Planted Clique, Sum-of-Squares and Pseudo-Calibration

Ankur Moitra (MIT)

joint work with Boaz Barak, Sam Hopkins, Jon Kelner, Pravesh Kothari and Aaron Potechin

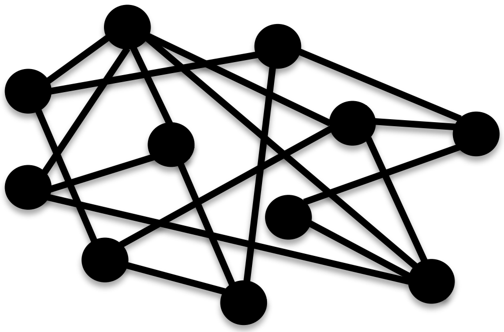
PLANTED CLIQUE

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

PLANTED CLIQUE

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

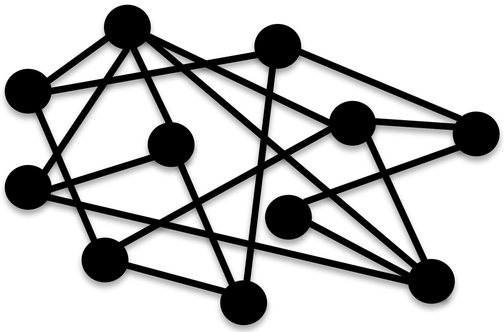
Step #1: Generate E-R random graph G(n, ½)



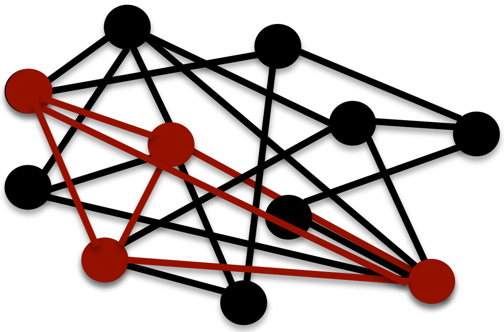
PLANTED CLIQUE

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

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



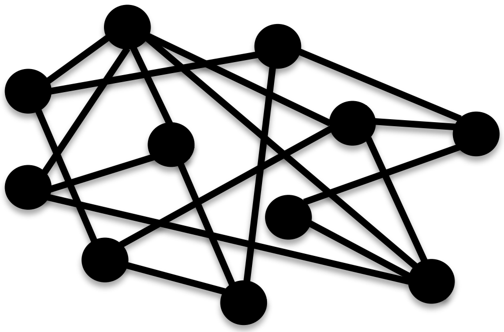
Step #2: Add a clique on random set of ω vertices



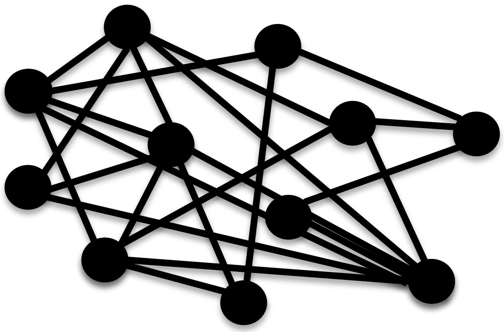
PLANTED CLIQUE

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

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



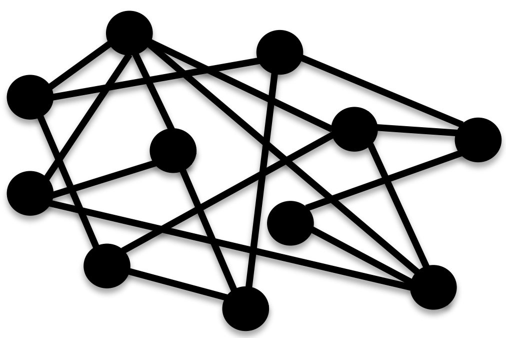
Step #2: Add a clique on random set of ω vertices



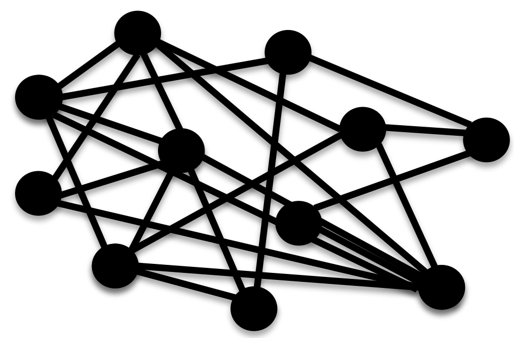
PLANTED CLIQUE

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?

