approximate” certificate of expansion. Jerrum and Sinclair [21] and then Leighton and Rao [22] previously showed how to use multicommodity flows to give “approximate” certificates; this technique was then clarified by Sinclair [32] and Diaconis and Saloff-Coste [11]. Their certificate was essentially an embedding of a complete graph into the underlying graph with minimal expansion. This certificate could certify expansion to within an  $\Theta(\log n)$  factor.

The results of this section represent a continuation of that work but with a better certificate: for any graph with  $\alpha(G) = \alpha$  we can exhibit a certificate to the effect that the expansion is at least  $\Omega(\alpha/\sqrt{\log n})$ . Furthermore, this certificate can be computed in polynomial time. The certificate involves using multicommodity flows to embed (weighted) expander graphs. The previous approaches were the same except that the expander was limited to be the complete graph. In our approach, we can choose the expander which is the easiest to embed in the underlying graph.

We remark that this view lead to recent faster algorithms of Arora, Kale and Hazad [AHK04] for approximating sparsest cuts. We note that this view essentially follows from SDP duality, in the sense that the certificate we use is a feasible solution to the dual of the semi-definite program that we used in previous sections.

In this discussion it will be more convenient to look at weighted graphs. For a weighted graph  $G = (V, W)$  in which  $c_{ij}$  denotes the weight on edge  $\{i, j\}$  the sparsest cut is defined as

$$\Phi(W) = \min_{S \subseteq V: |S| \leq n/2} \frac{\sum_{i \in S, j \in \bar{S}} c_{i,j}}{|S| |\bar{S}|}.$$

We similarly define  $\alpha(W)$ .

A word on convention. Weighted graphs in this section will be symmetric, i.e.,  $c_{ij} = c_{ji}$  for all node pairs  $i, j$ . We call  $\sum_j c_{ij}$  the *degree* of node  $i$ . We say that a weighted graph is  $d$ -*regular* if all degrees are exactly  $d$ . We emphasize that  $d$  can be a fraction.

## 7.1 Multicommodity flows as graph embeddings

A *multicommodity flow* in an unweighted graph  $G = (V, E)$  is an assignment of a *demand*  $d_{ij} \geq 0$  to each node pair  $i, j$  such that we can route  $d_{ij}$  units of flow from  $i$  to  $j$ , and can do this simultaneously for all pairs while satisfying capacity constraints. More formally, for each  $i, j$  and each path  $p \in \mathcal{P}_{ij}$  there exists  $f_p \geq 0$  such that

$$\forall i, j \in V \quad \sum_{p \in \mathcal{P}_{ij}} f_p = d_{ij} \quad (15)$$

$$\forall e \in E \quad \sum_{p \ni e} f_p \leq 1. \quad (16)$$

Note that every multicommodity flow in  $G$  can be viewed as an *embedding* of a weighted graph  $G' = (V, E', d_{ij})$  on the same vertex set such that the weight of edge  $\{i, j\}$  is  $d_{ij}$ . We assume the multicommodity flow is symmetric, i.e.,  $d_{ij} = d_{ji}$ . The following inequality is immediate from definitions, since flows do not exceed edge capacities.

$$\alpha(G) \geq \alpha(G') \quad (17)$$

The following is one way to look at the Leighton-Rao result where  $K_n$  is the complete graph on  $n$  nodes. The embedding mentioned in the theorem is, by (17), a certificate showing that expansion is  $\Omega(\alpha/\log n)$ .

**THEOREM 20 (LEIGHTON-RAO [22])**

*If  $G$  is any  $n$ -node graph with  $\alpha(G) = \alpha$ , then it is possible to embed a multicommodity flow in it with each  $f_{ij} = \alpha/n \log n$  (in other words, a scaled version of  $K_n$ ).*

**REMARK 10** The same theorem is usually stated using all  $f_{ij} = 1$  (i.e., an unweighted copy of  $K_n$ ) and then bumping up the capacity of each edge of  $G$  to  $O(n \log n/\alpha)$ . A similar restatement is possible for our theorem about expander flows (Theorem 24).

We note that the embedding of Theorem 20 can be found in polynomial time using a multicommodity flow computation, and that this embedding is a “certificate” via (17) that  $\alpha(G) = \Omega(\alpha/\log n)$ .

## 7.2 Expanders and Eigenvalues

*Expanders* can be defined in more than one way. We use the following definition.

**DEFINITION 11 (EXPANDERS)** *For any  $c > 0$ , a  $d$ -regular weighted graph  $(w_{ij})$  is a  $\beta$ -expander if for every set of nodes  $S$ ,  $w(S, \bar{S}) = \sum_{i \in S, j \in \bar{S}} w_{ij}$  is at least  $\beta d |S|$ . That is, its conductance is at least  $\beta$ .*

(We remark that this is sometimes called a *conductor* since the number of edges across the cut is proportional to the number of edges that stays on one side.)

If a weighted graph is a  $\beta$ -expander, then the second eigenvalue of its Laplacian lies in the interval  $[\beta^2/2, 2\beta]$  (see, for example, [Alo86, 32].) In fact, we have the following algorithmic version from Alon[Alo86] as well as Alon-Milman[3].

**LEMMA 21**

*There is a polynomial time algorithm that given a weighted graph with conductance  $\gamma$ , finds a cut of conductance  $\sqrt{\gamma}/2$ .*

The next lemma extends the algorithm to output a *balanced* cut of small conductance.

