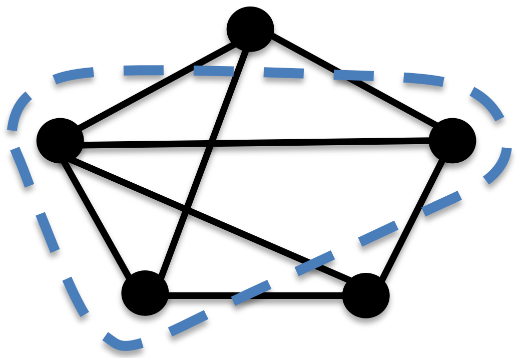
Sum-of-Squares, with a View Towards Average-case Complexity

Ankur Moitra (MIT)

KITP Tutorial, January 11, 2019

A CLASSIC HARD PROBLEM: MAXCUT

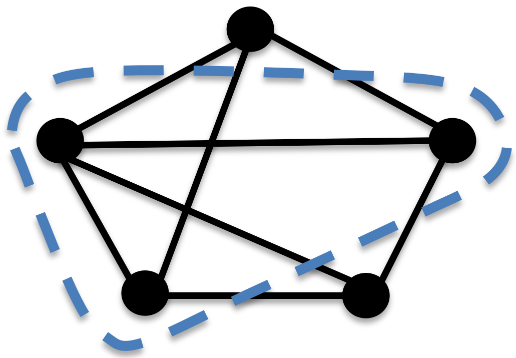
Goal: given a graph $G = (V, E)$:



find a cut $U \subseteq V$ that maximizes the number of crossing edges

A CLASSIC HARD PROBLEM: MAXCUT

Goal: given a graph $G = (V, E)$:

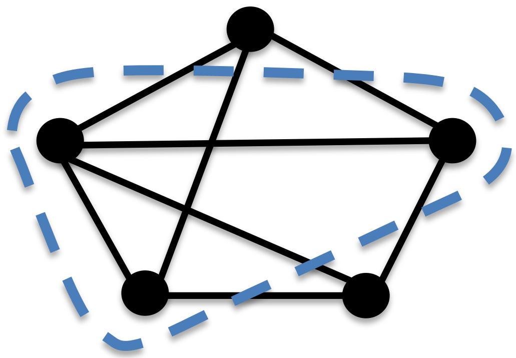


find a cut $U \subseteq V$ that maximizes the number of crossing edges

NP-hard to maximize exactly, one of **[Karp, '72]**'s 21 problems

A CLASSIC HARD PROBLEM: MAXCUT

Goal: given a graph $G = (V, E)$:



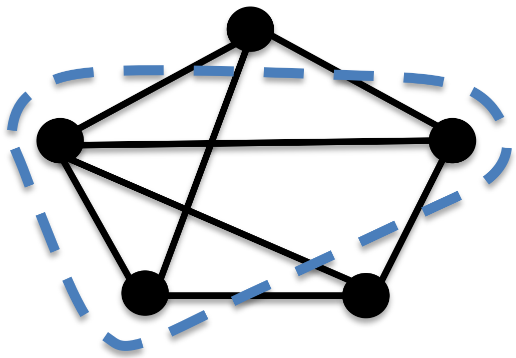
find a cut $U \subseteq V$ that maximizes the number of crossing edges

NP-hard to maximize exactly, one of **[Karp, '72]**'s 21 problems

How well can we approximate MAXCUT?

A CLASSIC HARD PROBLEM: MAXCUT

Goal: given a graph $G = (V, E)$:



find a cut $U \subseteq V$ that maximizes the number of crossing edges

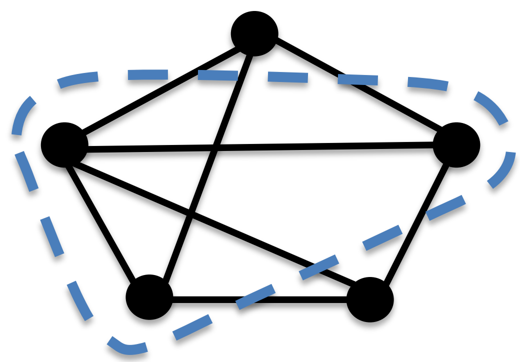
NP-hard to maximize exactly, one of **[Karp, '72]**'s 21 problems

How well can we approximate MAXCUT?

Simple ½-approximation algorithm: Choose U randomly.

A CLASSIC HARD PROBLEM: MAXCUT

Goal: given a graph $G = (V, E)$:



find a cut $U \subseteq V$ that maximizes the number of crossing edges

NP-hard to maximize exactly, one of **[Karp, '72]**'s 21 problems

How well can we approximate MAXCUT?

Simple ½-approximation algorithm: Choose U randomly. **But can we do better?**

MAXCUT AS A QUADRATIC PROGRAM

We can also formulate MAXCUT as optimizing a polynomial, subject polynomial constraints:

$$\max \sum_{(i,j)\in E} (x_i - x_j)^2$$

$$\text{s.t. } x_i^2 = x_i \text{ for all } i$$

MAXCUT AS A QUADRATIC PROGRAM

We can also formulate MAXCUT as optimizing a polynomial, subject polynomial constraints:

$$\max \sum_{(i,j)\in E} (x_i - x_j)^2$$

$$\text{s.t. } x_i^2 = x_i \text{ for all } i$$

x_i's are 0/1 valued

MAXCUT AS A QUADRATIC PROGRAM

We can also formulate MAXCUT as optimizing a polynomial, subject polynomial constraints:

$$\max \sum_{(i,j)\in E} (x_i - x_j)^2$$

counts the number of edges crossing the cut

$$\text{s.t. } x_i^2 = x_i \text{ for all } i$$

x_i's are 0/1 valued

MAXCUT AS A QUADRATIC PROGRAM

We can also formulate MAXCUT as optimizing a polynomial, subject polynomial constraints:

$$\max \sum_{(i,j)\in E} (x_i - x_j)^2$$

counts the number of edges crossing the cut

$$\text{s.t. } x_i^2 = x_i \text{ for all } i$$

x_i's are 0/1 valued

Now we can leverage the **Sum-of-Squares (SOS) Hierarchy**…

SUM-OF-SQUARES HIERARCHY

Introduced by [Parrilo ‘00], [Lasserre ‘01]

- strengthens **Sherali-Adams**, **Lovasz-Schrijver**, **LS+**
- breaks integrality gaps for other hierarchies [Barak et al, ‘12]
- highly successful convex relaxation

  sparsest cut [ARV ’04]

  unique games [ABS ’10], [BRS ‘12], [GS ‘12]
- optimal among all poly. sized SDPs for random CSPs [LRS ‘15]
- best known algorithm for several **average-case** problems

  planted sparse vector, dictionary learning [BKS ’14, ‘15]

  noisy tensor completion [BM ’15], tensor PCA [HSS ‘15]

OUTLINE

Part I: Introduction

- MAXCUT and the Sum-of-Squares Hierarchy
- A Dual View via Pseudo-expectation

Part II: Rounding SOS

Part III: Fooling SOS

- Planted Clique and its Applications
- The MPW Moments and Corrections
- Pseudo-calibration and Fourier Analysis

Part IV: Sparse PCA and Computational vs. Statistical Tradeoffs

Part V: Equivalence with Spectral Methods

OUTLINE

Part I: Introduction

- MAXCUT and the Sum-of-Squares Hierarchy
- **A Dual View via Pseudo-expectation**

Part II: Rounding SOS

Part III: Fooling SOS

- Planted Clique and its Applications
- The MPW Moments and Corrections
- Pseudo-calibration and Fourier Analysis

Part IV: Sparse PCA and Computational vs. Statistical Tradeoffs

Part V: Equivalence with Spectral Methods

A DUAL VIEW

(Usually introduced as proof system related to Hilbert's 17th prob.)

A DUAL VIEW

(Usually introduced as proof system related to Hilbert's 17th prob.)

Goal: Find operator that behaves like the expectation over a distribution on solutions

$$\widetilde{\mathbb{E}} : \underbrace{\mathcal{P}_n^{\leq d}}_{\text{degree} \leq d \text{ polynomials in } n \text{ variables}} \to \mathbb{R}$$

A DUAL VIEW

(Usually introduced as proof system related to Hilbert's 17th prob.)

Goal: Find operator that behaves like the expectation over a distribution on solutions

$$\widetilde{\mathbb{E}} : \underbrace{\mathcal{P}_n^{\leq d}}_{\text{degree} \leq d \text{ polynomials in n variables}} \to \mathbb{R}$$

Called a **Pseudo-expectation**

A DUAL VIEW

(Usually introduced as proof system related to Hilbert's 17th prob.)

