Expander Flows, Geometric Embeddings and Graph Partitioning

Sanjeev Arora* Satish Rao† Umesh Vazirani‡

ABSTRACT

We give a $O(\sqrt{\log n})$-approximation algorithm for SPARSEST CUT, 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 "certificate" for a graph's expansion, by embedding an n-node expander in it with appropriate dilation and congestion. We call this an expander flow.

Categories and Subject Descriptors

F.2 [**Theory of Computation**]: Analysis of Algorithms

General Terms

Algorithms,Theory

Keywords

graph partitioning,sparsest cuts,normalized cuts, embeddings,semidefinite programming,approximation algorithms,expander,clustering,conductance, spectral methods, eigenvalues

*Computer Science Department, Princeton University. 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-009818.

†Computer Science Division, UC Berkeley. www.cs.berkeley.edu/∼satishr. Partially supported by NSF award CCR-0105533

‡Computer Science Division, UC Berkeley.www.cs.berkeley.edu/∼vazirani Partially supported by NSF ITR Grant CCR-0121555

STOC'04, June 13–15, 2004, Chicago, Illinois, USA.

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 [29]).

Here we give new approximation algorithms for some of the important problems in this class. In a graph $G = (V, E)$, for any cut $(S, \overline{S})$ where $|S| \leq |V|/2$, the *edge expansion* of the cut is $\left|E(S, \overline{S})\right| / |S|$. In the SPARSEST CUT problem we wish to determine the cut with the smallest edge expansion:

$$\alpha(G) = \min_{S \subseteq V, |S| \leq |V|/2} \frac{\left|E(S, \overline{S})\right|}{|S|}. \tag{1}$$

A cut $(S, \overline{S})$ is *c-balanced* if both $S, \overline{S}$ have at least $c|V|$ vertices. In the c-BALANCED-SEPARATOR problem we wish to determine $\alpha_c(G)$, the minimum expansion of c-balanced cuts. In the GRAPH CONDUCTANCE problem we wish to determine

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

where $E(S)$ denotes the set of edges incident to nodes in S. We can reduce each of these problems to constant degree graphs and moreover for this class, edge expansion and conductance are related by a constant factor.

A weak approximation for GRAPH CONDUCTANCE follows from the connection —first discovered in context of Riemannian manifolds [7]—between conductance and the eigenvalue gap of the Laplacian: $2\Phi(G) \geq \lambda \geq \Phi(G)^2/2$ [3, 2, 19]. The approximation factor is $1/\Phi(G)$, and hence $\Omega(n)$ in the worst case, however for constant $\phi(G)$ it is an excellent bound. This connection between eigenvalues and expansion has had enormous influence in a variety of fields (see e.g. [8]).

Leighton and Rao [20] designed the first true approximation by giving $O(\log n)$-approximations for SPARSEST CUT and GRAPH CONDUCTANCE and $O(\log n)$-pseudo-approximations for c-BALANCED SEPARATOR. They used a linear programming relaxation of the problem based on multicommodity flow proposed in [28]. This led to approximation algorithms for numerous NP-hard problems, see [29]. However, the integrality gap of the LP is $\Omega(\log n)$, and crossing this $\log n$ barrier therefore calls for new techniques.

In this paper we give $O(\sqrt{\log n})$-approximations for SPARSEST CUT and GRAPH CONDUCTANCE and $O(\sqrt{\log n})$-pseudo-approximation to c-BALANCED SEPARATOR.

Now we give a quick overview of our results. Our algorithm uses *semidefinite programming* (SDP). These are mathematical programs in which each vertex i is assigned some point v_i on the unit sphere in $\Re^n$. In our case the goal is to find an assignment such that the average distance between all pairs of points is "large" whereas the average distance between endpoints of edges is minimized.

The complexity of finding such embeddings depends crucially on the notion of distance. Under the standard Euclidean norm (or even ℓ_1 norm) the problem is NP hard: the optimum cut can be efficiently recovered from the optimum vectors. The notion of distance that is more tractable (and used in SDPs) is the *square* of the Euclidean norm, the so-called ℓ_2^2 norm. With this distance function, the embedding problem is related to finding eigenvectors of the adjacency matrix of the graph and thus yields only weak approximations. To tighten the relaxation, one can ask that ℓ_2^2 distances between the v_i's form a metric: every triple i, j, k satisfies the triangle inequality, i.e, $|v_i - v_j|^2 + |v_j - v_k|^2 \geq |v_i - v_k|^2$. The general SDP framework allows such constraints. Furthermore, these constraints correspond exactly to the linear constraints in the Leighton-Rao LP relaxation and therefore this ℓ_2^2 embedding subsumes both the eigenvalue as well as the $O(\log n)$ linear programming bound. The conjectured integrality gap of the resulting relaxation to the cut problem is $O(1)$ [14], and is known to be at least 10/9 [32].

Our $O(\sqrt{\log n})$-approximation relies on a new result about the geometric structure of such embeddings: 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 ℓ_2^2 distance. (We also present a randomized algorithm to find such sets.) This result is tight for an n-vertex hypercube —whose natural embedding into $\Re^{\log n}$ defines an ℓ_2^2 metric— where any two large sets are within $O(1/\sqrt{\log n})$ distance.

Finding such a well-separated subset pair suffices for a good approximation. Since the sum of the ℓ_2^2 distances between endpoints of edges is small in the embedding, the sets S and T cannot have many edges between them, and this is the basis of finding a small cut. Formally, finding a good separator involves shrinking S to a point and performing a breadth first search from it and outputting the level with fewest edges (Section 2.1.1).

The algorithm for finding the above-mentioned well-separated pair S, T is complicated (Section 5) but we also describe a simpler algorithm (Section 3) that works for a somewhat smaller separation $\Delta = \Omega(1/\log^{2/3} n)$. In that case the idea is to partition the vectors with a randomly oriented hyperplane slice of prescribed thickness. Points that fall inside the slice are discarded. The sets of points on the two sides of the slice are our first candidates for S, T. However, they can contain a few pairs of points $v_i \in S, v_j \in T$ whose squared distance is less than Δ, which we discard. The technically hard part in the analysis is to prove that not too many points get discarded. This makes essential use of a phenomenon called *measure concentration*, a cornerstone of modern convex geometry [5].

Graph embeddings and expander flows: Our ideas also imply a new structural result in graph theory: an embedding of expander graphs in any arbitrary graph that is more efficient (in terms of maximum edge congestion, which is the number of expander edges routed though a single graph edge) than any known before. This result is proved using techniques similar to the ones used to prove the existence of the Δ-separated sets (though it is not an immediate corollary and requires some work). To understand the connection to approximation algorithms, 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 to embed a complete graph into the given graph with minimum congestion. This can be accomplished fractionally by routing a single unit of flow between each pair of vertices while minimizing the maximum congestion, μ of any edge. Clearly, every cut must have expansion at least $1/n\mu$. (See Section 7.) This is exactly the certificate used in the Leighton-Rao paper, where it is shown that congestion $O(\alpha/n \log n)$ suffices (and this amount of congestion is required on some graphs.)