**LEMMA 22**

*For any constant  $c > 0$  there is a polynomial time algorithm that, given any weighted graph and a number  $\gamma > 0$  behaves as follows. If the graph has a  $c$ -balanced cut of conductance less than  $\gamma$  then the algorithm outputs a  $c/2$ -balanced cut of conductance  $2\sqrt{\gamma}$ . If the graph does not have such a cut, the algorithm finds a set of at least  $(1 - c/2)n$  nodes such that the induced subgraph (with deleted edges replaced by self-loops) on them has conductance at least  $2\gamma$ .*

PROOF: First, we find a non-expanding cut in the weighted graph using the algorithm of Lemma 21 with this value of  $\gamma$ . It either outputs a cut with conductance at most  $\sqrt{\gamma}$  or else fails, in which case we have verified that the graph has conductance at least  $2\gamma$ . In the former case, we delete the nodes on the small side of the cut and repeat on the remaining graph, where a self-loop is added to each node with weight equal to the removed degree for that node. We continue until the number of deleted nodes first exceeds a  $c/2$  fraction of the nodes. Now there are three cases to consider.

First, if the process deletes less than  $c/2$  fraction of the vertices, then the remaining graph (which has at least  $(1 - c/2)n$  vertices) has conductance  $2\gamma$ , and thus in the original graph every  $c$ -balanced cut has conductance at least  $\gamma$ .

Second, if the process deletes between  $c/2$  and  $1/2$  of the nodes, then the union of the deleted sets gives a cut with conductance at most  $\sqrt{\gamma}$  and we can therefore just output this union.

Finally, if the process deletes more than half the nodes in total then the final removed set had size greater than  $c/2$  and has conductance  $2\sqrt{\gamma}$ , so the algorithm can output this final removed set.  $\square$

Usually the word “expander” is reserved for the case when  $\beta > 0$  is a fixed constant independent of the graph size, in which case the second eigenvalue is bigger than some fixed positive constant. Thus an expander family can be recognized in polynomial time by an eigenvalue computation, and this motivates our definition of expander flows.

### 7.3 Expander flows

An *expander flow* is a multicommodity flow that is a  $\beta$ -expander for some constant  $\beta$ . Such flows can be used to certify the expansion of a graph.

LEMMA 23

If a graph  $G$  contains a multicommodity flow  $(f_{ij})$  that is  $d$ -regular and is a  $\beta$ -expander, then  $\alpha(G) \geq \beta d$ .

PROOF: For every set  $S$  of nodes, the amount of flow leaving  $S$  is at least  $\beta d$ , and hence this is a lower bound on the number of graph edges leaving  $S$ .  $\square$

The previous Lemma is most useful if  $\beta$  is constant, in which case an eigenvalue computation can be used to verify that the given flow is indeed an expander. Thus a  $d$ -regular expander flow may be viewed as a certificate that the expansion is  $\Omega(d)$ . The following Theorem says that such flows exist for  $d = \alpha/\sqrt{\log n}$  in every graph  $G$  satisfying  $\alpha(G) = \alpha$ . This is an interesting structural property of every graph, in the same spirit as Theorem 20. It also yields a certificate that  $\alpha(G) = \Omega(\alpha/\sqrt{\log n})$ .

THEOREM 24

There is a constant  $\beta > 0$  such that the following is true for every unweighted graph  $G$ .

If  $\alpha(G) = \alpha$ , then  $G$  contains a  $d$ -regular multicommodity flow in it that is a  $\beta$ -expander, where  $d = \alpha/\sqrt{\log n}$ .

The theorem above is implied by the following algorithmic version of the theorem.

THEOREM 25

There is a  $\beta_0 > 0$  and a polynomial-time algorithm that, given a graph  $G = (V, E)$  and a degree bound  $d$  either finds a  $d$ -regular  $\beta_0$ -expander flow in  $G$  or else finds a cut of expansion  $O(d\sqrt{\log n})$ .

The following lemma states that given an flow that expands on large sets, one can extend this flow to expand on all sets or find a small cut.

LEMMA 26

There is a polynomial time algorithm with the following property. Given a  $d$ -regular flow in a graph  $G = (V, E)$  where for all  $S \subset V$ ,  $|S| \geq cn$ , satisfies

$$\sum_{i \in S, j \in \bar{S}} f_{ij} \geq \beta d |S|,$$

the algorithm either finds a cut of expansion  $O(d)$  or finds a  $d$ -regular  $\Omega(\beta^2)$  expander flow.

PROOF:

We apply Lemma 22 to the weighted graph induced by the flow and obtain a subgraph  $B$  (with self loops corresponding to the removed edges) which has  $3n/4$  nodes and conductance at least  $\beta^2$ .

We now augment the given flow by embedding a single flowpath to each of the deleted nodes,  $A$ , as follows. We form a flow network, by attaching each deleted vertex to a source and each vertex in the remaining expander to the sink with capacity  $d$  arcs. We compute the feasible flow that minimizes the maximum congestion (congestion of an edge being the multiplicative factor by which the capacity is exceeded). If the congestion of the resulting flow is greater than 1, the corresponding minimum cut has expansion  $O(d)$ . Otherwise, the union of this flow and the previous  $d$ -regular flow can be embedded with constant congestion in  $G$ .

