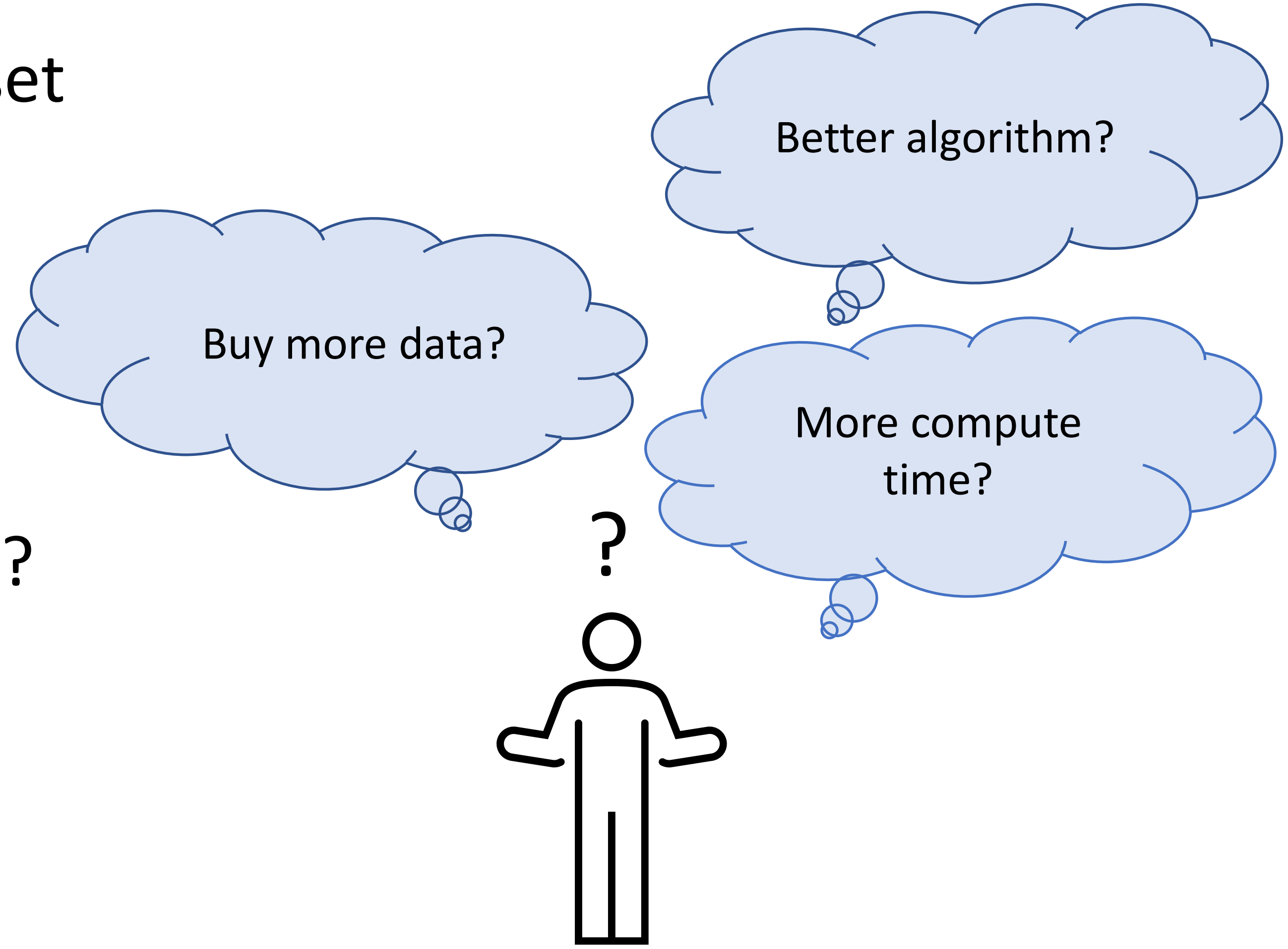
Understanding Statistical-vs-Computational Tradeoffs via Low-Degree Polynomials

Alex Wein

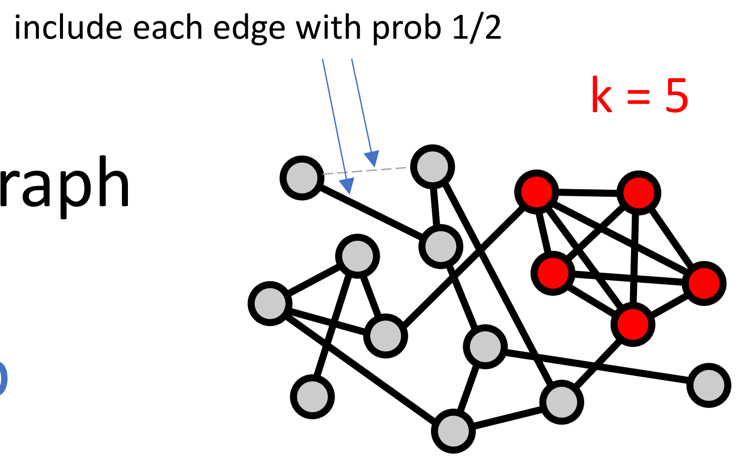
How Much Can We Learn From Data?

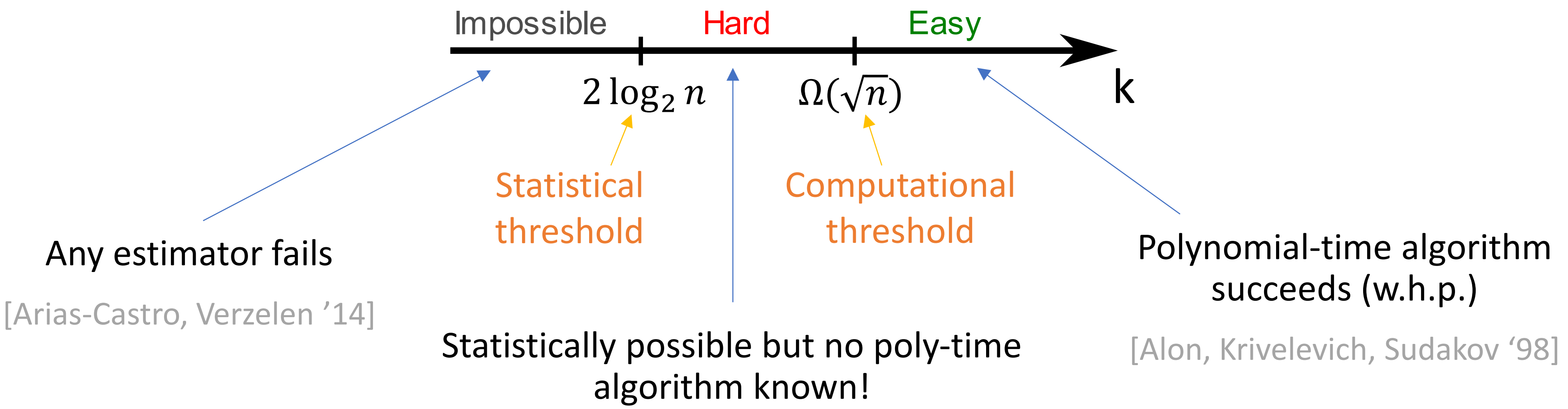
- Hidden “signal” in a large dataset
- Criteria:
  - Statistical efficiency
  - Computational efficiency
- Fundamental limits?
- Which tradeoffs are achievable?