This paper considers a generalization of this approach, where we embed not the complete graph but some flow that is an expander. We show how this idea can be used to derive a certificate to the effect that the expansion is $\Omega(\alpha/\sqrt{\log n})$ (see Section 7). The conjectures presented in the full version imply that this approach can be improved to certify that the expansion is $\Omega(\alpha)$.

Expander flows also provide a different and possibly more efficient (though the current writeup ignores efficiency issues besides polynomiality) $O(\sqrt{\log n})$-approximation algorithm for graph separators that uses multicommodity flows combined with eigenvalue computations. A polynomial bound follows by observing that embedding a particular graph (the expander flow) with minimum congestion is a multicommodity flow problem. The condition that the embedded graph is an expander can be imposed by exponentially many linear constraints, one for each cut. A violated (within a constant factor) constraint can be efficiently found by an eigenvalue conmputation, and thus the linear program can be solved by the Ellipsoid method. More details appear in the full version. In fact, the algorithms of this paper (including the SDP rounding) were discovered in this setting.

Related Work.

Semidefinite programming and approximation algorithms: Semidefinite programs (SDPs) have numerous applications in optimization. They are solvable in polynomial time via the ellipsoid method [16], and more efficient interior point methods are now known [1, 25]. In a seminal paper, Goemans and Williamson [15] used SDPs to design good approximation algorithms for MAX-CUT and MAX-k-SAT. Researchers soon extended their techniques to other problems [18, 17, 14], but lately progress in this direction has 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 main theorem about ℓ_2^2 spaces (and the "rounding" technique that follows from it) takes a more global view of the metric space. The ideas may prove useful for other problems where triangle inequality constraints are conjectured to tighten SDP relaxations.

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 [10] 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 [30] suggested a similar technique and also noted that the Leighton-Rao multicommodity flow can be viewed as a generalization of the Jerrum-Sinclair [19] 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 ℓ_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 ℓ_1 (hopefully, with low distortion), and then derive an approximation algorithm. This approach was pioneered in [21] and [4]; see Shmoys [29] for a survey. A major open problem in this area is to show that ℓ_2^2 metrics (i.e., solutions to the SDP with triangle inequality constraints) embed into ℓ_1 with $O(1)$ distortion. Showing this would prove an integrality gap of $O(1)$ not only for SPARSEST CUT but also for a more general version of the problem involving nonuniform demands between vertex pairs. The current paper does not address this conjecture. However, James Lee pointed out to us that our results (specifically, Theorem 1 which was earlier implicit in our paper and now explicit thanks to his observation) do represent partial progress: they give an embedding of ℓ_2^2 metrics into ℓ_1 in which the *average* edge distorts by at most $\sqrt{\log n}$ factor. Furthermore, we do note in the full version of the paper that for the version of SPARSEST CUT considered here, the embedding conjecture is overkill. Instead we present weaker conjectures that are sufficient to prove an $O(1)$ integrality gap. We mention a related conjecture about ℓ_1 spaces that is also of interest.

2. DEFINITIONS AND RESULTS

Throughout the paper we will assume that we are dealing with constant degree unweighted graphs, since the general case can be reduced to this case, as is well-known. Furthermore, GRAPH CONDUCTANCE also reduces to SPARSEST CUT on constant degree graphs.

DEFINITION 1 (ℓ_2^2 REPRESENTATION) *A* vector representation *of a graph is an assignment of a vector to each node, say v_i assigned to node i. It is called an ℓ_2^2-*representation *if for all i, j, k:*

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

An ℓ_2^2-representation is called a unit-ℓ_2^2 representation *if all its vectors have unit length.*

REMARK 1 Equivalently, one can say that the unit-ℓ_2^2-representation associates a positive semidefinite $n \times n$ matrix M with the graph with diagonal entries 1 and $\forall i, j, k, M_{ij} + M_{jk} - M_{ik} \leq 1$. The vector representation $v_1, v_2, \ldots, v_n$ is the Cholesky factorization of M, namely $M_{ij} = \langle v_i, v_j \rangle$.

Every cut $(S, \overline{S})$ gives rise to a natural unit-ℓ_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.)

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

$$\forall i \, |v_i|^2 = 1 \tag{5}$$

$$\forall i, j, k \quad |v_i - v_j|^2 + |v_j - v_k|^2 \geq |v_i - v_k|^2 \tag{6}$$

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

This SDP motivates the following definition.

DEFINITION 2 *An ℓ_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).

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

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

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

As we mentioned before the SDPs subsume both the eigenvalue approach and the Leighton-Rao approach [14]. 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 ℓ_2^2 representations

In general, ℓ_2^2-representations are not well-understood[1]. This is not surprising since in $\Re^d$ the representation can have at most 2^d distinct vectors [9], 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 ℓ_2^2 representations. Note that

[1] A well-known —but alas, wide-open—conjecture says that they are closely related to the better-understood ℓ_1 metrics.

we assume the dimension $d \gg \log n$; this is without loss of generality since we could always embed the vectors in a higher dimensional space.

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

THEOREM 1 (MAIN)
For every $c > 0$, any c-spread unit-ℓ_2^2 representation with n points contains Δ-separated subsets S, T of size $\Omega(n)$, where $\Delta = \Omega(1/\sqrt{\log n})$. Furthermore, there is a randomized polynomial-time algorithm for finding these subsets S, T.

REMARK 2 The natural embedding of the boolean hypercube $\{-1, 1\}^d$ (appropriately scaled) shows that this theorem is tight to within a constant factor. This follows from the isoperimetric inequality for hypercubes.

2.1.1 Immediate corollary: $\sqrt{\log n}$-approximation

Let c' be the constant in the $\Omega(n)$ bound on the sizes of sets S and T in theorem 1. Let $W = \sum_{\{i,j\} \in E} |v_i - v_j|^2$ be the optimum value for the SDP defined by equations (4)–(7). (Specifically, it is the objective scaled by 4.) 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 whose expansion is $O(\sqrt{\log n} W/n)$.

COROLLARY 2
There is a randomized polynomial-time algorithm that finds with high probability a cut $(S_{obs}, \overline{S_{obs}})$ that is c'-balanced, and has expansion $\alpha_{obs} = O(W\sqrt{\log n}/n)$.

