A Ihara-Bass Formula for Non-Boolean Matrices and Strong Refutations of Random CSPs

Tommaso d'Orsi

ETH Zurich

Luca Trevisan

Bocconi

SAT solvers and average-case complexity

- SAT solvers work well on very large-scale instances coming from program verification, VLSI, etc
- For most applications, it is important to be able to *certify unsatisfiability of unsatisfiable formulas*

- Average-case complexity does not provide an explanation for feasibility of solving SAT in practice

Refuting random k-SAT

- Pick a random k-SAT formula with n variables, m clauses
- If $m > c_k n$, formula is unsatisfiable whp
- Seems hard to find proof of unsatisfiability when m is, say, O(n log n)
- Feige proposed it as a complexity assumption
- Problem becomes easier for larger m. When is it poly-time?

Refuting random k-SAT

- Easy to see: if $m > c_k n^{k-1}$ there is, whp, an efficiently constructable refutation by *tree-like resolution*
  - More work: same if $m > c_k n^{k-1} / \log n$

- By *spectral methods:* whp efficiently constructable refutation if $m > c_k n^{\lceil k/2 \rceil + o(1)}$ [Goerdt, Krivelevich 2001]

- By more sophisticated spectral methods: whp *strong refutation* if

  $m > \frac{1}{\epsilon^2} c_k n^{\frac{k}{2}}$ if k is even

  $m > \frac{1}{\epsilon^2} c_k n^{\frac{k}{2}} \text{ polylog } n$ if k is odd

  [Friedman, Goerdt 2001] . . . [Allen, O'Donnell, Witmer 2015]

Our result

- Efficiently computable strong refutation if

  $m > \frac{1}{\epsilon^2} c_k n^{k/2}$ if k is even

  $m > \frac{1}{\epsilon^2} c_k n^{k/2} \operatorname{polylog} n$ if k is odd

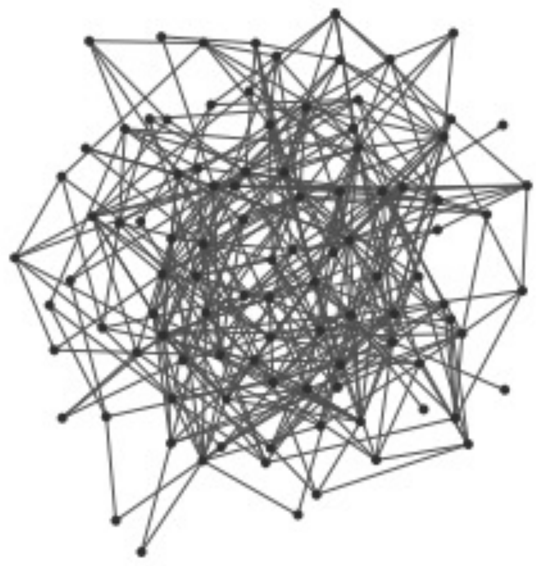
  [Friedman, Goerdt 2001] . . . [Allen, O'Donnell, Witmer 2015]

- Our result:

$m > \frac{1}{\epsilon^2} c_k n^{k/2}$ even if k odd

"Refuting" the existence of a large max cut

- Sample $G \sim \mathcal{G}_{n,\frac{d}{n}}$



- Whp, max cut $\leq \frac{1}{2} + \frac{c}{\sqrt{d}}$

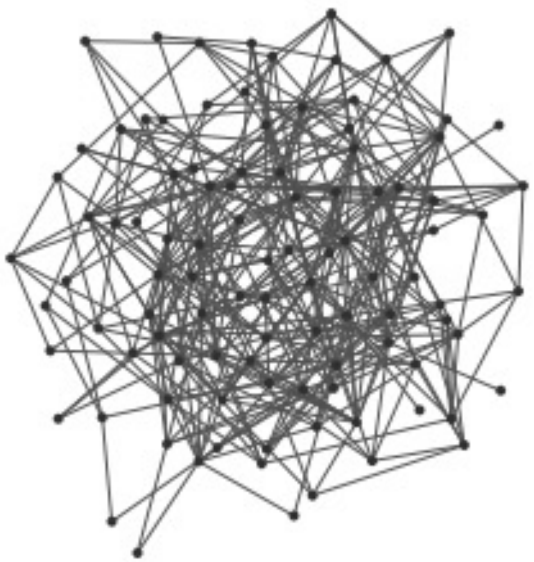
  Proof: Chernoff bounds + union bound

- Whp, there is efficiently computable proof that max cut $\leq \frac{1}{2} + \frac{c\prime}{\sqrt{d}}$

