On the optimality of the random hyperplane rounding technique for MAX CUT

Uriel Feige and Gideon Schechtman
Faculty of Mathematics and Computer Science
Weizmann Institute, Rehovot 76100, Israel
feige,gideon@wisdom.weizmann.ac.il

January 13, 2002

Abstract

MAX CUT is the problem of partitioning the vertices of a graph into two sets, maximizing the number of edges joining these sets. This problem is NP-hard. Goemans and Williamson proposed an algorithm that first uses a semidefinite programming relaxation of MAX CUT to embed the vertices of the graph on the surface of an n dimensional sphere, and then uses a random hyperplane to cut the sphere in two, giving a cut of the graph. They show that the expected number of edges in the random cut is at least $\alpha \cdot \mathrm{sdp}$, where $\alpha \simeq 0.87856$ and sdp is the value of the semidefinite program, which is an upper bound on opt, the number of edges in the maximum cut.

This manuscript shows the following results:

1. The integrality ratio of the semidefinite program is α. The previously known bound on the integrality ratio was roughly 0.8845.
2. In the presence of the so called "triangle constraints", the integrality ratio is no better than roughly 0.891. The previously known bound was above 0.95.
3. There are graphs and optimal embeddings for which the *best* hyperplane approximates opt within a ratio no better than α, even in the presence of additional valid constraints. This strengthens a result of Karloff that applied only to the expected number of edges cut by a random hyperplane.

1 The Algorithm of Goemans and Williamson

For a graph $G(V, E)$ with $|V| = n$ and $|E| = m$, MAX CUT is the problem of partitioning V into two sets, such that the number of edges connecting the two sets is maximized. This problem is NP-hard to approximate within ratios better than 16/17 [10]. Partitioning the vertices into two sets at random gives a cut whose expected number of edges is $m/2$, trivially giving an approximation algorithm with expected ratio at least 1/2. For many years, nothing substantially better was known. In a major breakthrough, Goemans and Williamson [8] gave an algorithm with approximation ratio of 0.87856. For completeness, we review their well known algorithm, which we call algorithm GW.

MAX CUT can be formulated as an integer quadratic program. With each vertex i we associate a variable $x_i \in \{-1, +1\}$, where -1 and $+1$ can be viewed as the two sides of the cut. With an edge $(i,j) \in E$ we associate the expression $\frac{1-x_i x_j}{2}$ which evaluates to 0 if its endpoints are on the same side of the cut, and to 1 if its endpoints are on different sides of the cut.

The integer quadratic program for MAX CUT:

Maximize: $\sum_{(i,j)\in E} \frac{1-x_i x_j}{2}$

Subject to: $x_i \in \{-1,+1\}$, for every $1 \le i \le n$.

This integer quadratic program is relaxed by replacing the x_i by unit vectors v_i in an n-dimensional space (the x_i can be viewed as unit vectors in a 1-dimensional space). The product $x_i x_j$ is replaced by an inner product $v_i v_j$. Geometrically, this corresponds to embedding the vertices of G on a unit n-dimensional sphere S_n, while trying to keep the images of vertices that are adjacent in G far apart on the sphere.

The geometric program for MAX CUT:

Maximize: $\sum_{(i,j)\in E} \frac{1-v_i v_j}{2}$

Subject to: $v_i \in S_n$, for every $1 \le i \le n$.

The geometric program is equivalent to a semidefinite program in which the variables y_{ij} are the inner products $v_i v_j$, and the n by n matrix Y whose i,j entry is y_{ij} is constrained to be positive semidefinite (i.e., the matrix of inner products of n vectors). The constraint $v_i \in S_n$ is equivalent to $v_i v_i = 1$, which gives the constraint $y_{ii} = 1$.

The semidefinite program for MAX CUT:

Maximize: $\sum_{(i,j)\in E} \frac{1-y_{ij}}{2}$

Subject to: $y_{ii} = 1$, for every $1 \le i \le n$,

and the matrix $Y = (y_{ij})$ is symmetric and positive semidefinite.

This semidefinite program can be solved up to arbitrary precision in polynomial time, and thereafter a set of vectors v_i maximizing the geometric program (up to arbitrary precision) can be extracted from the matrix Y (for more details, see [8]).

The value of the objective function of the geometric program is at least that of the MAX CUT problem, as any ± 1 solution for the integer quadratic program is also a solution of the geometric program, with the same value for the objective function.

One approach to convert a solution of the geometric program to a feasible cut in the graph is to partition the sphere S_n into two halves by passing a hyperplane through the origin of the sphere, and labeling vertices on one half by -1 and on the other half by $+1$. The choice of hyperplane may affect the quality of solution that is obtained. Surprisingly, a random hyperplane is expected to give a cut that is not far from optimal.

Consider an arbitrary edge (i,j). Its contribution to the value of the objective function is $\frac{1-v_i v_j}{2}$. The probability that it is cut by a random hyperplane is directly proportional to the angle between v_i and v_j, and can be shown to be exactly $\frac{\cos^{-1}(v_i v_j)}{\pi}$. Hence the ratio between the expected contribution of the edge (i,j) to the final cut and its contribution to the objective function of the geometric program is $\frac{2\cos^{-1}(v_i v_j)}{\pi(1-v_i v_j)}$. This ratio is minimized when the angle between v_i and v_j is $\theta \simeq 2.33$ radians, giving a ratio of $\alpha \simeq 0.87856$. By linearity of expectation, the expected number of edges cut by the random hyperplane is at least α times the value of geometric program, giving an α approximation for MAX CUT.

We remark that a random hyperplane can be chosen algorithmically by choosing a random unit vector r, which implicitly defines the hyperplane $\{x | xr = 0\}$. See details in [8].

2 Research questions

In general, our research goal is the following.

Research goal: Improve the approximation ratio of MAX CUT beyond $\alpha \simeq 0.87856$.

As an intermediate step, we wish to obtain a better understanding of algorithm GW. Let us note that there are four values that interest us:

1. The value of the SDP, denoted by sdp.
2. The value of the true maximum cut, denoted by opt.
3. The expected value of the cut produced by the randomized algorithm, denoted by exp.
4. The actual value of the cut produced by the algorithm, denoted by alg.

We note that we may assume that for the GW algorithm, alg $\geq$ exp (perhaps up to a negligibly small additive term). There are two ways of establishing such a result.

1. **Independent repetitions**. Repeat the random hyperplane rounding technique polynomially many times independently, and take the best of these cuts. As the number of edges in a cut is bounded by at most n^2, the variance of the random variable that denotes the number of edges that we get in a particular cut is also bounded. Hence with high probability, at least one of the random cuts has value not much lower than exp.
2. **Derandomization**. Using the method of conditional expectations one can deterministically select a hyperplane that gives a cut with value at least that of exp, up to perhaps a negligibly small additive term that is a result of performing numeric computations with finite precision [13].

We can summarize the relations between the four values as follows:

$$\text{exp} \leq \text{alg} \leq \text{opt} \leq \text{sdp}.$$

The questions that we shall study are the following.

- How close is the value of the SDP to the value of the maximum cut? The worst case ratio (taken over all possible graphs) of opt/sdp will be called the *integrality ratio*. An integrality ratio close to 1 means that we can estimate the value of the maximum cut within a small error term, without necessarily exhibiting a cut that is nearly optimal.
- How close is the value of the cut found by the GW algorithm (after the rounding) to the value of the maximum cut? The worst case ratio of alg/opt is called the *approximation ratio*.

The analysis of Goemans and Williamson shows that the expected number of edges cut by the random hyperplane rounding technique is at least an α fraction of the value of the semidefinite relaxation. That is, they show that for every graph

$$\frac{\text{exp}}{\text{sdp}} \geq \alpha.$$

As there exists at least one cut achieving this expectation, the integrality ratio of the SDP is no worse than $\alpha \simeq 0.87856$. The worst integrality ratio known for the SDP is roughly 0.8845. It is obtained for the 5-cycle. Associate with each vertex of the 5-cycle a unit vector, and lay all these vectors as a symmetric star in the plane, in the (cyclic) order $1, 3, 5, 2, 4$. Any two vectors that correspond to adjacent vertices make an angle of $360 \cdot 2/5 = 144$, with cosine -0.8090 and contribution 0.9045 to SDP. Hence the value of the SDP is roughly $5 \cdot 0.9045 = 4.5225$, whereas the maximum cut cuts 4 edges, giving an integrality ratio of roughly 0.8845. In this manuscript we show:

Theorem 1 *For every $\epsilon > 0$, there is a graph for which the integrality ratio of the SDP of Goemans and Williamson is no better than $\alpha + \epsilon$.*

We now turn to the approximation ratio for the GW algorithm. Recall that we may assume that alg $\geq$ exp. Using also the fact that opt $\leq$ sdp, the analysis of Goemans and Williamson implies that the approximation ratio is at least α. But is it better than α? This question was studied by Karloff [12]. For every $\epsilon > 0$, he constructed a graph that has the following properties:

1. sdp = opt.

2. The graph has an optimal embedding on the sphere for which exp $\leq (\alpha + \epsilon)$sdp.

Thus Karloff was able to show that the ratio exp/opt can be arbitrarily close to α. However, this does not show the approximation ratio alg/opt is as bad as α, because of the possibility that for these graphs alg $>$ exp. Indeed, it was observed that in Karloff's graphs there are hyperplanes that produce the optimal cut, and the possibility that either independent repetitions or derandomization will produce such a cut has not been ruled out.

In this manuscript we show how Karloff's examples can be extended to show that the approximation ratio is indeed arbitrarily close to α.

Theorem 2 *For every $\epsilon > 0$, there is a graph and an optimal embedding of this graph on the sphere such that for* every *hyperplane, alg $\leq (\alpha + \epsilon)$opt.*

Experience shows that for linear programming relaxations and semidefinite programming relaxations, one does not expect to find a rounding technique that gives a better approximation ratio than the integrality ratio. Hence we are lead to seeking ways of improving the integrality ratio of the SDP of GW. This can be done by adding to the semidefinite program constraints that are valid for every true cut. Some specific constraints, the so called *triangle constraints*, were suggested and analysed in [6], where it was shown that they lead to better approximation algorithms for problems related to MAX CUT, such as MAX DICUT and MAX 2SAT. However, it is not known whether the triangle constraints help in improving the integrality ratio for MAX CUT beyond α. For the 5-cycle, the integrality ratio of the SDP with triangle constraints is 1. We show that if the triangle constraints do indeed improve the integrality ratio for MAX CUT, this improvement is limited.

Theorem 3 *For the SDP for MAX CUT with triangle constraints, there are graphs with integrality ratio no better than roughly* 0.891.

As for the ratio exp/opt (with the random hyperplane rounding technique), Karloff shows that it remains arbitrarily close to α even when arbitrary valid constraints are added to the SDP. Our Theorem 2 that shows alg/opt $\simeq \alpha$ also extends to the case that arbitrary valid constraints are added (even when the best hyperplane is used).

2.1 Organization of this paper

In Section 3 we prove that the integrality ratio of the GW SDP is α, and that with triangle constraints the integrality ratio is no better than roughly 0.891. In Section 4 we show that the approximation ratio is no better than α. Section 4 is composed of several parts. In Section 4.1 we review the triangle constraints and the consequences of adding them to the SDP. In Section 4.2 we describe other valid constraints, and methods of obtaining SDP solutions that satisfy them. In Section 4.3 we produce examples similar in spirit to Karloff's examples. The proofs in this section are elementary and do not involve the dual of the SDP, nor the notion of eigenvalues. In Section 4.4 we show how these and Karloff's examples can be extended to prove Theorem 2. This extension involves graphs that are not connected, a technicality that is dealt with in Section 4.5.

2.2 Notation and conventions

We shall use the convention (unless explicitly stated otherwise) that the distance between two vectors is the angle between them. Likewise, the distance between two points on a sphere is the angle they make as viewed from the center of the sphere.

We shall often not care about precise calculations, and may neglect terms that are insignificant to our main results. Hence we shall sometimes use notation such as $\simeq$, or terms such as "roughly", to denote the fact that the results presented are not strictly accurate, but are close enough to the true results for our purposes. Likewise, we use the term *high probability* in the sense of "probability high enough to make the rest of the arguments work", without exactly quantifying it. In most cases (but perhaps not all), this means probability that can be made arbitrarily close to 1 by taking appropriate values (large enough or small enough, as the case may be) for the parameters involved.

3 The Integrality Ratio

In this section we show that the integrality ratio opt/sdp may be arbitrarily close to $\alpha \simeq 0.87856$. For every $\epsilon > 0$, we construct a graph (that depends on ϵ) for which opt/sdp $\leq \alpha + \epsilon$. Unlike the 5-cycle example for which the integrality ratio was computed exactly, we will not be able to exactly compute opt and sdp. Instead, we shall bound sdp from below and opt from above, thus bounding the integrality ratio from above.

Fix an arbitrary $\epsilon > 0$. We now explain how to construct a graph G satisfying opt/sdp $\leq \alpha + \epsilon$. For our graph we will need a lower bound on sdp. This will be done by exhibiting a feasible (though not necessarily optimal) solution to the geometric program for MAX CUT. Rather than first construct a graph G and then try to find good feasible solutions to the SDP, we shall have a geometric solution in mind, and then build G around it.

Recall that $\theta \simeq 2.33$ radians is the worst angle for the GW algorithm. That is, if the unit vectors associated with the endpoints of an edge make an angle of θ, then the probability

θ/π that a random hyperplane cuts the edge is α times the contribution $(1-\cos\theta)/2$ of the edge to the value of the SDP.

$$\frac{2\theta}{(1-\cos\theta)\pi} = \alpha.$$

Let $\theta_1 < \theta < \theta_2$ be such that the interval $[\theta_1, \theta_2]$ is the largest satisfying

$$\frac{2\theta'}{(1-\cos\theta')\pi} \leq \alpha + \epsilon \tag{1}$$

for every $\theta_1 \leq \theta' \leq \theta_2$.

Consider a unit sphere S^{d-1} in R^d. Distribute n points uniformly on the sphere, where $n >> d$. When $d = 2$ (namely, the sphere is a circle), the n points can be placed at intervals of size $2\pi/n$. When $d > 2$, they can be distributed uniformly at random. The n points represent the n vertices of G, and their placement on the sphere represents a feasible solution to the geometric program. Now join two vertices by an edge if the angle that the corresponding unit vectors makes is between θ_1 and θ_2. This completes the description of the graph G, and the description of a feasible solution for the geometric program.

Equation (1) implies that for the graph G and its geometric embedding, exp/sdp $\leq \alpha+\epsilon$. Moreover, if the points of G are dense and uniform on the sphere, all hyperplanes cut roughly the same number of edges, and hence the number of edges cut by any hyperplane is at most a fraction of $(\alpha + \epsilon)$ of the value of the SDP. This last statement is true regardless of the question of whether the feasible solution that we exhibited for the SDP is optimal or not. Hence, we would be done if we could show that for G, the maximum cut is indeed a cut obtained by a hyperplane. Showing this is the crux of the proof. There are certain pitfalls that should be avoided, and they are described next.

