## Overlap Gap Property: an Algorithmic Barrier to Optimizing Over Random Structures

David (Dato) Gamarnik

MIT (2005-Present)

Tbilisi State University (1986-1990)

September 1, 2022

#### XII INTERNATIONAL CONFERENCE OF THE GEORGIAN MATHEMATICAL UNION

 Survey paper with the same title in Proceedings of National Academy of Science (PNAS), 2021

 Erdös-Rényi random graph G(N, p). N nodes. Every edge present with probability p, i.i.d.

- Erdös-Rényi random graph G(N, p). N nodes. Every edge present with probability p, i.i.d.
- The largest clique (fully connected subgraph) is  $\sim 2 \log_{\frac{1}{p}} N$ .

- Erdös-Rényi random graph G(N, p). N nodes. Every edge present with probability p, i.i.d.
- The largest clique (fully connected subgraph) is  $\sim 2 \log_{\frac{1}{2}} N$ .
- A trivial greedy algorithm finds a clique of size  $\sim \log_{\frac{1}{p}} N$  (half optimum).

- Erdös-Rényi random graph G(N, p). N nodes. Every edge present with probability p, i.i.d.
- The largest clique (fully connected subgraph) is  $\sim 2 \log_{\frac{1}{2}} N$ .
- A trivial greedy algorithm finds a clique of size  $\sim \log_{\frac{1}{p}} N$  (half optimum).
- Karp [1976] Improve half-optimality?

- Erdös-Rényi random graph G(N, p). N nodes. Every edge present with probability p, i.i.d.
- The largest clique (fully connected subgraph) is  $\sim 2 \log_{\frac{1}{2}} N$ .
- A trivial greedy algorithm finds a clique of size  $\sim \log_{\frac{1}{p}} N$  (half optimum).
- Karp [1976] Improve half-optimality?
- Still open. This is embarrassing...

• Sparse Erdös-Rényi graph  $\mathbb{G}(N, d/N)$ .

- Sparse Erdös-Rényi graph  $\mathbb{G}(N, d/N)$ .
- The largest independent set is  $\sim 2(1 + o_d(1))\frac{\log d}{d}N$ . Frieze [1990].

- Sparse Erdös-Rényi graph  $\mathbb{G}(N, d/N)$ .
- The largest independent set is  $\sim 2(1 + o_d(1))\frac{\log d}{d}N$ . Frieze [1990].
- A trivial greedy algorithm finds an independent set of size  $\sim (1 + o_d(1)) \frac{\log d}{d} N$ . (half optimum)

- Sparse Erdös-Rényi graph  $\mathbb{G}(N, d/N)$ .
- The largest independent set is  $\sim 2(1 + o_d(1))\frac{\log d}{d}N$ . Frieze [1990].
- A trivial greedy algorithm finds an independent set of size  $\sim (1 + o_d(1)) \frac{\log d}{d} N$ . (half optimum)
- Nothing better known.

### Statistics-to-Computation gap

### Statistics-to-Computation gap

Problems exhibiting a similar *statistics-to-computation* gap:

Random K-SAT Proper coloring of a random graphs MaxCut on random hypergraphs Ground state of a spin glass model Stochastic Block Model LDPC Codes **Planted Clique** Spiked Tensor problem Sparse Regression and Phase Retrieval Sparse Covariance Estimation problem Graph alignment Binary perceptron Mixture of Gaussians etc, etc.

## What is the obstruction to algorithms?

### What is the obstruction to algorithms?

 Change in the geometry of solutions at the onset of hardness, Overlap Gap Property (OGP)

## What is the obstruction to algorithms?

- Change in the geometry of solutions at the onset of hardness, Overlap Gap Property (OGP)
- Originating in the theory of spin glass. Giorgio Parisi (Nobel Prize Physics 2021)



(a) Emerges in most models known to exhibit an apparent algorithmic hardness

- (a) Emerges in most models known to exhibit an apparent algorithmic hardness
- (b) Consistent with the hardness/tractability phase transition for many models analyzed to the day

- (a) Emerges in most models known to exhibit an apparent algorithmic hardness
- (b) Consistent with the hardness/tractability phase transition for many models analyzed to the day
- (c) Allows to mathematically rigorously rule out a large class of algorithms as potential contenders, specifically the algorithms which exhibit the input stability (noise insensitivity), such as **Boolean circuits** (this talk).

