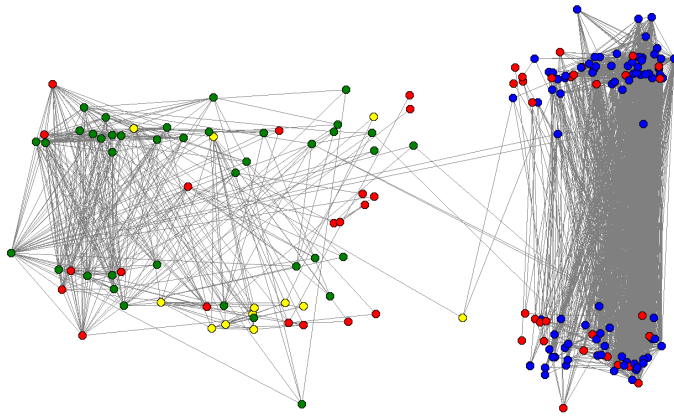## Maximum cut and related problems

Figure 1: The graph of Renato Paes Leme's friends on the social network Orkut, the partition to top and bottom parts is an approximation of the maximum cut while the partition to left and right sides is an approximation of the sparsest cut

In this lecture, we will discuss three fundamental NP-hard optimization problems that turn out to be intimately related to each other. The first is the problem of finding a maximum cut in a graph. This problem is among the most basic NP-hard problems. It was among the first problems shown to be NP-hard (Karp [1972]).[1] The second problem is estimating the expansion[2] of a graph. This problem is related to isoperimetric questions in geometry, the study of manifolds, and a famous inequality by Cheeger [1970]. The third problem is estimating mixed norms of linear operators. This problem is more abstract than the previous ones but also more versatile. It is closely related to a famous inequality by Grothendieck [1953].

How are these seemingly different problems related? First, it turns out that all three of them can be phrased in terms of optimizing a quadratic polynomial over the hypercube. Second, all three have beautiful approximation algorithms based on degree-2 sum-of-squares which give the best[3] known guarantees for the respective problem.[4] Third, any improvement in the approximation guarantee of these algorithms would refute the *Unique Games Conjecture* (Khot [2002]) or the closely related *Small-Set Expansion Hypothesis* (Raghavendra and Steurer [2010]).

[1] This problem also shows that small syntactic changes in the problem definition can make a big difference for its computational complexity. The problem of finding the minimum cut in a graph has a polynomial-time algorithm, whereas a polynomial-time algorithm for finding a maximum cut is unlikely because it would imply P=NP.

[2] In the literature, several different terms, e.g., expansion, conductance, and sparsest-cut value, are used to describe closely related parameter of graphs. In these notes, we will not distinguish between these parameters and stick to the term expansion.

[3] for a suitable notion of "best"

[4] One interesting feature of these algorithms also is that we don't know how to achieve the same approximation guarantees with different algorithmic techniques. In this sense, sum-of-squares brings something unique to the table for these problems.
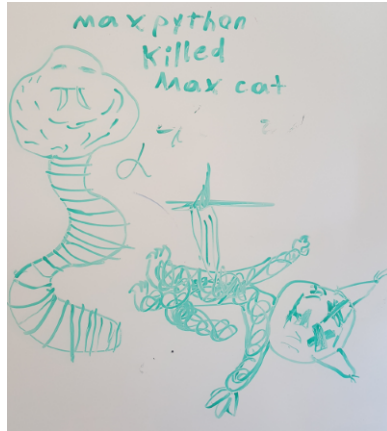
Approximating the maximum cut

We now define the Max Cut problem:

1. Problem (Max Cut). Given a graph, find a bipartition of the vertex set that cuts as many edges as possible.

For every graph G, we let $\text{maxcut}(G)$ denote the fraction of edge cut by an optimal Max Cut solution for G. Max cut is a natural graph problem, motivated by applications such as efficient design of electric circuits or communication networks, but has also found uses to questions in statistical physics such as finding an energy minimizing configuration in the ising model.[5] But more than this, the Max Cut problem has served as a prototypical testbed for algorithmic techniques that have found extensive and sometimes surprising applications. For example, a variant of the algorithm for Max Cut that we present in this lecture has been used for the *Phase Retrieval* problem that appears in X-ray crystallography, diffraction imaging, astronomical imaging and other applications (see (Candès et al. [2013]; Waldspurger et al. [2015])).