Consider d, the dimension of the embedding. The simplest choice is to take $d = 2$. Unfortunately, this does not work. Observe that $\theta \simeq 2.33$ is very close to $3\pi/4$. If θ was exactly $3\pi/4$ and we would take $\theta_1 = \theta_2 = \theta$ and n divisible by 8, then the graph G would just decompose into a collection of 8-cycles, and hence the maximum cut would include *all* edges. As $\theta \neq 3\pi/4$, G is not a bipartite graph, but is very close to it. By cutting the cycle into 8 equal segments, each of size $\pi/4$, and placing the vertices of even numbered segments on one side of the cut and those of odd numbered segments on the other side, we cut much more than a θ/π fraction of the edges. Hence cuts generated by hyperplanes are far from optimal for this graph. (This observation was apparently first made by Goemans.)

To overcome this problem, we must take $d > 2$. In our proof, d is taken to be some function that grows with $1/\epsilon$. It is an interesting open question whether some fixed d (e.g., $d = 3$?) suffices for all values of ϵ.

Another issue of importance is the distribution of points on the sphere. Graphs that are composed of the $n = 2^d$ points represented as the vectors $\frac{1}{\sqrt{d}}\{\pm 1\}^d$ were considered in [12, 1]. Corollary 15 in Section 4.2 shows that these so called "hypercubes embeddings" cannot be used to demonstrate an integrality ratio other than 1. (This fact was known prior to this paper.) For our construction the number of points will also be exponential in d, but they will be distributed at random on the sphere.

3.1 A Dense Version

In this section we show a construction with a large number of points (more than a simple exponential in the dimension d). Later, in Section 3.2 we will show a construction with fewer points. The latter construction is not needed in order to analyse the integrality ratio of the SDP used in the GW algorithm, but will be used in analysing the integrality ratio for the SDP with triangle constraints.

To simplify notation and not deal with fractions of ϵ, we shall show an integrality ratio of $\alpha + O(\epsilon)$, rather than $\alpha + \epsilon$.

We measure distances between points on the unit sphere by the angle between their respective unit vectors.

The parameters of our construction are chosen as follows.

- The angles θ_1 and θ_2 are chosen so as to satisfy Equation (1). (Using a Taylor expansion of (1) around θ shows that taking them in the range $\theta \pm O(\sqrt{\epsilon})$ suffices.)
- The dimension d is chosen as the smallest such that in S^{d-1} the $(d-1)$-dimensional volume of a cap of radius $(\pi - \theta_2)$ is an ϵ-fraction of the volume of a cap of radius $(\pi - \theta_1)$. (From the values of θ_1 and θ_2 and from Lemma 9 it will follow that taking $d \simeq \sqrt{1/\epsilon}\log(1/\epsilon)$ suffices.)
- We also use a small enough parameter γ (e.g., $\gamma < \epsilon/d$) satisfying:
  - The volume of a cap of radius $\pi - \theta_1 - \gamma$ is at least a $1-\epsilon$ fraction of the volume of a cap of radius $\pi - \theta_1 + \gamma$.
  - For every θ', $\cos(\theta')$ and $\cos(\theta' \pm \gamma)$ differ by an additive factor of at most $\pm\epsilon$.

We are now ready to describe the graph G for which opt/sdp $\leq \alpha + O(\epsilon)$.

1. Choose the parameters θ_1, θ_2, d and γ as described above.
2. Partition S^{d-1} into $n = (\frac{O(1)}{\gamma})^d$ equal volume cells, where each cell has diameter at most γ. (See Lemma 21 in the appendix for more details.)
3. In each cell pick one arbitrary point to represent a vertex of G.
4. Edges connect those pairs of vertices whose distance is between θ_1 and θ_2.

Theorem 4 *For the graph G described above, the integrality ratio opt/sdp of the GW SDP is at most $\alpha + O(\epsilon)$.*

To prove Theorem 4 we first consider G_c, a continuous analogue of the graph G. We may think of G_c as follows. The vertices of G_c are all points of S^{d-1}. The edges of G_c are pairs of vertices of distance at least θ_1. (At the moment we ignore θ_2. It will be considered again later.) Because the graph has an infinite (and uncountable) number of vertices and edges, we shall not use terms such as the "number" of edges, but rather the "measure" of edges.

Denote by μ the normalized $(d-1)$-dimensional natural measure on S^{d-1} and by μ^2 the induced product measure $\mu \times \mu$ on $S^{d-1} \times S^{d-1}$. Given a (measurable) subset A of S^{d-1} and a ρ between 0 and π, we also denote

$$\mu_\rho(A) = \mu^2(\{(x,y);\ x \in A,\ y \notin A, \text{ the distance between } x \text{ and } y \text{ is at least } \rho\}).$$

For G_c, we want to partition it into A and $\bar{A}$ while maximizing the measure of edges cut. This corresponds to finding A that maximizes $\mu_\rho(A)$.

Theorem 5 *Fix an a between 0 and 1 and a ρ between 0 and π. Then the maximum of $\mu_\rho(A)$ where A ranges over all (measurable) subsets of S^{d-1} of measure a is attained for a(ny) cap of measure a.*

Theorem 5 can be deduce from results of Baernstein and Taylor [3]. We indicate the proof in the appendix. It is very much motivated by a proof of the isoperimetric inequality on the sphere (see for example [4] or [17]).

Corollary 6 *Fix a ρ between 0 and π. Then the maximum of $\mu_\rho(A)$ where A ranges over all (measurable) subsets of S^{d-1} is attained for a(ny) cap of measure $1/2$ (namely, a half-sphere).*

Proof: Theorem 5 implies that the cut that maximizes the measure of edges cut is a cap. We show that of all caps, the cap that cuts the highest measure of edges is a half sphere. Consider any other cap C. Without loss of generality we may assume that $\mu(C) < 1/2$, as otherwise we consider $\bar{C}$, the complement of C. Let z be the center of C, and consider a cap H of measure $1/2$ (half sphere) centered at z. Every point in $H \setminus C$ has a higher measure of neighbors in $\bar{H} = \bar{H} \bigcap \bar{C}$ then in C. This implies that by joining $H \setminus C$ with C the measure of edges cut does not decrease, proving $\mu_\rho(H) \geq \mu_\rho(C)$. $\square$

We have seen that for G_c, a half sphere (namely, a cap produced by a hyperplane through the center of the sphere) maximizes $\mu_\rho(A)$. Let us now estimate the measure of edges cut by a hyperplane. Note that by symmetry, all hyperplanes cut the same measure of edges. Hence we can use the analysis of Goemans and Williamson regarding the fraction of edges cut by a random hyperplane. For edges with endpoints at distance ρ apart, this fraction is exactly ρ/π.

Let us denote by opt_c the measure of edges cut by a random hyperplane, divided by the total measure of edges. Similarly, let sdp_c denote the value of the SDP given by the identity embedding of G_c on S^{d-1} divided by the total measure of edges. This value is the expectation of $(1-uv)/2$ over the choice of a random edge (u,v). (We are not assuming here that this is the optimal value for the SDP on G_c, but it is a lower bound.)

Let us now modify G_c by removing all edges that connect pairs of vertices of distance more than θ_2. Recall that the dimension d is chosen as the smallest such that in S^{d-1} the volume of a cap of radius $(\pi - \theta_2)$ is an ϵ-fraction of the volume of a cap of radius $(\pi - \theta_1)$. Hence the fraction of edges removed from G_c is at most ϵ. This implies that neither opt_c nor sdp_c change by more than ϵ. In particular, a random hyperplane is within at most $O(\epsilon)$ of the maximum cut for the new G_c. Observe that all edges now make angles in the range $[\theta_1, \theta_2]$, and recall that in this range, a random hyperplane is expected to cut at most

$(\alpha+\epsilon)$sdp edges. Hence for the modified G_c, the integrality ratio is at most $\alpha + O(\epsilon)$, as desired.

For simplicity in the rest of the presentation, we let sdp_c and opt_c refer to the modified G_c rather than the original one. This introduces an $O(\epsilon)$ error in the analysis, which is negligible.

We now return to our graph G, which is a discrete version of the modified G_c.

Proof of Theorem 4: Recall that to obtain G, we partition S^{d-1} into n equal size cells, where each cell has diameter at most γ, and place a single vertex in each cell. For the discrete version G, let opt_d denote the ratio between the number of edges in the maximum cut and the total number of edges. Let sdp_d denote the ratio between the value of the SDP given by the identity embedding of G on S^{d-1} and the total number of edges. We shall show that $\text{sdp}_d \geq \text{sdp}_c - \epsilon$ and that $\text{opt}_d \leq \text{opt}_c + \epsilon$. From this and from the fact that $\text{sdp}_c > 1/2$ it will follow that the integrality ratio in the discrete case is larger than that of the continuous case by at most $O(\epsilon)$, proving Theorem 4.

To analyse the relation between sdp_d and sdp_c, partition pairs of cells into five types:

1. *Near pairs.* These are pairs of cells such that every point in one cell is within distance less than θ_1 from every point in the other cell. These pairs contribute 0 both to sdp_c and to sdp_d.

2. *Distant pairs.* These are pairs of cells such that every point in one cell is within distance more than θ_2 from every point in the other cell. These pairs contribute 0 both to sdp_c and to sdp_d.

3. *Contributing pairs.* These are pairs of cells such that every point in one cell is within distance between θ_1 and θ_2 from every point in the other cell. Let v_1 be the vertex in one cell of such a pair and let v_2 be the vertex in the other cell. These vertices are joined by an edge in G, and contribute $(1 - v_1 v_2)/2$ to sdp_d. Let $\theta' = \cos^{-1} v_1 v_2$, and let $1/n$ be the volume of each cell (where n is the number of vertices in G). Then every pair of points in the cells makes an angle of $\theta' \pm \gamma$, and the measure that these cells contribute to sdp_c is in the range $(1 - \cos(\theta' \pm \gamma))/2n^2$. Using the fact that $\cos\theta'$ differs from $\cos(\theta' \pm \gamma)$ by at most $\pm\epsilon$ (because γ was chosen to be small), we get that the contribution of contributing pairs to sdp_c and sdp_d are within an additive term of $O(\epsilon)$ of each other.

4. *Near mixed pairs.* These are pairs of cells that have pairs of points within distance less that θ_1, and pairs of points within distance at least θ_1. At worst they contribute 0 to sdp_d. To upper bound the contribution to sdp_c, consider an arbitrary cell. It participates in mixed pairs only with cells entirely contained in the difference of two caps – one of radius $\pi - \theta_1 + \gamma$ and the other of radius $\pi - \theta_1 - \gamma$. As the volume of this region is at most an ϵ fraction of the volume of a cap of radius $\pi - \theta_1$ (and the volume of a cap of radius $\pi - \theta_2$ can be neglected), the influence of the near mixed pairs on the value of sdp_c is at most an ϵ fraction of sdp_c.

5. *Distant mixed pairs.* These are pairs of cells that have pairs of points within distance less than θ_2, and pairs of points within distance at least θ_2. At worst they contribute 0 to sdp_d. Their contribution to sdp_c can be bounded by ϵ, as was done for the case of near mixed pairs.

The above implies $\text{sdp}_d \geq \text{sdp}_c - O(\epsilon)$. We now consider opt_d. Consider the maximum cut C in G, and let opt_d be the fraction of edges that it cuts. Color by red one side of the cut and by green the other side. Consider now a cut C_c of the continuous sphere, for which cells are monochromatic and get the color that the cut C gave to the vertex that lies within the cell. Partitioning pairs of cells into five types as above, it can be shown that the fraction of the measure of the edges cut by C_c is at least $\text{opt}_d - O(\epsilon)$. But as this fraction is at most opt_c, we get that $\text{opt}_d \leq \text{opt}_c + O(\epsilon)$.

We conclude that

$$\frac{\text{opt}_d}{\text{sdp}_d} \leq \frac{\text{opt}_c + O(\epsilon)}{\text{sdp}_c - O(\epsilon)} \leq \frac{\text{opt}_c}{\text{sdp}_c} + O(\epsilon) \leq \alpha + O(\epsilon),$$

completing the proof of Theorem 4. □

3.2 A Sparse Version

We have seen an integrality ratio arbitrarily close to α when vertices are densely packed on the sphere. Namely, we partitioned the sphere into equal volume cells of small diameter, placed a single vertex in each cell, and joined two vertices by an edge if their distance was close to θ. In this section we wish to show a similar result when vertices are placed uniformly at random on the sphere. Moreover, we want the number of vertices to be much smaller than the number of cells. Specifically, the graph G' that we consider in this section is generated at random as follows. We let the dimension d be the same as it was for the dense graph G of the previous section. We select n' points uniformly at random on the sphere S^{d-1}, where each such point will represent a vertex of G'. We connect two points by an edge if the distance (angle) between them is roughly θ (namely, between θ_1 and θ_2, where θ_1 and θ_2 are chosen as for the dense graph G). The number of vertices n' is chosen to be such that each vertex is expected to have degree $O(\log n')$, where the O notation hides some large constant. We call the resulting graph G' the *sparse* version. We note that G' is a random graph, and hence its properties (such as the desired integrality ratio) will not hold with certainty, but only with high probability (over the choice of G'). We assume throughout that G' has at least one edge.

Recall that in the dense version we had $n = (O(1)/\gamma)^d$ vertices, where $\gamma > 0$ is very small and tends to 0 as d grows. In the sparse version the number of vertices will be $n' = c^d$, where $c > 1$ is some constant for which the expected degree becomes roughly $\log n' = \Theta(d)$. For our choice of θ, we will have $c < 2$. Hence the sparse graph G' has much fewer vertices than the dense graph G, giving a sparser configuration of points on S^{d-1}. It can be verified that for our choice of parameters, the average degree in the dense graph is $n^{1-o(1)}$ whereas the average degree in the sparse graph is $(n')^{o(1)}$, further justifying the terms "dense" and "sparse".

We wish to show that the integrality ratio of the sparse version is similar to that of the dense version (in particular, arbitrarily close to α). Recall that sdp_d and opt_d denote the fractional values of the SDP and of the maximum cut on the dense graph. Similarly, we denote by sdp_s and opt_s the fractional values for the sparse graphs. With slight abuse of notation, we consider sdp_d and sdp_s to be the fractional values for the SDP under a particular embedding of the respective graph on S^{d-1}, namely, the identity embedding. Hence they do not necessarily correspond to optimal solutions to the SDP, but they do bound the optimum value of the SDP from below. We have seen that also under this interpretation, $\text{opt}_d/\text{sdp}_d \simeq \alpha$. We shall show that $\text{opt}_s/\text{sdp}_s \simeq \alpha$. To do this we show that $\text{opt}_s \leq \text{opt}_d + \epsilon$ and that $\text{sdp}_s \geq \text{sdp}_d - \epsilon$.

