Is Planted Coloring Easier than Planted Clique?

Alex Wein

UC Davis

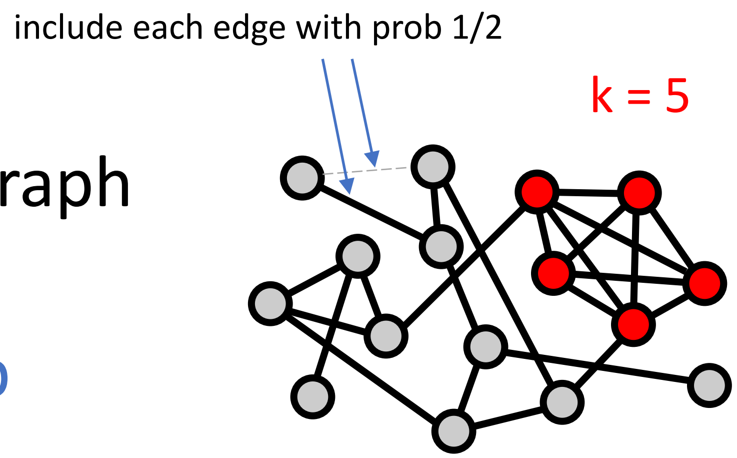
Joint work with: Pravesh Kothari, Santosh Vempala, Jeff Xu

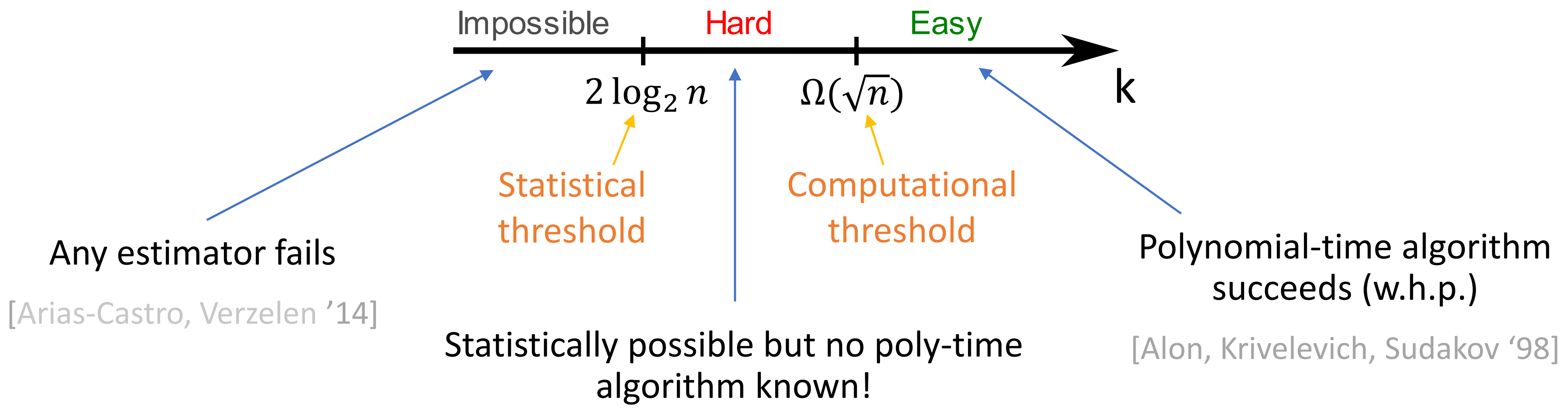
I. Planted Clique & Planted Coloring

Detection, Recovery, Refutation

Planted Clique Problem

- 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





Algorithmic Tasks

- **Detection**: distinguish $\mathbb{P}$ vs $\mathbb{Q}$ w.h.p.
  - $\mathbb{Q}$: G(n,1/2)
  - $\mathbb{P}$: G(n,1/2) + {k-clique}
- **Recovery**: given G $\sim \mathbb{P}$, identify the clique vertices (exactly, w.h.p.)
- **Refutation**: given G $\sim \mathbb{Q}$, *prove* there is no k-clique

Alg: count total edges

Alg: max degree

Alg: spectral (next slide)

- All have poly-time algorithms when $k \gg \sqrt{n}$ (ignoring log factors)
- No poly-time algorithms known when $k \ll \sqrt{n}$

Refuting a Large Clique

Why?

- A – adjacency matrix (±1 valued, 1's on diagonal)
- If there is a k-clique $S \subseteq [n]$,

$$\lambda_{\max}(A) \geq \frac{\mathbb{1}_S^\top A \mathbb{1}_S}{||\mathbb{1}_S||^2} = \frac{k^2}{k} = k$$

- Under $\mathbb{Q}$ = G(n,1/2),

$$\lambda_{\max}(A) \leq 3\sqrt{n} \qquad \text{w.h.p.}$$

- Refutation alg: output NO if $\lambda_{max}(A) < k$, MAYBE otherwise
- Succeeds when $k \gg \sqrt{n}$:
    - If graph has a k-clique, output is *always* MAYBE
    - If graph is drawn from $\mathbb{Q}$, output is NO w.h.p.