[5] For more applications, see also the survey (Deza and Laurent [1994]).

Since a random bipartition cuts half the edges in expectation, there is an efficient randomized[6] algorithm to find a bipartition that cuts at least $\frac{1}{2} \cdot \text{maxcut}(G)$ of the edges. We say that this algorithm achieves an approximation factor of $\frac{1}{2}$. In fact, this algorithm for Max Cut was suggested by Erdős in 1967, and is one of the first analyses of any approximation algorithm.

[6] If you haven't encountered randomized algorithms before, it's a useful exercise to convince yourself that this algorithm can be derandomized: Show that there exists an efficient deterministic algorithm that given an undirected graph outputs a bipartition that cuts at least half of the edges.

A priori, it is not clear how to beat this approximation ratio or if that is possible at all. In a random d-regular graph (which is an excellent expander), one cannot cut more than a $\frac{1}{2} + \varepsilon$ fraction of the edges (where ε goes to zero as d goes to infinity). But locally, it is hard to distinguish a random d-regular graph from a random d-regular "almost bipartite" graph, where we split the vertex sets into two parts of equal size and each edge is with probability ε inside one of those parts and with probability $1 - \varepsilon$ between them. Such a graph G obviously has $\text{maxcut}(G) \geq 1 - \varepsilon$ but every neighborhood of it looks like a d-regular tree, just as in the case of a random d-regular graph. For this reason, "combinatorial" (or even linear programming) algorithms

have a hard time getting an approximation factor better than $\frac{1}{2}$ for Max Cut. Indeed, Chan et al. [2013] show that no linear programming relaxation of polynomial-size can have an approximation factor smaller than $\frac{1}{2}$ factor for Max Cut.

As alluded to before, sum-of-squares techniques allow us to go beyond approximation factor $\frac{1}{2}$ for Max Cut (Goemans and Williamson [1994]).

2. Theorem (Max Cut approximation). *There exists a polynomial time algorithm that given a graph, outputs a bipartition such that the number of cut edges is at least* 0.878 *times the maximum number possible.*

Max cut as a quadratic optimization problem over the hypercube

In order to apply the sum-of-squares framework we formulate Max Cut as a polynomial optimization problem. Let G be a graph of n vertices. Identify the vertex set of G with $[n]$. We let a vector $x \in \{0,1\}^n$ represent the bipartition of $[n]$ such that one side consists of every vertex i with $x_i = 1$. Let f_G be the function that assigns to every $x \in \{0,1\}^n$ the number $f_G(x)$ of edges cut by the bipartition that x represents. Then, f_G agrees with the following quadratic polynomial for every $x \in \{0,1\}^n$,

$$f_G(x) = \sum_{(i,j) \in E(G)} (x_i - x_j)^2 . \tag{1}$$

Thus the task of computing $\mathrm{maxcut}(G)$ is identical to the task of computing $\max_{x \in \{0,1\}^n} f_G(x)$.[7]

[7] In the previous lecture our focus was on *minimizing* a function over the cube. However note that maximizing a function f is equivalent to minimizing $-f$, and hence we will freely use whichever formulation is more natural for a given problem.

Sum-of-squares certificates for Max Cut

A simple way to describe the power of sum-of-squares for Max Cut is in terms of certificates. In order to certify that some graph G and some value c satisfy $\max f_G \geq c$ it is enough to exhibit a single bipartition $x \in \{0,1\}^n$ of G such that $f_G(x) \geq c$. In contrast, it is unclear how to efficiently certify statements of the $\max f_G \leq c$ because any certificate has to rule out an exponential number of possible bipartitions x. The following theorem shows that sum-of-squares provides efficient certificates for the inequality $f_G \leq c$ as long as $c \geq \max f_G / 0.878$.

3. Theorem (degree-2 sos certificates for Max Cut). *For every graph G with n vertices, there exist linear functions $g_1, \ldots, g_n \colon \{0,1\}^n \to \mathbb{R}$ such that for every $x \in \{0,1\}^n$,*

$$\max(f_G) - 0.878 \cdot f_G(x) = g_1(x)^2 + \cdots + g_n(x)^2. \tag{2}$$

The rest of this section is devoted to proving Theorem 2 and Theorem 3.