Goal: Find operator that behaves like the expectation over a distribution on solutions

$$\underbrace{\widetilde{\mathbb{E}} : \mathcal{P}_n^{\leq d}}_{} \to \mathbb{R}$$

degree ≤ d polynomials in n variables

Called a **Pseudo-expectation**

Let's see what it looks like for MAXCUT…

Degree d relaxation for MAXCUT:

$$\max \quad \widetilde{\mathbb{E}}[\textstyle\sum_{(i,j)\in E}(x_i - x_j)^2]$$

such that:

(1) $\widetilde{\mathbb{E}}$ is linear

(2) $\widetilde{\mathbb{E}}[1] = 1$

(3) $\widetilde{\mathbb{E}}[p^2] \geq 0$ for all deg(p) ≤ d/2

(4) $\widetilde{\mathbb{E}}[x_i^2 p] = \widetilde{\mathbb{E}}[x_i p]$ for all deg(p) ≤ d-2

Degree d relaxation for MAXCUT:

$$\max \quad \widetilde{\mathbb{E}}[\sum_{(i,j)\in E}(x_i - x_j)^2]$$

such that:

(1) $\widetilde{\mathbb{E}}$ is linear

(2) $\widetilde{\mathbb{E}}[1] = 1$

(3) $\widetilde{\mathbb{E}}[p^2] \geq 0$ for all deg(p) ≤ d/2

(4) $\widetilde{\mathbb{E}}[x_i^2 p] = \widetilde{\mathbb{E}}[x_i p]$ for all deg(p) ≤ d-2

(1) – (3) are the usual constraints that say Ẽ behaves like it is taking the expectation under some distribution on assignments to the variables

Degree d relaxation for MAXCUT:

$$\max \quad \widetilde{\mathbb{E}}[\textstyle\sum_{(i,j)\in E}(x_i - x_j)^2]$$

such that:

(1) $\widetilde{\mathbb{E}}$ is linear

(3) $\widetilde{\mathbb{E}}[p^2] \geq 0$ for all deg(p) ≤ d/2

(2) $\widetilde{\mathbb{E}}[1] = 1$

(4) $\widetilde{\mathbb{E}}[x_i^2 p] = \widetilde{\mathbb{E}}[x_i p]$ for all deg(p) ≤ d-2

(1) – (3) are the usual constraints that say Ẽ behaves like it is taking the expectation under some distribution on assignments to the variables

(4) is because we want the distribution to be supported on 0/1 valued assignments

Degree d relaxation for MAXCUT:

$$\max \ \widetilde{\mathbb{E}}[\sum_{(i,j)\in E}(x_i - x_j)^2]$$

such that:

(1) $\widetilde{\mathbb{E}}$ is linear

(2) $\widetilde{\mathbb{E}}[1] = 1$

(3) $\widetilde{\mathbb{E}}[p^2] \geq 0$ for all deg(p) ≤ d/2

(4) $\widetilde{\mathbb{E}}[x_i^2 p] = \widetilde{\mathbb{E}}[x_i p]$ for all deg(p) ≤ d-2

But why is this a relaxation for MAXCUT?

Degree d relaxation for MAXCUT:

$$\max \quad \widetilde{\mathbb{E}}[\textstyle\sum_{(i,j)\in E}(x_i - x_j)^2]$$

such that:

(1) $\widetilde{\mathbb{E}}$ is linear

(2) $\widetilde{\mathbb{E}}[1] = 1$

(3) $\widetilde{\mathbb{E}}[p^2] \geq 0$ for all deg(p) ≤ d/2

(4) $\widetilde{\mathbb{E}}[x_i^2 p] = \widetilde{\mathbb{E}}[x_i p]$ for all deg(p) ≤ d-2

Claim: If there is a cut that has at least k edges crossing, there is a feasible solution to (1) – (4) with objective value ≥ k

Degree d relaxation for MAXCUT:

$$\max \quad \widetilde{\mathbb{E}}[\textstyle\sum_{(i,j)\in E}(x_i - x_j)^2]$$

such that:

(1) $\widetilde{\mathbb{E}}$ is linear

(2) $\widetilde{\mathbb{E}}[1] = 1$

(3) $\widetilde{\mathbb{E}}[p^2] \geq 0$ for all deg(p) ≤ d/2

(4) $\widetilde{\mathbb{E}}[x_i^2 p] = \widetilde{\mathbb{E}}[x_i p]$ for all deg(p) ≤ d-2

Claim: If there is a cut that has at least k edges crossing, there is a feasible solution to (1) – (4) with objective value ≥ k

Proof: if $a_1, a_2, \ldots, a_n$ is the indicator vector of the cut U, set

$$\widetilde{\mathbb{E}}[p(x_1, x_2, \cdots, x_n)] = p(a_1, a_2, \cdots, a_n) \quad \blacksquare$$

Can we efficiently solve this relaxation?

Can we efficiently solve this relaxation?

Theorem: There is an $n^{O(d)}$-time algorithm for finding such an operator, if it exists

Can we efficiently solve this relaxation?

Theorem: There is an $n^{O(d)}$-time algorithm for finding such an operator, if it exists

It is a semidefinite program on a $n^{O(d)} \times n^{O(d)}$ matrix whose entries are the pseudo-expectation applied to monomials

Can we efficiently solve this relaxation?

Theorem: There is an $n^{O(d)}$-time algorithm for finding such an operator, if it exists

It is a semidefinite program on a $n^{O(d)}$ x $n^{O(d)}$ matrix whose entries are the pseudo-expectation applied to monomials

How well does SOS approximate MAXCUT?

OUTLINE

Part I: Introduction

- MAXCUT and the Sum-of-Squares Hierarchy
- A Dual View via Pseudo-expectation

Part II: Rounding SOS

Part III: Fooling SOS

- Planted Clique and its Applications
- The MPW Moments and Corrections
- Pseudo-calibration and Fourier Analysis

Part IV: Sparse PCA and Computational vs. Statistical Tradeoffs

Part V: Equivalence with Spectral Methods

OUTLINE

Part I: Introduction

- MAXCUT and the Sum-of-Squares Hierarchy
- A Dual View via Pseudo-expectation

Part II: Rounding SOS

Part III: Fooling SOS

- Planted Clique and its Applications
- The MPW Moments and Corrections
- Pseudo-calibration and Fourier Analysis

Part IV: Sparse PCA and Computational vs. Statistical Tradeoffs

Part V: Equivalence with Spectral Methods

APPROXIMATION ALGORITHMS FOR MAXCUT

Revolutionary work of **[Goemans, Williamson]**:

Theorem: There is a α_{GW}-approximation algorithm for

$$\alpha_{GW} = \min_{-1 \leq \rho \leq 1} \frac{2 \arccos \rho}{(1-\rho)\pi} \geq 0.878$$

for MAXCUT

APPROXIMATION ALGORITHMS FOR MAXCUT

Revolutionary work of **[Goemans, Williamson]**:

Theorem: There is a α_{GW}-approximation algorithm for

$$\alpha_{GW} = \min_{-1 \leq \rho \leq 1} \frac{2 \arccos \rho}{(1-\rho)\pi} \geq 0.878$$

for MAXCUT

We will give an alternate proof by rounding the degree two Sum-of-Squares relaxation

Main Question: How do you round a pseudo-expectation to find a cut?

I.e. if I give you $\widetilde{\mathbb{E}}$ how do you find a cut with at least

$$\alpha_{GW}\widetilde{\mathbb{E}}[\sum_{(i,j)\in E}(x_i - x_j)^2]$$

edges crossing (in expectation)?

Main Question: How do you round a pseudo-expectation to find a cut?

I.e. if I give you $\widetilde{\mathbb{E}}$ how do you find a cut with at least

$$\alpha_{GW}\widetilde{\mathbb{E}}[\sum_{(i,j)\in E}(x_i - x_j)^2]$$

edges crossing (in expectation)?

Main Idea: Use a sample from a Gaussian distribution whose moments match the pseudo-moments

Main Question: How do you round a pseudo-expectation to find a cut?

I.e. if I give you $\widetilde{\mathbb{E}}$ how do you find a cut with at least

$$\alpha_{GW}\widetilde{\mathbb{E}}[\sum_{(i,j)\in E}(x_i - x_j)^2]$$

edges crossing (in expectation)?

Main Idea: Use a sample from a Gaussian distribution whose moments match the pseudo-moments

---

Aside: Rounding higher degree relaxations is **much** harder b/c you cannot necc. find a r.v. whose moments match the pseudo-moments

Claim: Without loss of generality, can assume for all i

$$\widetilde{\mathbb{E}}[x_i] = \frac{1}{2}$$

Claim: Without loss of generality, can assume for all i

$$\widetilde{\mathbb{E}}[x_i] = \frac{1}{2}$$

Intuition: You can always change U to V\U without changing the value of the cut, so WLOG x_i has probability 1/2 of being in U

GAUSSIAN ROUNDING

Let y be a Gaussian vector with mean μ and covariance Σ for

$$\mu = \widetilde{\mathbb{E}}[x] \text{ and } \Sigma = \widetilde{\mathbb{E}}[(x-\mu)(x-\mu)^T]$$

GAUSSIAN ROUNDING

Let y be a Gaussian vector with mean μ and covariance Σ for

$$\mu = \widetilde{\mathbb{E}}[x] \text{ and } \Sigma = \widetilde{\mathbb{E}}[(x-\mu)(x-\mu)^T]$$

Now set $a_i = 0$ if $y_i \leq \frac{1}{2}$ and otherwise $a_i = 1$

GAUSSIAN ROUNDING

Let y be a Gaussian vector with mean μ and covariance Σ for

$$\mu = \widetilde{\mathbb{E}}[x] \text{ and } \Sigma = \widetilde{\mathbb{E}}[(x-\mu)(x-\mu)^T]$$

Now set $a_i = 0$ if $y_i \leq \frac{1}{2}$ and otherwise $a_i = 1$

We will show that for each (i, j) we have

$$\mathbb{E}[(a_i - a_j)^2] \geq \alpha_{GW} \widetilde{\mathbb{E}}[(x_i - x_j)^2]$$

which, by linearity of expectation, will complete the proof

For each edge (i,j), calculate contribution to **objective value**:

$$\widetilde{\mathbb{E}}[(x_i - x_j)^2] = \widetilde{\mathbb{E}}[x_i^2] - 2\widetilde{\mathbb{E}}[x_i x_j] + \widetilde{\mathbb{E}}[x_j^2]$$