#### Theorem (Informal, G, Jagannath & Wein [2022])

If polynomial size Boolean circuit C with depth  $p_n$  finds better than half-optimum ind set in  $\mathbb{G}(n, d/n)$ , then its depth is at least

$$p_n \geq \Omega\left(\frac{\log n}{\log\log n}\right)$$

#### Theorem (Informal, G, Jagannath & Wein [2022])

If polynomial size Boolean circuit C with depth  $p_n$  finds better than half-optimum ind set in  $\mathbb{G}(n, d/n)$ , then its depth is at least

$$p_n \ge \Omega\left(\frac{\log n}{\log\log n}\right)$$

• Half-optimum ind sets can be found by depth O(1) circuits.

#### Theorem (Informal, G, Jagannath & Wein [2022])

If polynomial size Boolean circuit C with depth  $p_n$  finds better than half-optimum ind set in  $\mathbb{G}(n, d/n)$ , then its depth is at least

$$p_n \ge \Omega\left(\frac{\log n}{\log\log n}\right)$$

- Half-optimum ind sets can be found by depth O(1) circuits.
- State of the art o(log n/ log log n), Rossman [2010], Li, Razborov & Rossman [2017], (though for the decision not the search problem).

Theorem (Informal, G, Jagannath & Wein [2022])

If Boolean circuit C with a constant depth O(1) finds better than half-optimum ind set in  $\mathbb{G}(n, d/n)$ , then its size is at least

 $\exp\left(n^{\Theta(1)}\right)$ 

#### Theorem (Informal, G, Jagannath & Wein [2022])

If Boolean circuit C with a constant depth O(1) finds better than half-optimum ind set in  $\mathbb{G}(n, d/n)$ , then its size is at least

 $\exp\left(n^{\Theta(1)}\right)$ 

• State of the art lower bound is  $\exp(\log^{\Theta(1)} n)$ , Rossman [2010].

#### Theorem (G & Sudan [2014])

Fix  $\frac{1}{2} + \frac{1}{2\sqrt{2}} < \alpha < 1$ . There exists  $0 \le \nu_1 < \nu_2 < 1$  such that with prob  $1 - \exp(-\Omega(n))$  for every two  $\alpha$ -optimum independent sets I, J in  $\mathbb{G}(n, d/n)$ 

$$\frac{|I \cap J|}{\mathcal{OPT}} \in [0, \nu_1] \cup [\nu_2, 1].$$

#### Theorem (G & Sudan [2014])

Fix  $\frac{1}{2} + \frac{1}{2\sqrt{2}} < \alpha < 1$ . There exists  $0 \le \nu_1 < \nu_2 < 1$  such that with prob  $1 - \exp(-\Omega(n))$  for every two  $\alpha$ -optimum independent sets I, J in  $\mathbb{G}(n, d/n)$ 

$$\frac{|I \cap J|}{\mathcal{OPT}} \in [0, \nu_1] \cup [\nu_2, 1].$$

Intersection of every two large enough ind sets is either "small" or "large" (gap in overlaps)

#### Theorem (G & Sudan [2014])

Fix  $\frac{1}{2} + \frac{1}{2\sqrt{2}} < \alpha < 1$ . There exists  $0 \le \nu_1 < \nu_2 < 1$  such that with prob  $1 - \exp(-\Omega(n))$  for every two  $\alpha$ -optimum independent sets I, J in  $\mathbb{G}(n, d/n)$ 

$$\frac{|I \cap J|}{\mathcal{OPT}} \in [0, \nu_1] \cup [\nu_2, 1].$$

Intersection of every two large enough ind sets is either "small" or "large" (gap in overlaps) This was used to rule out local (Factor of IID) algorithms.