  Proof: [Feige, Ofek 2005] or Grothendieck's inequality + Chernoff bounds

“Refuting” the existence of a large max cut

- ~~Sample $G \sim \mathcal{G}_{n,\frac{d}{n}}$~~
- Sample G so that each edge has probability $\frac{d}{n}$ and edges are polylogn-wise independent



- Can we certify whp that max cut $\leq \frac{1}{2} + \frac{c}{\sqrt{d}}$?
- Is it even true whp?

(If distribution has entropy $o(n)$, we cannot take union bounds!)

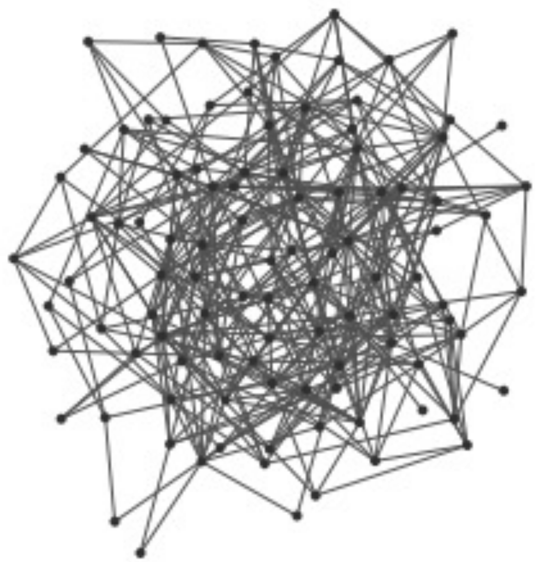
“Refuting” the existence of a large max cut

- Sample G so that each edge has probability $\frac{d(n)}{n}$ and edges are polylogn-wise independent

- By trace methods, whp non-trivial eigenvalues of adjacency matrix $\leq \sqrt{d(n)\log n}$ in magnitude

- Trace calculation needs only $O(\log n)$-wise independence of edges

- Whp, max cut is certifiably $\frac{1}{2} + c\frac{\sqrt{\log n}}{\sqrt{d(n)}}$



Strong refutations of k-SAT

- Refuting random 4-SAT formula with n variables, m clauses reduces to a problem similar to
  - Find a certificate that a given random graph with n^2 vertices and m independent random edges has a max cut $\leq \frac{1}{2} + \epsilon$

- Refuting random 3-SAT formula with n variables, m clauses reduces to a problem similar to
  - Find a certificate that a given random graph with n^2 vertices and $\frac{m^2}{n}$ random-but-correlated edges has a max cut $\leq \frac{1}{2} + \epsilon$

Strong refutations of k-SAT, k even

- Refuting random 4-SAT formula with n variables, m clauses reduces to a problem similar to
  - Find a certificate that a given random graph with n^2 vertices and m independent random edges has a max cut $\leq \frac{1}{2} + \epsilon$

- Refuting random k-SAT (k even) formula with n variables, m clauses reduces to a problem similar to
  - Find a certificate that a given random graph with $n^{k/2}$ vertices and m independent random edges has a max cut $\leq \frac{1}{2} + \epsilon$
  - Can do if $m > \frac{c}{\varepsilon^2} n^{k/2}$

Strong refutations of k-SAT

- Refuting random 3-SAT formula with n variables, m clauses reduces to a problem similar to
  - Find a certificate that a given random graph with n^2 vertices and $\frac{m^2}{n}$ random-but-correlated edges has a max cut $\leq \frac{1}{2} + \epsilon$

- Refuting random k-SAT formula (k odd) with n variables, m clauses reduces to a problem similar to
  - Find a certificate that a given random graph with $n^{(k+1)/2}$ vertices and $\frac{m^2}{n^{(k-1)/2}}$ random-but-correlated edges has a max cut $\leq \frac{1}{2} + \epsilon$