Aside: sum-of-squares certificates and eigenvalues

There is a fairly straightforward way to certify *some* non-trivial bound on the maximum cut. For a graph d-regular G with vertex set $[n]$ and edge set E, let $A = A(G)$ be its adjacency matrix of G, i.e.,s the n-by-n matrix

$$A_{i,j} = \begin{cases} 1 & \text{if } \{i,j\} \in E, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Let λ be the smallest eigenvalue (i.e., most negative) of A. Note that λ is in $[-d,0)$ (can you see why?) and, as you will show in the exercise below, $\lambda = -d$ if and only if G contains a bipartite connected component. Now, for every bipartition $x \in \{0,1\}^n$ corresponding to some set $S \subseteq [n]$, if $y \in \{\pm 1\}^n$ is the vector $y = 1 - 2x$ then $y^t A y = 2(|E(S,S)| + |E(V \setminus S)|) - 2|E(S, V \setminus S)| = 2|E| - 4|E(S, V \setminus S)|$.[8] Hence, since $y^\top A y \geq \lambda \|y\|^2 = \lambda n$, we get that $|E(S, V \setminus S)| \leq |E|/2 - n\lambda/4$. In particular, since $|E| = nd/2$, if $\lambda \geq -(1-\epsilon)d$ we get that $|E(S, V \setminus S)| \leq (1 - \epsilon/2)|E|$.

[8] The edges inside S and $V \setminus S$ get counted with a positive sign and the ones between S and its complement get counted with a negative sign. The factor of two is because every edge is counted twice.

The exercises below ask you to show that this bound can in fact be proven via a sum-of-squares proof and moreover that this bound is non-trivial (i.e., smaller than $|E|$) for every non-bipratite graph. (The first exercise is closely related to this exercise in the notes for lecture 1).

4. Exercise (Smallest eigenvalue bound for max cut). Let G be a d-regular graph with n vertices. Suppose the adjacency matrix of G has smallest eigenvalue λ. Show that for $c = \frac{1}{2} \cdot (1 - \lambda/d)$ the function $c \cdot |E| - f_G$ has a degree-2 sos certificate.[9]

[9] This certificate shows that $\text{maxcut}(G) \leq c$

5. Exercise (non triviality of eigenvalue bound). Prove that for every d-regular connected non-bipartite graph G, the minimum eigenvalue of $A(G)$ is larger than $-d$.

6. Exercise (Limits of eigenvalue based approximation). Show that for every $\epsilon > 0$ there exists some d and a d regular connected graph

G such that $\mathrm{maxcut}(G) \leq 1/2 + \epsilon$ but the minimum eigenvalue of G is at most $-(1-\epsilon)d$.

7. Exercise (Eigenvalue bound for random graphs). Show that for a random d-regular graph the eigenvalue bound certifies that $\mathrm{maxcut}(G) \leq 1/2 + o(1)$, where $o(1)$ goes to 0 as d goes to infinity.

For more about the relationship between the minimum eigenvalue and the max cut value, see Trevisan [2009].

Pseudo-distributions for Max Cut

Recall the definition of pseudo-distribution over the hypercube. For functions $\mu, f\colon \{0,1\}^n \to \mathbb{R}$, we denote the *pseudo-expectation of f under μ* as

$$\tilde{\mathbb{E}}_\mu f = \sum_{x \in \{0,1\}^n} \mu(x) \cdot f(x). \tag{4}$$

(We extend the above definition also to vector-valued functions f.) We say that μ is a degree-$2k$ pseudo-distribution if $\mathbb{E}_\mu 1 = 1$ and $\mathbb{E}_\mu g^2 \geq 0$ for all $g\colon \{0,1\}^n \to \mathbb{R}$ with $\deg g \leq k$.

By the duality between sos certificates and pseudo-distribution, the theorem below implies Theorem 3. Furthermore, since we can optimize over low-degree moments of pseudo-distributions over the hypercube up to arbitrary accuracy, the theorem below also implies Theorem 2.

8. Theorem (Rounding pseudo-distributions for Max Cut). *For every graph G and degree-2 pseudo-distribution μ over the hypercube, there exists a probability distribution μ' over the hypercube such that*

$$\mathbb{E}_{\mu'} f_G \geq 0.878 \cdot \tilde{\mathbb{E}}_\mu f_G. \tag{5}$$

Furthermore, there exists a randomized polynomial-time algorithm that given the pseudo-distribution μ (say represented by its degree-2 moments $\mathbb{E}_{\mu(x)} xx^\mathsf{T}$) outputs a sample from μ'.