## Landscape of the OGP





## Landscape of the OGP





## Landscape of the OGP



### Comparison with clustering



Note: OGP is stronger than clustering!
### References on OGP based results

### References on OGP based results

- G, Jagannath & Wein [2022] Boolean circuits (this talk)
- G, Jagannath & Wein [2020] Low degree polynomials, Langevin dynamics
- ◊ G & Jagannath [2020] AMP algorithms
- Coja-Oghlan, Haqshenas & Hetterich [2017] Random Walk (WAKLSAT)
- G & Sudan [2017] Survey Propagation algorithms
- Farhi, G & Gutmann [2017] Quantum (QAOA) algorithms
- G, Kizildag, Perkins & Xu [In progress] Kim-Roche algorithm for Binary perceptron
- A Rahman & Virag [2017], Wein [2020]
- Bresler & Huang [2021], Huang & Sellke [2021]

• Generate independent copies  $\mathbb{G}, \tilde{\mathbb{G}} \stackrel{d}{=} \mathbb{G}(n, d/n).$ 

- Generate independent copies  $\mathbb{G}, \tilde{\mathbb{G}} \stackrel{d}{=} \mathbb{G}(n, d/n)$ .
- Generate an arbitrary order on <sup>n</sup><sub>2</sub> and interpolate
  G<sub>0</sub> = G, G<sub>1</sub>, G<sub>2</sub>, ..., G<sub><sup>n</sup><sub>2</sub></sub> = G̃.

- Generate independent copies  $\mathbb{G}, \tilde{\mathbb{G}} \stackrel{d}{=} \mathbb{G}(n, d/n).$
- Generate an arbitrary order on  $\binom{n}{2}$  and interpolate  $\mathbb{G}_0 = \mathbb{G}, \mathbb{G}_1, \mathbb{G}_2, \dots, \mathbb{G}_{\binom{n}{2}} = \tilde{\mathbb{G}}.$

#### Theorem

Fix  $\frac{1}{2} + \frac{1}{2\sqrt{2}} < \alpha < 1$ . For all large enough *d*, there exists  $0 \le \nu_1 < 1/2 < \nu_2 < 1$  such that with prob  $1 - \exp(-\Omega(n))$  for every  $0 \le t \le {n \choose 2}$  and every  $\alpha$ -optimum independent sets  $I_0$  in  $\mathbb{G}_0$  and  $I_t$  in  $\mathbb{G}_t$ 

$$\frac{|I_0 \cap I_t|}{\mathcal{OPT}} \in [0, \nu_1] \cup [\nu_2, 1].$$

Furthermore, when  $t = \binom{n}{2}$  only  $\in [0, \nu_1]$  is possible.

- Generate independent copies  $\mathbb{G}, \tilde{\mathbb{G}} \stackrel{d}{=} \mathbb{G}(n, d/n).$
- Generate an arbitrary order on  $\binom{n}{2}$  and interpolate  $\mathbb{G}_0 = \mathbb{G}, \mathbb{G}_1, \mathbb{G}_2, \dots, \mathbb{G}_{\binom{n}{2}} = \tilde{\mathbb{G}}.$

#### Theorem

Fix  $\frac{1}{2} + \frac{1}{2\sqrt{2}} < \alpha < 1$ . For all large enough *d*, there exists  $0 \le \nu_1 < 1/2 < \nu_2 < 1$  such that with prob  $1 - \exp(-\Omega(n))$  for every  $0 \le t \le {n \choose 2}$  and every  $\alpha$ -optimum independent sets  $I_0$  in  $\mathbb{G}_0$  and  $I_t$  in  $\mathbb{G}_t$ 

$$\frac{|I_0 \cap I_t|}{\mathcal{OPT}} \in [0, \nu_1] \cup [\nu_2, 1].$$

Furthermore, when  $t = \binom{n}{2}$  only  $\in [0, \nu_1]$  is possible.