Strong refutations of k-SAT

- Refuting random 3-SAT formula with n variables, m clauses reduces to a problem similar to
  - Find a certificate that a given random graph with n^2 vertices and $\frac{m^2}{n}$ random-but-correlated edges has a max cut $\leq \frac{1}{2} + \epsilon$

- Refuting random k-SAT formula (k odd) with n variables, m clauses reduces to a problem similar to
  - Find a certificate that a given random graph with $n^{(k+1)/2}$ vertices and $\frac{m^2}{n^{(k-1)/2}}$ random-but-correlated edges has a max cut $\leq \frac{1}{2} + \epsilon$
  - Can do if $m > \frac{c}{\varepsilon^2} n^{k/2} \operatorname{polylog} n$

Strong refutations of random 4-SAT

- Enough to provide strong refutation of random 4-XOR [Feige 2002] +...
- To find strong refutation of random 4-XOR problem, we can apply trivial (and seemingly not useful) reduction to 2-XOR:

Max # satisfiable constraints in

$$x_1x_3x_5x_7 = 1$$
$$x_2x_3x_6x_7 = -1$$
$$x_1x_4x_5x_7 = 1$$
$$\dots$$

$$x_1, \dots, x_n \in \{1, -1\}^n$$

$\leq$

Max # satisfiable constraints in

$$y_{1,3}y_{5,7} = 1$$
$$y_{2,3}y_{6,7} = -1$$
$$y_{1,4}y_{5,7} = 1$$
$$\dots$$

$$y_{1,1}, \dots, y_{n,n} \in \{1, -1\}^{n^2}$$

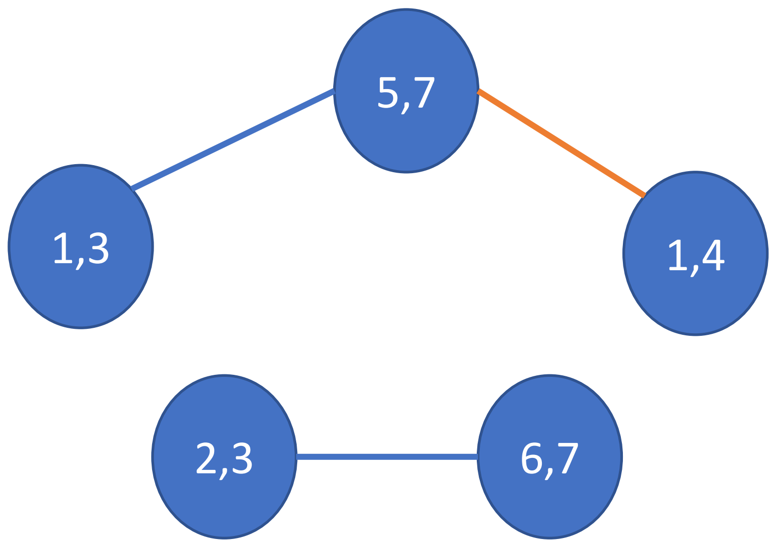
Reduction to random 2-XOR

- Strong refutation of random 4-XOR with n variables, m constraints reduces to proving that the optimum is small in
- A random 2-XOR problem with n^2 clauses, m constraints
- Equivalently, a random correlation clustering problem in a graph with n^2 vertices, m random edges

Max # satisfiable constraints in

$$y_{1,3}y_{5,7} = 1$$
$$y_{2,3}y_{6,7} = -1$$
$$y_{1,4}y_{5,7} = 1$$
$$\dots$$

$$y_{1,1}, \dots, y_{n,n} \in \{1,-1\}^{n^2}$$



Reduction to random 2-XOR

- Strong refutation of random 4-XOR with n variables, m constraints reduces to proving that

$$\max_{y_{1,1},\dots,y_{n,n} \in \{-1,1\}^{n^2}} y^T M y \leq \varepsilon m$$

where

$$M_{i,j,h,k} = \begin{cases} 1 \text{ if } x_i x_j x_h x_k = 1 \text{ is a constraint} \\ -1 \text{ if } x_i x_j x_h x_k = -1 \text{ is a constraint} \\ 0 \text{ otherwise} \end{cases}$$