For each edge (i,j), calculate contribution to **objective value**:

$$\widetilde{\mathbb{E}}[(x_i - x_j)^2] = \widetilde{\mathbb{E}}[x_i^2] - 2\widetilde{\mathbb{E}}[x_i x_j] + \widetilde{\mathbb{E}}[x_j^2]$$

$$= \widetilde{\mathbb{E}}[x_i] - 2\widetilde{\mathbb{E}}[x_i x_j] + \widetilde{\mathbb{E}}[x_j]$$

For each edge (i,j), calculate contribution to **objective value**:

$$\widetilde{\mathbb{E}}[(x_i - x_j)^2] = \widetilde{\mathbb{E}}[x_i^2] - 2\widetilde{\mathbb{E}}[x_i x_j] + \widetilde{\mathbb{E}}[x_j^2]$$

$$= \widetilde{\mathbb{E}}[x_i] - 2\widetilde{\mathbb{E}}[x_i x_j] + \widetilde{\mathbb{E}}[x_j]$$

$$= \boxed{\frac{1}{2}(1 - \rho)} \text{ for } \rho = 4\widetilde{\mathbb{E}}[x_i x_j] - 1$$

For each edge (i,j), calculate contribution to **objective value**:

$$\widetilde{\mathbb{E}}[(x_i - x_j)^2] = \widetilde{\mathbb{E}}[x_i^2] - 2\widetilde{\mathbb{E}}[x_i x_j] + \widetilde{\mathbb{E}}[x_j^2]$$

$$= \widetilde{\mathbb{E}}[x_i] - 2\widetilde{\mathbb{E}}[x_i x_j] + \widetilde{\mathbb{E}}[x_j]$$

$$= \boxed{\frac{1}{2}(1 - \rho)} \text{ for } \rho = 4\widetilde{\mathbb{E}}[x_i x_j] - 1$$

And its contribution to the **expected number of edges crossing**:

For each edge (i,j), calculate contribution to **objective value**:

$$\widetilde{\mathbb{E}}[(x_i - x_j)^2] = \widetilde{\mathbb{E}}[x_i^2] - 2\widetilde{\mathbb{E}}[x_i x_j] + \widetilde{\mathbb{E}}[x_j^2]$$

$$= \widetilde{\mathbb{E}}[x_i] - 2\widetilde{\mathbb{E}}[x_i x_j] + \widetilde{\mathbb{E}}[x_j]$$

$$= \boxed{\tfrac{1}{2}(1 - \rho)} \text{ for } \rho = 4\widetilde{\mathbb{E}}[x_i x_j] - 1$$

And its contribution to the **expected number of edges crossing**:

$$\mathrm{Var}(y_i) = \widetilde{\mathbb{E}}[(x_i - \tfrac{1}{2})^2] = \widetilde{\mathbb{E}}[x_i] - \tfrac{1}{4} = \tfrac{1}{4}$$

For each edge (i,j), calculate contribution to **objective value**:

$$\widetilde{\mathbb{E}}[(x_i - x_j)^2] = \widetilde{\mathbb{E}}[x_i^2] - 2\widetilde{\mathbb{E}}[x_i x_j] + \widetilde{\mathbb{E}}[x_j^2]$$

$$= \widetilde{\mathbb{E}}[x_i] - 2\widetilde{\mathbb{E}}[x_i x_j] + \widetilde{\mathbb{E}}[x_j]$$

$$= \boxed{\tfrac{1}{2}(1-\rho)} \text{ for } \rho = 4\widetilde{\mathbb{E}}[x_i x_j] - 1$$

And its contribution to the **expected number of edges crossing**:

$$\mathrm{Var}(y_i) = \widetilde{\mathbb{E}}[(x_i - \tfrac{1}{2})^2] = \widetilde{\mathbb{E}}[x_i] - \tfrac{1}{4} = \tfrac{1}{4} \text{ and } \mathrm{Cov}(y_i, y_j) = \tfrac{\rho}{4}$$

For each edge (i,j), calculate contribution to **objective value**:

$$\widetilde{\mathbb{E}}[(x_i - x_j)^2] = \widetilde{\mathbb{E}}[x_i^2] - 2\widetilde{\mathbb{E}}[x_i x_j] + \widetilde{\mathbb{E}}[x_j^2]$$

$$= \widetilde{\mathbb{E}}[x_i] - 2\widetilde{\mathbb{E}}[x_i x_j] + \widetilde{\mathbb{E}}[x_j]$$

$$= \boxed{\tfrac{1}{2}(1 - \rho)} \text{ for } \rho = 4\widetilde{\mathbb{E}}[x_i x_j] - 1$$

And its contribution to the **expected number of edges crossing**:

$$\mathrm{Var}(y_i) = \widetilde{\mathbb{E}}[(x_i - \tfrac{1}{2})^2] = \widetilde{\mathbb{E}}[x_i] - \tfrac{1}{4} = \tfrac{1}{4} \text{ and } \mathrm{Cov}(y_i, y_j) = \tfrac{\rho}{4}$$

Now we can compute:

$$\mathbb{P}[a_i \neq a_j] = \mathbb{P}[\mathrm{sgn}(s) \neq \mathrm{sgn}(\rho s + \sqrt{1 - \rho^2} t)]$$

independent std Gaussians

For each edge (i,j), calculate contribution to **objective value**:

$$\widetilde{\mathbb{E}}[(x_i - x_j)^2] = \widetilde{\mathbb{E}}[x_i^2] - 2\widetilde{\mathbb{E}}[x_i x_j] + \widetilde{\mathbb{E}}[x_j^2]$$

$$= \widetilde{\mathbb{E}}[x_i] - 2\widetilde{\mathbb{E}}[x_i x_j] + \widetilde{\mathbb{E}}[x_j]$$

$$= \boxed{\tfrac{1}{2}(1-\rho)} \text{ for } \rho = 4\widetilde{\mathbb{E}}[x_i x_j] - 1$$

And its contribution to the **expected number of edges crossing**:

$$\mathrm{Var}(y_i) = \widetilde{\mathbb{E}}[(x_i - \tfrac{1}{2})^2] = \widetilde{\mathbb{E}}[x_i] - \tfrac{1}{4} = \tfrac{1}{4} \text{ and } \mathrm{Cov}(y_i, y_j) = \tfrac{\rho}{4}$$

Now we can compute:

$$\mathbb{P}[a_i \neq a_j] = \mathbb{P}[\mathrm{sgn}(s) \neq \mathrm{sgn}(\rho s + \sqrt{1-\rho^2}\, t)]$$

$$= \boxed{\frac{\arccos \rho}{\pi}}$$

independent std Gaussians (s, t)

Putting it all together, we have for every edge (i, j):

$$\mathbb{P}[a_i \neq a_j] \geq \frac{2\arccos\rho}{(1-\rho)\pi}\widetilde{\mathbb{E}}[(x_i - x_j)^2] \geq \alpha_{GW}\widetilde{\mathbb{E}}[(x_i - x_j)^2]$$

which completes the proof ■

OUTLINE

Part I: Introduction

- MAXCUT and the Sum-of-Squares Hierarchy
- A Dual View via Pseudo-expectation

Part II: Rounding SOS

Part III: Fooling SOS

- Planted Clique and its Applications
- The MPW Moments and Corrections
- Pseudo-calibration and Fourier Analysis

Part IV: Sparse PCA and Computational vs. Statistical Tradeoffs

Part V: Equivalence with Spectral Methods

OUTLINE

Part I: Introduction

- MAXCUT and the Sum-of-Squares Hierarchy
- A Dual View via Pseudo-expectation

Part II: Rounding SOS

Part III: Fooling SOS

- **Planted Clique and its Applications**
- The MPW Moments and Corrections
- Pseudo-calibration and Fourier Analysis

Part IV: Sparse PCA and Computational vs. Statistical Tradeoffs

Part V: Equivalence with Spectral Methods

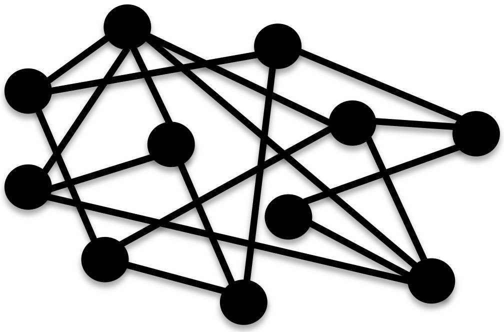
PLANTED CLIQUE

Introduced by **[Jerrum, ‘92]**, **[Kucera, ’95]**:

PLANTED CLIQUE

Introduced by **[Jerrum, '92]**, **[Kucera, '95]**:

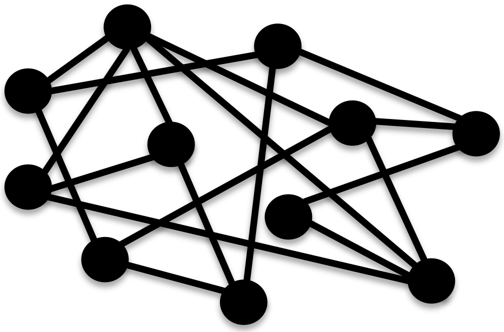
Step #1: Generate E-R random graph G(n, ½)