• Generate an arbitrary total order on  $\binom{n}{2}$  pairs  $e(1), \ldots, e(n(n-1)/2)$ .

- Generate an arbitrary total order on  $\binom{n}{2}$  pairs  $e(1), \ldots, e(n(n-1)/2)$ .
- $\mathbb{G}_t \stackrel{d}{=} \mathbb{G}(n, d/n), t \leq Tn^{O(1)}$ , where  $G_{t+1}$  is obtained from  $\mathbb{G}_t$  by resampling edge  $e(1 + t \mod \binom{n}{2})$ .

- Generate an arbitrary total order on  $\binom{n}{2}$  pairs  $e(1), \ldots, e(n(n-1)/2)$ .
- $\mathbb{G}_t \stackrel{d}{=} \mathbb{G}(n, d/n), t \leq Tn^{O(1)}$ , where  $G_{t+1}$  is obtained from  $\mathbb{G}_t$  by resampling edge  $e(1 + t \mod \binom{n}{2})$ .

#### Theorem (Wein 2020)

For every  $\epsilon > 0$ ,  $K \ge 1 + 5/\epsilon^2$  and d large enough the following holds with probability at least  $1 - \exp(-\Omega(n))$ : there does not exist a sequence of times  $t_1, \ldots, t_K$  with  $0 \le t_k \le T$  and  $1/2 + \epsilon$ -optimal ind sets sets  $l_1, \ldots, l_K \subset [n]$  in  $\mathbb{G}_{t_1}, \ldots, \mathbb{G}_{t_K}$  such that

$$|I_k \setminus (\cup_{1 \le \ell < k} I_\ell)| \in \left[\frac{\epsilon}{4} \frac{\log d}{d} n, \frac{\epsilon}{2} \frac{\log d}{d} n\right], \qquad 2 \le k \le K.$$

### e-OGP – obstruction to stable (noise-insensitive) algs

## e-OGP - obstruction to stable (noise-insensitive) algs

#### Theorem (Meta-theorem)

Suppose an algorithm  $\mathcal{A} : G \to \{0,1\}^n$  is stable: a "small" change in input G results in a small change on the output  $\mathcal{A}(G)$ . Then  $\mathcal{A}$  cannot overcome the e-OGP barrier.

## e-OGP - obstruction to stable (noise-insensitive) algs

#### Theorem (Meta-theorem)

Suppose an algorithm  $\mathcal{A} : G \to \{0,1\}^n$  is stable: a "small" change in input G results in a small change on the output  $\mathcal{A}(G)$ . Then  $\mathcal{A}$  cannot overcome the e-OGP barrier.

Proof by picture:



## Boolean circuits. Background

### Boolean circuits. Background

- Boolean circuit *C* a mapping {0, 1}<sup>(n)</sup><sub>2</sub> → {0, 1}<sup>n</sup> encoded by a directed graph with logical gates ¬, ∨, ∧
- Size s(n)- number of gates. Depth p(n) length of the longest path



Poly-size Boolean circuits for computing the *n*-PARITY function has depth ⊖(log n/ log log n). Hastad [1986].
 Hastad's Switching Lemma

- Poly-size Boolean circuits for computing the *n*-PARITY function has depth ⊖(log n/ log log n). Hastad [1986].
  Hastad's Switching Lemma
- A lot of research on circuit depth for *Decision version* (single output node) of optimization problems (CLIQUE). Lynch [1986], Beame [1990], Rossman [2010], Li, Razborov & Rossman [2017]

- Poly-size Boolean circuits for computing the *n*-PARITY function has depth Θ(log *n*/ log log *n*). Hastad [1986].
  Hastad's Switching Lemma
- A lot of research on circuit depth for *Decision version* (single output node) of optimization problems (CLIQUE). Lynch [1986], Beame [1990], Rossman [2010], Li, Razborov & Rossman [2017]
- Poly-size Circuits computing cliques of size k(n) require depth

$$\Omega\left(\log n/(k^2(n)\log\log n)\right) = o\left(\log n/\log\log n\right)$$