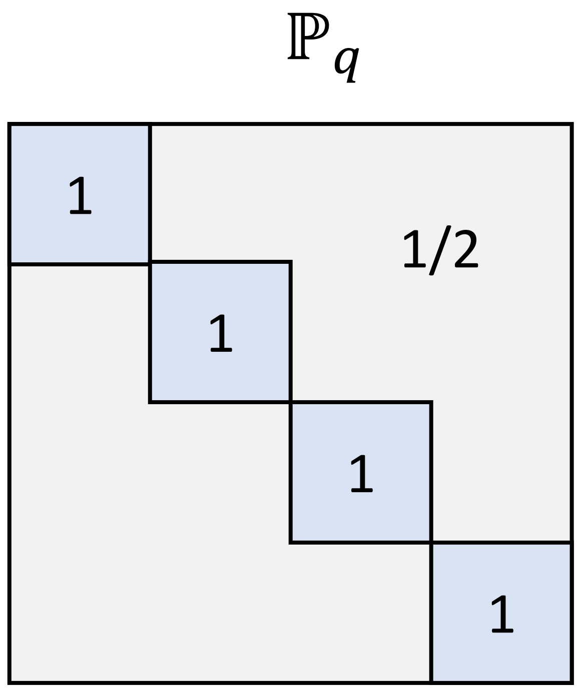
refutation task

Recall: for planted clique, all three tasks (detection, recovery, refutation) have the same computational threshold $k \approx \sqrt{n}$

This is not true in general…

Many Planted Cliques / Planted Coloring

- $\mathbb{P}$: q disjoint planted cliques of size k=n/q
  - Complement graph has a planted q-coloring
- **Detection**: distinguish $\mathbb{P}_q$ versus $\mathbb{Q}$ = G(n,1/2)
  - Easy when $k \gg 1$ (count total edges)
- **Recovery**: given G $\sim \mathbb{P}_q$, recover the cliques exactly
  - Easy when $k \gg \sqrt{n}$ (common neighbors)
- **Refutation**: given G $\sim \mathbb{Q}$, *prove* there is no q-coloring
  - Easy when $k \gg \sqrt{n}$ (spectral)
- Are these optimal? Is coloring easier than clique?

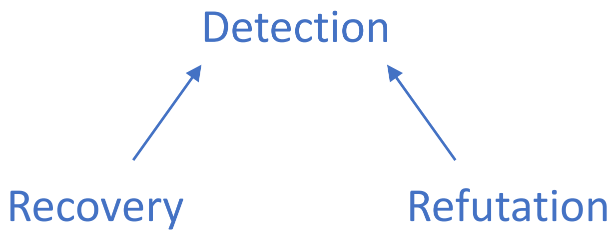


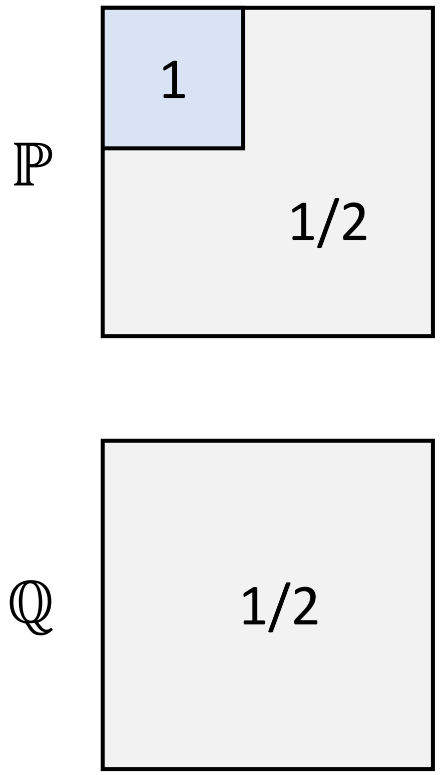
Our Perspective

- **Goal**: understand computational complexity of (1) recovery in $\mathbb{P}_q$ and (2) refutation of q-colorability in $\mathbb{Q}$ = G(n,1/2)
- Forget detection for now… but we will introduce various testing problems as proof constructs
- No formal relation between recovery and refutation
- Refutation can be strictly harder [Bandeira, Banks, Kunisky, Moore, **W** ’20]

Hardness of Recovery/Refutation (Clique)