Lemma 7 *For sdp_s and sdp_d defined as above, $sdp_s \geq sdp_d - \epsilon$.*

Proof: The values of sdp_d and sdp_s are averages over edges of quantities $(1 - \cos\theta')/2$. Both for G and G' all edges make angles between θ_1 and θ_2. When θ_1 and θ_2 are sufficiently close, the contribution of each edge is $(1 - \cos\theta)/2$, up to an error of ϵ. So in both cases the average is roughly $(1 - \cos\theta)/2$. □

The more difficult part of the proof is to show that $\text{opt}_s \leq \text{opt}_d + \epsilon$. The idea is to view the sparse graph as a large enough induced subgraph of the dense graph, chosen at random. Thereafter, we shall prove a general theorem that for every graph, the value of the maximum cut (as a fraction of the total number of edges) is preserved in random large enough subgraphs of the graph (up to a negligible error, with high probability over the choice of the random subgraph). The sparse graph G' can indeed be viewed as a random subgraph of the dense graph G. View the selection of vertices to the sparse graph as first selecting a cell at random, and then selecting a point at random in the cell. This is identical to the process of selecting a random vertex from the dense graph. (Recall that a vertex in the dense graph is located arbitrarily in a cell, and hence can be taken also as a random point in its cell.) For the process of selecting the whole sparse graph, there is also the issue of the effect of sampling with replacement, but it can be neglected because $n' \ll \sqrt{n}$. Hence the sparse graph is just a random sample of the dense graph. To simplify the analysis, rather than select exactly n' vertices, we select each vertex of the dense graph independently with probability $p = n'/n$. Hence the actual number of vertices in the sparse graph is a random variable with expectation n', and is strongly concentrated around this expectation.

Now we wish to show that because the sparse graph is a random sample of the dense graph, the relative size of the maximum cut in the sparse graph is with high probability similar to that of the dense graph. The dense graph G has very special structure. However, in the rest of the analysis we shall ignore most of this structure and treat G as an arbitrary graph with n vertices and m edges. The only property of G that we shall use is that G is *nearly regular* – all vertices in G have roughly the same degree. We denote the maximum degree by Δ, and assume that $\Delta \leq 4m/n$, i.e., at most twice the average degree. It is easy to see that this property indeed holds for the dense graph. As for the sparse graph G', we shall view it as being generated from G by sampling each vertex independently with probability p. The only restriction that we use here is that p is large enough so that (with high probability) the average degree in G' is significantly more than $\log n'$.

The proof of the following lemma uses ideas from [9] and appears in the appendix.

Lemma 8 *Let G be an arbitrary graph with n vertices and m edges and maximum degree Δ bounded by twice the average degree. Select each vertex of G independently with probability p, and consider the random graph G' which is the subgraph of G induced on the selected vertices. Then for every $\epsilon > 0$ there is some constant $c > 0$ such that if $p = c\log(n/\Delta)/\Delta$, then $|opt_d - opt_s| > \epsilon$ with probability at most ϵ. (The probability is taken over the random choice of G'.)*

Using Lemma 7 and Lemma 8 we conclude that the integrality ratio for the sparse graph is similar to that of the dense graph, namely, arbitrarily close to α.

3.3 Adding Triangle Constraints

In this section we show an upper bound of roughly 0.891 on the integrality ratio of the SDP for MAX CUT with triangle constraints. The triangle constraints themselves are discussed in Section 4.1.

In valid solutions to the GW SDP, the unit vectors are not constrained, in the sense that they can be arbitrary points on the unit sphere. But when triangle constraints are added, not all point configurations are feasible. We say that three unit vectors, u, v and w, form a *forbidden triple* if the sum of their inner products $uv + vw + wu < -1$. With triangle constraints, only point configurations without forbidden triples are allowed.

Remark: With the triangle constraints, a triple is also forbidden if $-uv-vw+wu < -1$. This is similar to viewing a point configuration as including also every antipodal point (for v we include also $-v$) and requiring that in the new point configuration no triple satisfies $uv + vw + wu < -1$. Note that a forbidden triple cannot contain two points antipodal to each other. Technically, we should prove Proposition 10 in the framework that includes antipodal points. It is easy to see that the proof extends to this case with virtually no change.

The integrality ratio examples shown for the GW SDP contain forbidden triples, and hence are infeasible when triangle constraints are present. The sparse version was introduced so as to overcome this problem. If a random point configuration is sparse enough, then it contains no forbidden triple. Unfortunately, in this case it will also not contain pairs of points at distance more than θ apart, and hence will not show an integrality ratio of α. For this reason, we shall consider a different angle β, where $\pi/2 < \beta < \theta$, and connect points if their distance is at least β. For the graph that we obtain we will show an integrality ratio of roughly 0.891 rather than $\alpha \simeq 0.878$.

We would like β to be as close as possible to θ. For this reason we will not take a very sparse point configuration with no forbidden triples, but rather a moderately sparse point configuration that contains a small number of forbidden triples. This will allow us to pack more points on the sphere, giving a higher value for β. We will still need to get rid of the forbidden triples, but this will turn out to be easy, because their number is small.

Before proceeding with a detailed proof, let us introduce a known technical lemma. Its proof is presented for completeness (and also because we could not find a convenient reference for this lemma).

Lemma 9 *For all $\pi/2 < \beta < \pi$ and two random unit vectors in R^d, the probability that their distance (namely, the angle between them) is at least β is roughly $(\sin\beta)^{d-1}$. Equivalently (using $\sin^2\beta + \cos^2\beta = 1$), for all $0 < a < 1$ and two random unit vectors in R^d, the probability that their inner product is at least a is roughly $(1-a^2)^{(d-1)/2}$. Fixing the direction of one of the two vectors to be along x_1, we state the lemma in yet another form (which also quantifies the term "roughly"), namely,*

$$\frac{c}{\sqrt{d}}(1-a^2)^{\frac{d-1}{2}} \le \mu(\{(x_1,\ldots,x_d) \in S^{d-1};\ x_1 > a\}) \le \frac{1}{2}(1-a^2)^{\frac{d-1}{2}} \qquad (2)$$

for some constant c.

Proof: Let β be the angle such that $\cos\beta = a$. We are interested in the ratio of the area of the cap of radius β to the area of the whole sphere. Cut off this cap from the sphere.

This cap now has a basis of radius $\sin\beta$ and we are interested in the surface area of the cap without the basis. This can be upper bounded by the surface area of a half sphere in R^d of radius $\sin\beta$. (Formally, this can be proved by considering the projection of the cap on the half sphere, where this projection is taken as a radial projection from the center of the sphere on which the cap lies. Details omitted.) It can be lower bounded by the surface area of the basis. (Formally, this can be proved by considering a parallel projection of the basis on the cap.) The surface area of the basis is the volume of a radius $\sin\beta$ ball in R^{d-1}. As is well known (see [18], for example) the volume of a unit ball in R^n is $2\pi^{n/2}/n\Gamma(n/2)$ and its surface area is $2\pi^{n/2}/\Gamma(n/2)$. Here Γ is the gamma function satisfying $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$ and $\Gamma(x+1) = x\Gamma(x)$. As we are dealing with balls of radius $\sin\beta$, we need to scale volumes by $(\sin\beta)^d$ and surface areas by $(\sin\beta)^{d-1}$. Straightforward calculations prove the Lemma (noting also that $\sin\beta = \sqrt{1-a^2}$, and that for large x, $\Gamma(x+1/2) = \Theta(\sqrt{x}\Gamma(x))$). □

We now compute the number of random points that we can place on the sphere without creating a violated triple.

Proposition 10 *For any $\epsilon > 0$ the probability that three independently chosen vectors on the sphere S^{d-1} form a forbidden triple is smaller than $(\frac{16}{27}+\epsilon)^{d/2}$, provided d is large enough.*

Proof: We note that for three unit vectors to form a forbidden triple, at least for one pair of the three vectors, their inner product must be negative. W.l.o.g., assume that $v_1v_2 < 0$. We first pick one vector v_1. Then we partition the range $[0,1]$ into k equal size segments, each of length $1/k$, where $k \simeq 1/\epsilon$. For each a, a multiple of $1/k$ such that $0 \le a < 1$, we perform the following computation. Choose at random the second vector v_2, and compute the probability $p_1(a)$ that $-a \ge v_1v_2 > -a-1/k$. Rather than compute $p_1(a)$ exactly, we upper bound it by the probability that $v_1v_2 \le -a$, or equivalently, the probability that $v_1v_2 > a$. From Lemma 9 we get that

$$p_1(a) \le \frac{1}{2}(1-a^2)^{\frac{d-1}{2}}.$$

Now conditioned on the inner product v_1v_2 being in the range as above, we choose the third vector v_3 at random and compute the probability $p_2(a)$ that $v_3(v_1+v_2) < -1+a+1/k$ (which corresponds to the event that the triple is forbidden). Taking k large enough, we will be able to neglect the term $1/k$ compared to the value of a, with an additive error of $O(1/k)$. (We will soon see that we may assume $a \simeq 1/3 \gg 1/k$, justifying this simplification.) Hence we consider the event $v_3(v_1+v_2) < -1+a$. This probability can be upper bounded as follows. The vector v_1+v_2 has norm at most $\sqrt{2-2a}$. Normalizing it by $1/\sqrt{2-2a}$, we ask for the probability that v_3 has inner product more negative than $(-1+a)/\sqrt{2-2a} = -\sqrt{(1-a)/2}$ with a unit vector. Using Lemma 9 we upper bound this probability by

$$p_2(a) \le \frac{1}{2}(1-\frac{1-a}{2})^{\frac{d-1}{2}} = \frac{1}{2}(\frac{1+a}{2})^{\frac{d-1}{2}}.$$

Now we need to sum $p_1(a)p_2(a)$ over all a. Let us first find the value of a that maximizes the upper bound on the product $p_1(a)p_2(a)$. The expression $(1-a^2)(1+a)$ has a unique maximum at $a = 1/3$. This gives:

$$p_1(a)p_2(a) \le \frac{1}{4}(\frac{8}{9}\cdot\frac{2}{3})^{\frac{d-1}{2}} = \frac{1}{4}(\frac{16}{27})^{\frac{d-1}{2}}$$

where the base of the exponent is acurate up to an additive error of $O(1/k)$. Summing over k values of a, and absorbing the leading term $k/4$ in the base of the exponent (which can be done with little effect when d is sufficiently large), the proposition is proved. $\square$

Our plan is to place n points at random on S^{d-1} and have that the number of forbidden triples be $O(\epsilon n)$. Later we will deal with the forbidden triples.

Corollary 11 *For an arbitrarily small constant $0 < \epsilon < 1$, for large enough d, if $n \simeq (\frac{\sqrt{27}}{4} - \epsilon)^{d/2}$ points are placed at random on S^{d-1}, the expected number of forbidden triples is at most ϵn.*

Proof: With n random points in S^{d-1}, the expected number of forbidden triples is $\binom{n}{3}(\frac{16}{27}+\epsilon')^{d/2}$, where ϵ' is carried over from Proposition 10 (where it was denoted by ϵ), and can be made much smaller than ϵ by making d sufficiently large. Taking $n \simeq \sqrt{\epsilon}(\frac{16}{27}+\epsilon')^{-d/4}$ gives the desired result. $\square$

Now we construct a graph G by placing n points at random on S^{d-1}, where the value of n is as indicated in Corollary 11, and joining by an edge pairs of points that are of distance at least β. (Remark: as argued for the dense graph, when the dimension d is large enough then with overwhelming probability all but a negligible fraction of the edges will happen to connect pairs of points whose distance is essentially equal to β.) We want the average degree in G to be at least $\log n$. Using the bounds of Lemma 9 and taking the limit when $\epsilon \longrightarrow 0$, we choose $(\sin\beta)^2$ to be arbitrarily close to $4/\sqrt{27}$, implying $\beta \simeq 118.67$ and $\cos\beta \simeq -0.4798$. The graph G can be viewed as a sparse version of a dense graph similar to the construction of Section 3.2, with the only difference that β replaces θ. Hence for the SDP of GW on G we have an integrality ratio of roughly

$$\frac{2\beta}{\pi(1-\cos\beta)} \simeq 0.891$$

To extend this result to the SDP with triangle constraints, we modify the embedding of the graph G in S^{d-1} so that no forbidden triple remains. As the number of forbidden triples is below $n/3$, there is some vertex, say v, that is not a member of any forbidden triple. Take one vertex from each forbidden triple and move the vector associated with it so that it coincides with the vector associated with v. Now the geometric embedding satisfies the triangle constraints. As the number of forbidden triples is small compared to the number of vertices, and as G was nearly regular, this process effects only an $O(\epsilon)$ fraction of the edges from G. Hence the integrality ratio changes only by $O(\epsilon)$. This completes the proof of Theorem 3.

Remark: Rather than modify the solution to the SDP, we could modify G by removing from each forbidden triple one vertex. The results regarding the integrality ratio remain unchanged, but the description of the graph G becomes a bit more complicated.

4 The Approximation Ratio

In this section we show that even the best hyperplane does not give an approximation ratio better than α. As the result holds also in the presence of additional constraints to the SDP, we review these constraints first. This review is somewhat lengthy. Part of the reason for this is so as later to produce a simple and easy to understand example showing a relatively large gap between exp and opt.

4.1 Review of the Triangle Constraints

To try to improve integrality ratio, one may strengthen the SDP by additional constraints that are valid for true ± 1 solutions. The triangle constraints described below seem promising. One way of viewing them is as the set of valid constraints that involve triples of vertices. Regardless of which ± 1 values the three vertices receive, the following constraint is always satisfied:

$$|v_i + v_j + v_k|^2 \geq 1.$$

For unit vectors, this gives:

$$v_i v_j + v_j v_k + v_k v_i \geq -1 \tag{3}$$

We may add a similar constraint when $-v_j$ is substituted for v_j, as the value of $-v_j$ is also still ± 1. This gives the following constraint.

$$-v_i v_j - v_j v_k + v_k v_i \geq -1 \tag{4}$$

Constraint 4 can be seen to be equivalent to

$$|v_i - v_j|^2 + |v_j - v_k|^2 \geq |v_i - v_k|^2 \tag{5}$$

using the fact that $|u - v|^2 = u^2 + v^2 - 2uv = 2 - 2uv$ for unit vectors. Constraint 5 has a convenient geometric interpretation, saying that the angle opened by three vertices is never more than $\pi/2$ (the case of $\pi/2$ gives equality in the constraint, by Pythagoras' theorem). We call this the *acute angle rule*. We note that from constraint 3 we can derive a triangle inequality similar to constraint 5, but with v_j replaced by $-v_j$.