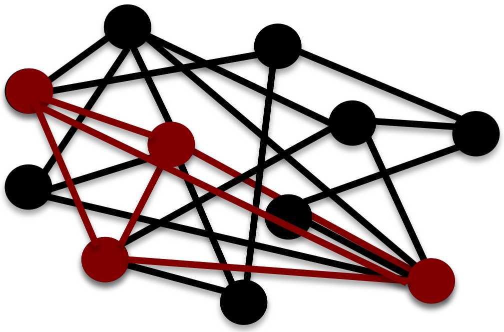
PLANTED CLIQUE

Introduced by **[Jerrum, '92]**, **[Kucera, '95]**:

Step #1: Generate E-R random graph G(n, ½)



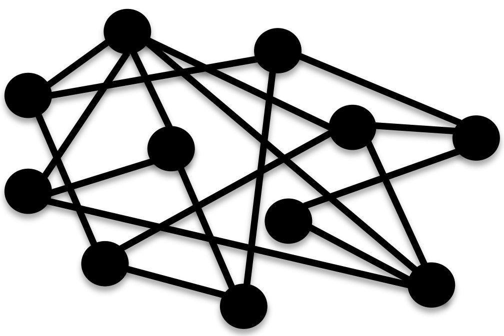
Step #2: Add a clique on random set of ω vertices



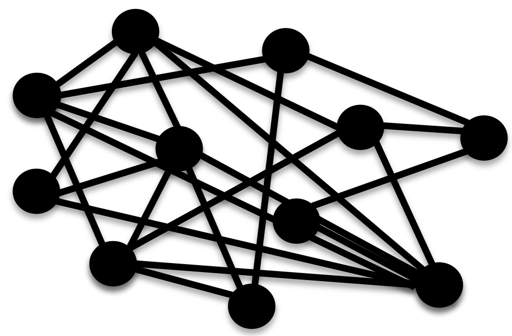
PLANTED CLIQUE

Introduced by **[Jerrum, '92]**, **[Kucera, '95]**:

Step #1: Generate E-R random graph G(n, ½)



Step #2: Add a clique on random set of ω vertices

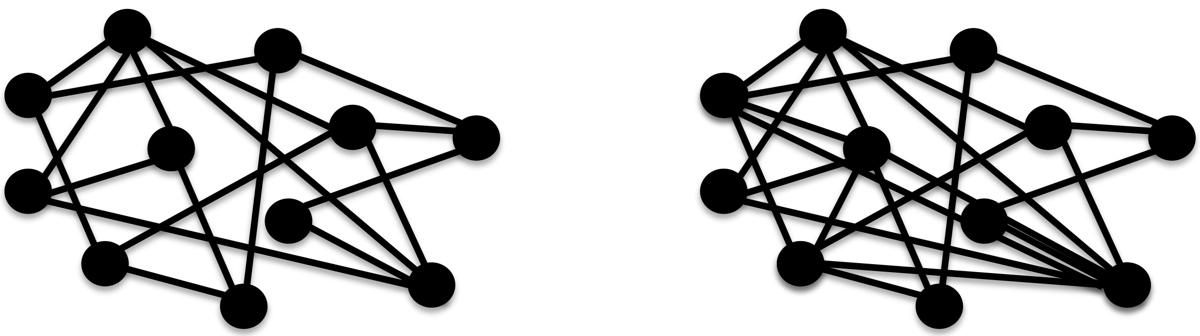


PLANTED CLIQUE

Introduced by **[Jerrum, '92]**, **[Kucera, '95]**:

Step #1: Generate E-R random graph G(n, ½)

Step #2: Add a clique on random set of ω vertices



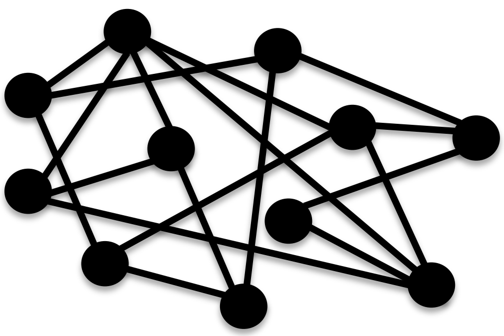
Can we find the planted clique?

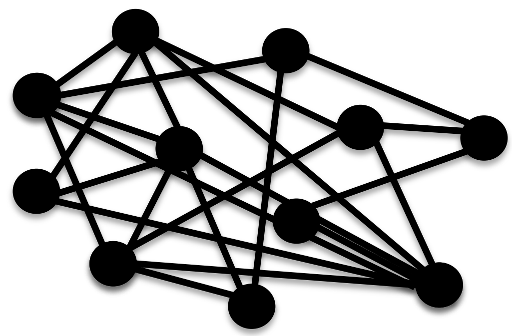
PLANTED CLIQUE

Introduced by **[Jerrum, '92]**, **[Kucera, '95]**:

Step #1: Generate E-R random graph G(n, ½)

Step #2: Add a clique on random set of ω vertices





Can we find the planted clique?

And how large does ω need to be?

Quasi-polynomial time:

Fact: There is an $n^{O(\log n)}$-time algorithm (brute-force) that can find planted cliques of size $\omega \geq C \log n$, for any $C > 2$

Quasi-polynomial time:

Fact: There is an $n^{O(\log n)}$-time algorithm (brute-force) that can find planted cliques of size $\omega \geq C \log n$, for any $C > 2$

Polynomial time:

Fact: There is a polynomial time algorithm that succeeds (whp) for $\omega \geq C \sqrt{n \log n}$ (**degree counting**)

Quasi-polynomial time:

Fact: There is an $n^{O(\log n)}$-time algorithm (brute-force) that can find planted cliques of size $\omega \geq C \log n$, for any $C > 2$

Polynomial time:

Fact: There is a polynomial time algorithm that succeeds (whp) for $\omega \geq C \sqrt{n \log n}$ (**degree counting**)

Theorem [Alon, Krivelevich, Sudakov]: There is a polynomial time algorithm that succeeds (whp) for $\omega \geq C \sqrt{n}$ (**spectral**)

Quasi-polynomial time:

Fact: There is an $n^{O(\log n)}$-time algorithm (brute-force) that can find planted cliques of size $\omega \geq C \log n$, for any $C > 2$

Polynomial time:

Fact: There is a polynomial time algorithm that succeeds (whp) for $\omega \geq C \sqrt{n \log n}$ (**degree counting**)

Theorem [Alon, Krivelevich, Sudakov]: There is a polynomial time algorithm that succeeds (whp) for $\omega \geq C \sqrt{n}$ (**spectral**)

Theorem [Deshpande, Montanari]: There is a nearly linear time algorithm that succeeds (whp) for $\omega \geq \sqrt{n/e}$

APPLICATIONS OF PLANTED CLIQUE

Planted Clique (and variants) are basic problems in **average-case complexity**, imply many other hardness results:

APPLICATIONS OF PLANTED CLIQUE

Planted Clique (and variants) are basic problems in **average-case complexity**, imply many other hardness results:

- Discovering motifs in biological networks **[Milo et al '02]**
- Computing the best Nash Equilibrium **[HK '11]**, **[ABC '13]**
- Property testing **[Alon et al '07]**
- Sparse PCA **[Berthet, Rigollet '13]**
- Compressed sensing **[Koiran, Zouzias '14]**
- Cryptography **[Juels, Peinado '00]**, **[Applebaum et al '10]**
- Mathematical finance **[Arora et al '10]**

LOWER BOUNDS?

Is it *actually* hard to find $n^{1/2-\varepsilon}$-sized planted cliques?

LOWER BOUNDS?

Is it *actually* hard to find $n^{1/2-\varepsilon}$-sized planted cliques?

Complexity-theoretic reasons lower bound are unlikely to be based on **P vs. NP**

e.g. **[Feigenbaum, Fortnow '93]**, **[Bogdanov, Trevisan '06]**

LOWER BOUNDS?

Is it *actually* hard to find $n^{1/2-\varepsilon}$-sized planted cliques?

Complexity-theoretic reasons lower bound are unlikely to be based on **P vs. NP**

e.g. **[Feigenbaum, Fortnow '93]**, **[Bogdanov, Trevisan '06]**

Our best evidence seems to Sum-of-Squares lower bounds

Sum-of-Squares for planted clique:

(1) $\widetilde{\mathbb{E}}$ is linear

(2) $\widetilde{\mathbb{E}}[1] = 1$

(3) $\widetilde{\mathbb{E}}[p^2] \geq 0$

for all deg(p) ≤ d/2

general

Sum-of-Squares for planted clique:

(1) $\widetilde{\mathbb{E}}$ is linear

(2) $\widetilde{\mathbb{E}}[1] = 1$

(3) $\widetilde{\mathbb{E}}[p^2] \geq 0$

for all deg(p) ≤ d/2

general

(4) $\widetilde{\mathbb{E}}[x_i^2 p] = \widetilde{\mathbb{E}}[x_i p]$

(**booleanity**)

Sum-of-Squares for planted clique:

(1) $\widetilde{\mathbb{E}}$ is linear

(2) $\widetilde{\mathbb{E}}[1] = 1$

(3) $\widetilde{\mathbb{E}}[p^2] \geq 0$

for all deg(p) ≤ d/2

general

(4) $\widetilde{\mathbb{E}}[x_i^2 p] = \widetilde{\mathbb{E}}[x_i p]$

(5) $\widetilde{\mathbb{E}}[\sum x_i] = \omega$

(**clique size**)

Constraints on the pseudo-expectation:

(1) $\widetilde{\mathbb{E}}$ is linear

(2) $\widetilde{\mathbb{E}}[1] = 1$

(3) $\widetilde{\mathbb{E}}[p^2] \geq 0$

for all deg(p) ≤ d/2

general

(4) $\widetilde{\mathbb{E}}[x_i^2 p] = \widetilde{\mathbb{E}}[x_i p]$

(5) $\widetilde{\mathbb{E}}[\sum x_i] = \omega$

(6) $\widetilde{\mathbb{E}}[x_i x_j p] = 0$

for all (i,j) not an edge

(**clique constraints**)

Constraints on the pseudo-expectation:

(1) $\widetilde{\mathbb{E}}$ is linear

(2) $\widetilde{\mathbb{E}}[1] = 1$

(3) $\widetilde{\mathbb{E}}[p^2] \geq 0$

for all deg(p) ≤ d/2

general

(4) $\widetilde{\mathbb{E}}[x_i^2 p] = \widetilde{\mathbb{E}}[x_i p]$

(5) $\widetilde{\mathbb{E}}[\sum x_i] = \omega$

(6) $\widetilde{\mathbb{E}}[x_i x_j p] = 0$

for all (i,j) not an edge

specific to planted clique

Constraints on the pseudo-expectation:

(1) $\widetilde{\mathbb{E}}$ is linear

(2) $\widetilde{\mathbb{E}}[1] = 1$

(3) $\widetilde{\mathbb{E}}[p^2] \geq 0$

for all deg(p) ≤ d/2

general

(4) $\widetilde{\mathbb{E}}[x_i^2 p] = \widetilde{\mathbb{E}}[x_i p]$

(5) $\widetilde{\mathbb{E}}[\sum x_i] = \omega$

(6) $\widetilde{\mathbb{E}}[x_i x_j p] = 0$

for all (i,j) not an edge

specific to planted clique

Can SOS find n^{ε}-sized planted cliques in polynomial time?

A STRONG LOWER BOUND

Nearly optimal lower bound against SOS, for the planted clique problem (via pseudo-Bayesian techniques):

Theorem [Barak, Hopkins, Kelner, Kothari, Moitra, Potechin]: The integrality gap of the level d Sum-of-Squares hierarchy is

$$n^{\frac{1}{2} - c\sqrt{d/\log n}}$$

for some constant $c > 0$

For any $d = o(\log n)$, the integrality gap is $n^{1/2-o(1)}$

A STRONG LOWER BOUND

Nearly optimal lower bound against SOS, for the planted clique problem (via pseudo-Bayesian techniques):

Theorem [Barak, Hopkins, Kelner, Kothari, Moitra, Potechin]:
The integrality gap of the level d Sum-of-Squares hierarchy is

$$n^{\frac{1}{2} - c\sqrt{d/\log n}}$$

for some constant c > 0

For any d = o(log n), the integrality gap is $n^{1/2-o(1)}$

Builds on **[Meka, Potechin, Wigderson '14]**, **[Deshpande Montanari '15]**, **[Hopkins, Kothari, Potechin, Raghavendra, Scrhamm '16]**

A STRONG LOWER BOUND

Our Approach: **Pseudo-calibration**

New insights into what makes SOS powerful, and how to fool it

A STRONG LOWER BOUND

Our Approach: **Pseudo-calibration**

New insights into what makes SOS powerful, and how to fool it

When our *recipe* fails, it often yields spectral algorithms

OUTLINE

Part I: Introduction

- MAXCUT and the Sum-of-Squares Hierarchy
- A Dual View via Pseudo-expectation

Part II: Rounding SOS

Part III: Fooling SOS

- Planted Clique and its Applications
- The MPW Moments and Corrections
- Pseudo-calibration and Fourier Analysis

Part IV: Sparse PCA and Computational vs. Statistical Tradeoffs

Part V: Equivalence with Spectral Methods

OUTLINE

Part I: Introduction

- MAXCUT and the Sum-of-Squares Hierarchy
- A Dual View via Pseudo-expectation

Part II: Rounding SOS

Part III: Fooling SOS

- Planted Clique and its Applications
- **The MPW Moments and Corrections**
- Pseudo-calibration and Fourier Analysis

Part IV: Sparse PCA and Computational vs. Statistical Tradeoffs

Part V: Equivalence with Spectral Methods

PSEUDO-MOMENTS

How can we fool the SOS algorithm into thinking there is a $n^{1/2-o(1)}$ sized clique in $G(n, ½)$?

PSEUDO-MOMENTS

How can we fool the SOS algorithm into thinking there is a $n^{1/2-o(1)}$ sized clique in $G(n, \frac{1}{2})$?

Usual Approach: Adapt integrality gaps from weaker hierarchies

PSEUDO-MOMENTS

How can we fool the SOS algorithm into thinking there is a $n^{1/2-o(1)}$ sized clique in $G(n, ½)$?

Usual Approach: Adapt integrality gaps from weaker hierarchies

This works for random CSPs

PSEUDO-MOMENTS

How can we fool the SOS algorithm into thinking there is a $n^{1/2-o(1)}$ sized clique in $G(n, ½)$?

Usual Approach: Adapt integrality gaps from weaker hierarchies

This works for random CSPs

Theorem [Feige, Krauthgamer]: The integrality gap of the level d LS+ hierarchy is

$$\sqrt{\frac{n}{2^d}}$$

Theorem [Meka, Potechin, Wigderson]: The integrality gap of the level d Sum-of-Squares hierarchy is

$$n^{1/d-o(1)}$$

Theorem [Meka, Potechin, Wigderson]: The integrality gap of the level d Sum-of-Squares hierarchy is

$$n^{1/d - o(1)}$$

In particular, set:

$$\widetilde{\mathbb{E}}_{MPW}[\overbrace{\prod_{i \in A} x_i}^{x_A}] = 2^{\binom{|A|}{2}} \left(\frac{\omega}{n}\right)^{|A|}$$

if A is clique, zero otherwise.

Theorem [Meka, Potechin, Wigderson]: The integrality gap of the level d Sum-of-Squares hierarchy is

$$n^{1/d-o(1)}$$

In particular, set:

$$\widetilde{\mathbb{E}}_{MPW}[\overbrace{\prod_{i\in A} x_i}^{x_A}] = 2^{\binom{|A|}{2}}\left(\frac{\omega}{n}\right)^{|A|}$$

if A is clique, zero otherwise. Extend by linearity to all deg(p) ≤ d

Theorem [Meka, Potechin, Wigderson]: The integrality gap of the level d Sum-of-Squares hierarchy is

$$n^{1/d - o(1)}$$

In particular, set:

$$\widetilde{\mathbb{E}}_{MPW}[\underbrace{\prod_{i \in A} x_i}_{x_A}] = 2^{\binom{|A|}{2}} \left(\frac{\omega}{n}\right)^{|A|}$$

if A is clique, zero otherwise. Extend by linearity to all deg(p) ≤ d

Approach: Spectral bounds on **locally random matrices**

Theorem [Meka, Potechin, Wigderson]: The integrality gap of the level d Sum-of-Squares hierarchy is

$$n^{1/d - o(1)}$$

Theorem [Meka, Potechin, Wigderson]: The integrality gap of the level d Sum-of-Squares hierarchy is

$$n^{1/d-o(1)}$$

Improved analysis due to **[Deshpande, Montanari]**, for d = 4

$$n^{1/3-o(1)}$$

Theorem [Meka, Potechin, Wigderson]: The integrality gap of the level d Sum-of-Squares hierarchy is

$$n^{1/d - o(1)}$$

Improved analysis due to **[Deshpande, Montanari]**, for d = 4

$$n^{1/3 - o(1)}$$

And due to **[Hopkins, Kothari, Potechin]** for any d

$$n^{1/(\lceil d/2 \rceil + 1) - o(1)}$$

Theorem [Meka, Potechin, Wigderson]: The integrality gap of the level d Sum-of-Squares hierarchy is

$$n^{1/d - o(1)}$$

Improved analysis due to **[Deshpande, Montanari]**, for d = 4

$$n^{1/3 - o(1)}$$

And due to **[Hopkins, Kothari, Potechin]** for any d

$$n^{1/(\lceil d/2 \rceil + 1) - o(1)}$$

But these bounds are *tight* (for *these* moments)

KELNER'S POLYNOMIAL

Do the MPW moments work beyond $n^{1/(\lceil d/2 \rceil + 1)}$?

KELNER'S POLYNOMIAL

Do the MPW moments work beyond $n^{1/(\lceil d/2 \rceil + 1)}$?

Set $G_{i,j}$ = $\begin{cases} +1 & \text{if (i,j) an edge} \\ -1 & \text{else} \end{cases}$

$$P_{G,i} = \Big(\sum_j G_{i,j} x_j \Big)^\ell$$

KELNER'S POLYNOMIAL

Do the MPW moments work beyond $n^{1/(\lceil d/2 \rceil + 1)}$?

Set $G_{i,j} = \begin{cases} +1 & \text{if (i,j) an edge} \\ -1 & \text{else} \end{cases}$

$$P_{G,i} = \Big(\sum_j G_{i,j} x_j \Big)^\ell$$

If there is an ω-sized planted clique:

$$\mathop{\mathbb{E}}_{(G,x) \leftarrow G(n,1/2,\omega)}[P_{G,i}^2] \geq \left(\frac{\omega}{n}\right)\omega^{2\ell}$$

KELNER'S POLYNOMIAL

Do the MPW moments work beyond $n^{1/(\lceil d/2 \rceil + 1)}$?