Example: Planted Clique

- Find a planted k-clique in an n-vertex random graph
  - G(n,1/2) + {random k-clique}
- Believed to have a statistical-computational gap





Not Just Planted Clique...

Sparse PCA
Community detection (SBM)
Tensor PCA
Random CSPs
Spiked Wigner model
Spiked Wishart model
Planted submatrix
Planted dense subgraph
Planted vector in a subspace
Dictionary learning

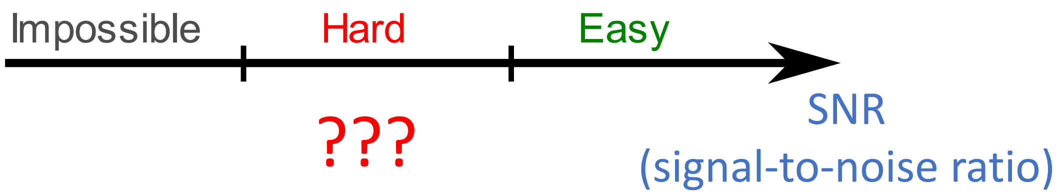
Non-gaussian component analysis
Independent component analysis
Tensor decomposition
Sparse linear regression
Phase retrieval
Group testing
Generalized linear models
Synchronization
Orbit recovery

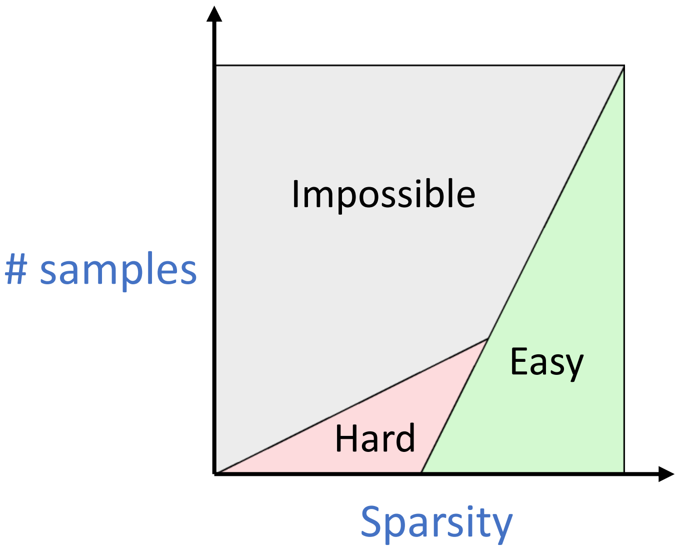
Gaussian clustering
Sparse clustering
Matrix completion
Tensor completion
Graph matching
Planted matching
Mixed membership SBM
Hypergraphic planted clique
Secret leakage planted clique
Continuous learning with errors

Robust sparse mean estimation
Certifying RIP
Spiked transport model
Hidden hubs
Planted coloring
Number partitioning
Nonnegative PCA
Cone-constrained PCA
Sparse tensor PCA
Robust sparse PCA

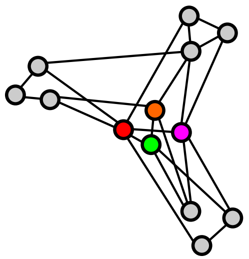
Learning neural networks
Sherrington-Kirkpatrick model
Spin glass optimization
...

Statistical-computational gaps are ubiquitous!

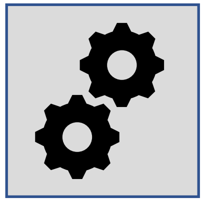




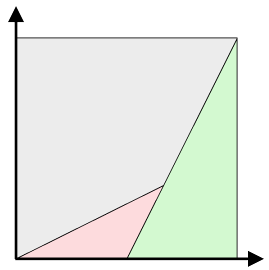
The Dream



New statistical problem



Systematic method
of analysis



Phase diagram

“Impossible” phase: classical statistics (Assouad, Fano, Le Cam, …)

“Easy” phase: algorithm design (spectral methods, message passing, …)

“Hard” phase: need evidence for computational hardness

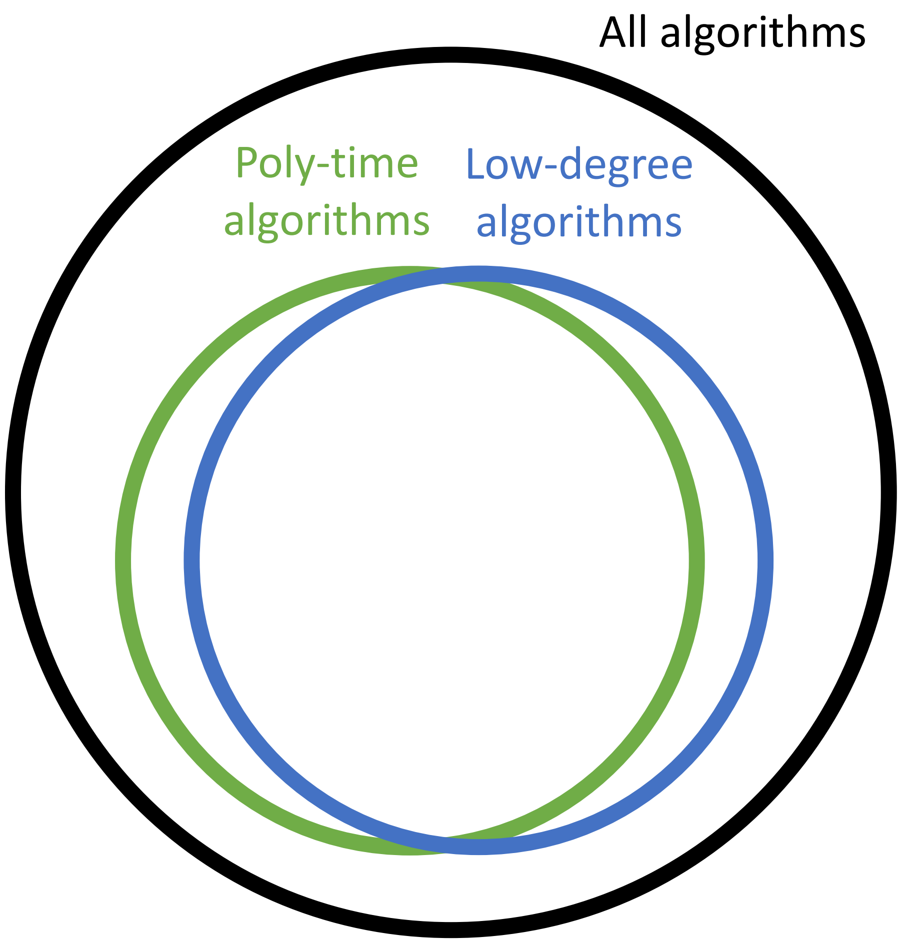
~~NP-hardness~~ “computational complexity of statistical inference”