Furthermore, the resulting union is an  $\Omega(\beta^2)$  expander flow: Any cut  $S$  where at least  $1/3$  of  $S$  is in  $B$  has expansion in the induced flow graph of  $\beta^2/3$ . Otherwise, at least  $2/3$  of  $S$  is in  $A$ , in which case the new flow from this subset sends at least  $d$  units of flow to at least  $|S|/3$  nodes not in  $S$ .

□

Thus, to prove Theorem 25 it suffices to prove the following lemma.

LEMMA 27

There is a positive constant  $\beta$ , where there is a polynomial time algorithm that computes a  $d$ -regular flow in a graph  $G = (V, E)$  where for all  $S \subset V$ ,  $|S| \geq cn$ ,

$$\sum_{i \in S, j \in \bar{S}} f_{ij} \geq \beta d |S|,$$

or finds a cut of expansion  $O(d\sqrt{\log n})$ .

PROOF: The algorithm tries to find a feasible solution to the following linear program, where the goal is to find a flow that expands on all large sets. Let  $\mathcal{P}_{ij}$  be the set of paths that go from  $i$  to  $j$ .

$$\forall i \quad \sum_j \sum_{p \in \mathcal{P}_{ij}} f_p = d \quad (18)$$

$$\forall e \quad \sum_j \sum_{e \in p} f_p \leq 1 \quad (19)$$

$$\forall S \subseteq V, |V|/3 \leq |S| \leq |V|/2 \quad \sum_{i \in S, j \in \bar{S}} \sum_{p \in \mathcal{P}_{ij}} f_p \geq \beta d |S|. \quad (20)$$

We note that if constraint (20) was enforced for all sets rather than just large sets, the resulting linear program would specify an expander flow.

Even though this linear program has exponentially many constraints, we can use the Ellipsoid algorithm to solve it approximately in the following sense. Using the algorithm of Lemma 22 as an approximate separation oracle in Ellipsoid algorithm, we can guarantee that if the LP is feasible for a certain  $\beta$ , then we can find a solution that satisfies all constraints for a somewhat smaller value number  $\Omega(\beta^2)$ . Furthermore, if the LP is infeasible for this value of  $\beta$ , then this algorithm finds a dual feasible solution of negative value to the closely related linear program where the constraints (20) apply to sets of size down to  $|V|/6$ .

Thus the heart of the proof is to show how to produce produce a cut with expansion  $O(d\sqrt{\log n})$  given such a dual feasible solution. We will use ideas from our earlier SDP rounding algorithm.

The dual has nonnegative variables  $z_S$  for each  $S$  of size between  $|V|/6$  and  $|V|/2$  and the following constraints.

$$\min \sum_e w_e + d \sum_i s_i - \beta d \sum_S z_S |S|$$

$$\sum_{e \in p} w_e + s_i + s_j \geq \sum_{S: i \in S, j \in \bar{S}} z_S \quad \forall i, j \forall p \in \mathcal{P}_{ij} \quad (21)$$

$$z_S, w_e \geq 0$$

We can assume that  $\sum_S z_S = 1$  by rescaling, since scaling does not affect the sign of the dual objective function.

To prove Lemma 27, we show how to use a negative valued solution to the dual program to produce a cut of expansion  $O(d\sqrt{\log n})$ .

Let us consider the weighted graph, with weights  $w_e$  on the edges, together with a distribution  $z_S$  on cuts in this graph. We will give a procedure that either finds a cut of expansion  $O(d\sqrt{\log n})$  or finds  $\Omega(n)$  disjoint pairs of vertices  $i, j$  where each pair  $(i, j)$  has

$$w_{ij} \leq W/nd \quad (22)$$

where  $W = \sum_e w_e$ , and

$$z_{ij} = \sum_{i \in S, j \in \bar{S}} z_S = \Omega(1) \quad (23)$$

(i.e. short paths that cross most cuts). In the second case, we argue that the dual value is positive which is a contradiction.

Summing over the constraints (21) for these pairs, we obtain

$$\frac{W}{d} + \sum_i s_i = \Omega(n),$$

or equivalently that

$$\sum_e w_e + d \sum_i s_i - \Omega(dn) \geq 0.$$

Observing that  $\sum_S z_S |S| \leq n$  and choosing a sufficiently small constant  $\beta$ , we note that the dual value is positive, thus showing that the LP is feasible (i.e., a  $\beta$  expander flow exists).

We start by using the  $z_S$  variables to define a  $c$ -spread  $\ell_2^2$  embedding of the graph in  $\mathbb{R}^M$  with a coordinate for each cut  $S$ . Each vertex  $x$  is mapped to a vector  $v_x$  where the  $S$ th coordinate is  $\pm\sqrt{z_S}$  depending on which side of  $S$  it lies. Moreover, the typical pair of vertices is  $\Omega(1)$  apart in  $\ell_2^2$  distance. This follows from the fact that each coordinate  $S$  contributes  $\Theta(z_S)$  to the  $\ell_2^2$  distance for constant fraction pairs since  $|S|$  is  $\Omega(n)$ . We let  $z_{ij}$  denote the distance between  $i$  and  $j$  in this metric.