Set $G_{i,j} = \begin{cases} +1 & \text{if (i,j) an edge} \\ -1 & \text{else} \end{cases}$

$$P_{G,i} = \Big(\sum_j G_{i,j} x_j \Big)^{\ell}$$

If there is an ω-sized planted clique:

$$\mathop{\mathbb{E}}_{(G,x) \leftarrow G(n,1/2,\omega)}[P_{G,i}^2] \geq \left(\frac{\omega}{n}\right) \omega^{2\ell}$$

But if G is sampled from G(n, ½):

$$\mathop{\mathbb{E}}_{G \leftarrow G(n,1/2)}[\widetilde{\mathbb{E}}_{MPW}[P_{G,i}^2]] \leq (n^{\ell}) \left(\frac{\omega}{n}\right)^{\ell} = \omega^{\ell}$$

KELNER'S POLYNOMIAL

Do the MPW moments work beyond $n^{1/(\lceil d/2 \rceil + 1)}$?

Set $G_{i,j} = \begin{cases} +1 & \text{if (i,j) an edge} \\ -1 & \text{else} \end{cases}$ $\quad P_{G,i} = \Big(\sum_j G_{i,j} x_j\Big)^{\ell}$

If there is an ω-sized planted clique:

$$\mathop{\mathbb{E}}_{(G,x) \leftarrow G(n,1/2,\omega)}[P_{G,i}^2] \geq \left(\frac{\omega}{n}\right)\omega^{2\ell}$$

But if G is sampled from G(n, ½):

$$\mathop{\mathbb{E}}_{G \leftarrow G(n,1/2)}[\widetilde{\mathbb{E}}_{MPW}[P_{G,i}^2]] \leq (n^{\ell})\left(\frac{\omega}{n}\right)^{\ell} = \omega^{\ell}$$

Need: $\omega \leq n^{1/(\ell+1)} = n^{1/(d/2+1)}$ otherwise something is wrong

KELNER'S POLYNOMIAL

Do the MPW moments work beyond $n^{1/(\lceil d/2 \rceil + 1)}$?

KELNER'S POLYNOMIAL

Do the MPW moments work beyond $n^{1/(\lceil d/2 \rceil + 1)}$?

This example can be used to find a squared polynomial whose pseudo-expectation is negative for $\omega > n^{1/(\lceil d/2 \rceil + 1)}$

$$\widetilde{\mathbb{E}}_{MPW}[P^2] < 0$$

KELNER'S POLYNOMIAL

Do the MPW moments work beyond $n^{1/(\lceil d/2 \rceil + 1)}$?

This example can be used to find a squared polynomial whose pseudo-expectation is negative for $\omega > n^{1/(\lceil d/2 \rceil + 1)}$

$$\widetilde{\mathbb{E}}_{MPW}[P^2] < 0$$

Intuition: A good pseudo-expectation attempts to **hide** info about what vertices participate in the planted clique

KELNER'S POLYNOMIAL

Do the MPW moments work beyond $n^{1/(\lceil d/2 \rceil + 1)}$?

This example can be used to find a squared polynomial whose pseudo-expectation is negative for $\omega > n^{1/(\lceil d/2 \rceil + 1)}$

$$\widetilde{\mathbb{E}}_{MPW}[P^2] < 0$$

Intuition: A good pseudo-expectation attempts to **hide** info about what vertices participate in the planted clique

But vertices with a **standard deviation higher degree**, should be a constant factor more likely to be in the p.c. (**soft constraint**)

FIXING THE MPW-MOMENTS

This family of polynomials is essentially the only thing that goes wrong at d = 4

Theorem [Hopkins et al.], [Raghavendra, Schramm]: The integrality gap of the level 4 Sum-of-Squares hierarchy is

$$n^{1/2-o(1)}$$

FIXING THE MPW-MOMENTS

This family of polynomials is essentially the only thing that goes wrong at d = 4

Theorem [Hopkins et al.], [Raghavendra, Schramm]: The integrality gap of the level 4 Sum-of-Squares hierarchy is

$$n^{1/2-o(1)}$$

Approach: Add an explicit correction term of fix all $P_{G,i}$'s, even more dependent random matrix theory

FIXING THE MPW-MOMENTS

This family of polynomials is essentially the only thing that goes wrong at d = 4

Theorem [Hopkins et al.], [Raghavendra, Schramm]: The integrality gap of the level 4 Sum-of-Squares hierarchy is

$$n^{1/2-o(1)}$$

Approach: Add an explicit correction term of fix all $P_{G,i}$'s, even more dependent random matrix theory

Is there a fix for higher degrees?

FIXING THE MPW-MOMENTS

This family of polynomials is essentially the only thing that goes wrong at d = 4

Theorem [Hopkins et al.], [Raghavendra, Schramm]: The integrality gap of the level 4 Sum-of-Squares hierarchy is

$$n^{1/2-o(1)}$$

Approach: Add an explicit correction term of fix all $P_{G,i}$'s, even more dependent random matrix theory

Is there a fix for higher degrees?

It turns out for d = 6, even the fixes need fixes, and on and on…

FIXING THE MPW-MOMENTS

This family of polynomials is essentially the only thing that goes wrong at d = 4

Theorem [Hopkins et al.], [Raghavendra, Schramm]: The integrality gap of the level 4 Sum-of-Squares hierarchy is

$$n^{1/2-o(1)}$$

Approach: Add an explicit correction term of fix all $P_{G,i}$'s, even more dependent random matrix theory

Is there a fix for higher degrees?

It turns out for d = 6, even the fixes need fixes, and on and on…

36 pgs → 40 pgs → 26 pgs → 69 pgs → ??? pgs

OUTLINE

Part I: Introduction

- MAXCUT and the Sum-of-Squares Hierarchy
- A Dual View via Pseudo-expectation

Part II: Rounding SOS

Part III: Fooling SOS

- Planted Clique and its Applications
- The MPW Moments and Corrections
- Pseudo-calibration and Fourier Analysis

Part IV: Sparse PCA and Computational vs. Statistical Tradeoffs

Part V: Equivalence with Spectral Methods

OUTLINE

Part I: Introduction

- MAXCUT and the Sum-of-Squares Hierarchy
- A Dual View via Pseudo-expectation

Part II: Rounding SOS

Part III: Fooling SOS

- Planted Clique and its Applications
- The MPW Moments and Corrections
- **Pseudo-calibration and Fourier Analysis**

Part IV: Sparse PCA and Computational vs. Statistical Tradeoffs

Part V: Equivalence with Spectral Methods

PSEUDO-CALIBRATION

Can we find pseudo-moments that satisfy the following:

$$\mathop{\mathbb{E}}_{G \leftarrow G(n,1/2)}[\widetilde{\mathbb{E}}[f(G,x)]] = \mathop{\mathbb{E}}_{(G,x) \leftarrow G(n,1/2,\omega)}[f(G,x)]$$

for all *simple* functions f?

PSEUDO-CALIBRATION

Can we find pseudo-moments that satisfy the following:

$$\mathop{\mathbb{E}}_{G \leftarrow G(n,1/2)}[\widetilde{\mathbb{E}}[f(G,x)]] = \mathop{\mathbb{E}}_{(G,x) \leftarrow G(n,1/2,\omega)}[f(G,x)]$$

for all polynomials f that are low-degree in $G_{i,j}$'s and x_i's?

Consider the pseudo-expectation of some monomial:

$$\widetilde{\mathbb{E}}[x_A] : G \to \mathbb{R}, \text{ and let } \chi_T(G) = \prod_{(i,j)\in T} G_{i,j}$$

Consider the pseudo-expectation of some monomial:

$\widetilde{\mathbb{E}}[x_A] : G \to \mathbb{R}$, and let $\chi_T(G) = \prod_{(i,j)\in T} G_{i,j}$

We can write any such function in terms of its **Fourier expansion**

$$\widetilde{\mathbb{E}}[x_A](G) = \sum_{T \subseteq \binom{[n]}{2}} \widehat{\widetilde{\mathbb{E}}[x_A]}(T)\chi_T(G)$$

Consider the pseudo-expectation of some monomial:

$$\widetilde{\mathbb{E}}[x_A] : G \to \mathbb{R}, \text{ and let } \chi_T(G) = \prod_{(i,j)\in T} G_{i,j}$$

We can write any such function in terms of its **Fourier expansion**

$$\widetilde{\mathbb{E}}[x_A](G) = \sum_{T \subseteq \binom{[n]}{2}} \widehat{\widetilde{\mathbb{E}}[x_A]}(T)\chi_T(G)$$

How should we set the Fourier coefficients?