PROOF: We use the algorithm of Theorem 1 to produce Δ-separated subsets S, T for $\Delta = g/\sqrt{\log n}$. Let V_0 denote the vertices whose vectors are in S. Associate with each edge $e = \{i, j\}$ a length $w_e = |v_i - v_j|^2$. (Thus $W = \sum_{e \in E} w_e$.) In the rest of the proof "distance" in the graph is measured with respect to this length function.

Denote by V_s the set of vertices within distance d of V_0 and by E_s the set of edges leaving V_s. We do breadth-first search and find $s \leq \Delta/2$ that minimizes $|E_s| / |V_s|$. We output the cut $(V_s, \overline{V_s})$; let α_{obs} denote the expansion of this cut. (We can assume without loss of generality that $|V_{\Delta/2}| \leq n/2$, since we could switch S and T otherwise.)

For any $s \leq \Delta/2$, we have

$$|E_s| \geq \alpha_{obs}|V_s| \geq \alpha_{obs}c'n,$$

since $|V_0| = |S| \geq c'n$.

The total length of the edges $W = \sum_e w_e$, is thus at least $\Delta/2$ times the minimum number of edges crossing at any point along this length $\Delta/2$ interval. More formally, the total length of the edges

$$W = \sum_e w_e \geq \int_{s=0}^{\Delta/2} |E_s| \, ds \geq \tfrac{\Delta}{2} \cdot \alpha_{obs}c'n.$$

The corollary follows by solving for α_{obs}. □

3. $\Delta = \Omega(\log^{-2/3} N)$-SEPARATED SETS

We now give an algorithm that given a c-spread ℓ_2^2 representation finds Δ-separated sets of size $\Omega(n)$ for $\Delta = \Theta(1/\log^{2/3} n)$. Our correctness proof assumes a key theorem (Theorem 5) whose proof appears in Section 4. The algorithm will be improved in Section 5 to allow $\Delta = \Theta(1/\sqrt{\log n})$.

The algorithm is given a c-spread ℓ_2^2-representation. We select constants $c', \sigma > 0$ depending on c.

> SET-FIND:
> Input: A c-spread unit vector representation $v_1, v_2, \ldots, v_n \in \Re^d$.
> Parameters: Desired separation Δ, desired balance c', and projection gap, σ.
>
> Pick a random line u passing through the origin, and let
>
> $$S_u = \{v_i : \langle v_i, u\rangle \geq \frac{\sigma}{\sqrt{d}}\},$$
>
> $$T_u = \{v_i : \langle v_i, u\rangle \leq -\frac{\sigma}{\sqrt{d}}\}.$$
>
> If $|S_u| < 2c'n$ or $|T_u| < 2c'n$, HALT , else proceed as follows. Pick any $v_i \in S_u, v_j \in T_u$ such that $|v_i - v_j|^2 \leq \Delta$, and delete i from S_u and j from T_u. Repeat until no such v_i, v_j can be found and output the remaining sets S, T.

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 [15]) to identify the sets S_u, T_u of vertices that project far apart. It then prunes these sets to find sets S, T.

Notice that if SET-FIND does not HALT prematurely, it returns a Δ-separated pair of sets. Thus, we need to show that in SET-FIND often both S_u and T_u are larger than $2c'n$ *and* that no more than $c'n$ points are deleted from S_u and T_u. The first claim is relatively easy, and we show this in the next subsection. Analysing the deletion process is much harder and forms the bulk of the paper. We state the formal claims about the process in the following subsestion, and proved it in Section 4.

3.1 Projection and S_u, T_u

We first remind the reader that in $\Re^d$, the projection of any unit vector on a random direction is distributed essentially like a Gaussian with expectation 0 and standard deviation $1/\sqrt{d}$.

LEMMA 3 (GAUSSIAN BEHAVIOR OF PROJECTIONS)
If v is a vector of length ℓ in $\Re^d$ and u is a randomly chosen unit vector then (i) for $x \leq 1$, $\Pr[|\langle v, u\rangle| \leq \frac{x\ell}{\sqrt{d}}] \leq 3x$. (ii) for $x \leq \sqrt{d}/4$, $\Pr[|\langle v, u\rangle| \geq \frac{x\ell}{\sqrt{d}}] \leq e^{-x^2/4}$.

If the projection length in a particular direction u is $t\ell/\sqrt{d}$, we say that t is the *stretch* of v in direction u. (This definition is motivated by the fact that $\ell/\sqrt{d}$ is the root mean square of the projection length of v in a random direction.) Lemma 3-(ii) implies that a vector v has stretch t in a random direction u with probability at most $e^{-t^2/4}$. We will this use the notion of stretch and this fact extensively in subsequent sections.

Now using part (i) of Lemma 3 and Goemans-Williamson, it is easy to prove that if the v_i's are c-spread then with constant probability, S_u and T_u are large.

LEMMA 4
For every positive $c < 1/3$, there are $c', \sigma > 0$ such that the probability (over the choice of u) is at least $c/8$ that the sets S_u, T_u defined in SET-FIND-(c', σ) *have size at least $2c'n$.*

PROOF: Goemans-Williamson show that for any two points x, y on a unit sphere,

$$\Pr[\text{a random hyperplane separates } x, y] \geq .878 \frac{|x-y|^2}{4}.$$

By definition of c-spread, the sum of the distances between the points is at least $c(1-c)n^2$. Therefore the expected number of pairs that are separated by a random hyperplane is at least an^2, where $a = .878c(1-c)$. By Markov's bound the probability that the number of separated pairs is less that $an^2/2$ is at most $(1-a)/(1-a/2) \leq 1 - a/2$.

Since these $an^2/2$ pairs of nodes are split by the hyperplane, there must be at least $an/2$ nodes on the smaller side.

By Lemma 3-(i), the probability that the projection of a point on the unit sphere falls within $\sigma/\sqrt{d}$ of the origin is at most 3σ. By choosing σ appropriately, and applying the Markov bound, we can ensure that the probability that more than $an/4$ points fall within $\sigma/\sqrt{d}$ is at most $a/4$. Now by the union bound both S_u and T_u have at least $a/4$ points with probability at least $a/4$.

The lemma follows by noting that $a/4 > c/8$, and choosing $2c' = a/4$. $\square$

3.2 Number of deletions

To analyse the number of deletions, we note that any deleted pair $v_i \in S_u, v_j \in T_u$ is such that the vector $v_i - v_j$ has stretch $t = 2\sigma/\sqrt{\Delta}$, since its length was at most $\sqrt{\Delta}$ and its projection length was at least $2\sigma/\sqrt{d}$. (Note: From now on, we will often say the pair v_i, v_j has a certain stretch when we mean $v_i - v_j$.) If Δ were $(16\sigma^2/\log n)$ then the analysis would be trivial since any deleted pair has stretch at least $4\sqrt{\log n}$; this event occurs with probability less than $e^{-4\log n} \ll 1/n^2$ by Lemma 3-(ii). Thus, we expect *no* pairs to be deleted. (Aside: This is an alternative version of Leighton-Rao.)