Our algorithm attempts first to find a small cut as follows:

1. Use dual  $z_S$  to create an embedding as described above.
2. Project onto a random direction  $\vec{u}$  and identify two sets  $A$  and  $B$  whose projections are  $\geq \sigma/\sqrt{n}$  ( $\leq -\sigma/\sqrt{n}$  respectively) on  $\vec{u}$ .
3. Connect a source vertex to each node in  $A$  and a sink to each node in  $B$  with unit capacity edges, assign each edge of  $G$  a capacity of  $C = \lceil 1/kd \rceil$ , and either compute a feasible flow or output the minimum cut.

If the algorithm outputs a cut, it is easy to see that it has small expansion, i.e., at most  $kd$  since there must be an  $l$  where  $l$  sources (or sinks) are disconnected from the sink (or source) by at most  $l/kd$  edges. So, choosing  $k = \Theta(\sqrt{\log n})$  gives the desired cut.

Otherwise, we wish to apply Lemma 7 to either find a cut or to produce a pair  $(i, j)$  satisfying conditions (22) and (23). To do so, we must form matching covers and define a proximity graph.

We form matching covers as follows. A feasible flow between  $A$  and  $B$  computed for a direction  $\vec{u}$  can be decomposed into a set of paths pairing nodes in  $A$  with nodes in  $B$ . Since each edge is used by at most  $C$  such paths, the typical flowpath between  $A$  and  $B$  has length of  $CW/n = \Theta(W/nkd)$ . For a direction  $\vec{u}$ , we output those pairs where the length is at most twice the typical distance, yielding a matching with  $\Omega(n)$  pairs.

We define a graph with edges between each pair of vertices with  $w_{i,j} \leq W/kdn$ . If there exists an edge  $(i,j)$  with  $z_{i,j} = \Omega(1)$ , then the pair  $(i,j)$  satisfies (22) and (23). Otherwise, this graph is a  $\zeta$ -proximate graph as required in Lemma 15.

If the application of the Lemma 15 terminates with a non-expanding set  $S$  in the  $\zeta$ -proximate graph a randomly chosen  $\delta$ -neighborhood of  $S$  for  $\delta \leq W/kdn$  yields a cut with sufficiently small expansion,  $O(kd)$  since the resulting cut has  $\Omega(n)$  nodes on each side and the cut has expected size  $O(W/\delta) = O(kd)$ . If the Lemma terminates with a pair of vertices  $i,j$ , then  $z_{ij} = \Omega(1)$  and  $w_{ij} = O(\sqrt{\log n}W/kdn) \leq W/dn$  for suitable choice of  $k = \Theta(\sqrt{\log n})$  as desired.

Deleting this pair, leaves a set of vertices that remains matching covered for slightly worse parameters. In particular, it removes 1 pair from the  $\Theta(n)$  pairs in each matching. Thus, the process can be repeated  $\Theta(n)$  times to yield  $\Theta(n)$  pairs. This contradicts the assumption that the dual is negative which implies that we would find a cut at some point in this process.

We note that as before the analysis can be simplified using Lee's techniques [Lee05]. In particular, cuts are only produced by the projection based procedure above.

□

## 7.4 Expander flows as dual solutions

Now we show that expander flows are actually a natural family of dual solutions to the SDP for sparsest cut. Denote by  $OPT(W)$  the value of the simplest SDP relaxation (with no triangle inequality) of this problem: find vectors  $v_1, v_2, \dots, v_n \in \mathbb{R}^n$  so as to:

$$\min \sum_{\{i,j\}} c_{i,j} |v_i - v_j|^2 \tag{24}$$

$$\sum_{i \neq j} |v_i - v_j|^2 = 1 \tag{25}$$

Let  $OPT_{\Delta}(W)$  denote the value of the relaxation when we also impose the triangle inequality constraints. Now we relate  $OPT(\cdot)$  and  $OPT_{\Delta}(\cdot)$ .

Recall from our discussion of expander flows that a multicommodity flow in  $G$  can be viewed as an *embedding* of the demand graph  $D = (d_{ij})$ . We will show that such a flow is a dual solution to the SDP relaxation with triangle inequality (we have learnt that this observation was known to others, including Sudan and Williamson).

LEMMA 28 (FLOWS AS DUAL SOLUTIONS)

If a demand graph  $D = (d_{ij})$  can be embedded in  $G = (V, W)$  then

$$OPT_{\Delta}(W) \geq OPT(D).$$

REMARK 11 Of course, when  $D$  is an expander, then the value  $OPT(D)$  is (upto scaling) essentially the same as the eigenvalue bound for sparsest cut.

To prove the lemma we need to write the duals. Let us do this for the simple SDP, first rewriting it using  $x_{ij}$  and  $z_{ij}$  to replace  $|v_i - v_j|^2$  and  $\langle v_i, v_j \rangle$  respectively. We use  $\text{diag}(a_i)$  to denote the  $n \times n$  matrix which has  $a_1, a_2, \dots, a_n$  on the diagonal and 0 everywhere else.

$$PRIMAL \qquad \qquad \qquad DUAL \qquad \qquad \qquad (26)$$

$$\min \sum_{\{i,j\}} c_{ij} x_{ij} \qquad \qquad \qquad \max \quad z \qquad \qquad \qquad (27)$$

$$\sum_{i,j} x_{ij} = 1 \qquad \qquad \qquad z + s_{ij} = c_{ij} \quad \forall i, j \qquad \qquad \qquad (28)$$

$$\forall i, j \quad x_{ij} + 2z_{ij} - z_{ii} - z_{jj} = 0 \qquad \qquad \qquad s_{ij} \text{ unrestricted} \qquad \qquad \qquad (29)$$

$$(z_{ij}) \text{ is psd} \qquad \qquad \qquad \text{diag}(\sum_{j \neq i} s_{ij}) - (2s_{ij})_{i \neq j} \text{ is psd} \qquad \qquad \qquad (30)$$