Reduction to random 2-XOR

- Want to prove that

$$\max_{y_{1,1},\ldots,y_{n,n} \in \{-1,1\}^{n^2}} y^T M y \leq \varepsilon m$$

Proof:

$$\max_{y_{1,1},\ldots,y_{n,n} \in \{-1,1\}^{n^2}} y^T M y$$

$$\leq \max_{\substack{y_{1,1},\ldots,y_{n,n} \in \{-1,1\}^{n^2} \\ z_{1,1},\ldots,z_{n,n} \in \{-1,1\}^{n^2}}} y^T M z$$

$$= ||M||_{\infty \to 1}$$

$$\leq \sqrt{mn^2} \text{ whp}$$

Strong refutations of random 4-SAT

- Enough to provide strong refutation of random 4-XOR [Feige 2002] +...
- Can write random 4-XOR formula with n variables and m constraints as
- $\max_{x_1 \dots x_n \in \{-1,1\}^n} \frac{m}{2} + \frac{1}{2} \sum_{i,j,k,h} b_{i,j,k,h} x_i x_j x_k x_h$
  - Where m of the $b_{i,j,k,h}$ are non-zero, and each is equally likely to be ± 1

How to deal with random 3-XOR

- Strong refutation of random 3-XOR with n variables, m constraints means proving that

$$\max_{x_1,\ldots,x_n \in \{-1,1\}^n} \sum T_{i,j,k} x_i x_j x_k \leq \varepsilon m$$

where

$$T_{i,j,k} = \begin{cases} 1 \text{ if } x_i x_j x_k = 1 \text{ is a constraint} \\ -1 \text{ if } x_i x_j x_k = -1 \text{ is a constraint} \\ 0 \text{ otherwise} \end{cases}$$

How to deal with random 3-XOR

$$\max_{x_1,\dots,x_n \in \{-1,1\}^n} \sum T_{i,j,k} x_i x_j x_k$$

$$\leq \max_{x_1,\dots,x_n \in \{-1,1\}^n} \sqrt{\sum_i x_i^2} \sqrt{\sum_i \Big(\sum_{j,k} T_{i,j,k} x_j x_k \Big)^2}$$

$$= \sqrt{n} \cdot \max_{x_1,\dots,x_n \in \{-1,1\}^n} \sqrt{\sum_{i,a,b,c,d} T_{i,a,b} T_{i,c,d} x_a x_b x_c x_d}$$

How to deal with random 3-XOR

Enough to prove

$$\max_{x_1,\ldots,x_n \in \{-1,1\}^n} \sum_{i,a,b,c,d} T_{i,a,b} T_{i,c,d} x_a x_b x_c x_d \leq \frac{\varepsilon^2 m}{n}$$

How to deal with random 3-XOR

$$\max_{x_1,\ldots,x_n \in \{-1,1\}^n} \sum_{i,a,b,c,d} T_{i,a,b} T_{i,c,d} x_a x_b x_c x_d$$

$$= \max_{x_1,\ldots,x_n \in \{-1,1\}^n} \sum_{a,b,c,d} x_a x_c \left(\sum_i T_{i,a,b} T_{i,c,d} \right) x_b x_d$$