Recall that the integrality gap for the 5-cycle was demonstrated for an embedding in which all five unit vectors make a star lying in the plane. These vectors define the five vertices of a symmetric pentagon. Each angle of the pentagon is $3\pi/5 > \pi/2$, and hence this is not an embedding that satisfies the triangle constraints.

The above argument can be generalized in two respects, shown in the next two propositions.

Proposition 12 *For arbitrary graphs, if the optimal vector solution to the SDP with triangle constraints lies in the plane, then $sdp = opt$.*

Proof: Consider any three unit vectors in the plane. Then either at least two of them lie on the same line (though possibly in opposite directions), or the endpoints of the three make a triangle in which one of the angles is obtuse, or we can negate one of the vectors and get a triangle with an obtuse angle. Hence any vector solution in the plane that obeys the acute angle rule must have all vectors lie on just two lines intersecting in the origin (though on each line the vectors can point in either direction). It can be shown that in this case the objective function of the SDP is linear in the cosine of angle between the two lines, implying that the maximum is achieved when the angle is either 0 or π. But in this case we have a one-dimensional ± 1 solution, or in other words, a true cut in the graph. The number of edges cut is exactly equal to the objective function of the SDP. $\square$

Proposition 13 *With the triangle constraints, for every cycle $sdp = opt$.*

Proof: Recall that sdp $\geq$ opt. Hence we need only show that sdp $\leq$ opt.

Even-length cycles are bipartite graphs. For every bipartite graph the maximum cut includes all the edges. As the value of the SDP cannot be more than all the edges, sdp = opt. (Moreover, in the optimal vector solution all edges must make an angle of π, and a random hyperplane cuts them with probability 1.)

It remains to consider odd-length cycles. For a k-cycle with k odd, the maximum cut contains $k-1$ edges. Hence we need to show that sdp $\leq k-1$. For 3-cycles (triangles), this is a direct consequence of constraint 3. For odd cycles $(v_1, \dots v_k)$ of length at least 5, sum up $v_1 v_j + v_j v_{j+1} + v_{j+1} v_1 \geq -1$ (constraint 3) over all even $2 \leq j \leq k-1$, and $-v_1 v_j + v_j v_{j+1} - v_{j+1} v_1 \geq -1$ (constraint 4) over all odd $3 \leq j \leq k-2$, giving (after cancellations) $\sum v_i v_{i+1} \geq -k+2$, implying that the value of the SDP is at most $k-1$. □

In fact, with triangle constraints the SDP gives optimal value for all planar graphs (because the triangle constraints define the cut polytope for planar graphs, and more generally, for graphs that do not contain K_5 as a minor [2]).

There is no known conjecture of which graph gives the worst integrality ratio with triangle constraints. In Section 3.3 we showed an integrality ratio of at most roughly 0.891. Prior to that, the bad examples were only known to have integrality ratio at most roughly 0.96, that is even worse than the hardness of approximation result for MAX CUT.

Determining the true integrality ratio for the SDP with triangle constraints is an important open question.

4.2 Other valid constraints

We describe a procedure for generating valid constraints for MAX CUT. This procedure generalizes the procedure used for generating the triangle constraints. For more information about valid constraints and their effects, see [5].

Select a subset of vertices $S \subset V$. Let $\{1, \dots, k\}$ denote this subset. Consider the complete graph on them. To each edge (i,j) in this complete graph, assign a weight w_{ij}. This gives a weighted graph that we denote by W. Let a be the weight of the maximum weight cut in W. Then the following is clearly a valid constraint.

$$\sum_{i,j \in S} w_{ij} \frac{1 - v_i v_j}{2} \leq a \tag{6}$$

This quadratic constraint of the geometric program for MAX CUT is linear in the variables y_{ij} of the SDP. We can derive other valid constraints by using the fact that in our intended solutions, the v_i are ± 1 variables. This means that negating a variable still defines a cut, and constraint 6 must still hold. Let $b_1, \dots, b_k$ be an arbitrary sequence of ± 1 values. Then we can generalize constraint 6 to:

$$\sum_{i,j \in S} w_{ij} \frac{1 - b_i b_j v_i v_j}{2} \leq a \tag{7}$$

Definition 1 *A* valid linear constraint for the SDP for MAX CUT *is any constraint that can be derived by the above process. For short, we call such constraints* valid constraints.

Remark: Our definition of valid constraints is not the most general possible, but it does seem to capture a large family of useful constraints. In particular, it includes the constraint sdp $\leq$ opt. There are other useful constraints that are not captured by our definition. For example, we can add the constraint $v_1 v_2 = 1$. This forces v_1 and v_2 to be on the same side of the cut. This constraint is not valid for all cuts but may be useful nevertheless, especially if we make another copy of the SDP and add to it the constraint $v_1 v_2 = -1$. One of these two constraints must be right. As another example, observe that for a true cut, the rank of matrix Y is 1. Hence we may wish to add this as a constraint. However, optimizing over rank restricted matrices is NP-hard (otherwise we could solve MAX CUT).

Having defined a very general notion of valid constraints, we now present methods for obtaining SDP solutions that satisfy *all* valid constraints. Every ± 1 solution necessarily satisfies all valid constraints, as it is a true cut. This can be generalized to what we call *hypercube embeddings.*

Definition 2 *A d-dimensional hypercube has the 2^d vertices $\frac{1}{\sqrt{d}}\{\pm 1\}^d$. A* hypercube embedding *is a mapping of the vertices of a graph to the vertices of a d-hypercube for some $d \geq 1$. The mapping need not be one-to-one or onto.*

Lemma 14 *Every hypercube embedding satisfies all valid constraints.*

Proof: The inner product of two vectors is the sum of the inner products of their coordinates. We shall use this simple fact.

Consider an arbitrary hypercube embedding, and ignore for the moment the scaling factor of $1/\sqrt{d}$. It embeds the vertices of the graph in a d-dimensional space, where each vertex has ± 1 entries in d coordinates. Consider now each coordinate separately. It defines a cut, as all vertices on it have ± 1 values. As such it satisfies the valid constraint. Summing up over all d coordinates, the valid constraint is satisfied with its righthand side a replaced by da. Now scaling each coordinate by $1/\sqrt{d}$, inner products are scaled by $1/d$, and we can scale back the righthand side to a. □

For the following corollary, we are interested in feasible solutions to the SDP, without requiring them to be optimal.

Corollary 15 *The value of a hypercube embedding solution to the MAX CUT SDP is no larger than opt.*

Proof: sdp $\leq$ opt is a valid constraint, and hypercube embeddings satisfy all valid constraints. □

Lemma 16 presents another method of obtaining SDP solutions that satisfy all valid constraints. We call this method the *disjoint union* method.

Lemma 16 *Let $G(V, E)$ be a graph, and disjointly partition its vertex set V to $V_1 \bigcup V_2$. Let $f_1(V_1)$ be an embedding of V_1 in R^{d_1} and $f_2(V_2)$ be an embedding of V_2 in R^{d_2}. Consider an embedding $f(V)$ of V in $R^{d_1+d_2}$ that maps the vertices of V_1 as f_1 does, and maps the vertices of V_2 as f_2 does, and the subspaces to which the two parts are mapped are orthogonal to each other (e.g., the set of nonzero coordinates is disjoint). Let $S \subset V$ be a set of vertices on which there is a valid constraint violated by f. Then either f_1 violates a valid constraint on $S \bigcap V_1$, or f_2 violates a valid constraint on $S \bigcap V_2$. (In particular, this implies that if f_1 and f_2 satisfy all valid constraints, so does f.)*

Proof: Recall that a valid constraint says that the value of the SDP on a hypothetical weighted subgraph involving vertices S is not more than the weight of the edges cut by the maximum cut for this subgraph. This subgraph involves vertices in $V_1 \bigcap S$ with internal edges, vertices of $V_2 \bigcap S$ with internal edges, and cross edges between $V_1 \bigcap S$ and $V_2 \bigcap S$. As the inner products $v_i v_j = 0$ for the cross edges, f gets exactly half the weight of the cross edges as a contribution to the value of the SDP. In addition it gets the contribution of f_1 from $V_1 \bigcap S$ and of f_2 from $V_2 \bigcap S$. Assume that neither f_1 nor f_2 violate a valid constraint. Then there are cuts in $V_1 \bigcap S$ and $V_2 \bigcap S$ with value at least as high as the contributions of f_1 and f_2 respectively. It is possible to combine these cuts in a way that adds at least half the weight of the cross edges, as there are two different ways of combining the cuts, and together they give all cross edges. Hence there is a true cut on S with value at least that of f on S. $\square$

4.3 Rounding by random hyperplanes

Karloff [12] shows families of graphs for which exp/opt approaches α from above. Though the graphs that Karloff considers and their embeddings on the sphere are quite easy to describe, Karloff's proof that the embeddings are optimal is quite involved. He does so by showing a solution of the same value to the dual of the semidefinite program. To analyse this solution, he provides expressions for the eigenvalues of certain matrices (luckily, these expressions were known), and analyses which of these expressions give the most negative eigenvalue. A proof along similar lines for a related family of graphs was given by Alon and Sudakov [1]. The manipulations in the proof of [1] are somewhat easier to follow.

Here we provide a simpler example than those of [12, 1]. The analysis is relatively straightforward, at a level that can be taught in class (and this indeed was done). Due to the simplicity of the example and its analysis, the main features of the example are easy to comprehend, and extensions of this example that lead to the proof of Theorem 2 come out naturally. However, the new example is somewhat weaker than that of Karloff in several respects that will be discussed later, and so to prove Theorem 2 we actually revert back to Karloff's examples.

Rather than consider algorithm GW, let us consider a related algorithm that we call GW$\triangle$. In this algorithm we solve the SDP with the triangle constraints, derive a vector solution, and cut it with a random hyperplane.

Proposition 17 *For the 7-cycle, there is an optimal vector solution to GW$\triangle$ for which exp $\leq 0.8788 \cdot$ sdp.*

Proof: Consider the 7-cycle. The value of the maximum cut is 6. There are seven cuts that achieve this value, where each such cut looses one particular edge (v_i, v_{i+1}). Proposition 13 shows that with triangle constraints, the optimal value for the SDP is also 6. A solution for the SDP attaining this value can be obtained by taking a hypercube embedding that places each of the seven optimal cuts in a different coordinate. By Lemma 14, such an embedding necessarily respects the triangle constraints (and any other valid constraint). The vectors are then:

$(+1, -1, +1, -1, +1, -1, +1)$
$(+1, +1, -1, +1, -1, +1, -1)$

$(-1, +1, +1, -1, +1, -1, +1)$
$(+1, -1, +1, +1, -1, +1, -1)$
$(-1, +1, -1, +1, +1, -1, +1)$
$(+1, -1, +1, -1, +1, +1, -1)$
$(-1, +1, -1, +1, -1, +1, +1)$
where each vector is normalized by $1/\sqrt{7}$ so as to make it a unit vector. The optimal cuts can be read along the coordinates (columns). The inner product between any two adjacent vectors is $-5/7$, giving an angle of 135.6, which is quite close to the worst angle for the random hyperplane rounding technique (which is 133.56). The probability that a random hyperplane cuts an edge is 0.7532, and the ratio between this and 6/7 is roughly 0.8788. □

Let us now compare between the 7-cycle example and those of Karloff. (The examples of Alon and Sudakov have properties similar to those of Karloff.) Karloff's example is stronger in several respects.

1. Karloff gives an infinite family of graphs for which the ratio exp/sdp becomes arbitrarily close to $\alpha \simeq 0.87856$. For the 7-cycle, this ratio is close ($\simeq 0.87878$), but not arbitrarily close.

2. The graphs in Karloff's example can become arbitrarily large and with arbitrarily large degree, making them robust against rounding techniques that exploit the fact that a graph is small (such as exhaustive search!) or that the degree is small (such as that in [7]).

3. Karloff's example applies both to algorithm GW and to GW△, whereas the 7-cycle example applies to algorithm GW△ but not to algorithm GW (where the optimal vector solution to the SDP is a star in the plane). We note that for any graph, whenever the value of the SDP solution is reduced by adding the triangle constraints, it makes sense to find a cut using the GW solution rather than GW△, as GW now has an integrality ratio strictly less than 1, and hence an approximation ratio strictly better than α.

Let us also note some similarities between the 7-cycle example and those of [12, 1]. The graphs involved can all be viewed as containing subsets of the vertices of a d-dimensional ± 1 hypercube. For the 7-cycle $d = 7$, for Karloff's graphs d can be arbitrarily large but must be even, and for the graphs of Alon and Sudakov d is arbitrary. An edge is placed between two vertices if their hamming distance is exactly q, where q is a carefully chosen parameter. For the 7-cycle, $q = 6$. The vector solution of the SDP is naturally taken as the hypercube embedding. By Lemma 14 it satisfies all valid constraints, and by Corollary 15 the graph contains a cut whose value is at least equal to that of this SDP solution.

In these graphs, all edges make angles whose inner product is $(d-2q)/d$. The parameter q is chosen so that the angle implied by this inner product is as close as possible to the worst case angle for the random hyperplane rounding technique. The main technical issue involved in these examples is showing that the hypercube embedding indeed maximizes the SDP. For the graphs of [12, 1] the proof is quite involved. To some extent, it is even coincidental. For example, in the graphs of Alon and Sudakov, if q is chosen carelessly to be odd, the resulting graphs are bipartite, and the true optimal cut gives a one dimensional vector solution of much better value than the high dimensional hypercube embedding. For the 7-cycle, the proof of the optimality of the hypercube embedding becomes straightforward, due to the use of the triangle constraints.

4.4 Rounding with arbitrary hyperplanes

The 7-cycle example shows that for GW$\triangle$, exp/opt < 0.8788. However, this example leaves much to be desired. It addresses only the expected size of the cut, but not the actual size of the cut found by the algorithm. In fact, it can be shown that with high probability, a random hyperplane produces an optimal cut.

Let p_i be the probability that the random hyperplane produces a cut with i edges. Then $\sum i \cdot p_i = \exp \simeq 0.8788\text{opt} \geq 5.27$. Observe that every cut of the 7-cycle contains an even number of edges. Hence $\sum_i i \cdot p_i \leq 6p_6 + 4(1 - p_6) = 2p_6 + 4$. It follows that p_6, the probability of producing a cut with six edges, is at least 0.63. The conclusion is that for algorithm GW$\triangle$, the 7-cycle is an example of a graph for which exp/opt is very close to α, but nevertheless the algorithm most likely does produce the optimal cut. That is, alg = opt. It is natural to ask whether a similar phenomenon happens for Karloff's example. Indeed, Karloff in his paper [12] is cautious enough to warn the reader that his results apply only to the ratio exp/opt, but not to the ratio alg/opt with the independent repetitions policy.

A simple way around this problem is as follows. Take n disjoint copies of 7-cycles, and consider a $7n$ dimensional vector solution where each cycle is embedded in 7-coordinates as described above for the 7-cycle, and the subspaces (nonzero coordinates) of each cycle are orthogonal (disjoint). This vector solution obeys the triangle constraints (and all valid constraints), by Lemma 16.