When $\Delta = \Omega(\log^{-2/3} n)$, it may be quite likely that many pairs are deleted. However, we observe that for a direction u the deleted pairs form a matching M_u. Moreover, if the procedure fails for a direction u the matching M_u is of size at least $c'n$. Thus if the procedure does not succeed with constant probabilty, we have large matchings M_u for most directions u where each matching edge has stretch $2\sigma/\sqrt{\Delta}$. We will show (Theorem 5) that this is impossible. Now we formalize the property of the matchings when SET-FIND often fails to produce a $\Delta = 1/t^2$-separated pair.

DEFINITION 4 ((t, γ, β)-STRETCHED) *An ℓ_2^2 set of points $v_1, v_2, \ldots, v_n \in \Re^d$ are (t, γ, β)*-stretched at scale l *if for at least γ fraction of directions u, there is a (partial) matching M_u with βn disjoint pairs $(i_1, j_1), (i_2, j_2), \ldots$, such that each i_m, j_m satisfies $|v_{i_m} - v_{j_m}|^2 \leq l^2$ and $\langle u, (v_{i_m} - v_{j_m})\rangle \geq tl/\sqrt{d}$. (In particular, pair v_{i_m}, v_{j_m} has stretch at least t in direction u.)*

THEOREM 5
For any $\gamma, \beta > 0$ there is a $C = C(\gamma, \beta)$ such that if $t > C(\log n)^{1/3}$ then a unit-ℓ_2^2 representation cannot be (t, γ, β)-stretched for any scale l.

Applying Theorem 5 with $l = \sqrt{\Delta}$ and $t = 2\sigma/\sqrt{\Delta}$ shows that there is some $\Delta = O(\log^{-2/3} n)$, such that the probability that SET-FIND removes a matching of size $c'n$ is $o(1)$. We conclude that SET-FIND outputs S, T of size $\geq c'n$ with probability $\Omega(1)$. This completes our analysis of SET-FIND.

4. PROOF OF THEOREM 5

The main idea in the proof is to show that if the large matchings M_u mentioned in Definition 4 exist for most directions u, then for $\Omega(1)$ fraction of directions we can string together $r = \Omega(t)$ pairs from these matchings to produce a vector whose projection is $\Omega(rtl/\sqrt{d})$. Triangle inequality implies that any such vector has squared length at most rl^2, which means that the stretch is is $\sqrt{r}t$.

Recall that for almost all directions, no pair of vectors has stretch more than $4\sqrt{\log n}$. Since the stringing together referred to above is possible for $\Omega(1)$ directions, we conclude that the stretch $O(\sqrt{r}t)$ cannot exceed $4\sqrt{\log n}$, which proves that t can only be $O(\log^{1/3} n)$.

4.1 Matching covers

The definition of (t, γ, β)-stretched pointsets suggests that for many direction there are many disjoint pairs of points which are stretched. We will work with a related notion.

DEFINITION 5 ((ϵ, δ)-MATCHING-COVERED POINT SET) *A set of points $V \subseteq \Re^d$ is (ϵ, δ)*-matching-covered at scale *l if for every unit vector $u \in \Re^d$, there is a (partial) matching M_u of V such that every $(v_i, v_j) \in M_u$ satisfies $|v_i - v_j|^2 \leq l^2$ and $|\langle u, v_i - v_j\rangle| \geq \epsilon$, 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.

LEMMA 6
If a set of n vectors is (t, γ, β)-stretched at some scale l, then they contain a subset X of $\Omega(n\gamma\beta)$ vectors that are (ϵ, δ)-matching covered at scale l, where $\delta = \Omega(\beta\gamma)$, $\epsilon \geq tl/\sqrt{d}$, and for every pair v_i, v_j in the matching cover, $|v_i - v_j|^2 \leq l^2$.

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 $\gamma\beta/2$ measure of directions (and remove the corresponding matched edges from the M_u's). Repeat. The aggregate measure of directions removed is $\gamma\beta n/2$. Thus at least $\gamma\beta n/2$ 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. $\square$

NOTATION From now on we restrict attention to the subset X mentioned in Lemma 6. Let H denote the multigraph on X formed by taking the union of all matchings M_u in the matching cover. For each $v_i \in X$, let $\text{Ball}(v_i, r)$ denote the set of v_j's whose distance from v_i in H is at most r.

We sometimes say that "v_j is r matching hops from v_i." Note that since the matching cover consists of edges of length $\leq l$, the triangle inequality for ℓ_2^2 representations implies that any such v_j, v_i satisfy $|v_i - v_j|^2 \leq rl^2$.

Now we define a related object.

DEFINITION 6 ((ϵ, δ)-COVER) *A set $\{w_1, w_2, \dots, \}$ of vectors in $\Re^d$ is an (ϵ, δ)-cover if every $|w_j| \leq 1$ and for at least δ fraction of unit vectors $u \in \Re^d$, there exists an j such that $\langle u, w_j \rangle \geq \epsilon$.*

REMARK 5 Since $-u$ is also a random unit vector, the probability is also $\geq \delta$ that there is a w_j that $\langle u, w_j \rangle \leq -\epsilon$. This will be important later in Lemma 10.

REMARK 6 Whenever we study these covers, we have a fixed $v_i \in X$ in mind and the vectors in the cover are of the form $v_j - v_i$. In such a case, we say that v_i is "centrally (ϵ, δ)-covered" by the v_j's in question.

Note that if $v_i \in X$, then the vectors $v_j - v_i$ for $v_j \in \text{Ball}(v_i, 1)$ form an (ϵ, δ)-cover. (The converse is not true: taking the union of such (ϵ, δ)-covers may not always give a matching cover.)

NOTATION: Let $S_r \subseteq X$ consist of all $v_i \in X$ such that the vectors $\{v_j - v_i : v_j \in \text{Ball}(v_i, r)\}$ form an $(\epsilon r/2, 1 - \delta/2)$-cover.

Note that thus far there is no reason to believe even that S_1 is nonempty, since we only know that for each $v_i \in X$, the set $\{v_j - v_i : v_j \in \text{Ball}(v_i, 1)\}$ is an (ϵ, δ)-cover, whereas in order for v_i to be in S_1 these vectors must form an $(\epsilon/2, 1 - \delta/2)$-cover.