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 '98]: 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 '98]: There is a polynomial time algorithm that succeeds (whp) for $\omega \geq C \sqrt{n}$ (**spectral**)

Theorem [Deshpande, Montanari '13]: 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 analysis**, many applications:

APPLICATIONS OF PLANTED CLIQUE

Planted Clique (and variants) are basic problems in **average-case analysis**, many applications:

- 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 come from **hierarchies**…

OUTLINE

Part I: Introduction

- Planted Clique and its Applications
- The Sum-of-Squares Hierarchy
- Our Results

Part II: Fooling SOS

- The Meka-Potechin-Wigderson Moments
- Kelner's Polynomial, and Corrections at d = 4
- Pseudo-Calibration and Fourier Analysis
- Symbolic Factorization and Intersection Terms

OUTLINE

Part I: Introduction

- Planted Clique and its Applications
- **The Sum-of-Squares Hierarchy**
- Our Results

Part II: Fooling SOS

- The Meka-Potechin-Wigderson Moments
- Kelner's Polynomial, and Corrections at $d = 4$
- Pseudo-Calibration and Fourier Analysis
- Symbolic Factorization and Intersection Terms

SUM-OF-SQUARES HIERARCHY

Powerful hierarchy of semidefinite programs, introduced by **[Shor ‘87]**, **[Nesterov ‘00]**, **[Parrilo ‘00]**, **[Lasserre ‘01]**

SUM-OF-SQUARES HIERARCHY

Powerful hierarchy of semidefinite programs, introduced by **[Shor '87]**, **[Nesterov '00]**, **[Parrilo '00]**, **[Lasserre '01]**

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

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

Called a **Pseudo-expectation**

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

$\underbrace{\hphantom{\text{for all deg(p) ≤ d/2 xxxxxx}}}_{\textbf{general}}$

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

(**booleanity**)

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

(**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

$\underbrace{\hspace{6em}}$

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

$\underbrace{\hspace{6em}}$

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

(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

E.g. if a_1, a_2, … a_n is the indicator vector of an ω-sized clique

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

meets (1) – (6)

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

(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

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

Called the level d **Sum-of-Squares Algorithm**

- 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]**

- 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]**

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

OUTLINE

Part I: Introduction

- Planted Clique and its Applications
- The Sum-of-Squares Hierarchy
- Our Results

Part II: Fooling SOS

- The Meka-Potechin-Wigderson Moments
- Kelner's Polynomial, and Corrections at $d = 4$
- Pseudo-Calibration and Fourier Analysis
- Symbolic Factorization and Intersection Terms

OUTLINE

Part I: Introduction

- Planted Clique and its Applications
- The Sum-of-Squares Hierarchy
- **Our Results**

Part II: Fooling SOS

- The Meka-Potechin-Wigderson Moments
- Kelner's Polynomial, and Corrections at d = 4
- Pseudo-Calibration and Fourier Analysis
- Symbolic Factorization and Intersection Terms

OUR RESULTS

We show a nearly optimal lower bound against SoS, for the planted clique problem:

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

OUR RESULTS

We show a nearly optimal lower bound against SoS, for the planted clique problem:

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

Improves upon [Meka, Potechin, Wigderson ‘14], [Deshpande Montanari ‘15], [Hopkins, Kothari, Potechin, Raghavendra, Scrhamm ‘16]

OUR RESULTS

Our Approach: **Pseudo-calibration**

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

OUR RESULTS

Our Approach: **Pseudo-calibration**

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

When our *recipe* fails, does it immediately yield algorithms?