Consider now a random hyperplane, which is defined by its normal: a vector r in a random direction. The projections of r on each of the n 7-dimensional subspaces are again in random directions, and moreover, independent. From each 7-cycle the random hyperplane it is expected to cut roughly 5.27 edges, and the actual number of edges cut is bounded between 0 and 6. The expected number of edges cut from all cycles is roughly $5.27 \cdot n$ (by the linearity of the expectation), and the probability of deviation from expectation by $k\sqrt{n}$ behaves like c^{k^2} for some $c < 1$ (from standard bounds on sums of independent random variables with bounded values).

Applying the same reasoning to Karloff's example rather than the 7-cycle example, we can show the following:

Theorem 18 *For algorithm GW and for any variant of it that uses additional valid constraints (such as the triangle inequalities), the approximation ratio is no better than α when the independent repetitions rounding policy is used. In more technical terms: for every $\epsilon > 0$ there is a constant $0 \leq c < 1$, such that for infinitely many values of n there are n-node graphs for which all but a c^n fraction of hyperplanes cut less than $(\alpha + \epsilon)$opt edges.*

Proof: Let k be a sufficiently large constant such that there is a k vertex graph (e.g., from Karloff's family of graphs) for which exp/opt $\leq \alpha + \epsilon/2$, and let m be the number of edges in such a graph. Observe that $m < k^2$. We call this k-vertex graph the *basic* graph. For a sufficiently large n divisible by k, the graph G is composed of n/k disjoint copies of the basic graph. To embed G, embed each copy of the basic graph with a hypercube embedding in a disjoint set of coordinates. Similarly to the argument given for disjoint unions of 7-cycles, standard bounds on large deviations for sums of bounded independent random variables show that the probability that the number of edges cut by a random hyperplane deviates from its expectation by more than $\epsilon mn/2k$ is at most c^n, for some $c < 1$ that depends only on k and ϵ.

We note that the embedding described above satisfies all valid constraints. For each basic graph, this follows from Lemma 14, and for the union of all basic graphs this follows from Lemma 16. Moreover, this embedding achieves a maximum value for the SDP. The objective function for G is the sum of objective functions for each of the copies of the basic graphs. Any solution for G induces solutions for the basic graphs in a natural way, and the values of these solutions is bounded by the optimum for the basic graph. $\square$

Theorem 18 shows that there are graphs and solutions to the SDP such that essentially every hyperplane gives a cut whose value is nearly as bad as $\alpha \cdot \text{opt}$. However, even for the graphs and embeddings considered in Theorem 18, there are hyperplanes that produce an optimal cut. (For example, a hyperplane whose normal is a vector with value 1 at every kth coordinate, and 0 elsewhere.) Keeping the same graphs as those of Theorem 18 but changing the embedding, we can derive examples in which *every* hyperplane gives a cut with value roughly $\alpha \cdot \text{opt}$.

Theorem 19 *For algorithms GW and GW△, any rounding technique that produces a cut induced by a hyperplane (such as the derandomized version of algorithm GW) has approximation ratio no better than α. In more technical terms: for every $\epsilon > 0$ there is a graph and an optimal solution to the SDP for which every hyperplane cuts less than $(\alpha + \epsilon)opt$ edges.*

Proof: The graphs considered are essentially those of Theorem 18. Let k be a sufficiently large constant such that there is a k vertex graph (e.g., from Karloff's family of graphs) for which $\text{exp}/\text{opt} \leq \alpha + \epsilon/3$, and let m be the number of edges in such a graph. Observe that $m < k^2$. For a sufficiently large n divisible by k, G is the graph composed of n/k disjoint copies of this basic graph.

Now rather than embedding each basic graph on a disjoint set of coordinates, we shall embed the graph in dimension d, where d will be chosen in the range $k \leq d < \epsilon n/3k$. The embedding will be randomized. In other words, we shall describe a large family of optimal embeddings, and will show that for a random member of this family, Theorem 19 holds. We will not exhibit a specific embedding, as this is not required by the Theorem.

For each of the copies of the graph, we choose at random a k-dimensional subspace of R^d, and embed this copy in that subspace using the optimal hypercube embedding. The value of d is chosen such that:

1. The number of "essentially different" hyperplanes in R^d is small, namely, below c^n where $c > 1$ is some constant that depends on d. This gives an upper bound on d. (d is chosen so that the constant c is smaller than the inverse of the constant c from Theorem 18.)

2. The triangle constraints are satisfied with high probability. This gives a lower bound on d. It is needed only for GW△, but not for GW. Hence for GW we may simply

choose $d = k$, meaning that all copies are embedded in the same subspace, but at random orientations.

We now upper bound the number of essentially different hyperplanes in R^d. A hyperplane is defined by its normal. Pick a small value of $\gamma > 0$ (e.g., $\gamma = \epsilon/d$). We say that two hyperplanes are *similar* if their normals make an angle of at most γ, and *different* otherwise. Picking normals greedily such that no two are of distance less than γ from each other, there are only $(O(1)/\gamma)^d$ normals in S^{d-1}. This defines a family F of $(O(1)/\gamma)^d$ hyperplanes, such that every other hyperplane is similar to some hyperplane in the family.

Now we argue that the cut produced by the best hyperplane H is approximated by a cut produced by a member of F. We will use the fact that G is nearly regular, and that the embedding on the sphere of the basic graphs composing G is independent. For each hyperplane $H' \in F$, consider a slice of width γ around it. It has measure at most ϵ (that depends on γ). Only an ϵ-fraction of the vertices of G is expected to lie there. Moreover, with probability $c^{-\epsilon n}$ (where $c > 1$ depends on k) there will not be more than a 2ϵ-fraction of the vertices there. As there are only $(O(1)/\gamma)^d$ hyperplanes in F, the same holds w.h.p. for all hyperplanes in F simultaneously. If H' is closest in F to H, then their symmetric difference lies in this slice. Hence the cuts that they induce on G differ by $O(\epsilon n)$ vertices, and by near regularity, by an $O(\epsilon)$ fraction of the edges.

Choosing the parameters such that $(O(1)/\gamma)^d c^{-\epsilon n} < 1/4$ ensures that for at least 3/4 of the random embeddings, every hyperplane cuts no more that $(\alpha + \epsilon)$opt edges. For n large enough, taking $d = o(n/\log n)$ suffices.

We now show that with high probability the triangle constraints are satisfied. For three vertices in the same copy, this follows from the hypercube embedding. For vertices in different copies, this follows from the fact that for random and independent unit vectors, the expected value of the absolute value of their inner product is roughly $1/\sqrt{d}$, and the probability of being larger than $\sqrt{c/d}$ is exponentially small in c. Hence if $d \gg k^2 \log n$ then with overwhelming probability the absolute value of the inner product of any two vertices belonging to two different copies is below $1/k$. A triangle constraint involving three vertices from three different copies is necessarily satisfied. For the case of two vertices from one copy and the third from another copy, we use the fact that if the two vertices of the same copy lie on the same line (in opposite directions), the triangle inequalities are satisfied. If these vertices lie on different lines, then their inner product is at least $-1 + 2/k$ (this is a consequence of the hypercube embedding) and the two other inner products involved in the triangle inequality have absolute value below $1/k$.

By choosing d such that $k^2 \log n \ll d \ll n/k$ our theorem is proved.

We note that we can accommodate valid constraints other than the triangle constraints, by giving a higher lower bound on d. However, if the number of terms in a constraint grows as a function of n, the lower bound on d might overtake the upper bound. □

4.5 Making the graph connected

In Theorem 19 we show that for every $\epsilon > 0$ there is a graph and an optimal solution to the SDP for which every hyperplane cuts less than $(\alpha + \epsilon)$opt edges. However, the graphs produced in the proof of Theorem 19 are not connected. This is an undesirable feature, because the "standard" approach to handling graphs that are not connected is first to separate them into connected components, and then run the MAX CUT algorithm (e.g., the GW algorithm) on each component separately. The cuts found for each component can easily be joined into one global cut. Hence inputs to the GW algorithm are typically assumed to be connected graphs.

In this section we show how to modify the graphs produced in the proof of Theorem 19 so as to make them connected. Recall that these graphs contained n/k components, each of size k, where k can be assumed to be arbitrarily large. The most natural approach to making these graphs connected is to pick one vertex in every connected component (call these vertices $w_1 \dots w_{n/k}$), and then add the edges (w_i, w_{i+1}) for $1 \leq i < n/k$. The graph then becomes connected. The number of edges that we add is negligible compared to the total number of edges in the graph (when k is large enough), and hence the effect of this modification on the value of the optimal cut is negligible. Likewise, the effect of this modification on the value of the SDP (either GW or GW$\triangle$) is negligible. However, this approach is not satisfactory, because it might force the vector configuration of the optimal solution to the SDP to be different than that assumed in the proof of Theorem 19. Indeed, the angle between the vector associated with w_i and the vector associated with w_{i+1} will be exactly π, rather than random. The change in the vector configuration may lead to the situation in which a hyperplane produces an optimal, or near optimal, cut.

To overcome this difficulty, rather than connecting w_i and w_{i+1} by a single edge, we connect them by a small subgraph that we call a "connecting gadget". The connection is done by identifying w_i with one specially chosen vertex of the connecting gadget, and identifying w_{i+1} with another specially chosen vertex of the connecting gadget. The key property of a connecting gadget is summarized in the following definition.

Definition 3 *A connected graph H is a* connecting gadget *(for SDPs for MAX CUT) if it has the following property. Let G be an arbitrary graph. Let G_H be a graph obtained from G and H by connecting two vertices of G via H, as explained above. Consider an arbitrary optimal vector configuration for the GW (or GW$\triangle$) SDP on G. Then there is an optimal vector configuration for the GW (or GW$\triangle$) SDP on G_H that is identical to the original configuration on all original vertices of G.*

We now describe a graph H that has 12 vertices and is a connecting gadget. Vertices $\{1,2,3\}$ and vertices $\{4,5,6\}$ make a complete bipartite graph $K_{3,3}$. Likewise, vertices $\{7,8,9\}$ and vertices $\{10,11,12\}$ make a complete bipartite graph. In addition include the four edges $(1,7)$, $(1,10)$ $(4,7)$, $(4,10)$. This completes the description of H.

The graph H has $9+9+4 = 22$ edges. It has a cut containing 20 edges. In fact, there are two such cuts. One of them is $\{1,2,3,7,8,9\}$ and the other is $\{1,2,3,10,11,12\}$. Likewise, the following also holds:

Lemma 20 *For the graph H described above, the optimal value of the GW SDP is 20.*

Proof: The optimal value of the GW SDP is at least 20, as H has a cut with 20 edges. We will now show that the optimal value of the SDP is at most 20. To do this, we shall use the fact that a semidefinite maximization problem has a dual minimization problem, and the optimal value of the minimization problem is at least as large as the optimal value of the maximization problem. Hence, to bound from above the value of the GW SDP, it suffices to show a feasible solution to the dual minimization problem of value 20.

The dual minimization problem to the GW SDP is as follows (for more details, see [8] and references therein). Let A be the adjacency matrix of a graph G with n vertices and m edges, and let $\lambda_1 \geq \ldots \geq \lambda_n$ denote its eigenvalues. The dual to the GW SDP on G is to minimize $m/2 + (\sum_{i=1}^{n} \gamma_i)/4$ subject to the matrix A_γ being positive semidefinite, where A_γ is the same as matrix A, except that it has the γ_i values on its diagonal. A feasible (though not necessarily optimal) solution to the dual is to take $\gamma_i = -\lambda_n$ for all i, thus adding $-\lambda_n$ to all eigenvalues and making the resulting A_γ positive semidefinite. Hence the value of the GW SDP is at most $m/2 - n\lambda_n/4$.

In our case of the graph H, $m = 22$. Hence the contribution of the term $m/2$ is 11. Now we compute λ_{12} for the adjacency matrix of H. The eigenvalues of $K_{3,3}$ are known to be $(3, 0, 0, 0, 0, -3)$. For the disjoint union of $K_{3,3}$, the only negative eigenvalue is -3, and it has multiplicity 2. The other edges that we add to G form a $K_{2,2}$ which has eigenvalues $(2, 0, 0, -2)$. For the sum of two matrices $A = A_1 + A_2$, $\lambda_i(A) \geq \lambda_i(A_1) + \lambda_n(A_2)$. It follows that for our matrix A, $\lambda_{10} \geq -2$. It can be seen that A has two eigenvalues of value -3, by considering the two linearly independent eigenvectors

$$(+1, +1, +1, -1, -1, -1, -1, -1, -1, +1, +1, +1)$$

and

$$(+1, +1, +1, -1, -1, -1, +1, +1, +1, -1, -1, -1).$$

These eigenvalues must correspond to λ_{11} and λ_{12}. Hence the term $n\lambda_{12}/4$ contributes $12 \cdot 3/4 = 9$, bounding the value of the SDP by $11 + 9 = 20$. □

Take arbitrarily one vertex from the first $K_{3,3}$ and one vertex from the second $K_{3,3}$. For concreteness we take vertices 6 and 7. Consider now the GW SDP on H. We claim that regardless of which unit vectors are assigned to 6 and 7, this can be completed to a vector configuration of optimal value for the SDP. Let v be the vector assigned to 6 and u the vector assigned to 7. Then we assign the vector $-v$ to 1,2,3, the vector v to 4,5 (and 6), the vector u to 8,9 (and 7) and the vector $-u$ to 10,11,12. The value of the SDP is then $9 + 9 + 4/2 = 20$ which is best possible.

The graph H can be used to connect two vertices w_i and w_{i+1} of a graph G. Identify w_i with vertex 6 of H and w_{i+1} with vertex 7 of H. Now any vector solution to the GW SDP that is optimal for G can be completed to a vector solution that is also optimal on the union of G and H. Moreover, this completion does not introduce any new vectors (other than $-v$ and $-u$), so it preserves the property that valid constraints (including the triangle constraints) are not violated.

It follows that the graph H that we described is indeed a connecting gadget. Needless to say, if we wish to connect several pairs of vertices in a graph G, we can use several disjoint copies of the graph H for this purpose. (However, if a vertex v of G is a member of several such pairs, then the respective connecting gadgets will eventually also share a vertex – via the process of identifying a vertex from the connecting gadget with v.)

From the above discussion (including the fact that the size k of a connected component of the original graph G can be arbitrarily large, whereas the number of vertices in H is a fixed constant) it follows that Theorem 19 holds also for connected graphs.