Rossman [2010],

- Poly-size Boolean circuits for computing the *n*-PARITY function has depth Θ(log *n*/ log log *n*). Hastad [1986].
  Hastad's Switching Lemma
- A lot of research on circuit depth for *Decision version* (single output node) of optimization problems (CLIQUE). Lynch [1986], Beame [1990], Rossman [2010], Li, Razborov & Rossman [2017]
- Poly-size Circuits computing cliques of size k(n) require depth

$$\Omega\left(\log n/(k^2(n)\log\log n)\right) = o\left(\log n/\log\log n\right)$$

Rossman [2010],

## Main result: circuit lower bound

### Main result: circuit lower bound

Let  $C(s(n), p(n), \alpha)$  be the class of size s(n) Boolean circuits with depth p(n) which produce factor  $\alpha$ -OPT independent set given a graph as an input.

### Main result: circuit lower bound

Let  $C(s(n), p(n), \alpha)$  be the class of size s(n) Boolean circuits with depth p(n) which produce factor  $\alpha$ -OPT independent set given a graph as an input.

Theorem (Circuit depth lower bound. G, Jagannath & Wein [2022])

Let  $\alpha \in (1/2, 1), \epsilon > 0$  and

$$p(n) \leq rac{\log n}{(1+\epsilon)\log\log n}$$

Then for every  $C \in C(s(n), p(n), \alpha)$  and all large enough n

$$s(n) \ge n^{(\log n)^{\frac{1}{3}}}$$

In particular, the size is super-polynomial.

## Stability (noise insensitivity) of circuits: LMN Theorem

## Stability (noise insensitivity) of circuits: LMN Theorem

#### Theorem (Linial-Mansour-Nisan' (LMN) Theorem, [1993])

For every circuit C with size s(N) and depth p(N) under the *i.i.d.* uniform distribution on  $\{0,1\}^N$ , the sum of Fourier coefficients associated with monomials of order

$$D_N \triangleq (\log s(N))^{O(p(N))}$$
 (That is  $(\log \text{Size})^{O(\text{Depth})}$ 

*is o*(1).

## Stability (noise insensitivity) of circuits: LMN Theorem

#### Theorem (Linial-Mansour-Nisan' (LMN) Theorem, [1993])

For every circuit *C* with size s(N) and depth p(N) under the *i.i.d.* uniform distribution on  $\{0,1\}^N$ , the sum of Fourier coefficients associated with monomials of order

$$D_N \triangleq (\log s(N))^{O(p(N))}$$
 (That is  $(\log \text{Size})^{O(\text{Depth})}$ )

*is o*(1).

From O'Donnell "Analysis of Boolean Functions"

28/34



# Stability (noise insensitivity) of circuits: Linial-Mansour-Nissan's Theorem

Informally, on  $\{0, 1\}^N$ , the circuit *C* can be approximated by an *N* variable polynomial of degree  $(\log \text{Size})^{O(\text{Depth})}$ .

# Stability (noise insensitivity) of circuits: Linial-Mansour-Nissan's Theorem

Informally, on  $\{0, 1\}^N$ , the circuit *C* can be approximated by an *N* variable polynomial of degree  $(\log \text{Size})^{O(\text{Depth})}$ .

Say, size  $s(N) = N^{\alpha}$ , depth  $p(N) = \beta \log N / (\log \log N)$ . Then

$$(\log \text{Size})^{\mathcal{O}(\text{Depth})} \approx (\log N)^{\beta \log N/(\log \log N)} = \exp(\beta \log N) = N^{\beta}.$$

When  $\beta < 1$  is small the "relevant" degree is sublinear in *N*.

# Stability (noise insensitivity) of circuits: Linial-Mansour-Nissan's Theorem
• Recall an interpolated sequence  $\mathbb{G}_0, \mathbb{G}_1, \dots, \mathbb{G}_{\frac{n(n-1)}{2}}$  of  $\mathbb{G}(n, d/n)$  graphs.