OUTLINE

Part I: Introduction

- Planted Clique and its Applications
- The Sum-of-Squares Hierarchy
- Our Results

Part II: Fooling SOS

- The Meka-Potechin-Wigderson Moments
- Kelner's Polynomial, and Corrections at d = 4
- Pseudo-Calibration and Fourier Analysis
- Symbolic Factorization and Intersection Terms

OUTLINE

Part I: Introduction

- Planted Clique and its Applications
- The Sum-of-Squares Hierarchy
- Our Results

Part II: Fooling SOS

- **The Meka-Potechin-Wigderson Moments**
- Kelner's Polynomial, and Corrections at d = 4
- Pseudo-Calibration and Fourier Analysis
- Symbolic Factorization and Intersection Terms

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 $G(n, ½)$?

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 $G(n, \frac{1}{2})$?

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 '03]: The integrality gap of the level d LS+ hierarchy is

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

Theorem [Meka, Potechin, Wigderson '14]: The integrality gap of the level d Sum-of-Squares hierarchy is

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

Theorem [Meka, Potechin, Wigderson '14]: 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 '14]: 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 '14]: 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

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

Theorem [Meka, Potechin, Wigderson '14]: The integrality gap of the level d Sum-of-Squares hierarchy is

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

Theorem [Meka, Potechin, Wigderson '14]: The integrality gap of the level d Sum-of-Squares hierarchy is

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

Improved analysis due to **[Deshpande, Montanari '15]**, for d = 4

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

Theorem [Meka, Potechin, Wigderson '14]: The integrality gap of the level d Sum-of-Squares hierarchy is

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

Improved analysis due to **[Deshpande, Montanari '15]**, for d = 4

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

And due to **[Hopkins, Kothari, Potechin '16]** for any d

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

Theorem [Meka, Potechin, Wigderson '14]: The integrality gap of the level d Sum-of-Squares hierarchy is

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

Improved analysis due to **[Deshpande, Montanari '15]**, for d = 4

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

And due to **[Hopkins, Kothari, Potechin '16]** for any d

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

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

OUTLINE

Part I: Introduction

- Planted Clique and its Applications
- The Sum-of-Squares Hierarchy
- Our Results

Part II: Fooling SOS

- The Meka-Potechin-Wigderson Moments
- Kelner's Polynomial, and Corrections at d = 4
- Pseudo-Calibration and Fourier Analysis
- Symbolic Factorization and Intersection Terms

OUTLINE

Part I: Introduction

- Planted Clique and its Applications
- The Sum-of-Squares Hierarchy
- Our Results

Part II: Fooling SOS

- The Meka-Potechin-Wigderson Moments
- **Kelner's Polynomial, and Corrections at d = 4**
- Pseudo-Calibration and Fourier Analysis
- Symbolic Factorization and Intersection Terms

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

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

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 ’16], [Raghavendra, Schramm ‘16]:
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 '16], [Raghavendra, Schramm '16]:
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 '16], [Raghavendra, Schramm '16]:
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 '16], [Raghavendra, Schramm '16]: 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 '16], [Raghavendra, Schramm '16]: 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

- Planted Clique and its Applications
- The Sum-of-Squares Hierarchy
- Our Results

Part II: Fooling SOS

- The Meka-Potechin-Wigderson Moments
- Kelner's Polynomial, and Corrections at d = 4
- Pseudo-Calibration and Fourier Analysis
- Symbolic Factorization and Intersection Terms

OUTLINE

Part I: Introduction

- Planted Clique and its Applications
- The Sum-of-Squares Hierarchy
- Our Results

Part II: Fooling SOS

- The Meka-Potechin-Wigderson Moments
- Kelner's Polynomial, and Corrections at $d = 4$
- **Pseudo-Calibration and Fourier Analysis**
- Symbolic Factorization and Intersection Terms

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:

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

Consider the pseudo-expectation of some monomial:

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

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

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

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:

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

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

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:

$$\mathop{\mathbb{E}}_{G \leftarrow G(n,1/2)}[\widetilde{\mathbb{E}}[x_A]\chi_T(G)] = \sum_{T' \subseteq \binom{[n]}{2}} \widehat{\widetilde{\mathbb{E}}[x_A]}(T') \underbrace{\mathop{\mathbb{E}}_{G \leftarrow G(n,1/2)}[\chi_T(G)\chi_{T'}(G)]}_{= \begin{cases} +1 & \text{if } T = T' \\ 0 & \text{else} \end{cases}}$$

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:

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

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:

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

pseudo-calibration

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

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

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

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

vertices of T

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

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

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

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

OUTLINE

Part I: Introduction

- Planted Clique and its Applications
- The Sum-of-Squares Hierarchy
- Our Results

Part II: Fooling SOS

- The Meka-Potechin-Wigderson Moments
- Kelner's Polynomial, and Corrections at d = 4
- Pseudo-Calibration and Fourier Analysis
- Symbolic Factorization and Intersection Terms

OUTLINE

Part I: Introduction

- Planted Clique and its Applications
- The Sum-of-Squares Hierarchy
- Our Results

Part II: Fooling SOS

- The Meka-Potechin-Wigderson Moments
- Kelner's Polynomial, and Corrections at d = 4
- Pseudo-Calibration and Fourier Analysis
- **Symbolic Factorization and Intersection Terms**

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**)

As is standard, it amounts to proving a certain matrix is PSD, whose entries are:

$$\mathcal{M}(I,J) = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |V(T) \cup I \cup J| \leq \tau}} \left(\frac{\omega}{n}\right)^{|V(T) \cup I \cup J|} \chi_T(G)$$

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

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

As is standard, it amounts to proving a certain matrix is PSD, whose entries are:

$$\mathcal{M}(I,J) = \sum_{\substack{T \subseteq \binom{[n]}{2} \\ |V(T) \cup I \cup J| \leq \tau}} \left(\frac{\omega}{n}\right)^{|V(T) \cup I \cup J|} \chi_T(G)$$

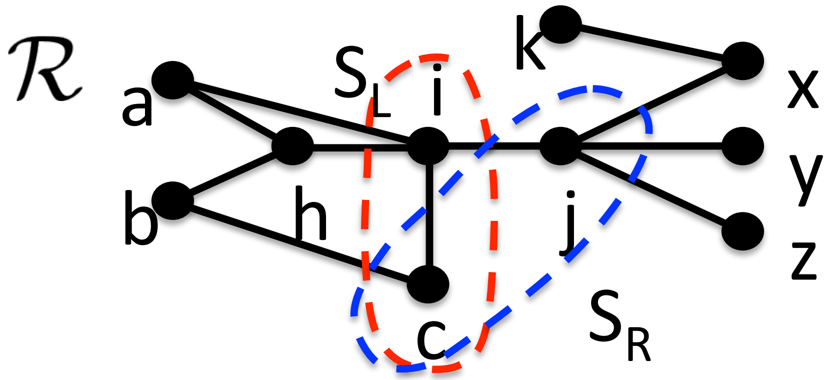
Goal: Write $\mathcal{M}$ as:

$$\mathcal{M} \approx \sum_{k} \mathcal{L}_k \mathcal{Q}_k \mathcal{L}_k^{+}$$

size of minimum vertex separator of T, btwn I and J

RIBBON DECOMPOSITION

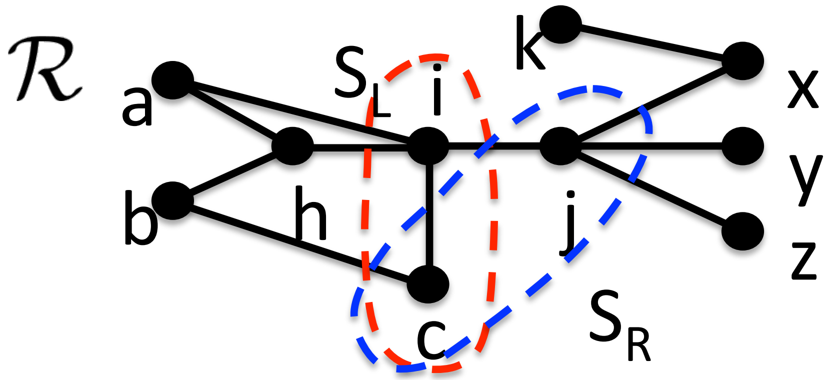
We call such graphs **(I,J)-Ribbons**, e.g.