Now we consider the SDP with triangle inequality. The usual formulation is to write for each triple  $i, j, k$  an additional constraint:  $x_{ij} + x_{jk} \geq x_{ik}$ . We use an equivalent but more complicated formulation: if  $\mathcal{P}_{ij}$  denotes the set of paths from  $i$  to  $j$  in the *complete* graph on  $n$  vertices, then

$$\forall i, j, \forall p \in \mathcal{P}_{ij}, \quad \sum_{(u,v) \in p} x_{uv} \geq x_{ij} \qquad \qquad \qquad (31)$$

Now let us write the primal and the dual.

$$PRIMAL \qquad \qquad \qquad DUAL$$

$$\min \sum_{\{i,j\}} c_{ij} x_{ij} \qquad \qquad \qquad \max \quad z$$

$$\sum_{i,j} x_{ij} = 1 \qquad \qquad \qquad z + s_{ij} - \sum_{p \in \mathcal{P}_{ij}} f_p + t_{ij} = c_{ij} \quad \forall i, j \qquad \qquad \qquad (32)$$

$$\forall i, j \quad x_{ij} + 2z_{ij} - z_{ii} - z_{jj} = 0 \qquad \qquad \qquad s_{ij} \text{ unrestricted} \qquad \qquad \qquad (33)$$

$$\forall i, j, \forall p \in \mathcal{P}_{ij}, \quad \sum_{(u,v) \in p} x_{uv} \geq x_{ij} \qquad \qquad \qquad (34)$$

$$(z_{ij}) \text{ is psd} \qquad \qquad \qquad \text{diag}(\sum_{j \neq i} s_{ij}) - (2s_{ij})_{i \neq j} \text{ is psd} \qquad \qquad \qquad (35)$$

Here  $t_{ij}$  is a shorthand for  $\sum_{p \ni \{i,j\}} f_p$ .

Now let  $D$  be the demand graph of any multicommodity flow that can be embedded in  $G$ . A flow is an assignment of nonnegative values to  $f_p$ 's such that  $t_{ij} \leq c_{ij}$  (indeed, the only nonzero  $f_p$ 's are for  $p$  that are actually paths in  $G$ ). Fixing these values of  $f_p$ 's, finding the best values to  $z$  and the  $s_{ij}$ 's is nothing but the same as finding  $OPT(W')$  where  $W'$  is the weighted graph in which  $W'_{ij}$  is  $c_{ij} - t_{ij} + D_{ij}$ , where  $D_{ij}$  is the total flow from  $i$  to  $j$ , namely  $\sum_{p \in \mathcal{P}_{ij}} f_p$ . Of course, since  $t_{ij} \leq c_{ij}$ , this is lowerbounded by  $OPT(D)$  where  $D$  is the demand graph of the flow. Thus the lemma is proved.

## 8 The open questions

Currently it is completely plausible that our techniques can be strengthened to achieve an approximation ratio much better than  $\sqrt{\log n}$  for UNIFORM SPARSEST CUT. Originally it was suggested in [15] that the SDP with triangle inequality has an integrality gap of  $O(1)$  but Khot and Vishnoi recently showed that the gap is  $\Omega(\log \log n)$ . As noted earlier, the example of the boolean hypercube shows that our main theorem (Theorem 1) about the existence of large  $\Delta$ -separated subsets cannot be improved, so some newer idea seems necessary.

Below we provide a suggested roadmap for attacking these questions. We three questions. All concern any constant degree graph  $G = (V, E)$  and any unit- $\ell_2^2$ - representation  $v_1, v_2, \dots, v_n$ , that is  $c$ -spread for some constant  $c$ .

Is there a function  $s(n) = o(\sqrt{\log n})$  such that:

**Question 1:** there is always an edge  $\{i, j\} \in E$  such that  $|v_i - v_j|^2 = \Omega(\alpha/s(n))$ ?

**Question 2:** there are pairs of vertices  $i, j$  such that  $d_G(i, j) = O(s(n)/\alpha)$  and  $|v_i - v_j|^2 = \Omega(1)$ ?

**Question 3:** Version of Question 2 whereby for some  $\tau > 0$ , an arbitrary set of  $\tau$  fraction of nodes are forbidden from being chosen as  $i, j$  and nevertheless these  $i, j$  exist.

Earlier in the paper we proved that the answer to these questions is Yes if  $s(n) = O(\sqrt{\log n})$ . The authors could not agree on a guess as to whether or not the answer is Yes to any of these questions for an  $s(n)$  that grows slower.

Now we summarise, without detailed proofs, the implications if the answers to these questions is Yes.

LEMMA 29

1. If the integrality gap of the SDP is  $O(s(n))$  then the answer to Question 1 is Yes.
2. A Yes answer to Question 2 suffices to prove that the integrality gap is  $O(s(n))$ .
3. A Yes answer to Question 3 implies the existence of expander flows in graphs that certify expansion of  $\alpha/s(n)$ . Namely, Theorem 24 is true with  $d = \alpha/s(n)$ .