A Restricted Class of Algorithms

Low-degree polynomial algorithms

- ☑ Good "proxy" for poly-time algorithms, for statistical problems
- ☑ Tractable to analyze
- ☑ Widely-applicable
- ☑ Sheds light on fundamental statistical questions

Other restricted classes: SoS, SQ, AMP, …



Low-Degree Polynomial Algorithms

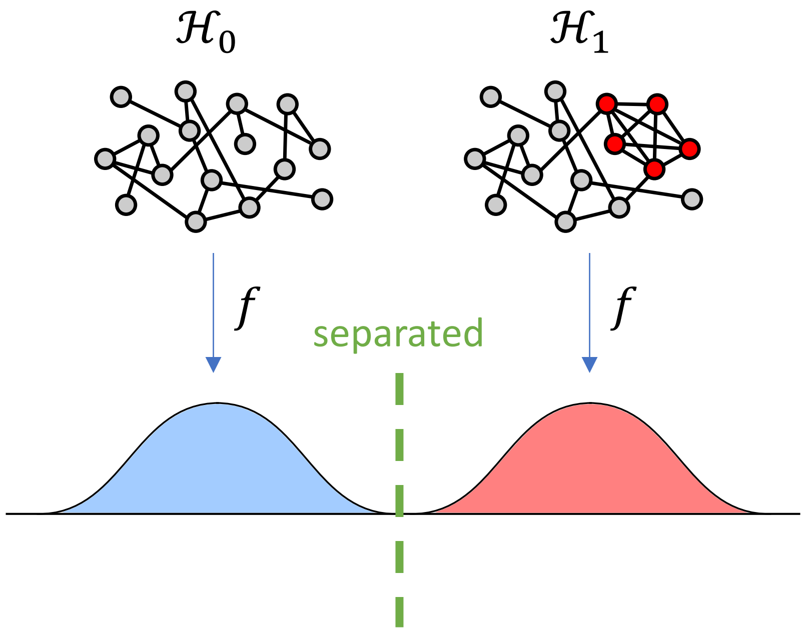
[Hopkins, Steurer '17; Hopkins, Kothari, Potechin, Raghavendra, Schramm, Steurer '17; Hopkins '18; Kunisky, **W**, Bandeira '19]

- **Hypothesis testing**: "is there a planted signal?"
  - Distinguish $\mathcal{H}_0$ (random graph) vs $\mathcal{H}_1$ (planted clique)
  - Goal: vanishing error probability
- **Low-degree polynomial algorithm**: multivariate polynomial of degree O(log n)

$$f : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}$$

Input: graph

Output: number



"Success": f's output is "small" under $\mathcal{H}_0$, "large" under $\mathcal{H}_1$

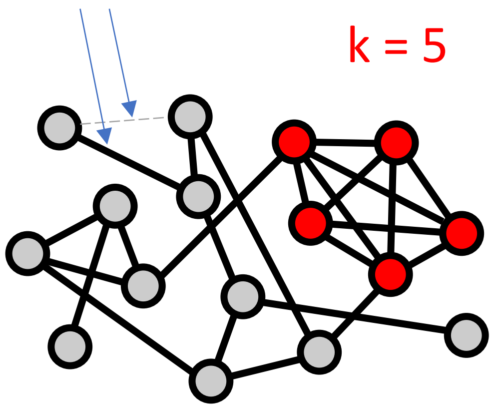
$$(\mathrm{E}_{\mathcal{H}_1}[f] - \mathrm{E}_{\mathcal{H}_0}[f])^2 \gg \max\{\mathrm{Var}_{\mathcal{H}_0}(f), \mathrm{Var}_{\mathcal{H}_1}(f)\}$$

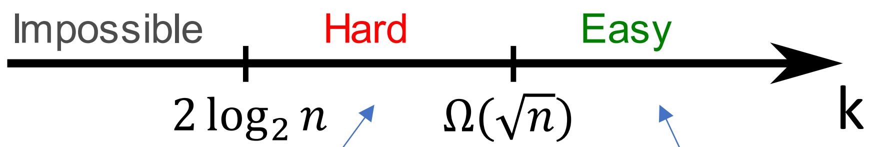
Back to Planted Clique

- ~~Find~~ Detect a planted k-clique in an n-vertex random graph
  - G(n,1/2) + {random k-clique}

include each edge with prob 1/2

k = 5





When $k \ll \sqrt{n}$, **all** low-degree algorithms fail
[Barak, Hopkins, Kelner, Kothari, Moitra, Potechin ’16; Hopkins ‘18]

When $k \gg \sqrt{n}$, some low-degree algorithm succeeds
(degree-1 polynomial: count total edges)

Low-Degree Algorithms Are Optimal??

Stellar track record for capturing the computational threshold

Planted clique [Barak, Hopkins, Kelner, Kothari, Moitra, Potechin '16; Hopkins '18]

Community detection [Hopkins, Steurer '17; Hopkins '18; Bandeira, Banks, Kunisky, Moore, **W** '21]

Mixed membership SBM [Hopkins, Steurer '17]

Tensor PCA [Hopkins, Kothari, Potechin, Raghavendra, Schramm, Steurer '17; Kunisky, **W**, Bandeira '19]

Spiked Wishart [Bandeira, Kunisky, **W** '20]

Spiked Wigner [Kunisky, **W**, Bandeira '19]

Sparse PCA [Ding, Kunisky, **W**, Bandeira '19]

Heavy tailed statistics [Cherapanamjeri, Hopkins, Kathuria, Raghavendra, Tripuraneni '20]

Max independent set [Gamarnik, Jagannath, **W** '20; **W** '20]

Secret leakage planted clique [Brennan, Bresler '20]

Hypergraphic planted clique [Luo, Zhang '20]

Sparse clustering [Löffler, **W**, Bandeira '20]

Certifying RIP [Ding, Kunisky, **W**, Bandeira '21]