with I = {a, b, c}, J = {c, x, y, z}.

RIBBON DECOMPOSITION

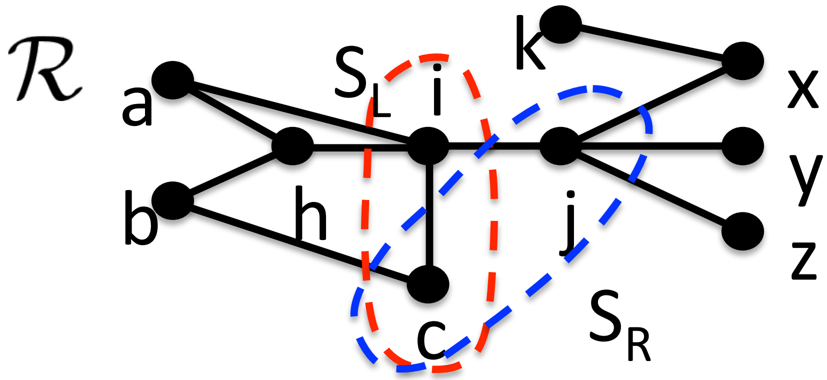
We call such graphs **(I,J)-Ribbons**, e.g.



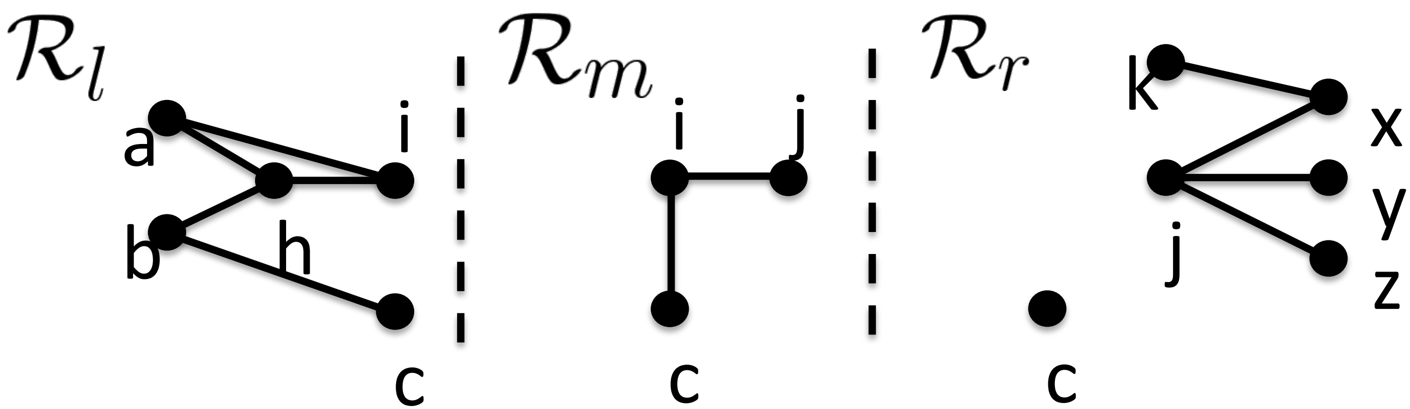
with I = {a, b, c}, J = {c, x, y, z}. Compute leftmost and rightmost minimum vertex separators S_L, S_R.