Proof. Let $\mathbf{1}$ denote the all-ones vector in $\{0,1\}^n$. For simplicity, assume $\tilde{\mathbb{E}}_{\mu(x)} x = \frac{1}{2} \cdot \mathbf{1}$. (Otherwise, consider the pseudo-distribution $\frac{1}{2}(\mu(x) + \mu(\mathbf{1} - x))$, which satisfies this property and has the same pseudo-expectation for f_G.) We define the probability distribution $\mu'(x')$ in terms of the following sampling algorithm:

- Choose ξ as a Gaussian vector with covariance $\tilde{\mathbb{E}}_{\mu(x)} xx^\mathsf{T}$ and mean $\tilde{\mathbb{E}}_{\mu(x)} x$ (see *quadratic sampling lemma* from lecture 1).

- Output $x' \in \{0,1\}^n$ with $x'_i = 0$ if $\xi_i < \frac{1}{2}$ and $x'_i = 1$ otherwise.

By the form of f_G, it is enough to show the following inequality for all $i, j \in [n]$.

$$\mathop{\mathbb{E}}_{\mu'(x')} (x'_i - x'_j)^2 \geq 0.878 \cdot \mathop{\tilde{\mathbb{E}}}_{\mu(x)} (x_i - x_j)^2 . \tag{6}$$

Fix the indices i and j in $[n]$. Since μ is a pseudo-distribution over the hypercube, x_i and x_j have variance $\tilde{\mathbb{E}}_{\mu(x)} x_i^2 - \frac{1}{4} = \tilde{\mathbb{E}}_{\mu(x)} x_j^2 - \frac{1}{4} = \frac{1}{4}$.[10] Let $\rho = 4\tilde{\mathbb{E}}_{\mu(x)} x_i x_j - 1$ be the covariance of x_i and x_j under μ (after rescaling them to have variance 1). Since ξ is Gaussian with the same first two moments as x,

$$(\xi_i, \xi_j) \sim \mathcal{N}\left(\frac{1}{2} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \frac{1}{4} \cdot \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right) . \tag{7}$$

Therefore, $2\xi_j - 1 = \rho \cdot (2\xi_i - 1) + \sqrt{1 - \rho^2} \cdot \xi_i^\perp$, where $\xi_i^\perp \sim \mathcal{N}(0,1)$ is a standard Gaussian independent of ξ_i. It follows that

$$\mathop{\mathbb{E}}_{D'(x')} (x'_i - x'_j)^2 = \mathop{\mathbb{P}}_{\xi_i, \xi_j} \left\{ \mathrm{sign}(2\xi_i - 1) \neq \mathrm{sign}(2\xi_j - 1) \right\}$$

$$= \mathop{\mathbb{P}}_{s,t \sim \mathcal{N}(0,1)} \left\{ \mathrm{sign}\, s \neq \mathrm{sign}\left(\rho \cdot s + \sqrt{1 - \rho^2} \cdot t \right) \right\} . \tag{8}$$

The event on the right-hand side consists of the set of all pairs (s, t) in two conic regions symmetric around the origin that each form an angle of $\arccos \rho$. By rotational symmetry of the Gaussian distribution, it follows that the event has probability $\frac{\arccos \rho}{\pi}$. □

At the same time,

$$\mathop{\mathbb{E}}_{\mu(x)} (x_i - x_j)^2 = \frac{1}{2} \cdot (1 - \rho) . \tag{9}$$

Therefore, it remains to verify that

$$\frac{2 \arccos \rho}{(1 - \rho)\pi} \geq 0.878 , \tag{10}$$

which we can do by single-variable calculus.

In the proof of Theorem 8 it may seem almost "accidental" that the final constant ends up being bigger than $1/2$. However, if we just wanted to see that the approximation ratio is bigger than $1/2$ a more direct calculation suffices. Note that if the optimal value $\mathrm{maxcut}(G)$ is smaller than say 0.99 then already the naive randomized algorithm achieves an approximation ratio of $0.99/0.5$, which is bigger than $1/2$. Therefore, the challenge is to find bipartitions for graphs with $\mathrm{maxcut}(G) \geq 0.99$ that cut significantly more than half of the edges. The following exercise gives a "calculus free" analysis of an algorithm with this property.