Planted submatrix [Schramm, **W** '20]

Planted dense subgraph [Schramm, **W** '20]

Multi-spiked Wigner/Wishart [Bandeira, Banks, Kunisky, Moore, **W** '21]

Planted affine planes [Ghosh, Jeronimo, Jones, Potechin, Rajendran '20]

Gaussian mixture models [Brennan, Bresler, Hopkins, Li, Schramm '21]

Gaussian graphical models [Brennan, Bresler, Hopkins, Li, Schramm '21]

Morris class of exponential families [Kunisky '20]

Robust sparse PCA [d'Orsi, Kothari, Novikov, Steurer '20]

Non-negative PCA [Bandeira, Kunisky, **W** '21]

Planted vector in a subspace [Mao, **W** '21]

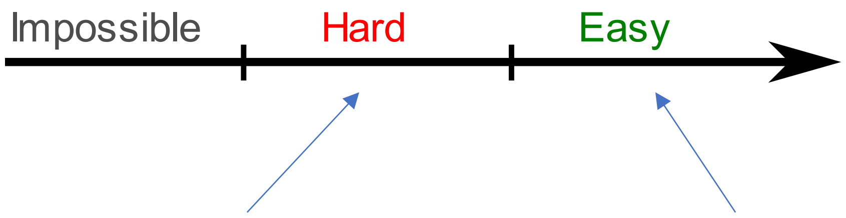
Random k-SAT [Bresler, Huang '21]

Sparse tensor PCA [Choo, d'Orsi '21]

Sparse linear regression [Arpino '21]

Gaussian clustering [Mao, **W** '21, Davis, Diaz, Wang '21]

Graph matching [Mao, Wu, Xu, Yu '21]



No known poly-time alg
All low-deg algs provably fail

Some poly-time alg provably succeeds
Some low-deg alg provably succeeds

"Low-Degree Conjecture" [Hopkins '18]: low-degree algorithms are optimal among all polynomial-time algorithms for *natural high-dimensional statistical problems**.

Outline

- I. Hypothesis testing
    - Planted sparse vector in a subspace
      Mao, **W**, “Optimal Spectral Recovery of a Planted Vector in a Subspace”
      Submitted
- II. Estimation
    - Planted submatrix
      Schramm, **W**, “Computational Barriers to Estimation from Low-Degree Polynomials”
      Annals of Statistics (to appear)
- III. Optimization
    - Max independent set in random graphs
      Gamarnik, Jagannath, **W**, “Low-Degree Hardness of Random Optimization Problems”
      FOCS 2020

      W, “Optimal Low-Degree Hardness of Maximum Independent Set”
      Mathematical Statistics and Learning, 2022

I. Hypothesis Testing

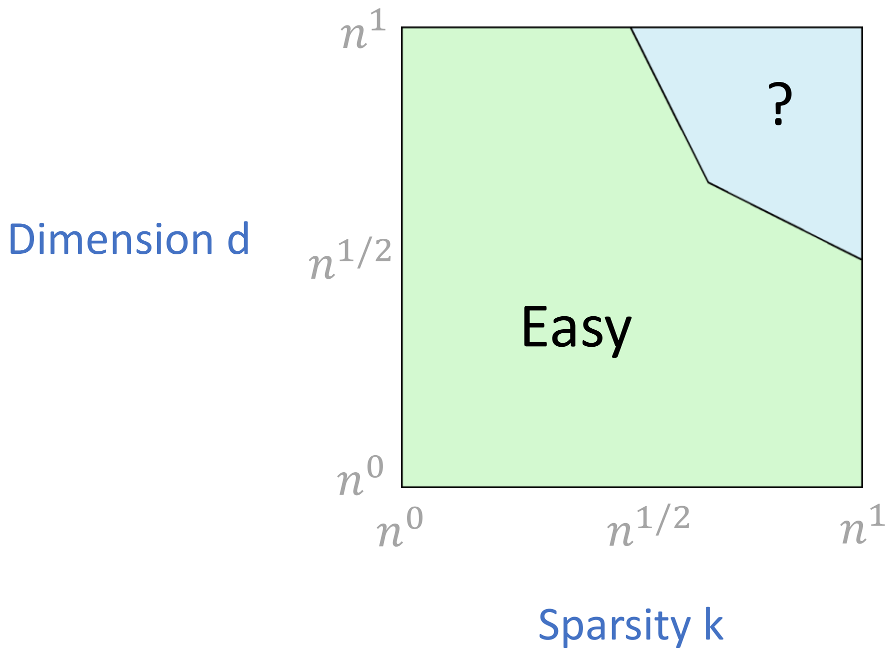
Planted Sparse Vector in a Subspace

(joint work with Cheng Mao)

Input: d-dimensional subspace of $\mathbb{R}^n$, $d \ll n$

- $\mathcal{H}_0$: random subspace
- $\mathcal{H}_1$: subspace containing a k-sparse vector (and otherwise random)

Prior work:
[DH13, BKS14, QSW14, HSSS16]



Applying the Low-Degree Framework

Compute low-degree chi-squared divergence

$$\chi^2_{\leq D}(\mathcal{H}_1 \| \mathcal{H}_0)$$

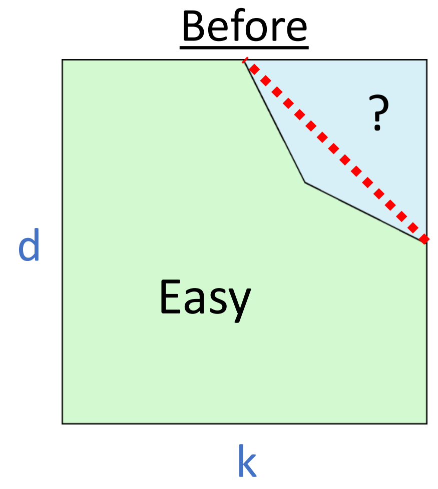
[BHKKMP16, HS17, HKPRSS17, Hop18]

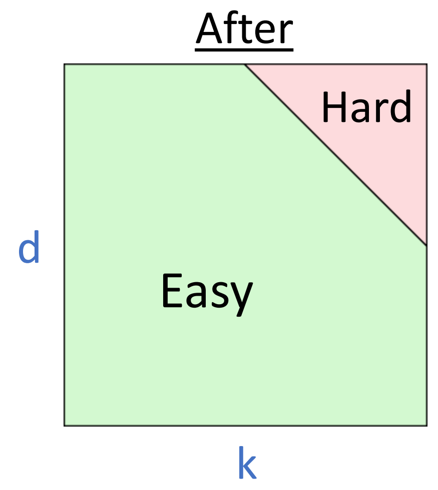
Prediction:

- "Easy" when $kd \ll n^{3/2}$
- "Hard" when $kd \gg n^{3/2}$

Poly-time algorithm succeeds when $kd \ll n^{3/2}$
[Mao, **W** '21]

All low-degree algorithms fail when $kd \gg n^{3/2}$
[Mao, **W** '21]





Spectral Methods

Input: $Y = \begin{array}{cc} & d \\ n & \begin{bmatrix} | & | & | & | \end{bmatrix} \end{array}$

- The algorithm: threshold leading eigenvalue $\lambda_{\max}$ of $d \times d$ matrix M

$$M_{ij} = \underbrace{\sum_{r=1}^{n} \left(\sum_{s=1}^{d} Y_{rs}^2 - \frac{d}{n} \right) Y_{ri} Y_{rj}}_{\text{Degree-4 polynomial in } Y}$$

[Hopkins, Schramm, Shi, Steurer ‘16]

- This is a low-degree algorithm:

$$f(Y) \coloneqq \mathrm{Tr}(M^{2p}) = \sum \lambda_i^{2p} \approx \lambda_{\max}^{2p}$$

need $p = \Theta(\log n)$

$\deg(f) = 4 \cdot 2p = \Theta(\log n)$

Applying the Low-Degree Framework

Compute low-degree chi-squared divergence

$\chi^2_{\leq D}(\mathcal{H}_1 \| \mathcal{H}_0)$

[BHKKMP16, HS17, HKPRSS17, Hop18]

Prediction:

- "Easy" when $kd \ll n^{3/2}$
- "Hard" when $kd \gg n^{3/2}$

<u>Poly-time algorithm</u> succeeds when $kd \ll n^{3/2}$

[Mao, **W** '21]

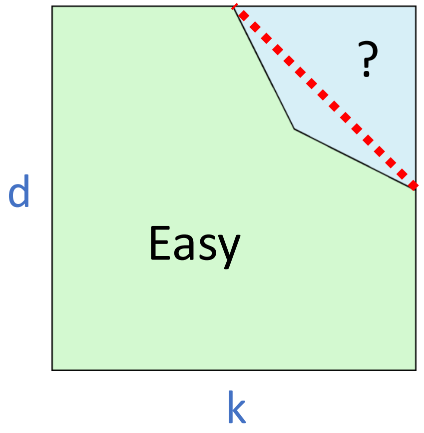
All <u>low-degree algorithms</u> fail when $kd \gg n^{3/2}$

[Mao, **W** '21]

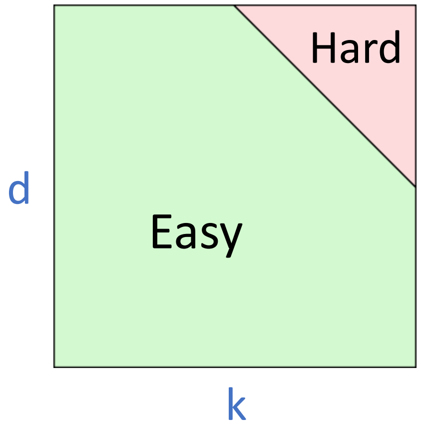
Spectral method

Includes <u>all</u> spectral methods

<u>Before</u>



<u>After</u>



II. Estimation

Two Fundamental Questions in Statistics

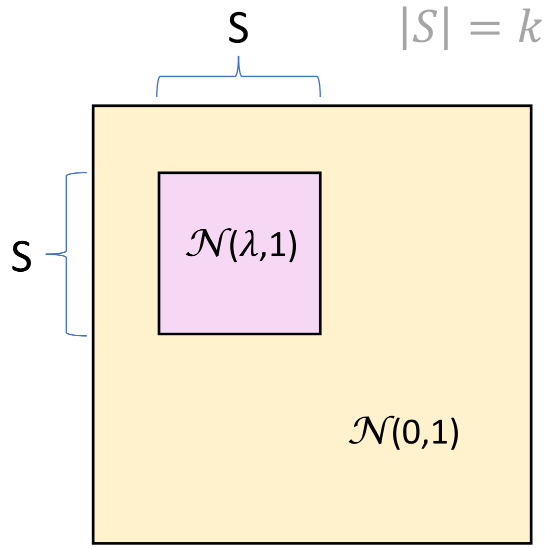
(joint work with Tselil Schramm)

Hypothesis Testing vs **Estimation**

"is there a planted signal?" — "find the planted signal"

- Are these tasks equally difficult?
- Not always!

Planted Submatrix

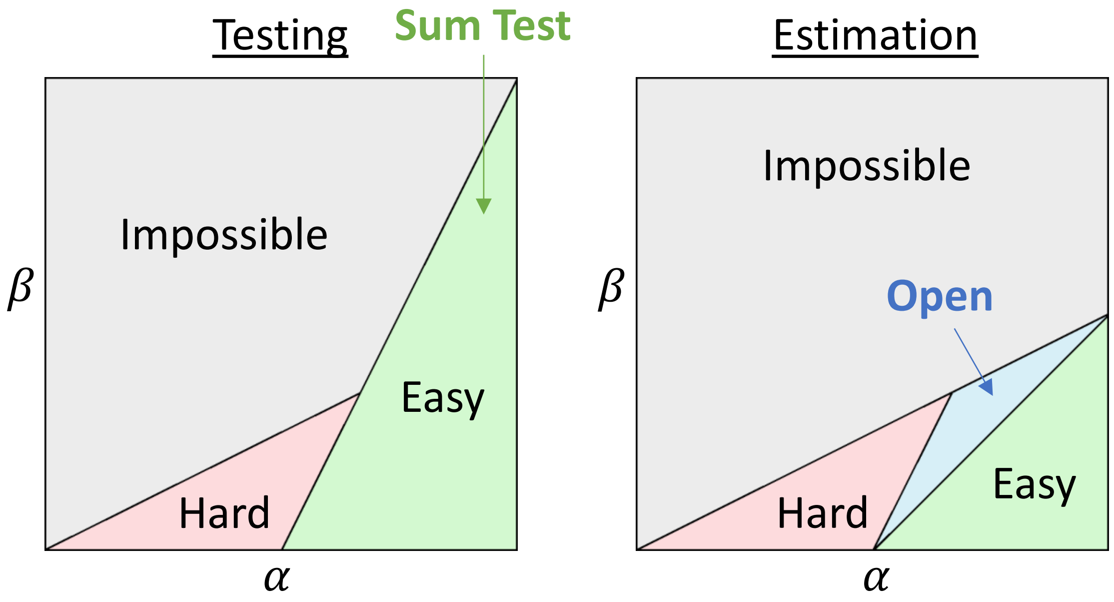


- $n \times n$ matrix with planted $k \times k$ submatrix
- <u>Testing</u>: distinguish vs $\mathcal{H}_0$: all entries $\mathcal{N}(0,1)$
- <u>Estimation</u>: find the "planted" indices S
- Regime: $k = n^{\alpha}, \lambda = n^{-\beta}$ where $\alpha, \beta \in [0,1]$