Thus the main technical step is the following. It assumes that stretch of the edges in the matching cover, namely, $t = \epsilon\sqrt{d}/l$, is larger than some fixed constant.

LEMMA 7 (MAIN)
(i) $S_1 = X$. (ii) There are constants $\eta = \eta(\delta), \rho = \rho(\delta)$ such that for $r \leq \eta t$, we have $|S_{r+1}| \geq \rho |S_r|$.

Theorem 5 follows immediately from Lemma 7.

PROOF:(Theorem 5) If the hypothesis of Theorem 5 is true, then Lemma 6 implies the existence of a set X of $\Omega(n)$ vectors v_i's that form an (ϵ, δ)-matching covered point set using edges of squared length at most l^2. Here $\epsilon = tl/\sqrt{d}$. Then Lemma 7 and a simple induction implies that for $r = \eta t$,

$$|S_r| \geq \rho^{r-1} |S| = \Omega(\rho^{r-1} n) \gg 1,$$

where we're using the fact that $r = o(\log n)$. Thus S_r is nonempty.

Let $v_i \in S_r$. Then for at least $1 - \delta/2$ fraction of directions u, some $v_j \in \text{Ball}(v_i, r)$ satisfies $|\langle v_i - v_j, u \rangle| \geq r\epsilon/2$. However, $|v_j - v_i| \leq \sqrt{r}l$, so we conclude that the stretch of $v_j - v_i$ is

$$\frac{r\epsilon/2 \times \sqrt{d}}{\sqrt{r}l} = \sqrt{r}t/2 = \sqrt{\eta}t^{3/2}/2 = \Omega(t^{3/2}).$$

But recall that for any set of n vectors, at most $1/n$ of the directions u are such that one of the $\binom{n}{2}$ pairs of vectors has stretch $> 4\sqrt{\log n}$. But since S_r is nonempty, we know that the probability is at least $1 - \delta/2$ that some stretch exceeds $\Omega(t^{3/2})$. We conclude that $t = O(\log^{1/3} n)$. $\square$

4.2 Proving Lemma 7

We prove Lemma 7 by induction. Recall that it was unclear even that S_1 is nonempty. In fact a phenomenon called *measure concentration* implies that $S_1 = X$. We first introduce this idea.

4.2.1 Measure concentration.

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

LEMMA 8 (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 ([5]) states that $\mu(A_\gamma)/\mu(A)$ is minimized for spherical caps[2] 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 is an immediate corollary.

LEMMA 9
Let $\{v_1, v_2, \dots, \}$ be a finite set of vectors that is an (ϵ, δ)-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. Thus by Lemma 8, $\mu(A_\gamma) \geq 1 - \exp(-t^2/2)$.

We argue that for each direction u in A_γ, there is a vector v_i in the (ϵ, δ) 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 θ is the angle between v_i and u. The projection length of v_i on u' is $|v_i| \cos\theta'$ where θ' is the angle between v_i and u'. The angle formed by u and u' is at most 2γ since $\alpha \leq 2\sin\alpha$, so $\theta - 2\gamma \leq \theta' \leq \theta + 2\gamma$. Since the absolute

[2]Levy's isoperimetric inequality is not trivial; see [27] for a sketch. However, results qualitatively the same — but with worse constants— as Lemma 8 can be derived from the more elementary Brunn-Minkowski inequality; this "approximate isoperimetric inequality" of Ball, de Arias and Villa also appears in [27].

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 $1 - \exp(-t^2/2)$. □

We can thus use Lemma 9 to boost δ to almost 1. However, we choose to do it only sometimes, not always. This explains the slightly strange hypothesis in the next Lemma. The proof of Lemma 7 will not use this lemma *per se* but will use the argument in it.

LEMMA 10
If $\{w_1, w_2, \ldots, w_k\} \subseteq \Re^d$ is an $(\epsilon_1, 1-\delta_1)$-cover and $\{w'_1, w'_2, \ldots, w'_l\}$ is an (ϵ_2, δ_2)-cover then the set $\{w_e - w'_f : 1 \leq e \leq k, 1 \leq f \leq l\}$ is a $(\epsilon_1+\epsilon_2, \delta_2 - \delta_1)$-cover.

PROOF: Let $u \in \Re^d$ be a random unit vector. The probability is at least $1-\delta_1$ that there is a w_e such that $\langle u, w_e\rangle \geq \epsilon_1$. The probability is at least δ_2 that there is a w'_f such that $\langle u, w'_f\rangle \leq -\epsilon_2$. Thus with probability at least $\delta_2 - \delta_1$, there exist w_e, w'_f such that $\langle u, w_e - w'_f\rangle \geq \epsilon_1 + \epsilon_2$. □

4.2.2 Proof of Lemma 7

Let $\sigma = \epsilon\sqrt{d}$, and assume the stretch $t = \sigma/l$ of the matching edges is larger than any desired constant.

We set $D(\sigma,\delta) = 8\sqrt{2\log(2/\delta)}/\sigma$, and $\rho(\delta) = \delta/4$.

First, we show $S_1 = X$. The hypothesis implies that every $v_i \in X$ is centrally (ϵ,δ)-covered by the set of $v_j \in \text{Ball}(v_i, 1)$. We apply Lemma 9 to each of these (ϵ,δ)-covers with $\gamma = \sigma/4l\sqrt{d}$. Note that

$$\gamma = \sigma/4l\sqrt{d} > \sqrt{2\log(2/\delta)}/\sqrt{d} + \sqrt{2\log(2/\delta)}/\sqrt{d},$$

so we conclude that v_i is also $(1-\delta/2, \epsilon - 2\gamma\ell)$ covered by $\text{Ball}(v_i, 1)$. Since $2\gamma\ell < \sigma/2\sqrt{d} \leq \epsilon/2$, we have thus shown that every $v_i \in X$ is also in S_1.

Assume the induction has worked for r steps and there is a set $S_r \subseteq T$ satisfying $|S_r| \geq \rho^{r-1}|X|$ such that every point $v_i \in S_r$ is centrally $(\epsilon_r, 1-\delta_0/2)$-covered by the vectors in $\text{Ball}(v_i, r)$, where $\epsilon_r \geq 0.5r\epsilon$.

For each $v_i \in S_r$ consider the set of all vectors $v_j - v_k$ where $v_j \in \text{Ball}(v_i, r)$ and $v_k \in \text{Ball}(v_i, 1)$. Lemma 10 implies that these vectors form an $(\epsilon_r + \epsilon, \delta/2)$ cover, but unfortunately this is no longer centered at v_i. Thus we are unable to prove in general that $v_i \in S_{r+1}$.