PROOF:

1. If the integrality gap of the SDP is  $O(s(n))$  then the value of the objective is  $\Omega(\alpha n/s(n))$  and hence at least one edge has value  $\Omega(\alpha/s(n))$ .
2. Uses a modification of our original proof that the integrality gap is  $O(\sqrt{\log n})$ .
3. Question 3 is a stronger version of statements proved earlier. One can mimic the proof of Theorem 24 using this stronger version of the corollary.

□

## 8.1 Bringing the Question “down” to $\ell_1$

Before explaining the title of this section we first recall the connection between cut problems and embeddings into  $\ell_1$ . A unit  $\ell_2^2$ -representation  $v_1, v_2, \dots, v_n$  of a graph is said to be  $\ell_1$  if there is a set of vectors  $u_1, u_2, \dots, u_n$  such that  $|v_i - v_j|^2 = |u_i - u_j|_1$ . We say that it is  $\ell_1$  upto distortion  $c$  if  $|u_i - u_j|_1 \leq |v_i - v_j|^2 \leq c \cdot |u_i - u_j|_1$ . The importance of embeddings to cut problems arises from the following equivalent characterization.

LEMMA 30 (WELL-KNOWN)

A unit  $\ell_2^2$ -representation  $v_1, v_2, \dots, v_n$  is  $\ell_1$  iff there is a  $\alpha_S \geq 0$  associated with each cut  $(S, \bar{S})$  such that

$$|v_i - v_j|^2 = \sum_S \alpha_S d_S(i, j), \quad (36)$$

where  $d_S(i, j) = 0$  if  $i, j$  are on the same side of the cut and 4 otherwise. (In other words,  $\ell_1$  representations correspond exactly to the cut cone.)

PROOF: One direction is trivial from convexity since every cut metric is  $\ell_1$ .

For a proof of the other direction see Shmoys’ survey [31] or Matousek’s book [26]. □

Thus minimization over  $\ell_1$  metrics is exactly equivalent to minimizing over cuts (and thus **NP**-hard). Goemans and Linial independently conjectured that every  $\ell_2^2$ -representation is  $\ell_1$  upto a distortion factor

$O(1)$ . If true, this would have implied an integrality gap of  $O(1)$  for NONUNIFORM SPARSEST CUT, but recently Khot and Vishnoi showed that distortion  $(\log \log n)^\epsilon$  is necessary. However the best upperbound on the distortion is  $O(\sqrt{\log n} \log \log n)$ , so a large gap remains.

One may reasonably question whether it makes much sense to focus attention exclusively on  $\ell_2^2$  metrics when these metrics merely happen to capture the properties of *one* family of SDPs. Shouldn't one expend more effort in identifying stronger SDP relaxations by specifying more constraints than the triangle inequality? After all, procedures such as Lovász-Schrijver [LS91] give us an infinite hierarchy of progressively tighter SDP relaxations to choose from. (The triangle inequality constraint is implied already in the relaxation at the third level of this hierarchy.)

Now we point out that at least in context of our Questions, adding other constraints (in addition to the triangle inequality) to the SDP may not help. We bring “down” the Questions to  $\ell_1$  so that it suffices to prove them for  $\ell_1$  metrics. Thus although adding other constraints could conceivably give a class of metrics that is a proper subset of  $\ell_2^2$ , proving our conjectures for that other class of metrics will not be any easier, since that class of metrics (being a relaxation of cut metrics) would always contain  $\ell_1$ .

**[Question 4:]** If the unit- $\ell_2^2$ -representation is actually  $\ell_1$ , then there is a vertex pair  $i, j$  such that  $d_G(i, j) = O(s(n)/\alpha)$  and  $|v_i - v_j|^2 = \Omega(1)$ .

LEMMA 31

*Questions 2 and 4 are equivalent.*

PROOF: Yes to Question 2  $\Rightarrow$  Yes to Question 4: Since  $\ell_1$  metrics are a subcase of  $\ell_2^2$  metrics, Question 4 is trivially a subcase of Question 2.

Yes to Question 4  $\Rightarrow$  Yes to Question 2: Given any unit- $\ell_2^2$  representation  $v_1, v_2, \dots, v_n$  of the graph, consider the uniform distribution on all hyperplane cuts (a la [GW]). Represent this distribution using Lemma 30 by an  $\ell_2^2$  representation that is  $\ell_1$ , namely, a set of unit vectors  $u_1, u_2, \dots, u_n$  where

$$|u_i - u_j|^2 = \Pr[i, j \text{ are separated in hyperplane cut produced using } v_1, \dots, v_n].$$

Let  $i, j$  be the pair of nodes in  $G$  whose existence follows if Question 4 has a Yes answer. Namely,  $|u_i - u_j|^2 = \Omega(1)$ . Then these also suffice for Question 2 since if

$$\Pr[i, j \text{ are separated in hyperplane cut on } v_1, \dots, v_n] = \Omega(1),$$

then  $|v_i - v_j|^2 = \Omega(1)$  also.  $\square$

REMARK 12 We can make a robust form of Question 4 whereby the pair exists even if we forbid  $\tau n$  nodes from playing the role of  $i, j$ . Giving a Yes answer to that question is *equivalent* to showing that expander flows can be used to certify an expansion of  $\alpha/s(n)$ .

## 9 Conclusions

In the three years since the initial distribution of this paper, the geometric techniques introduced here have found many uses in designing other SDP-based approximation algorithms as well as resolving open questions in metric space embeddings. We see several promising avenues of research in the near future.