[10] This step is the only one that uses the fact that we are interested in the maximum value of f_G over the hypercube as opposed to the sphere.
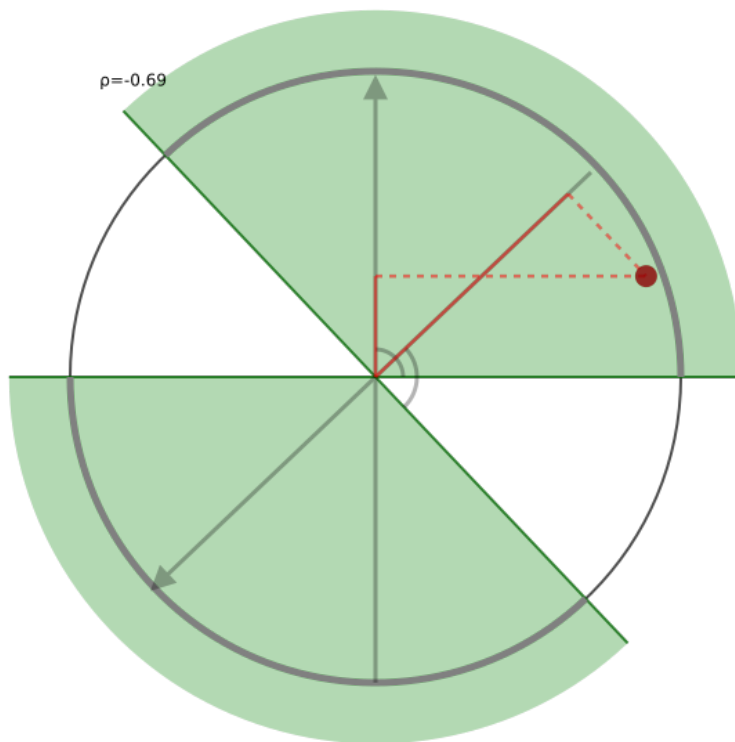
Figure 3: To generate two ρ-correlated Gaussian random variables we choose a standard two-dimensional Gaussian vector (s, t) (the red point) and project it onto the normals of two lines at an angle of $\arccos \rho$. The signs differ if and only if the two dimensional Gaussian falls between these two lines, which happens with probability $\frac{\arccos \rho}{\pi}$.
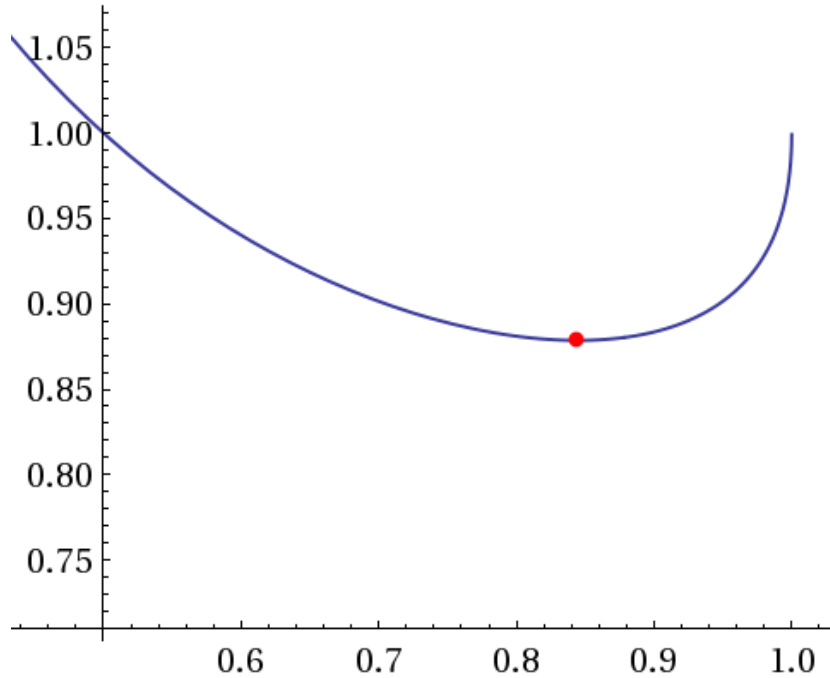
9. Exercise (rounding for near-bipartite graphs). Show every degree-2 pseudo-distribution μ over the hypercube, there exists an efficiently sampleable probability distribution μ' over the hypercube such that for every $\varepsilon > 0$ and every graph G,

$$\tilde{\mathbb{E}}_{\mu} f_G \geq (1 - \varepsilon) \cdot |E(G)| \implies \mathbb{E}_{\mu'} f_G \geq (1 - 2\sqrt{\varepsilon}) \cdot |E(Gs)|. \qquad (11)$$

(In particular, when $\text{maxcut}(G) \geq 0.99$, we can take $\varepsilon = 0.01$ and get bipartitions with expected value at least $1 - 2 \cdot 0.1 = 0.8 > 0.5$ under μ'.) [11]

[11] **Hint:** You can use the same construction for μ' as in the proof of Theorem 8.