RIBBON DECOMPOSITION

We call such graphs **(I,J)-Ribbons**, e.g.



with I = {a, b, c}, J = {c, x, y, z}. Compute leftmost and rightmost minimum vertex separators S_L, S_R. Decompose



SYMBOLIC FACTORIZATION

Now we can write:

$$\mathcal{M}(I,J) \approx \text{ sum over k of}$$

$$\underbrace{\Big(\sum_{\text{valid }\mathcal{R}_l}\left(\frac{\omega}{n}\right)^{|V(\mathcal{R}_l)|}\Big)}_{\mathcal{L}_k}\underbrace{\Big(\sum_{\text{valid }\mathcal{R}_m}\left(\frac{\omega}{n}\right)^{|V(\mathcal{R}_m)|-2k}\Big)}_{\mathcal{Q}_k}\underbrace{\Big(\sum_{\text{valid }\mathcal{R}_r}\left(\frac{\omega}{n}\right)^{|V(\mathcal{R}_r)|}\Big)}_{\mathcal{L}_k^T}$$

SYMBOLIC FACTORIZATION

Now we can write:

$$\mathcal{M}(I,J) \approx \text{ sum over k of}$$

$$\underbrace{\Big(\sum_{\text{valid } \mathcal{R}_l} \Big(\frac{\omega}{n}\Big)^{|V(\mathcal{R}_l)|}\Big)}_{\mathcal{L}_k}\underbrace{\Big(\sum_{\text{valid } \mathcal{R}_m} \Big(\frac{\omega}{n}\Big)^{|V(\mathcal{R}_m)|-2k}\Big)}_{\mathcal{Q}_k}\underbrace{\Big(\sum_{\text{valid } \mathcal{R}_r} \Big(\frac{\omega}{n}\Big)^{|V(\mathcal{R}_r)|}\Big)}_{\mathcal{L}_k^T}$$

Major issue: $\mathcal{R}_l, \mathcal{R}_m, \mathcal{R}_r$ were assumed to be **disjoint** except at S_L, S_R, $I \cap J$ which leads to substantial **error terms**

SYMBOLIC FACTORIZATION

Now we can write:

$$\mathcal{M}(I,J) \approx \text{ sum over k of}$$

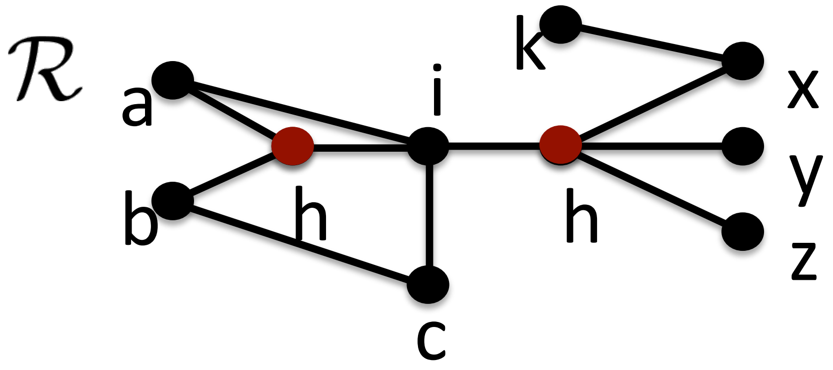
$$\underbrace{\Big(\sum_{\text{valid } \mathcal{R}_l} \left(\frac{\omega}{n}\right)^{|V(\mathcal{R}_l)|}\Big)}_{\mathcal{L}_k}\underbrace{\Big(\sum_{\text{valid } \mathcal{R}_m} \left(\frac{\omega}{n}\right)^{|V(\mathcal{R}_m)|-2k}\Big)}_{\mathcal{Q}_k}\underbrace{\Big(\sum_{\text{valid } \mathcal{R}_r} \left(\frac{\omega}{n}\right)^{|V(\mathcal{R}_r)|}\Big)}_{\mathcal{L}_k^T}$$

Major issue: $\mathcal{R}_l, \mathcal{R}_m, \mathcal{R}_r$ were assumed to be **disjoint** except at S_L, S_R, $I \cap J$ which leads to substantial **error terms**