$$\leq \max_{y_{1,1},\ldots,y_{n,n} \in \{-1,1\}^{n^2}} y^T M y$$

where $M_{a,c,b,d} = \sum_i T_{i,a,b} T_{i,c,d}$

How to deal with random 3-XOR

$$\max_{y_{1,1},\ldots,y_{n,n}\in\{-1,1\}^{n^2}} y^T M y$$

where $M_{a,c,b,d} = \sum_i T_{i,a,b} T_{i,c,d}$

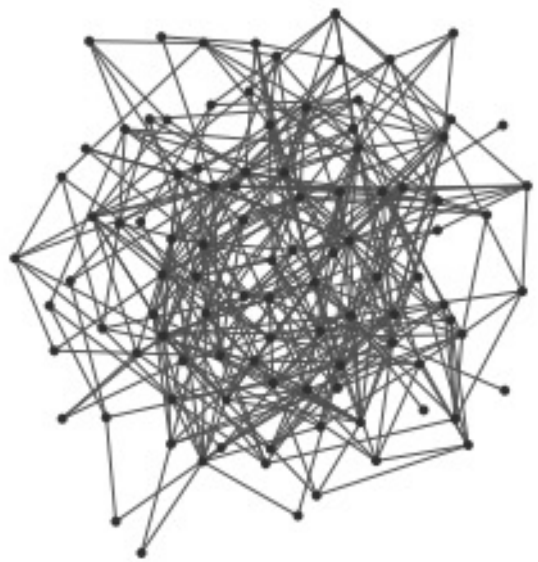
M is an $n^2 \times n^2$ matrix where we expect to see $\approx \frac{m^2}{n}$ non-zero entries

With trace methods, possible to prove spectral bounds sufficient for our goal when m is $n^{1.5} \text{poly} \log n$

[Allen, O'Donnell, Witmer 2015]

“Refuting” the existence of a large max cut

- ~~Sample $G \sim \mathcal{G}_{n,\frac{d}{n}}$~~
- Sample G so that each edge has probability $\frac{d}{n}$ and edges are polylogn-wise independent



- Can we certify whp that max cut $\leq \frac{1}{2} + \frac{c}{\sqrt{d}}$?
- Is it even true whp?

Yes, implicit in [Bordenave, Lelarge, Massoulié 2015] + [Fan, Montanari 2017]

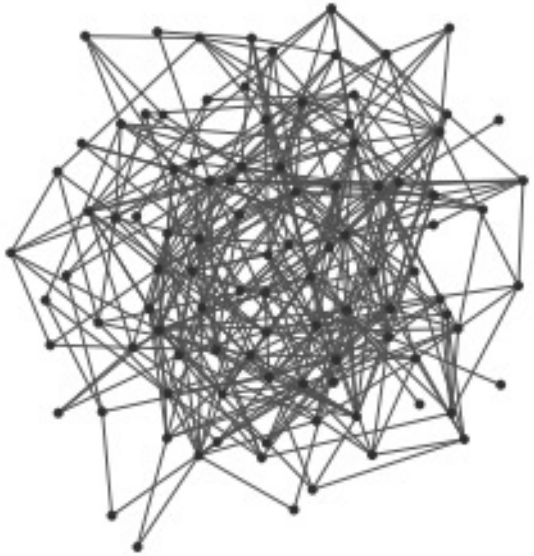
Non-backtracking operator

- Given undirected graph $G = (V, E)$

- Non-backtracking operator B is a $2|E| \times 2|E|$ Boolean 0/1 matrix such that

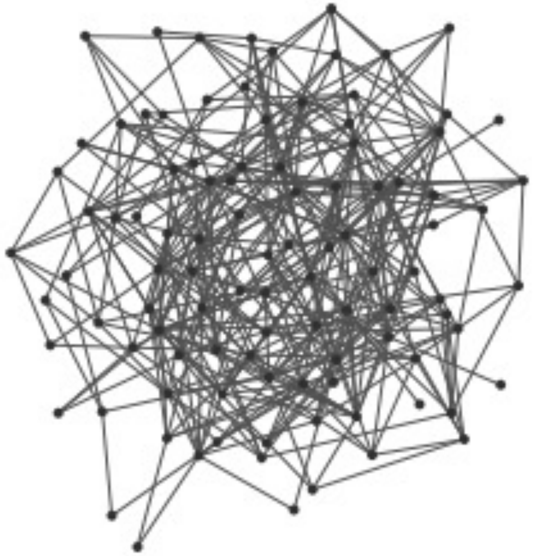
$B_{(u,v),(v,z)} = 1$ iff

$$(u, v) \in E,$$
$$(v, z) \in E,$$
$$u \neq z$$



Non-backtracking operator

- Sample $G \sim \mathcal{G}_{n,\frac{d}{n}}$



- Whp:
  - Largest real e-value of B is $\left(1 + o(1)\right) \cdot d$
  - All others are $\leq (1 + o(1)) \cdot \sqrt{d}$ in magnitude

[Bordenave, Lelarge, Massoulié 2015]