5 Conclusions

The work of Goemans and Williamson shows that for algorithm GW

$$\alpha \cdot \text{sdp} \leq \text{exp} \leq \text{alg} \leq \text{opt} \leq \text{sdp}.$$

Hence the ratio between any two of the above quantities is at least $\alpha \simeq 0.87856$. But can it be as low as α (or arbitrarily close to it)? Karloff showed graphs with exp/opt approaching α from above. In this work we show similar results for opt/sdp and alg/opt. The discussion in the beginning of Section 4.4 regarding the 7-cycle example illustrates that the ratio exp/alg can also be significant, and may sometimes be the reason for much of the ratio exp/opt in examples such as those of Karloff. Furthermore, interpreting alg as the result of taking the cut produced by the best hyperplane, Karloff's example shows a ratio of α for exp/alg, with equality between alg and opt. In the context of Karloff's example, taking the best hyperplane is a plausible possibility, because the embedding is in $O(\log n)$ dimensions and it suffices to consider one of $n^{O(\log n)}$ hyperplanes when searching for the best hyperplane. This can be done in moderately superpolynomial time.

When additional constraints are added to the SDP, a ratio of α between alg and opt still holds. However, our results regarding the integrality gap deteriorate. With triangle constraints we could only prove an integrality ratio of 0.891. When all valid constraints involving up to k vertices are added, our proof technique deteriorates further, and as k grows the integrality ratio that we can prove tends to 1. This situation is not satisfactory, because by the hardness of approximation results for MAX CUT [10], the integrality ratio must be below 16/17 unless P=NP.

An important research direction is to see whether the approximation ratio for MAX CUT can be improved beyond α. Specifically, this can be asked for algorithms based on the GW SDP with triangle constraints. Our results imply that neither a random hyperplane nor the best hyperplane suffice for this purpose. We can go a bit further and rule out a larger class of rounding techniques. Let us elaborate on this.

Using SDP we get an embedding of the vertices of the graph G on a sphere S^{d-1}. A rounding technique uses this embedding to produce a cut in G. We call a rounding technique *oblivious* if it divides the sphere into two parts (of arbitrary shape) without looking at the embedding of the graph on the sphere, identifies these parts as two sides of a cut, and places vertices of G in the side of the cut that corresponds to the part in which their respective vectors lie in the solution to the SDP. The random hyperplane rounding technique is oblivious. A consequence of Corollary 6 is that for every $\epsilon > 0$, when the dimension d is large enough, then for every $\pi/2 < \beta < \pi - \epsilon$ the following holds: if a graph containing a single edge is embedded on the sphere S^{d-1} such that the endpoints of the edge are at distance between β and $\beta + \epsilon$ from each other, then a random hyperplane is the oblivious rounding technique with *highest* probability of cutting the edge (up to low order terms). Using this observation, it can be shown that Theorem 18 extends to every oblivious rounding technique.

We now discuss some nonoblivious rounding techniques that were used for MAX CUT or similar problems.

Feige and Goemans [6] suggested to use rotations in cases where the sphere has a distinct "north pole" (and "south pole"). Rotating each vector somewhat towards its nearest pole prior to cutting the sphere with a random hyperplane can lead to improved solutions. The usefulness of this approach was demonstrated for problems such as MAX 2SAT, where there is a unique additional vector v_0 that is interpreted as the $+1$ direction and can serve as the north pole. Trying to apply a similar idea to MAX CUT, one should first choose a direction that can serve as the north pole of the sphere. The proof of Theorem 19 can be extended to show that no choice of north pole and rotation function gives an approximation ratio better than α. This is based on the observation that choosing a north pole at random gives an oblivious rounding technique.

Zwick [19] has used a notion of *outward rotations* for several problems. For MAX CUT, Zwick observes that there are two "bad" angles for which the random hyperplane fails to give an expectation above α. One is the angle θ mentioned in Section 1. The other is the trivial angle 0, for which the contribution to the value of the geometric program is 0, and so is the contribution to the cut produced by a hyperplane. Hence worst case instances for the GW algorithm may have an arbitrary mixture of both types of angles for pairs of vertices connected by edges. In the extreme case, where all angles are 0 (though this would never be the optimal geometric embedding) it is clear that a random hyperplane would not cut any edge, whereas ignoring the geometric embedding and giving the vertices ± 1 values independently at random is expected to cut roughly half the edges. This latter rounding technique is equivalent to first embedding the vertices as mutually orthogonal unit vectors, and then cutting with a random hyperplane. *Outward rotation* is a technique of averaging between the two embeddings: the optimal geometric embedding on one set of coordinates and the mutually orthogonal embedding on another set of coordinates. It can be used in order to obtain approximation ratios better than α whenever a substantial fraction of the edges have angles 0, showing that essentially the only case when the geometric embedding (perhaps) fails to have integrality ratio better than α is when all edges have angle θ. We note that the cuts obtained by first applying outward rotations are not oblivious cuts, because two vertices that were mapped to the same unit vector may end up at different sides of the cut. Theorem 2 does not apply in this case. Nevertheless, we note that the idea of outer rotation does not seem to provide any advantage when the embedding is such that all edges make angles larger than $\pi/2$, which is the case of greatest interest.

Given a cut in a graph, a *misplaced* vertex is one that is on the same side as most of its neighbors. The number of edges cut can be increased by having misplaced vertices change sides. This local heuristic was used by Feige, Karpinski and Langberg [7] on top of the random hyperplane rounding technique to obtain an approximation ratio significantly better than $\alpha \simeq 0.87856$ for MAX CUT on graphs with bounded degree (for graphs of degree at most 3 the approximation ratio is better than 0.92). The best hyperplane cannot reproduce results which are as strong, because the proof technique of Theorem 19 applies also to graphs of very low degree, such as the 7-cycle.

Finally, let us note that the graphs used in the proof of Theorems 18 and 19 are not connected. They were later made connected via special gadgets in Section 4.5. This suggest that perhaps prior to applying an SDP, a graph should be decomposed into components of strong connectivity, or more generally, cut along sparse cuts. It would be interesting to see whether the idea of first finding sparse cuts is helpful in finding dense cuts.

Acknowledgements

The first author is the Incumbent of the Joseph and Celia Reskin Career Development Chair. Part of his work is supported by a Minerva grant, project number 8354 at the Weizmann Institute. The second author is the William Petschek Professor of Mathematics. His research is partially supported by a grant of the Israel Science Foundation.

We thank Howard Karloff, Michael Langberg, Guenter Rote and the referees for *Random Structures and Algorithms* for their comments on earlier versions of the paper.

References

[1] N. Alon and B. Sudakov. "Bipartite subgraphs and the smallest eigenvalue. *Combinatorics, Probability & Computing*, 9:1–12, 2000.

[2] F. Barahona and A. Mahjoub. "On the cut polytope". *Mathematical Programming*, 36:157–173, 1986.

[3] A. Baernstein, B. A. Taylor. "Spherical rearrangements, subharmonic functions, and $*$-functions in n-space". *Duke Math. J.* 43:245–268, 1976.

[4] Y. Benyamini. "Two point symmetrization, the isoperimetric inequality on the sphere and some applications". *Longhorn Notes, Univ. of Texas, Texas Funct. Anal. Seminar,* 53–76, 1983-1984.

[5] M. Deza and M. Laurent. *Geometry of Cuts and Metrics.* Springer, 1997.

[6] U. Feige and M. Goemans. "Approximating the value of two prover proof systems, with applications to MAX 2-SAT and MAX DICUT". In *Proceedings of third Israel Symposium on Theory of Computing and Systems*, 1995, 182–189.

[7] U. Feige, M. Karpinski and M. Langberg. "Improved Approximation of MAX-CUT on Graphs of Bounded Degree". *Electronic Colloquium on Computational Complexity* ECCC 021, 2000.

[8] M. Goemans and D. Williamson. "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming". *Journal of the ACM*, 42, 1115-1145, 1995.

[9] O. Goldreich, S. Goldwasser, D. Ron. "Property testing and its connection to learning and approximation". *JACM* 45(4), 653–750, 1998.

[10] J. Hastad. "Some optimal inapproximability results". In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing*, 1997, 1-10.

[11] W. Hoeffding. "Probability inequalities for sums of bounded random variables". *Journal of the American Statistical Association*, 58:13–30, 1963.

[12] H. Karloff. "How good is the Goemans-Williamson MAX CUT algorithm?". *SIAM J. Comput.*, Vol. 29, No. 1, pp.336–350, 1999.

[13] S. Mahajan and H. Ramesh. "Derandomizing semidefinite programming based approximation algorithms". In *Proceedings of the 36th Annual IEEE Symposium of Foundations of Computer Science*, 162–169, 1995.

[14] P. Mattila. *Geometry of sets and measures in Euclidean spaces. Fractals and rectifiability.* Cambridge Studies in Advanced Mathematics, 44. Cambridge University Press, Cambridge, 1995

[15] C. McDiarmid. "On the method of bounded differences". In J. Siemons, editor, *Surveys in Combinatorics: Invited Papers at the 12th British Combinatorial Conference*, 148–188. Cambridge University Press, 1989.

[16] V.D. Milman, G. Schechtman. "Asymptotic theory of finite-dimensional normed spaces". Lecture Notes in Mathematics, 1200. Springer-Verlag, Berlin-New York, 1986.

[17] G. Schechtman. "Concentration, results and applications". Handbook on Banach spaces, to appear.

[18] R. Schneider. Convex Bodies: The Brunn-Minkowski Theory. Encyclopedia of Mathematics and its Applications, Volume 44, Cambridge University Press, 1993.

[19] U. Zwick. "Outward rotations: a tool for rounding solutions of semidefinite programming relaxations, with applications to MAX CUT and other problems". In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing*, 1999, 679–687.

6 Appendix – proofs for integrality ratio

6.1 The dense version

Lemma 21 *For each $0 < \gamma < \pi/2$ the sphere S^{d-1} can be partitioned into $N = (O(1)/\gamma)^d$ equal volume cells, each of diameter at most γ.*

Proof: Let $\{u_i\}_{i=1}^m$ be a set of vectors on the sphere S^{d-1} of maximal cardinality among all sets with the property that the distance between any two vectors is at least $\gamma/4$. Consider the graph whose vertices are the numbers $1, \dots, m$ and (i, j) is an edge if and only if the distance between u_i and u_j is exactly $\gamma/4$. By moving some of the u_i-s we can and shall assume that the graph is connected (this is done only to slightly improve the numerical constants below). Let T be a spanning tree of that graph. We shall consider it as a directed graph with edges pointing towards the root. Divide the sphere into the Voronoi cells $\{V_i\}$ induced by the u_i-s; i.e., V_i is the set of all points closer to u_i than to any other u_j (if a point is equidistant from two or more of the u_i-s put it in the cell with lowest index, say). Note that V_i contains a ball of radius $\gamma/8$ around u_i and is contained in a ball of radius $\gamma/4$ around u_i.

Let C be a constant (bounded by a universal constant) such that $N = (C/\gamma)^d$ is an integer and $1/N$ is smaller than the volume of a ball of radius $\gamma/8$ on the sphere. We are now going to change the Voronoi cells slightly starting with the cells which correspond to the "leaves" of T (i.e, the vertices which have only outgoing edges). For each such cell V_i pick a subset U_i of volume at most $1/N$ such that the volume of $C_i = V_i \setminus U_i$ is divisible by $1/N$. Look now at the next level of the tree (the vertices which are connected to leaves) and for each of them define tentatively $C_i' = V_i \cup \cup U_j$ where the last union is over all leaves j such that (j,i) is an edge. let $U_i \subset V_i$ be of measure at most $1/N$ such that the measure of $C_i = C_i' \setminus U_i$ is divisible by $1/N$. Continue in that manner towards the next level and on. Note that in changing each V_i to C_i we only borrow from neighbors and consequently the diameter of each C_i is at most γ. Also, the measure of each C_i is divisible by $1/N$. This is clear for all i except maybe the root, by construction. Since N is an integer it must hold also for the root. Finally, divide each C_i to subcells of measure exactly $1/N$ then each of the diameters is at most γ and their number is N which is $(O(1)/\gamma)^d$. $\square$

Remark: Through the use of Lemma 21 we prove the existence of a graph G. To algorithmically construct a graph G, we can follow the proof of Lemma 21. The fact that the sphere is continuous whereas computation is discrete might prevent us from achieving a partition of the sphere into exactly equal cells. However, with "sufficient" numerical precision the cells can be made equal up to an error that is negligible compared to ϵ. This will suffice for producing a graph G with integrality ratio below $\alpha + \epsilon$.

Proof of Theorem 5: Given a hyperplane H through the origin we denote by S_0 its intersection with the sphere S^{d-1}, by S_+ one of the two half spheres remaining in $S^{d-1} \setminus S_0$ and by S_- the other half sphere. We also denote by $\sigma = \sigma_H$ the reflection with respect to H. (For example, if H consists of all vectors whose first component is zero, then σ is the map which changes the sign of the first component of each vector not in H and is the identity on H.) Note that for any (measurable) set $A \subset S^{d-1}$, $\mu(A) = \mu(\sigma(A))$ and also that if x is in S_- and y in S_+ then the distance between x and $\sigma(y)$ is smaller than the distance between x and y.

Given a subset A of S^{d-1} we define its symmetrization A^* with respect to H in the following way: If x is a point in $A \cap S_-$ such that $\sigma(x)$ is not in A then we include $\sigma(x)$ in A^* and do not include x in A^*. All other points of A will be also in A^* and no other points will be added to A^*. Formally,

$$A^* = [A \cap (S_+ \cup S_0)] \cup [A \cap S_- \cap \sigma(A \cap S_+)] \cup [\sigma(A \cap S_- \setminus \sigma(A \cap S_+))].$$

The point is that by "pushing points of A up" whenever possible, A becomes closer to being a cap. Note that $\mu(A^*)$ is equal to $\mu(A)$. We shall next show that $\mu_\rho(A^*)$ is larger than or equal to $\mu_\rho(A)$.

We present the basic idea of the proof that $\mu_\rho(A^*) \geq \mu_\rho(A)$. We say that a pair of points x, y contributes to $\mu_\rho(A)$ (respectively, to $\mu_\rho(A^*)$) if exactly one of the points is in A (respectively, A^*) and $d(x,y) \geq \rho$. We shall show a one-to-one mapping from pairs of points contributing to $\mu_\rho(A)$ to pairs of points contributing to $\mu_\rho(A^*)$. Intuitively, this means that $\mu_\rho(A^*)$ is larger. Technically this is not a rigorous argument, because the number of pairs of points in infinite (and uncountable). However, in our case it can be made rigorous (by partitioning pairs of points into a finite number of sets and arguing with respect to the measure of each set separately), though details are omitted.