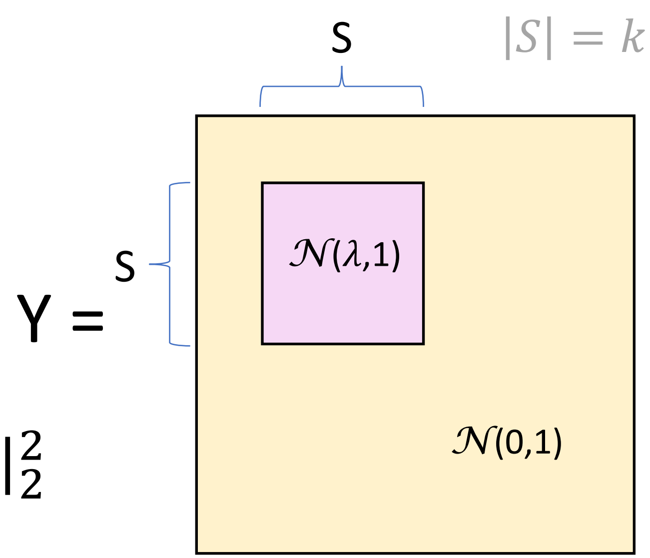
Phase diagram:

(assuming hardness of planted clique)

[KBRS11, BI13, BIS15, MW15, CLR17, BBH18]



Low-Degree Estimation



Degree-D mean squared error: $\mathrm{MMSE}_{\leq D} := \min_f \mathrm{E}\,\|f(Y) - \mathbf{1}_S\|_2^2$

degree-D multivariate polynomial $f: \mathbb{R}^{n^2} \to \mathbb{R}^n$

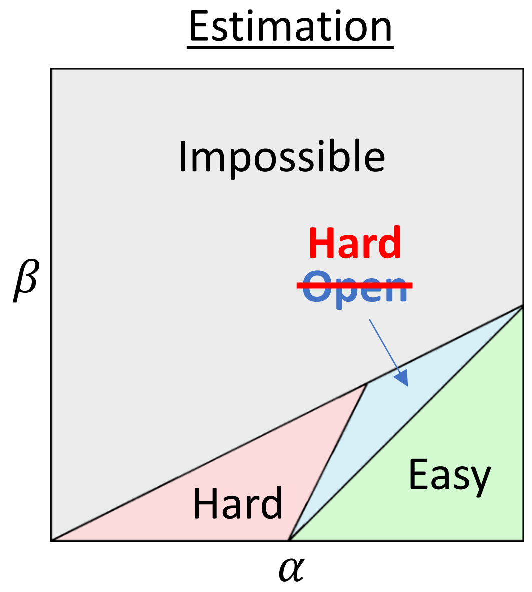
Theorem (Schramm, **W** '20)

- For α, β in the green region, $\mathrm{MMSE}_{\leq O(1)}$ is "small" (perfect estimation)
- For α, β in the blue, red, gray regions, $\mathrm{MMSE}_{\leq \mathrm{polylog}(n)}$ is "large" (trivial)

The "open" region is "hard" for low-degree polynomials

Resolves open question of [Hopkins, Steurer '17]

- Hypothesis testing has "closed form" solution, but estimation does not…



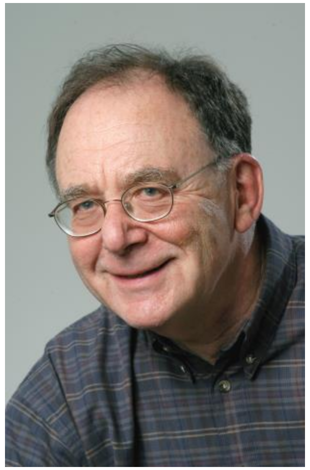
III. Optimization

Problems Without a Planted Signal

(joint work with David Gamarnik & Aukosh Jagannath)

- Imagine an airline wants to find the optimal schedule…
- Many good solutions exist
- Hope: low-degree framework for random optimization problems?

Maximum Independent Set in G(n,d/n)



Richard Karp
(Turing Award 1985)

- Sparse Erdős–Rényi graph G(n,d/n) — n vertices; average degree d
  - Double limit $n \to \infty$, then $d \to \infty$ — "d is a large constant"
- Goal: find a large independent set — $S \subseteq [n]$ with no internal edges
  - With high probability 1 – o(1)
- OPT = 2(log d/d)n [Frieze '90]
- ALG = (log d/d)n [Karp '76] greedy algorithm
- Is there a better poly-time algorithm?
  - Local algorithms achieve ALG and no better
    [Gamarnik, Sudan 14; Rahman, Virág '14]
  - Low-degree algorithms achieve ALG and no better
    [Gamarnik, Jagannath, **W** '20; **W** '20]

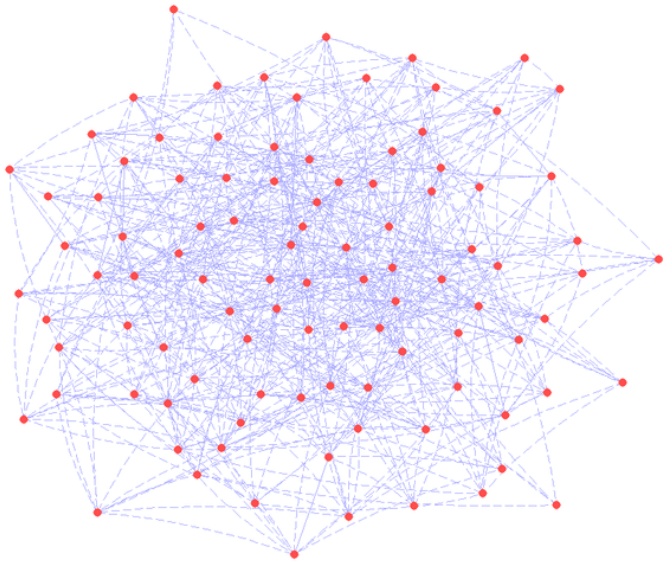


Image: [Frey, Hledik '18]

Main Result

- Max independent set in G(n,d/n): OPT = 2(log d/d)n, ALG = (log d/d)n
- Degree-D algorithm: degree-D polynomial $f:\{0,1\}^{\binom{n}{2}} \to \mathbb{R}^n$
- "Success": for $Y \sim$ G(n,d/n), output $f(Y)$ is an approximate indicator vector for an indep. set