First, there appears to be plenty of scope to use more sophisticated geometric arguments in analysing SDPs. Procedures such as Lovász-Schrijver [LS91] or Sherali-Adams [SA90] in principle give an infinite family of SDP relaxations that have thus far resisted all attempts at analysis but may have low integrality gaps. Since the triangle inequality constraints are obtained trivially in these frameworks, perhaps our analysis of the triangle inequality constraints gives a hint how to analyse more complicated SDPs.

Second, as already mentioned even the SDPs with triangle inequality are not fully understood as yet: there is a big gap between the lowerbounds and upperbounds known on the integrality gap. A similar

observation holds for questions about geometric embeddings such as the distortion required to embed  $\ell_2^2$  metrics into  $\ell_1$ .

Finally, there is a need to design efficient approximation algorithms, since solving SDPs is fairly inefficient (though polynomial time). Solving the SDP of (4)–(7) takes about  $n^{4.5}$  time for an  $n$ -node graph using interior point methods. As mentioned, a more efficient  $\tilde{O}(n^2)$ -time algorithm is now known [AHK04] that uses expander flows. Recently a more general *primal-dual* framework was invented [AK] for replacing SDPs with faster approximation algorithms for problems other than SPARSEST CUT. Our Structure Theorem plays an important role in the analysis of the running time. It would be interesting to continue to improve the running times, and to extend the framework to even more problems.

Finally, we feel expander flows ought to have other applications besides estimating the conductance/expansion. Some applications to Banach space theory have already been found [NRS].

## Acknowledgements

This project evolved over several years. A partial list of people who gave us useful feedback about our ideas and/or thought about our conjectures (apologies to people we forgot): Farid Alizadeh, Dorit Aharonov, Noga Alon, Moses Charikar, Michel Goemans, Eran Halperin, Mohammed Hajiaghayi, Subhash Khot, Robi Krauthgamer, Tom Leighton, Laci Lovász, Ran Raz, Madhu Sudan, Vijay Vazirani, Santosh Vempala, David Williamson.

We are very grateful to James Lee for comments on our first manuscript. He drew our attention to the geometric core of our argument, which helped us greatly improve the presentation.

## References

- [ACMM05] Amit Agarwal, Moses Charikar, Konstantin Makarychev, and Yury Makarychev.  $o(\sqrt{\log n})$  approximation algorithms for min uncut, min 2cnf deletion, and directed cut problems. In *STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 573–581, New York, NY, USA, 2005. ACM Press.
- [AHK04] Sanjeev Arora, Elad Hazan, and Satyen Kale.  $O(\sqrt{\log n})$  approximation to sparsest cut in  $\tilde{O}(n^2)$  time. In *FOCS '04: Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS'04)*, pages 238–247, Washington, DC, USA, 2004. IEEE Computer Society.
- [AK] Sanjeev Arora and Satyen Kale. A combinatorial, primal-dual approach to semidefinite programming.
- [ALN05] Sanjeev Arora, James R. Lee, and Assaf Naor. Euclidean distortion and the sparsest cut. In *STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 553–562, New York, NY, USA, 2005. ACM Press.
- [Alo86] Noga Alon. Eigenvalues and expanders. In *Combinatorica*, volume 6, pages 83–96, 1986.
- [CGR05] Shuchi Chawla, Anupam Gupta, and Harald R&#228;cke. Embeddings of negative-type metrics and an improved approximation to generalized sparsest cut. In *SODA '05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 102–111, Philadelphia, PA, USA, 2005. Society for Industrial and Applied Mathematics.
- [CHKR06] Moses Charikar, Mohammad Taghi Hajiaghayi, Howard Karloff, and Satish Rao.  $\ell_2$  spreading metrics for vertex ordering problems. In *SODA '06: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 1018–1027, New York, NY, USA, 2006. ACM Press.

- [Enf69] P. Enflo. On the nonexistence of uniform homeomorphisms between lp-spaces. *Ark. Mat.*, 8(13):103–105, 1969.
- [FHL05] Uriel Feige, MohammadTaghi Hajiaghayi, and James R. Lee. Improved approximation algorithms for minimum-weight vertex separators. In *STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 563–572, New York, NY, USA, 2005. ACM Press.
- [FL] Uriel Feige and James Lee. An improved approximation ratio for the minimum linear arrangement. *To appear in IPL*.
- [Kar05] George Karakostas. A better approximation ratio for the vertex cover problem. In *ICALP*, pages 1043–1050, 2005.
- [KLMN04] Robert Krauthgamer, James R. Lee, Manor Mendel, and Assaf Naor. Measured descent: A new embedding method for finite metrics. In *FOCS '04: Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS'04)*, pages 434–443, Washington, DC, USA, 2004. IEEE Computer Society.
- [KRV06] Rohit Khandekar, Satish Rao, and Umesh Vazirani. Graph partitioning using single commodity flows. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 385–390, New York, NY, USA, 2006. ACM Press.
- [KV05] Subhash A. Khot and Nisheeth K. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into  $\ell_1$ . In *FOCS '05: Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 53–62, Washington, DC, USA, 2005. IEEE Computer Society.
- [Lee05] James R. Lee. On distance scales, embeddings, and efficient relaxations of the cut cone. In *SODA '05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 92–101, Philadelphia, PA, USA, 2005. Society for Industrial and Applied Mathematics.
- [LR04] Kevin Lang and Satish Rao. A flow-based method for improving the expansion or conductance of graph cuts. In *IPCO*, pages 325–337, 2004.
- [LS91] L. Lovasz and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM Journal on Optimization*, 2(1):166–190, May 1991.
- [NRS] A. Naor, Y. Rabani, , and A. Sinclair. Quasisymmetric embeddings, the observable diameter, and expansion properties of graphs. Preprint, 2004.
- [SA90] H. D. Sherali and W. P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3:411–430, 1990.