Consider first a point $x \in S_0$. We claim that the contribution of pairs of points in which it participates to $\mu_\rho(A^*)$ is identical to the contribution to $\mu_\rho(A)$. Note first that $x \in A$ iff $x \in A^*$. As this holds for any other point $y \in S_0$, pairs x, y both in S_0 contribute the same to $\mu_\rho(A^*)$ and $\mu_\rho(A)$. Consider now a point $y \notin S_0$. The idea is to consider also $\sigma(y)$ at the same time. Now $d(x,y) = d(x,\sigma(y))$, and $|\{y,\sigma(y)\} \bigcap A| = |\{y,\sigma(y)\} \bigcap A^*|$. This implies that the two pairs x, y and $x, \sigma(y)$ contribute the same to $\mu_\rho(A^*)$ and $\mu_\rho(A)$. Going over all y we have the same claim for x.

Consider now a point $x_+ \in S_+$ and $x_- = \sigma(x_+)$. As $|\{x_+, x_-\} \bigcap A| = |\{x_+, x_-\} \bigcap A^*|$, the pair x_+, x_- contributes the same to $\mu_\rho(A^*)$ and $\mu_\rho(A)$.

It remains only to consider pairs x, y where neither point is in S_0 and $x \neq \sigma(y)$. Also here the idea is to consider them together with their reflections. So consider $x_+ \in S_+$ and $x_- = \sigma(x_+)$, and $y_+ \in S_+$ and $y_- = \sigma(y_+)$. They contribute 4 previously unconsidered pairs, namely x_+, y_+ and x_+, y_- and x_-, y_+ and x_-, y_-. Of these four pairs, it is easy to see that the number of pairs cut by A is equal to the number of pairs cut by A^*. Note also that $d(x_+, y_+) = d(x_-, y_-) < d(x_+, y_-) = d(x_-, y_+)$. A simple case analysis shows that of the four pairs, A^* always cuts pairs that are at least as far apart as the pairs cut by A. Hence the contribution to $\mu_\rho(A^*)$ cannot be smaller than that to $\mu_\rho(A)$. This finishes the (sketch of) proof that $\mu_\rho(A^*) \geq \mu_\rho(A)$.

The rest of the proof is very similar to the one exposed in [17]. The idea is to show that a repetitive operations of the kind above with many different hyperplanes reduces the set A to a cap. Some classical Mathematics needs to be used here.

Consider the metric space $\mathcal{C}$ of all closed subsets of S^{d-1} with the Hausdorff metric. (Two sets A and B are of distance at most ϵ if for all $a \in A$ there is a $b \in B$ whose distance from a is at most ϵ and the same holds when interchanging the roles of A and B.) Fix $A \in \mathcal{C}$ and consider the set $\mathcal{B} \subseteq \mathcal{C}$ of all sets $B \in \mathcal{C}$ satisfying:

- $\mu(B) = \mu(A)$
- For all $0 < \rho < \pi \quad \mu_\rho(B) \geq \mu_\rho(A)$, and
- for all $\epsilon > 0 \;\; \mu(B_\epsilon) \leq \mu(A_\epsilon)$.

Here B_ϵ denotes the set of all points whose distance to some point in B is at most ϵ. Note that $\mathcal{B}$ is not empty, as it contains A. Moreover, the symmetrization process preserves membership in $\mathcal{B}$. This was already proved for the first two properties of $\mathcal{B}$. For the third property we need to show that for every hyperplane, every set A and every $\epsilon > 0$, $\mu(A^*_\epsilon) \leq \mu(A_\epsilon)$. This is known, and used for example in the proof of the isoperimetric inequality that is reviewed in [17]. We added the condition $\mu(B_\epsilon) \leq \mu(A_\epsilon)$ to the definition of $\mathcal{B}$ in order to ensure that $\mathcal{B}$ is closed in $\mathcal{C}$; i.e., if B_n is a sequence of sets in $\mathcal{B}$ which tends in the Hausdorff metric to B then B is also in $\mathcal{B}$. We state it as a lemma and delay the proof.

Lemma 22 *$\mathcal{B}$ is closed in $\mathcal{C}$*

Fix a point $x_0 \in S^{d-1}$ and let C be the closed cap centered at x_0 with measure $\mu(A)$. To conclude the proof it is enough to prove that $C \in \mathcal{B}$. For any hyperplane H with $x_0 \notin H$ we denote by S_+ the open half sphere determined by H, containing x_0. $B \to \mu(B \cap C)$ is upper semi continuous on $\mathcal{C}$; i.e., if B_n is a sequence of sets in $\mathcal{C}$ which tends in the Hausdorff metric to B then $\mu(B \cap C)$ is at least $\limsup \mu(B_n \cap C)$.

$\mathcal{C}$ with the topology induced by the Hausdorff metric is a compact set and since $\mathcal{B}$ is a closed subset, $\mu(B \cap C)$ achieves its maximum for some $B \in \mathcal{B}$. It is enough to show that B can be taken to be equal to C. If this is not the case then $\mu(B \setminus C) = \mu(C \setminus B) > 0$.

Denote by $B(z,r)$ the ball of radius r centered at z. By Lebesgue's density theorem (see for example [14], page 40) there are points $x \in B \setminus C$ and $y \in C \setminus B$ such that, for some small $r > 0$, $\mu(B(x,r) \cap (B \setminus C)) > 0.99\mu(B(x,r))$ and $\mu(B(y,r) \cap (C \setminus B)) > 0.99\mu(B(y,r))$. Let H be the hyperplane perpendicular to the segment $[x,y]$ and crossing it at the midpoint $(x+y)/2$. Note that y is necessarily closer to x_0 then x is, because C is a cap centered at x_0, and the neighborhood of y is mostly in C whereas the neighborhood of x is mostly not in C. Let us denote by S_+ the half sphere containing x_0. Then $x \in S_-$ and $y \in S_+$. Applying the symmetrization $B \to B^*$ with respect to this hyperplane, a large fraction of $B(x,r)$ will be transferred into $B(y,r)$ while no point of $C \cap B$ is transferred to a point which is not in C (observe that the symmetrization of C is C). Thus, $\mu(B^* \cap C) > \mu(B \cap C)$. Since B^* also belongs to $\mathcal{B}$ we get a contradiction. □

Proof of Lemma 22: Note that the function $\mu(\cdot)$ is not continuous on $\mathcal{C}$. (Consider for example a sequence of finite sets $\{A_n\}_{n=1}^{\infty}$ such that $A_n \to S^{n-1}$. Then $\mu(A_n) = 0$ for all n but $\mu(S^{n-1}) = 1$.) This shows, for example, that the first condition in the definition of $\mathcal{B}$ is not a closed condition and explains the reason for the somewhat cumbersome proof below. What saves the day is the easily established fact that, in spite of this discontinuity, for all closed B, $\mu(B_\epsilon) \to \mu(B)$.

Let B_n be a sequence in $\mathcal{B}$ tending to B in the Hausdorff metric. For every $\epsilon, \delta > 0$ $(B_n)_{\epsilon+\delta} \supseteq B_\epsilon$ if n is large enough and $\mu((B_n)_{\epsilon+\delta}) \leq \mu(A_{\epsilon+\delta})$. Sending δ to zero we get $\mu(B_\epsilon) \leq \mu(A_\epsilon)$. This shows that B satisfies the last requirement and also (sending ϵ to zero) that $\mu(B) \leq \mu(A)$. That $\mu(B) \geq \mu(A)$ follows from $\mu(B_\epsilon) \geq \mu(B_n) = \mu(A)$ for any $\epsilon > 0$ provided n is large enough. This shows that B satisfied the first requirement of belonging to $\mathcal{B}$. To show that it satisfies the second requirement, note first that, since $\mu(B_{2\epsilon} \setminus B)$ tends to zero as ϵ tends to zero,

$$a(\epsilon) = \mu^2(\{(x,y);\ x \in B_{2\epsilon},\ y \notin B, \text{ the distance between } x \text{ and } y \text{ is at least } \rho\})$$

tends, as ϵ tends to zero, to $\mu_\rho(B)$. On the other hand, if n is large enough to ensure $B \subseteq (B_n)_\epsilon \subseteq B_{2\epsilon}$, then $a(\epsilon) \geq \mu_\rho((B_n)_\epsilon)$. Finally, since $\mu((B_n)_\epsilon \setminus B_n) \leq \mu(A_\epsilon \setminus A)$ and this tends to zero as ϵ tends to zero, we get that $\mu_\rho((B_n)_\epsilon)$ tends, as ϵ tend to zero, to $\mu_\rho(B_n)$ *uniformly* in n. So, given a $\delta > 0$ find an $\epsilon > 0$ such that $\mu_\rho((B_n)_\epsilon) \geq \mu_\rho(B_n) - \delta$ for all n and also $\mu_\rho(B) \geq a(\epsilon) - \delta$. Then if n is large enough,

$$\mu_\rho(B) \geq a(\epsilon) - \delta \geq \mu_\rho((B_n)_\epsilon) - \delta \geq \mu_\rho(B_n) - 2\delta \geq \mu_\rho(A) - 2\delta.$$

Since this holds for any $\delta > 0$ we get that B satisfies the second requirement as well. □

6.2 The sparse version

We shall use standard bounds on large deviations of random variables. One bound that we use is Chebyschev's inequality.

Lemma 23 *For a random variable X, let $\mu = E[X]$ denote its expectation, let $\sigma^2 = E[(X - E[X])^2]$ denote its variance, and let σ denote the standard deviation (the positive square root of σ^2). Then for every $\lambda > 0$,*

$$P[|X - \mu| \geq \lambda\sigma] \leq \frac{1}{\lambda^2}.$$

The other bound that we use deals with independent random variables. It is taken from [15], and the proof is based on [11] and is omitted.

Lemma 24 *Let $X_1, \ldots, X_n$ be independent random variables, with $0 \leq X_i \leq 1$ for each i, let $\bar{X} = (\sum X_i)/n$ and let $\mu = E[\bar{X}]$. Then for every $0 < \epsilon < 1$*

$$P[\bar{X} - \mu \geq \epsilon\mu] \leq e^{-\frac{1}{3}\epsilon^2 n\mu}.$$

The proof of the following proposition illustrates the use of the large deviation bounds. Future proofs will be given in less detail.

Lemma 25 *Let G be a graph with n vertices and m edges, and let G be nearly regular in the sense that its highest degree Δ is bounded by some absolute constant times the average degree. Select each vertex of G into G' independently with probability $p > 1/\Delta$. Let n' and m' denote the number of vertices and edges in G'. Then the expectation of n' is pn and of m' is p^2m. The probability that n' deviates by an ϵ fraction from its expectation is at most $e^{-\frac{1}{3}\epsilon^2 pn}$, and the probability that m' deviates by an ϵ fraction from its expectation is $O(1/\epsilon^2 pn)$.*

Proof: Let X_i be the random variable that is 1 if vertex i is chosen into G' and 0 otherwise. Then $n' = \sum_{i=1}^{n} X_i$. For every i, $E[X_i] = p$. By linearity of expectation, $E[\sum X_i] = pn$. The bound on the probability that n' deviates by ϵpn from its expectation follows from Lemma 24, substituting $\mu = p$.

Now we turn to m'. For an edge (i,j) in G, let Y_{ij} be the random variable that is 1 if edge (i,j) is chosen into G' and 0 otherwise. Then $m' = \sum_{(i,j)} Y_{ij}$. As each edge is selected only if its two endpoints are selected, we have $E[Y_{ij}] = p^2$. By linearity of expectation, $\mu = E[m'] = p^2m$. Let us now bound the variance of m'. For two edges in G, their selection into G' is independent, unless they share a vertex. Independent pairs of edges do not contribute to the variance. This gives:

$$E[(m' - E[m'])^2] = E[(\sum_{(i,j)}(Y_{ij} - p^2))^2] = E[\sum_{(i,j),(k,l)}(Y_{ij} - p^2)(Y_{kl} - p^2)]$$

$$= \sum_{(i,j)} E[(Y_{ij} - p^2)^2] + \sum_{(i,j),(i,l)} E[(Y_{ij} - p^2)(Y_{il} - p^2)] + 0 \leq p^2m + O(p^3\Delta m) = O(p^3\Delta^2 n).$$

(The last equality holds because we assumed that $p > 1/\Delta$.) Hence for the random variable m' we have $\sigma = O(p\Delta\sqrt{pn})$. Comparing σ and μ we get that

$$\frac{\sigma}{\mu} \leq O(\frac{p\Delta\sqrt{pn}}{p^2m}) \leq O(\frac{\sqrt{pn}}{pn}) = O(1/\sqrt{pn})$$

Now the concentration of m' around its expectation follows from Chebyschev's inequality. □

Lemma 25 implies that with high probability, $n' \simeq pn$ and $m' \simeq p^2m$, meaning in particular that the average degree in G' is $\Theta(p\Delta)$. Proposition 26 shows that G' is nearly regular, a property that significantly simplifies the proofs that follow.

Proposition 26 *Let $p = c\log(n/\Delta)/\Delta$ for a sufficiently large $c > 0$. Then with high probability, the maximum degree Δ' of G' is at most $2p\Delta$.*

Proof: Consider an arbitrary vertex $v \in G$. The probability that it is selected into G' is p. The expectation of the number of its neighbors that are selected into G' is at most $p\Delta$. By Lemma 24, the probability that the number of neighbors is twice as much is $e^{-p\Delta/3}$. Hence the probability that v gives a vertex in G' with degree above $2p\Delta$ is at most $pe^{-p\Delta/3}$. Using $p = c\log(n/\Delta)/\Delta$ for a sufficiently large $c > 0$, this probability is much smaller than $1/n$. Using the union bound over all vertices of G, no vertex of G' will have large degree. $\square$

Summarizing properties of G' that we established in Lemma 25 and Proposition 26, we have:

Corollary 27 *Let $\epsilon > 0$ be an arbitrarily small constant and let c be a sufficiently large constant. Given a nearly regular graph G and $p = c\log(\frac{n}{\Delta})/\Delta$, then a random vertex induced subgraph G' that contains every vertex of G independently with probability p is* typical *with high probability. Here typical means that $(1-\epsilon)pn \le n' \le (1+\epsilon)pn$, $(1-\epsilon)p^2m \le m' \le (1+\epsilon)p^2m$, and $\Delta' \le 2p\Delta$.*

Proof of Lemma 8: We use the notation (X, Y) to denote the set of edges between disjoint sets X and Y, and $|(X, Y)|$ to denote the number of these edges. Typically, $X \bigcup Y$ will be all vertices of the respective graph, in which case we say that (X, Y) is a cut. Let G have the vertex set V and let the random subgraph G' have the vertex set V'. For a cut (A, B) in G, its *projection* on G' is the cut (A', B') where $A' = A \bigcap V'$ and $B' = B \bigcap V'$.

Consider a maximum cut (A, B) in G that cuts $\text{opt}_d \cdot m$ edges. Then the projected cut (A', B') is expected to cut $p^2\text{opt}_d \cdot m$ in G'. By Chebyschev's inequality, with probability $1 - o(1)$, $|(A', B')|$ does not deviate from the expectation by more than $o(p^2\text{opt}_d \cdot m)$. Using Lemma 25 we get that $m' \simeq p^2m$. Hence (A', B') cuts roughly an opt_d fraction of the edges of G'. As the maximum cut of G' cuts at least as much, we have shown that $\text{opt}_s \ge \text{opt}_d - \epsilon$ with high probability.