The Fourier coefficients are chosen for us, by pseudo-calibration

The Fourier coefficients are chosen for us, by pseudo-calibration

Utilizing the expression

$$\widetilde{\mathbb{E}}[x_A](G) = \sum_{T \subseteq \binom{[n]}{2}} \widehat{\widetilde{\mathbb{E}}[x_A]}(T)\chi_T(G)$$

we can calculate:

$$\mathbb{E}_{G \leftarrow G(n,1/2)}[\widetilde{\mathbb{E}}[x_A \chi_T(G)]]$$

The Fourier coefficients are chosen for us, by pseudo-calibration

Utilizing the expression

$$\widetilde{\mathbb{E}}[x_A](G) = \sum_{T \subseteq \binom{[n]}{2}} \widehat{\widetilde{\mathbb{E}}[x_A]}(T)\chi_T(G)$$

we can calculate:

$$\mathop{\mathbb{E}}_{G \leftarrow G(n,1/2)}[\widetilde{\mathbb{E}}[x_A]\chi_T(G)]$$ (by linearity)

The Fourier coefficients are chosen for us, by pseudo-calibration

Utilizing the expression

$$\widetilde{\mathbb{E}}[x_A](G) = \sum_{T \subseteq \binom{[n]}{2}} \widehat{\widetilde{\mathbb{E}}[x_A]}(T)\chi_T(G)$$

we can calculate:

$$\mathbb{E}_{G \leftarrow G(n,1/2)}[\widetilde{\mathbb{E}}[x_A]\chi_T(G)] = \sum_{T' \subseteq \binom{[n]}{2}} \widehat{\widetilde{\mathbb{E}}[x_A]}(T')\mathbb{E}_{G \leftarrow G(n,1/2)}[\chi_T(G)\chi_{T'}(G)]$$

The Fourier coefficients are chosen for us, by pseudo-calibration

Utilizing the expression

$$\widetilde{\mathbb{E}}[x_A](G) = \sum_{T \subseteq \binom{[n]}{2}} \widehat{\widetilde{\mathbb{E}}[x_A]}(T)\chi_T(G)$$

we can calculate:

$$\mathbb{E}_{G \leftarrow G(n,1/2)}[\widetilde{\mathbb{E}}[x_A]\chi_T(G)] = \sum_{T' \subseteq \binom{[n]}{2}} \widehat{\widetilde{\mathbb{E}}[x_A]}(T') \underbrace{\mathbb{E}_{G \leftarrow G(n,1/2)}[\chi_T(G)\chi_{T'}(G)]}$$

$$= \begin{cases} +1 & \text{if } T = T' \\ 0 & \text{else} \end{cases}$$

The Fourier coefficients are chosen for us, by pseudo-calibration

Utilizing the expression

$$\widetilde{\mathbb{E}}[x_A](G) = \sum_{T \subseteq \binom{[n]}{2}} \widehat{\widetilde{\mathbb{E}}[x_A]}(T)\chi_T(G)$$

we can calculate:

$$\mathop{\mathbb{E}}_{G \leftarrow G(n,1/2)}[\widetilde{\mathbb{E}}[x_A \chi_T(G)]] = \widehat{\widetilde{\mathbb{E}}[x_A]}(T)$$

The Fourier coefficients are chosen for us, by pseudo-calibration

Utilizing the expression

$$\widetilde{\mathbb{E}}[x_A](G) = \sum_{T \subseteq \binom{[n]}{2}} \widehat{\widetilde{\mathbb{E}}[x_A]}(T)\,\chi_T(G)$$

we can calculate:

$$\mathop{\mathbb{E}}_{G \leftarrow G(n,1/2)}[\widetilde{\mathbb{E}}[x_A \chi_T(G)]] = \widehat{\widetilde{\mathbb{E}}[x_A]}(T)$$

pseudo-calibration

$$\triangleq \mathop{\mathbb{E}}_{(G,x) \leftarrow G(n,1/2,\omega)}[x_A \chi_T(G)]$$

The Fourier coefficients are chosen for us, by pseudo-calibration

Utilizing the expression

$$\widetilde{\mathbb{E}}[x_A](G) = \sum_{T \subseteq \binom{[n]}{2}} \widehat{\widetilde{\mathbb{E}}[x_A]}(T)\chi_T(G)$$

we can calculate:

$$\mathop{\mathbb{E}}_{G \leftarrow G(n,1/2)}[\widetilde{\mathbb{E}}[x_A \chi_T(G)]] = \widehat{\widetilde{\mathbb{E}}[x_A]}(T)$$

$$\triangleq \mathop{\mathbb{E}}_{(G,x) \leftarrow G(n,1/2,\omega)}[x_A \chi_T(G)] = \left(\frac{\omega}{n}\right)^{|\overbrace{V(T)}^{\text{vertices of T}} \cup A|}$$

pseudo-calibration

The Fourier coefficients are chosen for us, by pseudo-calibration

Utilizing the expression

$$\widetilde{\mathbb{E}}[x_A](G) = \sum_{T\subseteq\binom{[n]}{2}} \widehat{\widetilde{\mathbb{E}}[x_A]}(T)\,\chi_T(G)$$

we can calculate:

$$\mathbb{E}_{G\leftarrow G(n,1/2)}[\widetilde{\mathbb{E}}[x_A\chi_T(G)]] = \widehat{\widetilde{\mathbb{E}}[x_A]}(T)$$

pseudo-calibration

$$\triangleq \mathbb{E}_{(G,x)\leftarrow G(n,1/2,\omega)}[x_A\chi_T(G)] = \left(\frac{\omega}{n}\right)^{|\overbrace{V(T)}^{\text{vertices of T}}\cup A|}$$

It turns out , we need to truncate but at what degree?

TRUNCATION

Our pseudo-moments are:

$$\widetilde{\mathbb{E}}[x_A] = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |V(T) \cup A| \leq \tau}} \left(\frac{\omega}{n}\right)^{|V(T) \cup A|} \chi_T(G)$$

TRUNCATION

Our pseudo-moments are:

$$\widetilde{\mathbb{E}}[x_A] = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |V(T) \cup A| \leq \tau}} \left(\frac{\omega}{n}\right)^{|V(T) \cup A|} \chi_T(G)$$

Lemma: With high probability,

$$|\widetilde{\mathbb{E}}[1] - 1| \leq \tau \max_{t \leq \tau} 2^{t^2} \left(\frac{\omega}{\sqrt{n}}\right)^t$$

TRUNCATION

Our pseudo-moments are:

$$\widetilde{\mathbb{E}}[x_A] = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |V(T) \cup A| \leq \tau}} \left(\frac{\omega}{n}\right)^{|V(T) \cup A|} \chi_T(G)$$

Lemma: With high probability,

$$|\widetilde{\mathbb{E}}[1] - 1| \leq \tau \max_{t \leq \tau} 2^{t^2} \left(\frac{\omega}{\sqrt{n}}\right)^t$$

(1) This is why we need to truncate

TRUNCATION

Our pseudo-moments are:

$$\widetilde{\mathbb{E}}[x_A] = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |V(T) \cup A| \leq \tau}} \left(\frac{\omega}{n}\right)^{|V(T) \cup A|} \chi_T(G)$$

Lemma: With high probability,

$$|\widetilde{\mathbb{E}}[1] - 1| \leq \tau \max_{t \leq \tau} 2^{t^2} \left(\frac{\omega}{\sqrt{n}}\right)^t$$

(2) Is small enough for any $\omega \leq n^{1/2-\epsilon}$ for $\tau \leq \frac{\epsilon}{2} \log n$

$n^{-\Omega(\epsilon)}$

TRUNCATION

Our pseudo-moments are:

$$\widetilde{\mathbb{E}}[x_A] = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |V(T) \cup A| \leq \tau}} \left(\frac{\omega}{n}\right)^{|V(T) \cup A|} \chi_T(G)$$

Lemma: With high probability,

$$|\widetilde{\mathbb{E}}[1] - 1| \leq \tau \max_{t \leq \tau} 2^{t^2} \left(\frac{\omega}{\sqrt{n}}\right)^t$$

(3) Can always renormalize pseudo-expectation so $\widetilde{\mathbb{E}}[1] = 1$

TRUNCATION

Our pseudo-moments are:

$$\widetilde{\mathbb{E}}[x_A] = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |V(T) \cup A| \leq \tau}} \left(\frac{\omega}{n}\right)^{|V(T) \cup A|} \chi_T(G)$$

Lemma: With high probability,

$$|\widetilde{\mathbb{E}}[1] - 1| \leq \tau \max_{t \leq \tau} 2^{t^2} \left(\frac{\omega}{\sqrt{n}}\right)^t$$

(4) Similar bound holds (again by standard concentration) for

$$\widetilde{\mathbb{E}}[\sum_i x_i] = \omega(1 \pm n^{-\Omega(\epsilon)})$$

TRUNCATION

Our pseudo-moments are:

$$\widetilde{\mathbb{E}}[x_A] = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |V(T) \cup A| \leq \tau}} \left(\frac{\omega}{n}\right)^{|V(T) \cup A|} \chi_T(G)$$

TRUNCATION

Our pseudo-moments are:

$$\widetilde{\mathbb{E}}[x_A] = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |V(T) \cup A| \leq \tau}} \left(\frac{\omega}{n}\right)^{|V(T) \cup A|} \chi_T(G)$$

Lemma: If A is not a clique then

$$\widetilde{\mathbb{E}}[x_A] = 0$$

TRUNCATION

Our pseudo-moments are:

$$\widetilde{\mathbb{E}}[x_A] = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |V(T) \cup A| \leq \tau}} \left(\frac{\omega}{n}\right)^{|V(T) \cup A|} \chi_T(G)$$

Lemma: If A is not a clique then

$$\widetilde{\mathbb{E}}[x_A] = 0$$

Follows from Fourier expansion of AND, and grouping terms

TRUNCATION

Our pseudo-moments are:

$$\widetilde{\mathbb{E}}[x_A] = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |V(T) \cup A| \leq \tau}} \left(\frac{\omega}{n}\right)^{|V(T)\cup A|} \chi_T(G)$$

Lemma: If A is not a clique then

$$\widetilde{\mathbb{E}}[x_A] = 0$$

Follows from Fourier expansion of AND, and grouping terms

This is why we use $|V(T) \cup A| \leq \tau$ for truncation

TRUNCATION

Our pseudo-moments are:

$$\widetilde{\mathbb{E}}[x_A] = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |V(T) \cup A| \leq \tau}} \left(\frac{\omega}{n}\right)^{|V(T) \cup A|} \chi_T(G)$$