- Recall an interpolated sequence  $\mathbb{G}_0, \mathbb{G}_1, \dots, \mathbb{G}_{\frac{n(n-1)}{2}}$  of  $\mathbb{G}(n, d/n)$  graphs.
- A circuit *C* produces a sequence of solutions  $I_t = C(G_t), t = 0, 1, \dots, {n \choose 2}.$

- Recall an interpolated sequence  $\mathbb{G}_0, \mathbb{G}_1, \dots, \mathbb{G}_{\frac{n(n-1)}{2}}$  of  $\mathbb{G}(n, d/n)$  graphs.
- A circuit *C* produces a sequence of solutions  $I_t = C(G_t), t = 0, 1, ..., {n \choose 2}.$
- We use LMN and a large deviations estimate to show that

$$n^{-1} \|I_t - I_{t+1}\|_2^2 \leq (\nu_2 - \nu_1)^2$$

for all *t* with probability at least  $exp(-\delta D_n)$  with controlled  $\delta$ .

 By the e-OGP, this can only happen on the "exception" event with probability exp(-Θ(n)).

- By the e-OGP, this can only happen on the "exception" event with probability exp(-Θ(n)).
- Thus  $\exp(-\delta D_n) \leq \exp(-\Theta(n))$ .

 By the e-OGP, this can only happen on the "exception" event with probability exp(-Θ(n)).

• Thus 
$$\exp(-\delta D_n) \leq \exp(-\Theta(n))$$
.

• Since 
$$p(n) = \frac{\log n}{(1+\epsilon) \log \log n}$$
 we obtain

 By the e-OGP, this can only happen on the "exception" event with probability exp(-Θ(n)).

• Thus 
$$\exp(-\delta D_n) \leq \exp(-\Theta(n))$$
.

• Since 
$$p(n) = \frac{\log n}{(1+\epsilon) \log \log n}$$
 we obtain

$$\exp\left(-\delta(\log s(n))^{\frac{\log n}{(1+\epsilon)\log\log n}}\right) \le \exp\left(-\Omega(n)\right)$$
$$\implies$$
$$s(n) \ge n^{(\log n)^{\Omega(1)}} \square$$

### Small set avoidance

#### Small set avoidance

Large deviation lower bound is based on the following lemma

Large deviation lower bound is based on the following lemma

Lemma (Small Set Avoidance Lemma G, Jagannath & Wein [2020])

Let *E* be any subset of edges in  $\{0,1\}^n$ . Let *e* be the fraction of the edges in *E*. Consider a random walk  $Z_t, 0 \le t \le T$  in  $\{0,1\}^n$  with  $Z_0$  chosen u.a.r. The walk never crosses the set *E* with probability at least  $2^{-Te}$ .

### $\text{OGP} \rightarrow \text{spin glass hardness}$

A very similar method shows failure of Boolean circuits in finding ground states of spin glass models https://arxiv.org/pdf/2109.01342.pdf and is likely to be applicable for all models exhibiting OGP.

Challenges with applying this to other combinatorial optimization problems

Challenges with applying this to other combinatorial optimization problems

• Cliques on  $\mathbb{G}(n, 1/2)$ . The e-OGP occurs but with probability  $1 - \exp(-\Theta(\log^2 n))$ . This is too weak for obtaining circuit lower bounds.

Challenges with applying this to other combinatorial optimization problems

- Cliques on G(n, 1/2). The e-OGP occurs but with probability 1 - exp(-Θ(log<sup>2</sup> n)). This is too weak for obtaining circuit lower bounds.
- Decision vs search lower bounds.

Challenges with applying this to other combinatorial optimization problems

- Cliques on G(n, 1/2). The e-OGP occurs but with probability 1 - exp(-Θ(log<sup>2</sup> n)). This is too weak for obtaining circuit lower bounds.
- Decision vs search lower bounds.
- Are there (evidently) hard problems not exhibiting OGP in the way we have defined it?