In general, we can ask about the curve between the value of the pseudo-distribution and the value achieved by a rounding algorithm. Concretely, for every value of c, what is the largest value of s such that a degree-2 pseudo-distribution with $\tilde{\mathbb{E}}_{\mu} f_G \geq c \cdot |E(G)|$ for a graph G always allows us to efficiently find a bipartition x with value $f_G(x) \geq s \cdot |E(G)|$. It turns out that the above algorithm does not achieve the best possible "approximation curve" but similar ideas work, see (O'Donnell and Wu [2008]).

The following exercise asks you to analyze the approximation curve of the Goemans–Williamson algorithm for a particular range of c.

10. Exercise (rounded value vs. pseudo-distribution curve). Let $c_{GW} \approx 0.845$ be the minimizer of $c \mapsto \arccos(1-2c)/(\pi c)$ and let $\alpha_{GW} \approx 0.878$ be the minimum value of this function. Show that for every $c \geq c_{GW}$, every graph G, and every degree-2 distribution μ over the hypercube such that $\tilde{\mathbb{E}}_\mu f_G = c \cdot |E|$, there is an actual distribution μ' such that $\tilde{\mathbb{E}}_{\mu'} f_G \geq \arccos(1-2c)/\pi$. [12]

[12] **Hint:** Prove first that the following function $g : [0,1] \to \mathbb{R}$ is *convex*: $g(c) = \alpha_{GW}c$ for $c < c_{GW}$ and $g(c) = \arccos(1-2c)/\pi$ for $c \geq c_{GW}$. s Then prove that for every edge i, j, the probability that μ' will cut the edge is at least $g\left(\tilde{\mathbb{E}}_\mu(x_i - x_j)^2\right)$.

References

Emmanuel J. Candès, Yonina C. Eldar, Thomas Strohmer, and Vladislav Voroninski. Phase retrieval via matrix completion. *SIAM J. Imaging Sciences*, 6(1):199–225, 2013.

Siu On Chan, James R. Lee, Prasad Raghavendra, and David Steurer. Approximate constraint satisfaction requires large LP relaxations. In *FOCS*, pages 350–359. IEEE Computer Society, 2013.

Jeff Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. In *Problems in analysis (Papers dedicated to Salomon Bochner, 1969)*, pages 195–199. Princeton Univ. Press, Princeton, N. J., 1970.

Michel Deza and Monique Laurent. Applications of cut polyhedra. I, II. *J. Comput. Appl. Math.*, 55(2):191–216, 217–247, 1994. ISSN 0377-0427. doi: 10.1016/0377-0427(94)90020-5. URL http://dx.doi.org/10.1016/0377-0427(94)90020-5.

Michel X. Goemans and David P. Williamson. .879-approximation algorithms for MAX CUT and MAX 2sat. In *STOC*, pages 422–431. ACM, 1994.

A. Grothendieck. Résumé de la théorie métrique des produits tensoriels topologiques. *Bol. Soc. Mat. São Paulo*, 8:1–79, 1953.

Richard M. Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, The IBM Research Symposia Series, pages 85–103. Plenum Press, New York, 1972.

Subhash Khot. On the power of unique 2-prover 1-round games. In *STOC*, pages 767–775. ACM, 2002.

Ryan O'Donnell and Yi Wu. An optimal sdp algorithm for max-cut, and equally optimal long code tests. In *STOC*, pages 335–344. ACM, 2008.

Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *STOC*, pages 755–764. ACM, 2010.

Luca Trevisan. Max cut and the smallest eigenvalue. In *STOC*, pages 263–272. ACM, 2009.

Irène Waldspurger, Alexandre d'Aspremont, and Stéphane Mallat. Phase recovery, maxcut and complex semidefinite programming. *Math. Program.*, 149(1-2):47–81, 2015.