Theorem (Gamarnik, Jagannath, **W** '20; **W** '20)
- Some degree-O(1) algorithm can find an indep. set of size (1-ε)ALG with probability $1 - \exp(-n^{1/3})$
- No degree-polylog(n) algorithm can find an indep. set of size (1+ε)ALG with probability $1 - \exp(-n^{\Omega(1)})$

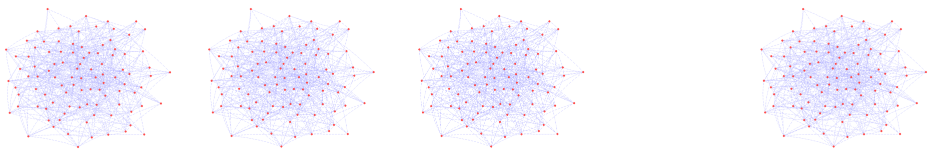
ALG is the algorithmic threshold for low-degree polynomial algorithms

Proof Overview (Hardness)

$(1+\varepsilon)\mathrm{ALG}$

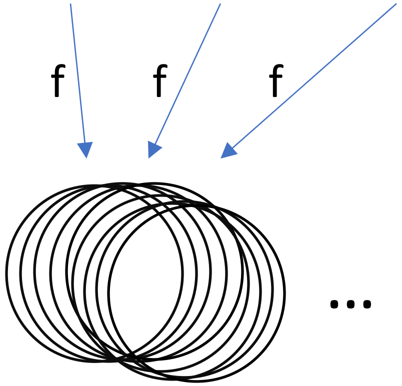
- Suppose (for contradiction) that low-degree f finds large independent sets

- At step i, resample edge i



- Since f is low degree, output is "stable"

$$Y^{(0)} \to Y^{(1)} \to Y^{(2)} \to \cdots \to Y^{(T)}$$



- But a long chain of large independent sets does not exist → contradiction

- Guaranteed to add an S_i after $m = \binom{n}{2}$ steps

Stability of Low-Degree Polynomials

$Y^{(0)} \in \{0,1\}^m$, i.i.d. Bernoulli(p) $\quad m = \binom{n}{2} \quad \mathrm{p} = \mathrm{d/n}$

$$Y^{(0)} \to Y^{(1)} \to Y^{(2)} \to \cdots \to Y^{(m)} \to Y^{(m+1)} \to \cdots \to Y^{(Km)}$$

resample entry 1 (edge 1) | resample entry 2 | resample entry m | resample entry 1

Theorem (Gamarnik, Jagannath, **W** '20)
Fix a degree-D polynomial $f: \{0,1\}^m \to \mathbb{R}^n$, and c > 0.
With probability at least $p^{4DK/c}$, *every* step t satisfies

$$\left\| f\left(Y^{(t)}\right) - f\left(Y^{(t-1)}\right) \right\|_2^2 \le c\, \mathrm{E} \left\| f\left(Y^{(0)}\right) \right\|_2^2.$$

With non-trivial probability, f's output is "smooth" along entire path

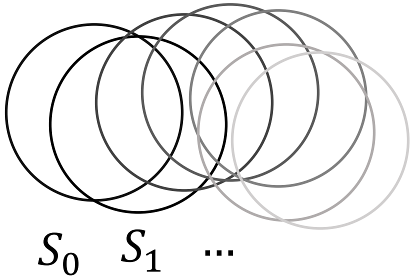
Forbidden Structure

Theorem (**W** ’20)
Fix $\varepsilon > 0$ and $K \geq 5/\varepsilon^2$.
With probability 1-exp(-Ω(n)) over $Y^{(0)} \to \cdots \to Y^{(Km)} \sim G(n, d/n)$, there does not exist a sequence of sets $S_0, S_1, \dots, S_K \subseteq [n]$ with the following properties:

- each S_i is an independent set in some $Y^{(t_i)}$,
- each S_i is “large”: $|S_i| \geq (1+\varepsilon)\mathrm{ALG}$,
- number of new vertices added at each step is “just right”:

$$\left|S_i \setminus \left(\cup_{j<i} S_j\right)\right| \in \left[\frac{\varepsilon}{4}\mathrm{ALG}, \frac{\varepsilon}{2}\mathrm{ALG}\right].$$

Never occurs:



Inspired by the **overlap gap property**
[Gamarnik, Sudan 14; Rahman, Virág ’14; CGPR17; GJ19; …]

Proof: first moment method

Proof Overview (Hardness)

$(1+\varepsilon)\mathrm{ALG}$

- Suppose (for contradiction) that low-degree f finds large independent sets

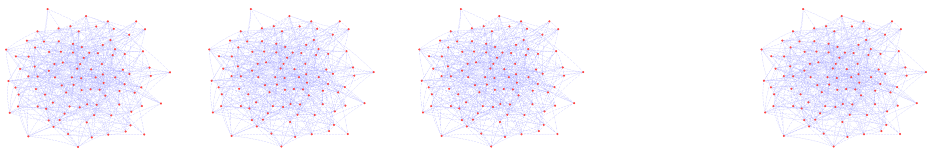
- At step i, resample edge i

- Since f is low degree, output is "stable"

  with non-trivial probability

- But a long chain of large independent sets does not exist → contradiction

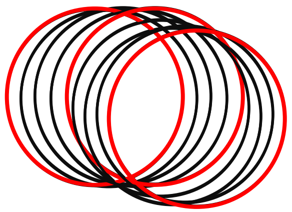
- Guaranteed to add an S_i after $m = \binom{n}{2}$ steps



$T = mK$

$$Y^{(0)} \to Y^{(1)} \to Y^{(2)} \to \cdots \to Y^{(T)}$$

f f f



…

Forbidden structure $S_0, S_1, \dots, S_K$

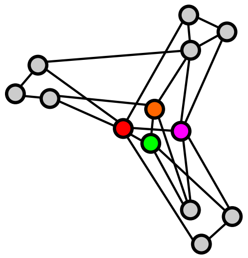
$$\left|S_i \setminus \left(\cup_{j<i} S_j\right)\right| \in \left[\frac{\varepsilon}{4}\mathrm{ALG}, \frac{\varepsilon}{2}\mathrm{ALG}\right]$$

Comments

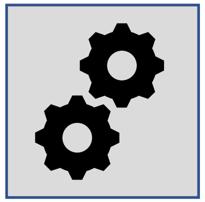
- Concrete evidence for hardness of Karp’s problem
  - Low-degree algorithms fail to surpass threshold ALG
- New techniques for proving failure of low-degree algorithms
  - Unifies “planted” and “non-planted” problems
- Not just for independent set
  - Random k-SAT [Bresler, Huang ’21]
- Can rule out any “stable” algorithm
  - Circuit lower bounds [Gamarnik, Jagannath, **W** ‘21]

Conclusion

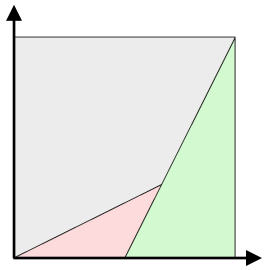
Achieving the Dream?