Idea: Keep iterating the decomposition, carefully charging

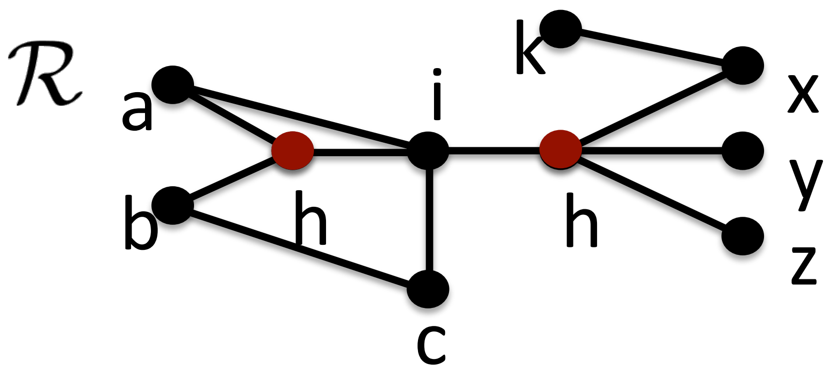
ITERATING THE DECOMPOSITION

Suppose h = j

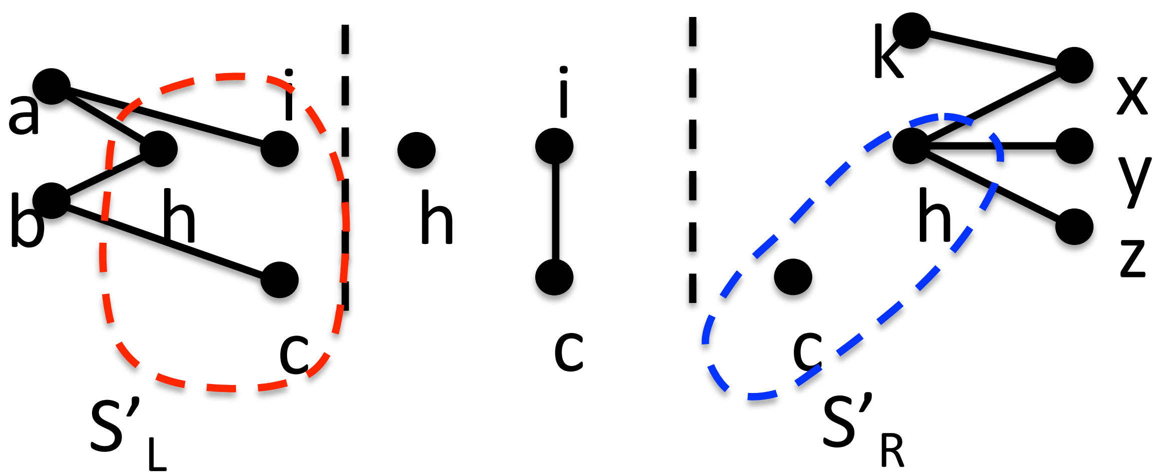


ITERATING THE DECOMPOSITION

Suppose h = j



Look for new leftmost, rightmost separators that separate I from J and intersection vertices



THE MAIN CHARGING ARGUMENT

Complications:

(1) Vertices can become isolated

(2) Separators not necessarily equal size

(3) Need to sum over all pre-images of ribbons, their contributions

Main Tradeoff Lemma: There is a way to tradeoff all these parameters, to charge error terms

Summary:

- Nearly optimal lower bounds against SoS, for the planted clique problem
- **Pseudo-calibration** as a recipe for constructing good pseudo-moments
- When the recipe fails, are there algorithms?
- Connections between **SoS-evidence** and **BP-evidence**?

Summary:

- Nearly optimal lower bounds against SoS, for the planted clique problem
- **Pseudo-calibration** as a recipe for constructing good pseudo-moments
- When the recipe fails, are there algorithms?
- Connections between **SoS-evidence** and **BP-evidence**?

Thanks! Any Questions?