The hard part of the Lemma is to show that $\text{opt}_s \le \text{opt}_d + \epsilon$. Each cut in G cuts at most $\text{opt}_d \cdot m$ edges. As above, we can show that its projection on G' is expected to cut at most roughly $\text{opt}_d \cdot m'$ edges. However, there is some probability that the projection cuts many more edges. The probability of a large deviation (as computed by Chebyschev's inequality) is small, but still much larger than 2^{-n}. Because the number of cuts in G is 2^n, we cannot use the union bound to argue that none of the cuts in G induces an abnormally large cut in G'.

To overcome this problem we do two things. One is to show that it suffices to consider projections of only a small number of *special cuts* in G, rather than all possible cuts in G. The other is for each special cut in G, to more carefully analyse the probability that the induced cut in G' is abnormally large, and prove bounds stronger than those implied by Chebyschev's inequality.

Definition 4 *For $0 \le \epsilon \le 1$, a cut in a graph is* ϵ-maximum *if the number of edges in the cut is at least $(1-\epsilon)$ times the number of edges in the maximum cut in the graph.*

Lemma 28 *There is a collection of* $2^{O(n/\epsilon^4\Delta)}$ special cuts *in* G*, such that with high probability over the choice of* G'*, the projection of one of these cuts is an* ϵ*-maximum cut in* G'*.*

In [9], a lemma similar to Lemma 28 is proved for the special case that $m = \Omega(n^2)$. The proof of [9] can be adapted to apply also to lemma 28. We assume Lemma 28 and continue with the proof of Lemma 8. The proof of Lemma 28 is sketched later, for completeness.

Intuitively, we would like to now prove the following:

- Consider an arbitrary cut (A, B) in G, and recall that $|(A, B)| \leq \text{opt}_d \cdot m$. With probability $2^{-\Omega(pn)}$ over the choice of G', $|(A', B')| \geq (1 + \epsilon)p^2\text{opt}_d \cdot m$.

When $m = \Omega(n^2)$, this follows from a proof in [9]. (Note: the $\Omega(pn)$ notation hides a constant that depends on ϵ.) But in our case, when m is small, this might not be true. To overcome this problem, we condition on G' being typical. Hence we prove the following modified property.

Proposition 29 *For an arbitrary cut* (A, B) *in a nearly regular graph* G*, the probability that* $|(A', B')| \geq (1 + \epsilon)p^2 \text{opt}_d \cdot m$ *is at most* $2^{-\Omega(\epsilon^2 pn)}$*, where probability is computed over the choice of* G' conditioned on G' being typical in the sense of Corollary 27.

Proof: As G' is assumed to be typical, the maximum degree in G' is within a factor of 4 of the average degree. This implies that if $|A'| < |B'|/16$ then (A', B') cuts at most $1/2 < \text{opt}_d$ the edges of G'. Hence we can assume that $|A'| \geq |B'|/16$. This in turn implies that we can assume that $|A| = \Omega(n)$, because otherwise, with probability at least $1 - 2^{-\Omega(pn)}$, $|A'| < |B'|/16$ (this can be proved as in Lemma 25). Similarly, we can assume that $|B| = \Omega(n)$.

To simplify the notation in the arguments that follow, we assume that $|(A, B)| = \text{opt}_d \cdot m = \Omega(m)$. This assumption can be made without loss of generality, because the larger $|(A, B)|$ is, the higher the probability that $|(A', B')| \geq (1 + \epsilon)p^2\text{opt}_d \cdot m$. (We waited with this assumption until we established $|A| = \Omega(n)$, because if $|A| = o(n)$ and $\Delta = O(m/n)$ the assumption does not make sense.)

Now consider the selection of A'. In expectation, $|(A', B)| = p|(A, B)|$. We claim that the probability of an $\epsilon \cdot p \cdot \text{opt}_d \cdot m$ deviation above the expectation is exponentially small in $\epsilon^2 pn$. The claim is proved using Lemma 24. Treat each vertex v of A as a random variable that has value $|(v, B)|/\Delta \leq 1$ if $v \in A'$ (which happens with probability p) and 0 otherwise. Hence in the terminology of Lemma 24, $\mu = p|(A, B)|/\Delta|A| = \Theta(p)$. Here we used the assumptions that $|(A, B)| = \Theta(m)$, $|A| = \Theta(n)$ and $\Delta = \Theta(m/n)$. Lemma 24 implies the claim.

Now we select vertices of B into B'. In expectation, $|(B', A')| = p|(B, A')|$. Again we claim that the probability of an ϵ-fraction deviation above the expectation is exponentially small in $\epsilon^2 pn$. To prove the claim we use the conditioning that G' is typical, which implies that the largest degree of a vertex from B' into A' is at most $2p\Delta$. View each vertex v of B as a random variable that has value $|(v, A')|/2p\Delta \leq 1$ if $v \in B'$ (which happens with probability p) and 0 otherwise. (Note that by our conditioning, if $|(v, A')| > 2p\Delta$ then necessarily $v \notin B'$, implying that the value of the random variable associated with

v is indeed never larger than 1.) In the terminology of Lemma 24, we have that $\mu \leq p|(B, A')|/2p\Delta|B| = \Theta(p)$. Using Lemma 24 the claim is proved.

The two claims above imply that conditioned on G' being typical, with probability at most $2^{-\Omega(\epsilon^2 pn)}$ we have $|(A', B')| > (1+\epsilon)p^2\text{opt}_d \cdot m$. $\square$

Let us summarize the properties that we have proved so far.

- With high probability over the choice of G', G' is typical, meaning that $n' \simeq pn$, $m' \simeq p^2 m$, and $\Delta' < 2p\Delta$ (Corollary 27).
- There are $2^{O(n/\epsilon^4\Delta)}$ *special* cuts in G. With high probability over the choice of a random subgraph G', a near maximal cut in G' can be obtained as a projection of one of the special cuts of G (Lemma 28).
- For an arbitrary special cut (A, B), the probability over the choice of G' that G' is typical and $|(A', B')| > (1+\epsilon)p^2\text{opt}_d \cdot m$ both hold is $2^{-\Omega(pn)}$.

The last property together with the upper bound on the number of special cuts imply (using the fact that we can choose $p > 1/\epsilon^6\Delta$) that it is highly unlikely that when G' is typical there is a projection of a special cut with more than $(1+\epsilon)p^2\text{opt}_d \cdot m$ edges. But as we showed that with high probability G' is typical and that a projection of a special cut is ϵ-maximum in G', we conclude that with high probability $\text{opt}_s \leq \text{opt}_d + O(\epsilon)$. This completes the proof of Lemma 8. $\square$

Proof of Lemma 28: Let $k = \Theta(1/\epsilon)$. Consider a partition of the vertices of G into k roughly equal size sets $V_1 \dots V_k$, where the number of edges in the subgraph induced by each set is $O(\epsilon^2 m)$. (Lemma 25 implies that with high probability, a random partition on V into k sets will result in subgraphs with roughly n/k vertices and m/k^2 edges.) In addition, select at random a sequence S of $\Theta(k^4 n/\Delta)$ distinct vertices from G, and partition in into k equal length consecutive subsequences $S_1 \dots S_k$. (Hence $|S_i| = \Theta(k^3 n/\Delta)$ for every i.) Consider all $2^{|S|}$ possibilities of two-coloring the vertices in S as *left* and *right*. Each such possibility will define a unique cut in G as follows. A vertex $v \in V_i$ is placed on the left side if more of its neighbors from S_i are colored *right* than *left*. Otherwise it is placed on the right side. We say that S is *good* for G and $V_1, \dots, V_k$ if at least one of the $2^{|S|}$ cuts generated above forms an ϵ-maximum cut.

Proposition 30 *For G and $V_1, \dots, V_k$ as above, almost all choices of S are good.*

Let us first note some subtleties involved in the proof of proposition 30. Intuitively, the two-coloring of S that induces a near maximum cut on G is when S is two-colored consistently with the maximum cut of G. However, this need not work. Consider for example a graph containing an independent set of size $n/2$, a clique of size $n/2$, and the two are connected by a complete bipartite graph. For a random S, roughly half the subsequences S_i will have a majority of vertices from the clique. If clique vertices are all colored *right* and the other vertices of S_i are colored *left* (which is consistent with the true maximum cut) then this induces a very bad cut on G, as all vertices of the respective V_i end up on the left side. To get a near maximum cut of G, we will need to two-color S in a way that colors some of its clique vertices *left*.

Proposition 30 is proved in [9] for the case $m = \Omega(n^2)$ (in which case $|S| = O(1/\epsilon^4)$). The proof extends to our case were m is smaller, because we make S larger. For completeness, we sketch the basic idea of the proof. Consider an optimal cut (A, B) in G. For a given S, we define inductively a sequence of cuts (A_i, B_i) for $0 \leq i \leq k$. Cut (A_0, B_0) is just (A, B). For $i > 0$, cut (A_i, B_i) is the cut one gets by placing the vertices of $V_1, \ldots, V_{i-1}$ and $V_{i+1}, \ldots, V_k$ exactly as in the cut (A_{i-1}, B_{i-1}), and the vertices of V_i according to a two-coloring of S_i. The two coloring of S_i that is used for this purpose is the one that maximizes the number of edges in the resulting cut (A_i, B_i). Observe that the cut (A_k, B_k) is a cut induced by a two-coloring of S. We shall show that with high probability, for every $1 \leq i \leq k$, $|(A_i, B_i)| \geq |(A_{i-1}, B_{i-1})| - \epsilon^2 m$. As $k = \Theta(1/\epsilon)$, this implies that $|(A_k, B_k)| \geq |(A, B)| - \epsilon m$, establishing that (A_k, B_k) is an ϵ-maximum cut.

The number of edges connecting vertices within a set V_i is $O(\epsilon^2 m)$, and the number of such sets is $O(1/\epsilon)$. Hence the total number of edges within all sets V_i is $O(\epsilon m)$, and their contribution to the final cut is negligible. For the current sketch of proof, we shall assume that there are no edges at all within the sets V_i. This simplifies the presentation of the proof. A more detailed proof that takes into account the effect of these edges can be found in [9].

We call a vertex $v \in V_i$ *misplaced* if it is in $A_{i-1} \bigcap B_i$, or in $B_{i-1} \bigcap A_i$. Note that only misplaced vertices contribute to the difference between (A_i, B_i) and (A_{i-1}, B_{i-1}). For a vertex $v \in V_i$, let $d_A(v)$ denote the number of neighbors that it has in A_{i-1} and let $d_B(v)$ denote the number of neighbors that it has in B_{i-1}. The vertices of V_i can be partitioned into three sets: *strongly-B* for which $d_A > d_B + \epsilon\Delta$, *strongly-A* for which $d_B > d_A + \epsilon\Delta$, and *neutral* for which $|d_A - d_B| \leq \epsilon\Delta$. Consider the sample S_i and a two-coloring of it consistent with (A_{i-1}, B_{i-1}). Let v be a vertex that is strongly-B. The number of B_{i-1}-neighbors of v in S_i is expected to be $k^3 d_B(v)/\Delta$. The number of A_{i-1}-neighbors of v in S_i is expected to be $k^3 d_A(v)/\Delta \geq k^3 d_B(v)/\Delta + k^2$. The difference of k^2 between the two is statistically significant because $k^3 d_B(v)/\Delta = O(k^3)$. Hence S_i will almost surely send v to side B_i, in which case the contribution of edges connected to v to $|(A_i, B_i)|$ is at least as large as their contribution to $|(A_{i-1}, B_{i-1})|$. The fraction of strongly-B vertices that are sent to the side A_i is below ϵ with high probability (over the choice of S_i), and the number of edges that are effected by these vertices is $O(\epsilon|V_1|\Delta) = O(\epsilon^2 m)$. A similar argument applies to strongly-A vertices in V_i. Hence we are left to deal with neutral vertices. If a neutral vertex v is misplaced, then in terms of edges adjacent to v, the number of edges cut in (A_i, B_i) differs from the number of edges cut in (A_{i-1}, B_{i-1}) by at most $\epsilon\Delta$, which totals $O(\epsilon\Delta|V_1|) = O(\epsilon^2 m)$ edges over all neutral vertices in V_i. Summing up the contribution of all misplaced vertices (strongly-B, strongly-A and neutral), we obtain that $|(A_i, B_i)| \geq |(A_{i-1}, B_{i-1})| - O(\epsilon^2 m)$, as desired. This completes the sketch of proof for Proposition 30.

We would now like to prove a proposition similar to Proposition 30 for G'. For this we need to partition the vertices of G' into $V'_1, \ldots, V'_k$. We do it in the unique way satisfying $V'_i \subset V_i$ for every i. We note that Lemma 25 can be used to show that with high probability over the choice of G', the V'_i will be of roughly equal cardinality, and the number of edges in subgraphs induced by each V'_i is $O(\epsilon^2 m')$. Choosing a random sequence S' of vertices in G' and two-coloring it in all $2^{|S'|}$ possible ways induces $2^{|S'|}$ cuts in G'. The proof of Proposition 30 shows that with high probability, one of them, say (A', B'), is ϵ-maximum for G'.

Now we reach an important point in the argument. We want the sequences S for G and S' for G' to be the same. Then, the two-coloring of S that induces the ϵ-maximum cut (A', B') in G' also induces some cut (A, B) in G. The important point is that (A', B') is precisely the projection on G' of the cut (A, B), which is the property needed to complete the proof of Lemma 28.

Why can we take S' to be the same as S? Note that $n'/\Delta' = \Theta(pn/p\Delta) = \Theta(n/\Delta)$. Hence the intended lengths of S and S' are similar. The main problem is that if S is fixed prior to the choice of G', then the vertices of S need not be contained at all in G'. However, the proof that (with high probability) S is good for G' works anyway. This can be formally shown by considering a new graph G'' whose vertex set V'' is the union over V' and S. We partition G'' into k parts $V_1'', \ldots, V_k''$ where V_i'' is the projection of V_i on G''. Note that V_i'' is identical to V_1' except for some vertices of S that enter V_i''. For G'' we can take $S' = S$. The sequence S' is indeed a random sequence of vertices in G'' (because both G' and S are random). Hence the proof of Proposition 30 does go through, and an ϵ-maximum cut of G'' is obtained as a projection of a cut of G. Now it remains to observe that G' and G'' differ in at most $|S|$ vertices, and this is a low order term compared to $\epsilon n'$ (assuming that $k^5 \ll \log(n/\Delta)$, which is indeed true because in our intended application we may take n/Δ to be arbitrarily large). Proposition 26 implies that G'' is also nearly regular. It follows that the maximum cuts in G' and G'' differ by $O(\epsilon m')$ edges, which can be neglected. $\square$