TRUNCATION

Our pseudo-moments are:

$$\widetilde{\mathbb{E}}[x_A] = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |V(T) \cup A| \leq \tau}} \left(\frac{\omega}{n}\right)^{|V(T) \cup A|} \chi_T(G)$$

Lemma: Let $f_G(x) = \sum_{|S| \leq 2d} c_A(G) x_A$ where $\deg(c_A) \leq \tau$, then

$$\mathop{\mathbb{E}}_{G \leftarrow G(n,1/2)}[\widetilde{\mathbb{E}}[f_G(x)]] = \mathop{\mathbb{E}}_{(G,x) \leftarrow G(n,1/2,\omega)}[f_G(x)]$$

What about proving positivity? e.g. $\widetilde{\mathbb{E}}[p^2] \geq 0$

What about proving positivity? e.g. $\widetilde{\mathbb{E}}[p^2] \geq 0$

This step is *by far* the most challenging (**as usual**)

What about proving positivity? e.g. $\widetilde{\mathbb{E}}[p^2] \geq 0$

This step is *by far* the most challenging (**as usual**)

Interestingly it is much easier to show that

$$\mathop{\mathbb{E}}_{G \leftarrow G(n,1/2)}[\widetilde{\mathbb{E}}[p^2]] \geq 0$$

OUTLINE

Part I: Introduction

- MAXCUT and the Sum-of-Squares Hierarchy
- A Dual View via Pseudo-expectation

Part II: Rounding SOS

Part III: Fooling SOS

- Planted Clique and its Applications
- The MPW Moments and Corrections
- Pseudo-calibration and Fourier Analysis

Part IV: Sparse PCA and Computational vs. Statistical Tradeoffs

Part V: Equivalence with Spectral Methods

OUTLINE

Part I: Introduction

- MAXCUT and the Sum-of-Squares Hierarchy
- A Dual View via Pseudo-expectation

Part II: Rounding SOS

Part III: Fooling SOS

- Planted Clique and its Applications
- The MPW Moments and Corrections
- Pseudo-calibration and Fourier Analysis

Part IV: Sparse PCA and Computational vs. Statistical Tradeoffs

Part V: Equivalence with Spectral Methods

SPARSE PRINCIPAL COMPONENT ANALYSIS

Goal: Given samples $X_1, X_2, \cdots, X_n \in \mathbb{R}^d$ from

$$\mathcal{N}(0, I + \theta v v^T)$$ **spiked covariance model**

where v is k-sparse and its nonzero entries are $\pm 1/\sqrt{k}$

SPARSE PRINCIPAL COMPONENT ANALYSIS

Goal: Given samples $X_1, X_2, \cdots, X_n \in \mathbb{R}^d$ from

$$\mathcal{N}(0, I + \theta v v^T)$$ **spiked covariance model**

where v is k-sparse and its nonzero entries are $\pm 1/\sqrt{k}$

How large does the signal parameter θ need to be to detect the spike?

Theorem: There is a $d^{O(k)}$-time algorithm (brute-force) that can detect the spike (with failure probability δ) when

$$\theta \gtrsim \sqrt{\frac{k \log d/\delta}{n}}$$

Theorem: There is a $d^{O(k)}$-time algorithm (brute-force) that can detect the spike (with failure probability δ) when

$$\theta \gtrsim \sqrt{\frac{k \log d/\delta}{n}}$$

Compute top eigenvalue of all k x k principal submatrices of the empirical covariance

Theorem: There is a $d^{O(k)}$-time algorithm (brute-force) that can detect the spike (with failure probability δ) when

$$\theta \gtrsim \sqrt{\frac{k \log d/\delta}{n}}$$

Compute top eigenvalue of all k x k principal submatrices of the empirical covariance

Theorem: There is a polynomial time algorithm that can detect the spike (with failure probability δ) when

$$\theta \gtrsim \sqrt{\frac{k^2 \log d/\delta}{n}}$$

Theorem: There is a $d^{O(k)}$-time algorithm (brute-force) that can detect the spike (with failure probability δ) when

$$\theta \gtrsim \sqrt{\frac{k \log d/\delta}{n}}$$

Compute top eigenvalue of all k x k principal submatrices of the empirical covariance

Theorem: There is a polynomial time algorithm that can detect the spike (with failure probability δ) when

$$\theta \gtrsim \sqrt{\frac{k^2 \log d/\delta}{n}}$$

Select the k largest entries along the diagonal of the empirical covariance matrix

LOWER BOUNDS FROM PLANTED CLIQUE

In an influential paper, **[Berthet, Rigollet]** showed:

LOWER BOUNDS FROM PLANTED CLIQUE

In an influential paper, **[Berthet, Rigollet]** showed:

Theorem: Assuming that there is no polynomial time algorithm for finding a planted clique of size

$$k = n^{1/2-\epsilon}$$

for any $\epsilon > 0$ then there is no polynomial time algorithm for **subgaussian** sparse PCA with

$$\sqrt{\frac{k^\alpha}{n}} \leq \theta \leq \sqrt{\frac{k^2 \log d}{n}}$$

for any $1 \leq \alpha < 2$ that succeeds with constant probability

DISCUSSION

Their reduction leaves open the following possibility:

Is there a quasi-polynomial time algorithm for detecting a spike in sparse PCA for much smaller values of θ?

DISCUSSION

Their reduction leaves open the following possibility:

Is there a quasi-polynomial time algorithm for detecting a spike in sparse PCA for much smaller values of θ?

But we can still hope to prove strong SOS lower bounds:

DISCUSSION

Their reduction leaves open the following possibility:

Is there a quasi-polynomial time algorithm for detecting a spike in sparse PCA for much smaller values of θ?

But we can still hope to prove strong SOS lower bounds:

> e.g. **[Hopkins et al.]** proved polynomial degree lower bounds for the Wigner variant

DISCUSSION

Their reduction leaves open the following possibility:

Is there a quasi-polynomial time algorithm for detecting a spike in sparse PCA for much smaller values of θ?

But we can still hope to prove strong SOS lower bounds:

e.g. **[Hopkins et al.]** proved polynomial degree lower bounds for the Wigner variant

Evidence for average-case complexity without reductions!

OUTLINE

Part I: Introduction

- MAXCUT and the Sum-of-Squares Hierarchy
- A Dual View via Pseudo-expectation

Part II: Rounding SOS

Part III: Fooling SOS

- Planted Clique and its Applications
- The MPW Moments and Corrections
- Pseudo-calibration and Fourier Analysis

Part IV: Sparse PCA and Computational vs. Statistical Tradeoffs

Part V: Equivalence with Spectral Methods

OUTLINE

Part I: Introduction

- MAXCUT and the Sum-of-Squares Hierarchy
- A Dual View via Pseudo-expectation

Part II: Rounding SOS

Part III: Fooling SOS

- Planted Clique and its Applications
- The MPW Moments and Corrections
- Pseudo-calibration and Fourier Analysis

Part IV: Sparse PCA and Computational vs. Statistical Tradeoffs

Part V: Equivalence with Spectral Methods

A SPECTRAL CHARACTERIZATION

Is SOS only as powerful as low degree polynomials?

A SPECTRAL CHARACTERIZATION

Is SOS only as powerful as low degree polynomials?

E.g. if low degree subgraph counts fail, then so does SOS:

A SPECTRAL CHARACTERIZATION

Is SOS only as powerful as low degree polynomials?

E.g. if low degree subgraph counts fail, then so does SOS:

Theorem [Hopkins et al.]: Suppose degree d SOS can distinguish between planted and unplanted instances and that the problem is **resilient to rerandomizing most coordinates.**

Then there is an $n^{O(d)}$ x $n^{O(d)}$ matrix Q whose entries are degree O(d) polynomials in the instance variables where

(1) $\mathbb{E}_{\mathcal{I}\sim \text{unplanted}}[\lambda^{+}(Q(\mathcal{I}))] \leq 1$

(2) $\mathbb{E}_{\mathcal{I}\sim \text{planted}}[\lambda^{+}(Q(\mathcal{I}))] \geq n^{10d}$

OPEN QUESTIONS

But how do you prove lower bounds against spectral methods whose entries are polynomials?

OPEN QUESTIONS

But how do you prove lower bounds against spectral methods whose entries are polynomials?

i.e. no spectral method on low degree subgraph counts succeeds would give new proof of SOS lower bound for planted clique

OPEN QUESTIONS

But how do you prove lower bounds against spectral methods whose entries are polynomials?

i.e. no spectral method on low degree subgraph counts succeeds would give new proof of SOS lower bound for planted clique

Can you prove SOS lower bounds for community detection beneath the Kesten-Stigum bound?

OPEN QUESTIONS

But how do you prove lower bounds against spectral methods whose entries are polynomials?

i.e. no spectral method on low degree subgraph counts succeeds would give new proof of SOS lower bound for planted clique

Can you prove SOS lower bounds for community detection beneath the Kesten-Stigum bound?

Can tools from random graph theory/statistics (e.g. **small subgraph conditioning method**, **contiguity**) be useful?

Summary:

- Sum-of-Squares hierarchy as a relaxation for **polynomial optimization**
- Upper bounds for **MAXCUT** and lower bounds for **planted clique**
- Lower bounds as a form of evidence for average-case hardness, **computational vs. statistical gaps**

Summary:

- Sum-of-Squares hierarchy as a relaxation for **polynomial optimization**
- Upper bounds for **MAXCUT** and lower bounds for **planted clique**
- Lower bounds as a form of evidence for average-case hardness, **computational vs. statistical gaps**

Thanks! Any Questions?