Instead, we argue differently and use an averaging argument to say that if S_r is large, so is S_{r+1}. Let $v_i \in S_r$. For $1-\delta/2$ fraction of directions u, there is a point $v_j \in \text{Ball}(v_i, r)$ such that $\langle v_j - v_i, u\rangle \geq \epsilon_r$. Also for δ fraction of directions u, there is a point in $v_k \in \text{Ball}(v_i, 1)$ such that $\langle v_k - v_i, u\rangle \leq -\epsilon$ and v_k is matched to v_i 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 $v_j \in \text{Ball}(v_k, r+1)$, 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 $(\epsilon_r + \epsilon, \rho^r\delta/2)$-cover.

For each point $v_i \in S_r$, 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 $\delta|S_r|/2$ vectors get assigned it by the process. Equivalently, for a random point in X, the expected measure of directions for which the point is assigned a vector is at least $\delta|S_r|/2\,|X|$. 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 $\delta\,|S_r|\,/4\,|X|$ fraction of the points in X must be assigned a vector for $\delta\,|S_r|\,/4\,|X|$ fraction of the directions.

We will define all such points of X to be the set S_{r+1} and note that for $\rho = \delta/4$ the size is at least $\rho\,|S_r|$ as required. However, we have to show another property for S_{r+1}. Thus far, since $\delta\,|S_r|\,/4\,|X| = \rho^r$, we have only shown that each point v_k in S_{r+1} is centrally $(\epsilon_r + \epsilon, \delta\rho^r)$-covered by $v_j \in \text{Ball}(v_k, r+1)$.

We now invoke measure concentration to show that these centered covers are also centered $(\epsilon_r + \epsilon/2, 1-\delta/2)$ covers, so long as $r = O(\sigma/l)$. Note that the vectors in the cover have squared length at most rl^2 due to the triangle inequality on the squared lengths. We apply Lemma 9 with $\ell = \sqrt{r}l$ and

$$\gamma = \epsilon/4\ell = \sigma/4\sqrt{d}\sqrt{r}l.$$

Now, we need

$$\gamma = \frac{\sigma}{4\sqrt{d}\sqrt{r}l} > \frac{2\sqrt{\log\frac{2}{\rho^r}} + \sqrt{\log(\frac{2}{\delta})}}{\sqrt{d}}, \tag{11}$$

to get that v_k is centrally $(\epsilon_r + \epsilon - 2\gamma\ell, 1-\delta/2)$ covered. The condition is satified when $r \leq \sigma/8l\sqrt{2\log(8/\delta)}$, since $\rho = \delta/4$.

By noting that $2\gamma\ell < \epsilon/2$, we now observe that each $v_k \in S_{r+1}$ is $((r+1)\epsilon/2, 1-\delta/2)$-covered by $v_j \in \text{Ball}(v_k, r+1)$ and our induction is complete. □

Now we state a corollary of the proof of Lemma 7 which will be useful in Section 5. As in Lemma 7, let $X \subseteq \Re^d$ be a pointset that is (ϵ,δ)-matching-covered using edges of squared length at most l^2.

The Corollary concerns, for some $s > 0$, a subset $T \subseteq X$ of size at least $|X|/2$ containing every v_i such that the set $\{v_j - v_i : |v_i - v_j|^2 \leq s\}$ is an $(\epsilon_1, 1-\delta/2)$-cover. Define T' to be the set of v_i's such that the set $\{v_j - v_i : |v_i - v_j|^2 \leq s + l^2\}$ is an $(\epsilon_1 + \epsilon/2, 1-\delta/2)$-cover. The corollary assumes $\epsilon\sqrt{d}$ is some constant, say σ.

COROLLARY 11 (COVER COMPOSITION)
There are constants ρ, f depending only on σ, δ such that if $s + l^2 \leq f$, then $|T'| \geq \rho|X|$.

PROOF: Straightforward from proof of Lemma 7; left to the reader. □

5. ACHIEVING $\Delta = \Omega(1/\sqrt{\log N})$.

To prove Theorem 1 with $\Delta = \Omega(1/\sqrt{\log n})$, we start by invoking SET-FIND with that Δ as separation parameter. If SET-FIND succeeds, we are done. Otherwise, as before we end up with matchings M_u in most directions u. Now, however, we cannot necessarily show that this leads to a contradiction. The bottleneck in our previous proof lies in the induction of Lemma 7, where the size of the set S_r decreases geometrically with r. To bypass that, we describe another algorithm finds a Δ-separated set. This uses the simple observation that if a point is well covered than all points that are close to it are also well covered. Now we formalize this.

LEMMA 12 (COVERING CLOSE POINTS)
Suppose $v_1, v_2, \ldots \in \Re^d$ are vectors such that for some point $v_0 \in \Re^d$ the vectors $v_1 - v_0, v_2 - v_0, \ldots$, form an (ϵ, δ)-cover. Then for every point v_0' such that $|v_0 - v_0'| = s$, the vectors $v_1 - v_0', v_2 - v_0', \ldots$ form an $(\epsilon - \frac{ts}{\sqrt{d}}, \delta - e^{-t^2/4})$-cover.

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

Armed with this lemma, we construct l^2-separated sets as follows. Our algorithm is very much like the inductive step in the proof lemma 7.

For each r, define S_r' to be the set of points v which is $(r\epsilon/4, 1 - 3\delta/4)$-covered by points that are within length $\ell_r = \sqrt{2r}/l$ of v. Consider the smallest r such that S_{r+1}' has cardinality less than $n/2$ (we argue below that such an r exists). We apply Corollary 11 to S_r', to get a set of $\delta n/4$ points T where each point $v \in T$ is $((r/4+1/2)\epsilon, 1-\delta/2)$ covered by points that are within squared Euclidean length $\ell_r^2 + l^2$ of v. It will follow using Lemma 12 that all points within length l of T are in S_{r+1}, and therefore the $\Omega(n)$ sized sets T, $S_1 - S_{r+1}$ are at least l^2-separated.

To argue correctness, we will show that for every $r \leq r_0$ where r_0 is $\theta(1/l^2)$, Corollary 11 implies $|T| \geq \rho n$ and Lemma 12 implies that the l-neighborhood of T is contained in S_{r+1}'. Now for some l which is $\theta(1/\log^{1/4} n)$ we argue that $|S_{r_0}'| = 0$. Recall that $v \in S_{r_0}'$ only if for most directions it participates in a stretched pair of points with stretch $\Omega(\sqrt{r_0}/l) = \Omega(1/l^2) = \Omega(\sqrt{\log n})$. In fact, for most directions there are no such stretched pairs.