Ihara-Bass formula

- If $G = (V, E)$ is an undirected graph
- A is the adjacency matrix
- D is the diagonal matrix such that $D_{v,v} = \text{degree}(v)$
- B is the non-backtracking operator

Then

$$\det(I - xB) = (1 - x^2)^{|E|-|V|} \det(I - xA + x^2(D - I))$$

Fan-Montanari

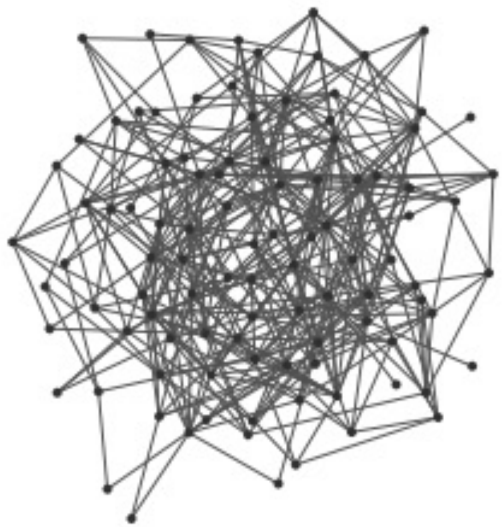
- If $G = (V, E)$ is an undirected graph
- A is the adjacency matrix of G
- B is the non-backtracking operator of G
- $\lambda_{\min}$ is the smallest (most negative) real eigenvalue of B

Then

$$A \succcurlyeq -|\lambda_{\min}| \cdot I - \frac{1}{|\lambda_{\min}|} \cdot (D - I)$$

“Refuting” the existence of a large max cut

- Sample $G \sim \mathcal{G}_{n,\frac{d}{n}}$
- By combining [Bordenave, Lelarge, Massoulié 2015] + [Fan, Montanari 2017]:



$$A \succcurlyeq -(1+o(1)\sqrt{d} \cdot I + (1+o(1) \cdot D/\sqrt{d}$$

- Enough to imply:

$$\text{max cut} \leq \frac{1}{2} + \frac{1+o(1)}{\sqrt{d}}$$

Goemans-Williamson relaxation can certify it

- [FM17] works for all graphs, [BLM15] works in random graphs with polylogn-wise independent edges and constant d

Our technical contributions

- Give a definition of non-backtracking operator B associated to an arbitrary symmetric matrix A (with arbitrary positive and negative entries)
- Prove a Ihara-Bass formula
- Prove a Fan-Montanari type result
- Prove a Bordenave-Leland-Massoulié type result for the matrices coming from the 3-XOR reduction

Our Ihara-Bass type formula

- We give a definition of a non-backtracking operator B associated to an arbitrary symmetric $n \times n$ matrix A with m non-zero entries (which can be arbitrary positive and negative numbers) such that

$$\det(I - xB + xL - xJ) = (1 - x^2)^{\frac{m}{2} - n} \cdot \det(I \ - xA + x^2(D - I))$$

- Where D is the analog of the matrix of degrees and L, J are matrices associated to A that are equal if A is Boolean

- A Fan-Montanari type result can be proved from the above formula

Our Bordenave-Leland-Massoulié type bound

- Take a random 3-XOR formula with n variables and m constraints
- Reduce bounding the max 3-XOR problem to a quadratic optimization problem defined by a $n^2 \times n^2$ matrix A with $\frac{m^2}{n}$ non-zero entries

- The non-backtracking operator B of A satisfies whp

$$||B - L + J|| \leq O\left(\frac{m}{n^{1.5}}\right)$$

- There is a certificate that in the 3-XOR, at most

$$\frac{m}{2} + c\sqrt{n^{1.5} \cdot m}$$

constraints can be simultaneously satisfied

Conclusions

- We give an algorithm that, whp, finds strong refutations of random 3XOR and random 3SAT problems where the number of constraints/clauses is order of $n^{1.5}$
- Breaks long-standing barrier
- Shows that one can analyze random matrices that have an expected constant number of non-zero entries per row, and such that the entries are non-independent
- Generalize the theory of non-backtracking operators to arbitrary matrices (graphs with arbitrary positive and negative weights) in a way that recovers both spectral bounds and algorithmic applications of the boolean/unweighted case