## References

- [1] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM J. Optimization*, 5:13–51, 1995.
- [2] N. Alon. Eigenvalues and expanders. *Combinatorica* 6:83–96, 1986.
- [3] N. Alon and V. Milman.  $\lambda_1$ , isoperimetric inequalities for graphs and superconcentrators. *J. Combin. Theory B* 38:73–88, 1985.

- [4] Y. Aumann and Y. Rabani. An  $O(\log k)$  approximate min-cut max-flow theorem and approximation algorithms. *SIAM J. Comp*
- [5] K. Ball. An elementary introduction to modern convex geometry, in *Flavors of Geometry*, S. Levy (ed.), Cambridge University Press, 1997.
- [6] M. Blum, R. Karp, O. Vornberger, C. Papadimitriou, M. Yannakakis. The complexity of testing whether a graph is a superconcentrator. *Inf. Proc. Letters* **13**:164-167, 1981.
- [7] J. Bourgain. On Lipschitz embeddings of finite metric spaces in Hilbert space. *Israel J. Mathematics* **52**:46-52, 1985.
- [8] J. Cheeger. A lower bound for the smallest eigenvalue of the Laplacian, in *Problem in Analysis*, 195-199, Princeton Univ. Press, (1970),
- [9] F. Chung. Spectral graph theory. *CBMS Regional Conference Series in Mathematics*, 92, American Mathematical Society, 1997.
- [10] L. Danzer and B. Grünbaum. On two problems of P. Erdős and V. L. Klee concerning convex bodies (in German). *Math. Zeitschrift* **79**:95-99, 1962.
- [11] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov chains. *Annals of Applied Probability*, **3**:696-730, 1993.
- [12] U. Feige and R. Krauthgamer. A polylogarithmic approximation of the minimum bisection. In *IEEE FOCS 2001* pp 105-115.
- [13] N. Garg and J. Köneman. Faster and Simpler Algorithms for Multicommodity Flow and other Fractional Packing Problems. In *IEEE FOCS 1997*.
- [14] N. Garg and V. V. Vazirani and M. Yannakakis. Approximate max-flow min-(multi)cut theorems and their applications. *SIAM J. Computing*, **25**(2):235-251, 1996. *Prelim. version in Proc. ACM STOC'93*.
- [15] M.X. Goemans. Semidefinite programming in combinatorial optimization. *Math. Programming*, **79**:143-161, 1997.
- [16] M. X. Goemans. *unpublished note*.
- [17] M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *JACM*, **42**(6):1115-1145, 1995.
- [18] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization*. Springer-Verlag, 1993.
- [19] H. Karloff and U. Zwick. A 7/8-approximation algorithm for MAX 3SAT? *Proc. of 38th IEEE FOCS* (1997), 406-415.
- [20] D. Karger, R. Motwani, and M. Sudan. Approximate graph coloring by semidefinite programming. *JACM*, **45**(2):246-265, 1998.
- [21] M. Jerrum and A. Sinclair. Approximating the permanent. *SIAM J. Comput.*, **18**(6):1149-1178, 1989.
- [22] T. Leighton and S. Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *JACM* **46** 1999. *Prelim. version in ACM STOC 1988*.
- [23] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica* **15**:215-246, 1995.

- [24] L. Lovász. On the Shannon capacity of a graph. *IEEE Trans. on Info. Theory* **IT-25**:1–7, 1979.
- [25] A. Lubotzky, R. Philips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, **8**:261-277, 1988.
- [26] J. Matousek. *Lectures on Discrete Geometry*. Springer Verlag, 2002.
- [27] Y. Nesterov and A. Nemirovskii. *Interior point polynomial methods in convex programming*. SIAM, Philadelphia, PA 1994.
- [28] S. Plotkin and D. B. Shmoys and E. Tardos. Fast approximation algorithms for fractional packing and covering problems. *Math. Operations Res.* **20**:257-301, 1995. *Prelim. version IEEE Foundations of Computer Science, 1991, 495-504.*
- [29] G. Schechtman. Concentration, results and applications. *Handbook of the Geometry of Banach Spaces*, volume 2, W.B. Johnson and J. Lindenstrauss (eds.), North Holland, 2003. Draft version available from Schechtman’s website.
- [30] F. Shahrokhi and D.W. Matula. The maximum concurrent flow problem. *Journal of the ACM*, **37**:318–334, 1990.
- [31] D. S. Shmoys. Cut problems and their application to divide and conquer. *Approximation Algorithms for NP-hard problems*, D.S. Hochbaum (ed.), PWS Publishing, 1995.
- [32] A. Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Prob., Comput.* **1**:351–370, 1992.
- [33] V. Vazirani. *Approximation algorithms*. Springer Verlag, 2002.
- [34] K. Zatloukal. *Personal communication*, November 2003.