To finish we argue that r_0 is $\theta(1/l^2)$. To use Lemma 12 as above, we need to ensure that the loss in projection $ts/\sqrt{d}$ (in our context $s = l$) is at most $\epsilon/4$, and that the loss in probability $e^{-t^2/4}$ is at most $\delta/4$. Choosing $t = 2\sqrt{\log 4/\delta}$, we see that l must be less than $\epsilon\sqrt{d}/t$ which is easily satisfied for sufficiently large n. Moreover, for some contant f, Corollary 11 can be applied as long as $\ell_r^2 = 2rl^2 \leq f$. This implies that the upper limit for r is $\theta(1/l^2)$.

REMARK 7 Simplifying slightly, here is a concrete algorithm.

For each r, find a set $\tilde{S}_r'$ where each point in $\tilde{S}_r'$ is approximately $(r\epsilon/4, 1 - 3\delta/4)$-covered by points that are within length $\sqrt{2r}/l$. This can be done by sampling $O(\log n)$ directions. For the first $\tilde{S}_r'$ with cardinality less than $|\tilde{S}_1'|/2$, we take the set $\tilde{S}_1' - \tilde{S}_r'$ to be S and take T to be all the points that are at least l^2 from S.

The sets S and T are l^2-separated by construction and the argument above shows that each set is large in the event SET-FIND usually fails.

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

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

THEOREM 13
There is a polynomial-time algorithm that, given a feasible SDP solution with value β, produces a cut $(S, \overline{S})$ satisfying $|E(S, \overline{S})| = O(\beta\,|S|\,n\sqrt{\log n})$.

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 14
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 $1/4n^2$ around v_k contains at least $n/2$ vectors, the algorithm finds a cut $(S, \overline{S})$ with expansion at most $O(\beta n)$.

PROOF: Let X be the subset of nodes that correspond to the vectors in the $1/4n^2$-geometric ball around v_k. Let $d(i,j) = |v_i - v_j|^2$ and when $\{i,j\}$ is an edge e we write $d(e)$.

To find a small cut, we shrink X to a point and perform a breadth first search from it traversing each outgoing edge e in time $d(e)$ (i.e. a bfs in the weighted graph). Let α_{obs} be the minimum expansion of any cut induced by this growing boundary. We will show that $\alpha_{obs} = O(\beta n)$. First we observe that when this bfs has grown to include all points within s of X, the number of points beyond the boundary, $N(s)$, is less than $n/2$. Therefore the number of edges crossing this cut is at least $\alpha_{obs} N(s)$. The total edge weight seen in this process is bounded by the sum of the weights of all edges and so by the lemma's hypothesis we obtain:

$$\int_{s>0}^{1} \alpha_{obs} N(s) \;\; ds \leq \beta.$$

We now show that there $\int_{s>0}^{l} N(s)$, which is the total distance of all vertices to X, is larger than $1/4n$, which implies that $\alpha_{obs} \leq 4n\beta$. First observe that since $\sum_{i<j} d(i,j) = 1$, the triangle inequality implies that node k also satisfies: $\sum_j d(k,j) \geq 1/2n$. Let $p(j)$ be the first point in X on the shortest path from j to k. Then,

$$\sum_{j \notin X} d(j, X) = \sum_j d(j,k) - \sum_j d(p(j), k) \geq \frac{1}{2n} - \frac{n}{4n^2} = \frac{1}{4n}.$$

$\square$

Thus we only need to consider the case where the hypothesis of Lemma 14 does not hold for any k. Namely, for each node k, less than $n/2$ vectors lie within a ball of squared radius less than $1/4n^2$. That is, the nodes are well spread out. Under this condition, the ideas from Corollary 2 and Section 5 can be used to produce c'-balanced cuts (where c' is some constant) of expansion $O(\beta n)$, thus showing that the integrality gap is $O(\sqrt{\log n})$. Now we sketch how this is done.

First, scale all vectors by $2n$ so that the squared-length of $1/4n^2$ becomes 1 and $\sum_{i<j} |v_i - v_j|^2 = 4n^2$. Now any sphere of radius 2 contains at most 1/2 the points. Furthermore, averaging shows that at least 9/10 fraction of points lie inside a sphere of radius 40. Thus $\Omega(1)$ fraction of nodes lie in a spherical annulus of inner radius 1 and outer radius 40. A version of Theorem 1 applies to such representations with constants appropriately modified for the diameters of the ball. As stated, the Theorem assumed the vector representation of the graph involves unit vectors, but looking over the proofs it is clear that the proofs go through (with the constants not as good) if 9/10 of the vectors have length lower-bounded and upperbounded by some constant. The reason is that by using the Goemans-Williamson analysis as before, we conclude that for some c', σ, the algorithm SET-FIND-(c', σ) outputs a c'-balanced cut with probability $\Omega(1)$. Then, the remainder of the proof just uses an upper bound on the vector length in various places.

7. EXPANDER FLOWS

7.1 Multicommodity flows

A *multicommodity flow* in an unweighted graph $G = (V, E)$ is an assignment of a *demand* $f_{ij} \geq 0$ to each node pair i, j such that we can route f_{ij} units of flow from i to j, and can do this simultaneously for all pairs while minimizing the maximum congestion β on any edge. Note that every multicommodity flow in G can be viewed as an *embedding* of a weighted graph $G' = (V, f_{ij})$ on the same vertex set such that the edge $\{i, j\}$ is f_{ij} fractionally present. Thus, the degree of node i is $\sum_{ij} f_{ij}$. If G' is a constant degree fractional expander (which therefore has constant conductance as well) then clearly the expansion of G is $\Omega(1/\beta)$.

In these terms, the Leighton-Rao approach embeds a K_n where each edge is $1/n$ fractionally present. They show that one can do this with congestion $\beta = O(\log n/\alpha)$, which gives a $O(\log n)$ approximate certificate for expansion. Moreover, they show this is tight in the case G is itself an expander.

Our original goal was to prove that for any G, there is a constant degree fractional expander G' that can be embedded with $\beta = O(1/\alpha)$. That is still open, but we sketch a proof of the following weaker theorem.

THEOREM 15
Given any graph G, there is a fractional constant degree expander that can be embedded in G with congestion $\beta = O(\sqrt{\log n}/\alpha(G))$.

For simplicity, in the following we will limit ourselves to certifying c-balanced $\alpha_c(G)$, by embedding a fractional graph (V, f_{ij}) where every c-balanced cut has good expansion. We call such a graph a c-balanced expander.