New statistical problem

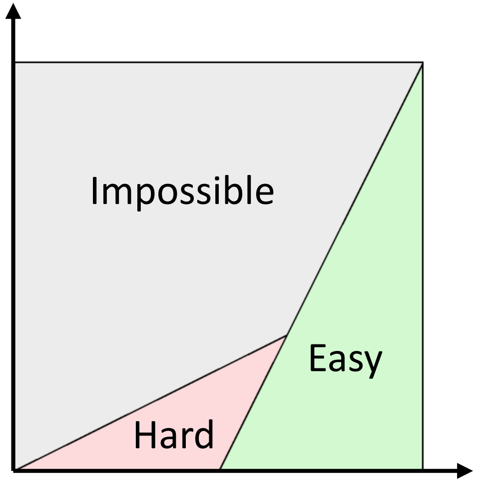


Systematic method
of analysis



Phase diagram

Achieving the Dream?



Low-Degree Hardness

- Hypothesis testing [KWB19, BKW20, DKWB19, LWB20, DKWB21, MW21]
- Estimation [SW20]
- Random optimization problems [GJW20, W20]
- Certification (via "quiet planting") [BKW20, DKWB21, BBKMW21, BKW21]

Statistical Impossibility

- [MPW16, PWBM18, PWB20]

Algorithm Design

New frameworks

- Kikuchi hierarchy [WAM19]
- Tensor networks [MW19]

Existing frameworks

- Spectral [PWBM18, MW19, WAM19, MW21]
- Approx. message passing [PWBM18]
- SDP [PW17, MPW16, PPWBAS19]
- LLL lattice basis reduction [ZSWB21]

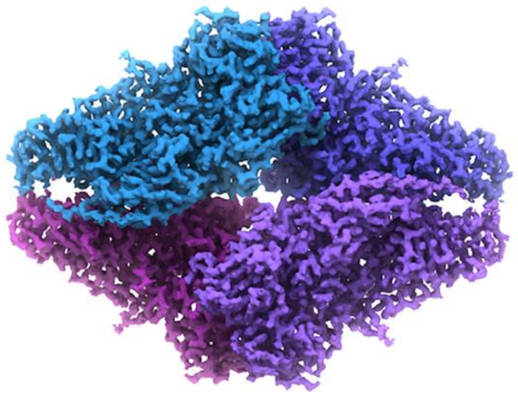
Application Domains

- Sparse PCA [DK**W**B19]
- Tensor PCA [**W**AM19]
- Robust community detection [P**W**17, MP**W**16]
- Sparse clustering [L**W**B20]
- Planted vector in a subspace [M**W**21]

Group actions

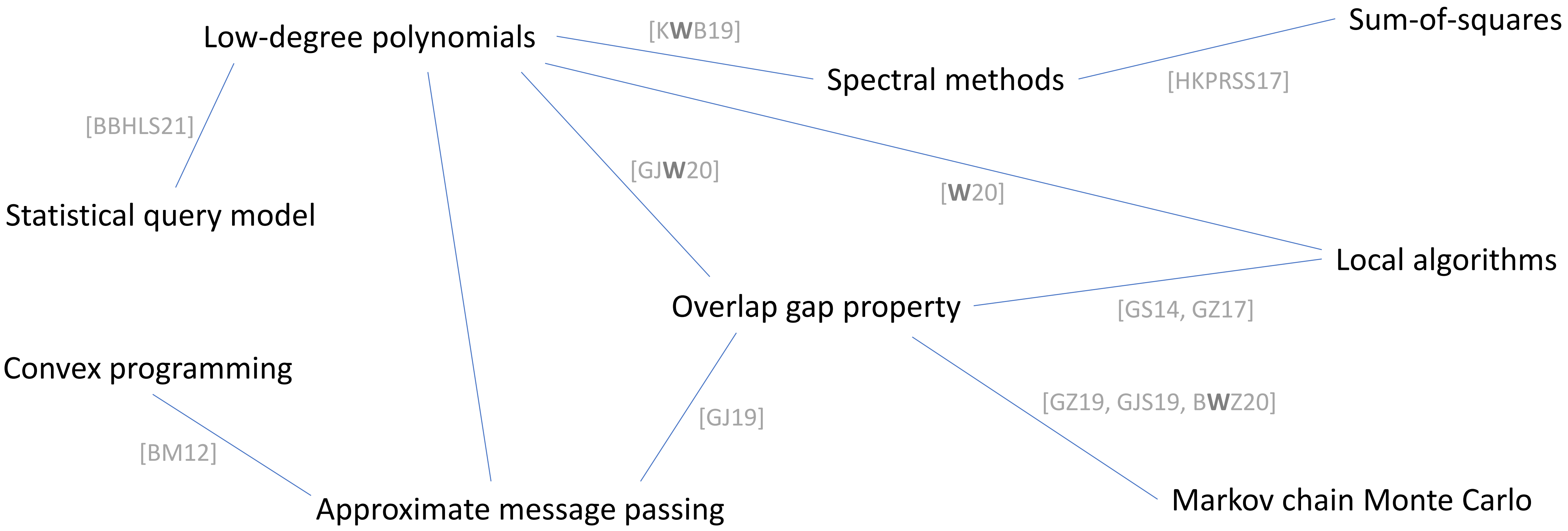
- Synchronization [P**W**BM18]
- Orbit recovery [BBKPWW**W**17, **W**18, M**W**19]

Cryo-electron microscopy



Bartesaghi et al., Science 348 (6239): 1147-1151

Rigorous Connections Between Frameworks



Some References

- Notes on Computational Hardness of Hypothesis Testing: Predictions using the Low-Degree Likelihood Ratio
  Kunisky, **W**, Bandeira
  arXiv 2019 (survey article on low-degree algorithms)
- Optimal Spectral Recovery of a Planted Vector in a Subspace
  Mao, **W**
  Submitted
- Computational Barriers to Estimation from Low-Degree Polynomials
  Schramm, **W**
  Annals of Statistics (to appear)
- Low-Degree Hardness of Random Optimization Problems
  Gamarnik, Jagannath, **W**
  FOCS 2020
- Optimal Low-Degree Hardness of Maximum Independent Set
  W
  Mathematical Statistics and Learning, 2022
- Spectral Planting and the Hardness of Refuting Cuts, Colorability, and Communities in Random Graphs
  Bandeira, Banks, Kunisky, Moore, **W**
  COLT 2021
- The Kikuchi Hierarchy and Tensor PCA
  W, El Alaoui, Moore
  FOCS 2019

Thanks!