7.2 Embeddings and Expander flows

NOTATION Let $\mathcal{F}_\beta$ be the set of fractional graphs (V, f_{ij}) that can be embedded into G with congestion β. Let $\mathcal{V}$ the set of c-spread unit-ℓ_2^2 representations on n points. Recall that c-balanced cuts may be viewed as c-spread unit-ℓ_2^2 representations

Thus to prove that a fractional constant degree c-balanced expander can be embedded in the graph G with congestion $\beta = \Omega(\sqrt{\log n}/\alpha)$ it suffices to prove that

$$\max_{(V,f_{ij})\in\mathcal{F}_\beta} \min_{(v_1,\ldots,v_n)\in\mathcal{V}} \sum_{ij} f_{ij} |v_i - v_j|^2 = \Omega(n).$$

By observing that $\mathcal{V}$ is a convex set and the payoff function can be written as a linear function (proof deferred to full paper), von Neumann's min-max principle shows that the next statement is equivalent to the previous one.

$$\min_{(v_1,\ldots,v_n)\in\mathcal{V}} \max_{(V,f_{ij})\in\mathcal{F}_\beta} \sum_{ij} f_{ij} |v_i - v_j|^2 = \Omega(n).$$

To prove the previous statement, we consider any choice of $v_1, v_2, \ldots, v_n$. The best such fractional graph (V, f_{ij}) is the solution to a linear program since $(v_1, \ldots, v_n)$ is fixed. To show that the optimum value of the LP is $\Omega(n)$ we consider its dual. Some manipulations (that appear in the complete version) show that it suffices to prove, for every graph H whose every c-balanced cut has expansion $\Omega(\alpha)$, that there exist $\Omega(n)$ pairs v_i, v_j for which $||v_i - v_j||^2 = \Omega(1)$ while $d_H(i,j) = O(\sqrt{\log n}/\alpha)$.

To prove the existence of such pairs, we follow the proof of our $O(\sqrt{\log n})$ result in Section 5. That proof could obtain points v_i and v_j where $|v_i - v_j|^2 = \Omega(1)$ by piecing together matched pairs from M_u's and from a ball growing procedure. A max-flow based argument shows that $d_H(x,y) = O(1/\alpha)$ for a constant fraction of the pairs $(x,y) \in M_u$ for any direction u. Moreover, the task accomplished by growing geometric balls in Lemma 12 can instead be accomplished by growing by $O(1/\alpha)$ in H. Piecng these pairs together at most $O(\sqrt{\log n})$ steps will find a pair v_i, v_j with $|v_i - v_j|^2 = \Omega(1)$, and $d_H(i,j) = O(\sqrt{\log n}/\alpha)$. Repeating $\Omega(n)$ times produces the desired result.

8. CONCLUSIONS

At the beginning of this project, we conjectured that it was possible to route expander flows with congestion $O(1/\alpha)$. This would imply that the integrality gap of the SDP would be $O(1)$. In the full version of the paper we present several conjectures which provide a roadmap for proving these results. Note that for hypercubes and related graphs, our rounding algorithm produces cuts whose value is $O(\sqrt{\log n})$ times the SDP value. Thus, a different rounding algorithm seems necessary.

We note that a conjecture of Goemans and Linial that ℓ_2^2 metrics can be embedded into ℓ_1 with constant distortion also implies a constant upper bound for the integrality gap of the SDP. (In fact it proves such a bound for the SDP for a more general version of SPARSEST

CUT.) As noted, our geometric results may be a starting point for making some progress on proving that ℓ_2^2 metrics can be embedded into ℓ_2 with $O(\sqrt{\log n})$ distortion. This would solve the long open question of whether ℓ_1 can be embedded into ℓ_2 with $O(\sqrt{\log n})$ distortion.

Our approximation algorithms are fairly inefficient (though polynomial time) because they use SDPs or related convex optimization. Do more efficient (combinatorial?) algorithms exist, possibly using expander flows? One loose analogy would be combinatorial versions of the Leighton-Rao multicommodity flow algorithm (see, e.g., [26, 12]), which may be useful in practice. Many practitioners continue to prefer eigenvalue methods over Leighton-Rao because the geometric meaning of eigenvalues (e.g., the connection to stretched rubberbands and such) has relevance in their application —computer vision, for example. Since the SDP relaxation may be viewed as a higher-dimensional analogue of eigenvalue computation, it may well turn out to share these properties of eigenvalues, and hence also their practical appeal.

Extending our ideas to other problems should be possible though it doesn't seem to be immediate. The problems in [29] would be a good list to try, especially minimum multicut, for which an $O(\log k)$-approximation was designed in [13].

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): 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.

9. REFERENCES

[1] F. Alizadeh. Interior point methods in semi- definite 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. λ_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. Cheeger. A lower bound for the smallest eigenvalue of the Laplacian, in *Problem in Analysis*, 195-199, Princeton Univ. Press, (1970),

[8] F. Chung. Spectral graph theory. *CBMS Regional Conference Series in Mathematics*, 92, American Mathematical Society, 1997.

[9] 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.

[10] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible Markov chains. *Annals of Applied Probability*, **3**:696–730, 1993.

[11] U. Feige and R. Krauthgamer. A polylogarithmic approximation of the minimum bisection. In *IEEE FOCS 2001* pp 105–115.

[12] N. Garg and J. Köneman. Faster and Simpler Algorithms for Multicommodity Flow and other Fractional Packing Problems. In *IEEE FOCS 1997.*

[13] 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.*

[14] M.X. Goemans. Semidefinite programming in combinatorial optimization. *Math. Programming,* **79**:143–161, 1997.

[15] 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.

[16] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization.* Springer-Verlag, 1993.

[17] H. Karloff and U. Zwick. A 7/8-approximation algorithm for MAX 3SAT? *Proc. of 38th IEEE FOCS* (1997), 406-415.

[18] D. Karger, R. Motwani, and M. Sudan. Approximate graph coloring by semidefinite programming. *JACM*, **45**(2):246–265, 1998.

[19] M. Jerrum and A. Sinclair. Approximating the permanent. *SIAM J. Comput.*, **18**(6):1149-1178, 1989.

[20] 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.*

[21] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica* **15**:215–246, 1995.

[22] L. Lovász. On the Shannon capacity of a graph. *IEEE Trans. on Info. Theory* **IT-25**:1–7, 1979.

[23] A. Lubotzky, R. Philips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, **8**:261-277, 1988.

[24] J. Matousek. Lectures on Discrete Geometry. Springer Verlag, 2002.

[25] Y. Nesterov and A. Nemirovskii. *Interior point polynomial methods in convex programming.* SIAM, Philadelphia, PA 1994.

[26] 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.*

[27] 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.

[28] F. Shahrokhi and D.W. Matula. The maximum concurrent flow problem. *Journal of the ACM*, 37:318–334, 1990.

[29] 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.

[30] A. Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Prob., Comput.* **1**:351–370, 1992.

[31] V. Vazirani. Approximation algorithms. Springer Verlag, 2002.

[32] K. Zatloukal. *Personal communication,* November 2003.