- Back to planted clique: assume detection is hard when $k \ll \sqrt{n}$
  - $\mathbb{P}$ (planted k-clique) vs $\mathbb{Q}$ = G(n,1/2)
- Recovery (in $\mathbb{P}$) must be hard when $1 \ll k \ll \sqrt{n}$
  - W.h.p., $\mathbb{Q}$ has no k-clique
  - If you could recover, you could distinguish $\mathbb{P}$ vs $\mathbb{Q}$
- Refuting a k-clique in $\mathbb{Q}$ must be hard when $k \ll \sqrt{n}$
  - W.h.p, $\mathbb{P}$ has a k-clique
  - If you could refute, you could distinguish $\mathbb{P}$ vs $\mathbb{Q}$

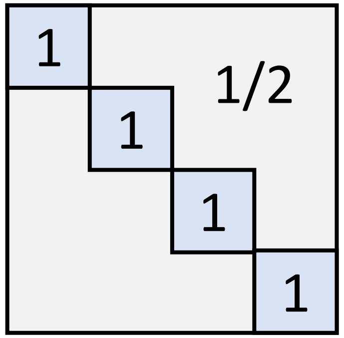




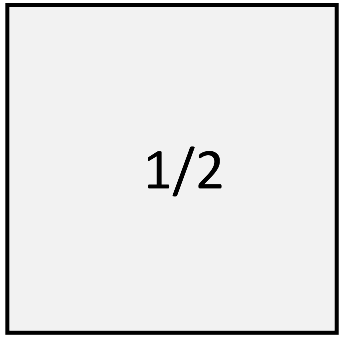
Hardness of Recovery/Refutation (Coloring)

- To show hardness of recovery in $\mathbb{P}_q$, construct $\widetilde{\mathbb{Q}}$ such that:
  - W.h.p., $\widetilde{\mathbb{Q}}$ is not q-colorable
  - Distinguishing $\mathbb{P}_q$ vs $\widetilde{\mathbb{Q}}$ is hard
  - Why: if you could recover, you could distinguish $\mathbb{P}_q$ vs $\widetilde{\mathbb{Q}}$
- To show hardness of refutation in $\mathbb{Q}$ = G(n,1/2), construct $\widetilde{\mathbb{P}}$ such that:
  - W.h.p., $\widetilde{\mathbb{P}}$ is q-colorable
  - Distinguishing $\widetilde{\mathbb{P}}$ vs $\mathbb{Q}$ is hard
  - Why: if you could refute, you could distinguish $\widetilde{\mathbb{P}}$ vs $\mathbb{Q}$

$\mathbb{P}_q$



$\mathbb{Q}$



Low-Degree Testing

[Hopkins, Steurer ‘17; Hopkins, Kothari, Potechin, Raghavendra, Schramm, Steurer ’17; Hopkins ’18; Kunisky, W, Bandeira ’19, …]

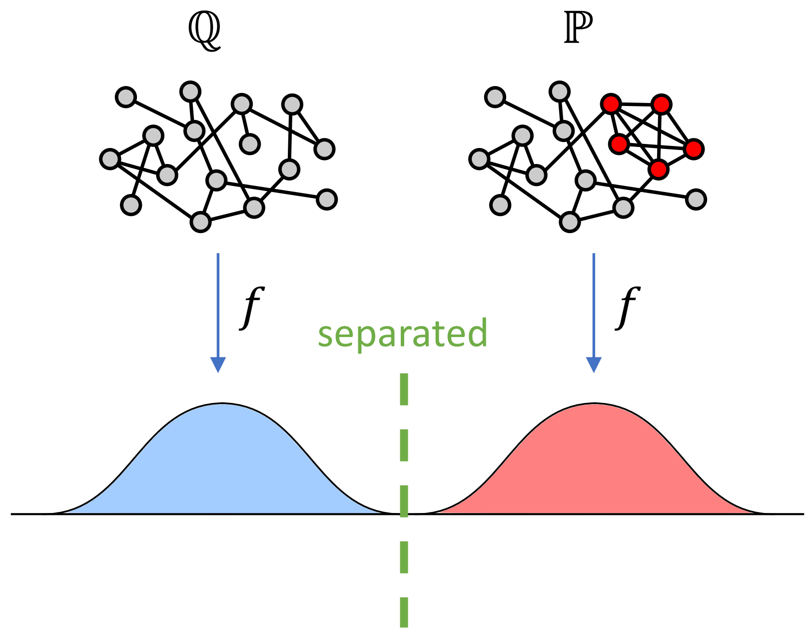
- **Low-degree test**: multivariate polynomial of degree O(log n)

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

Input: graph   Output: number

- E.g. count edges, triangles, …
- “Success”: f *strongly separates* $\mathbb{P}$ and $\mathbb{Q}$ if

$$\sqrt{\mathrm{Var}_{\mathbb{P}}(f) \vee \mathrm{Var}_{\mathbb{Q}}\,(f)} = o(|\mathrm{E}_{\mathbb{P}}[f] - \mathrm{E}_{\mathbb{Q}}[f]|)$$

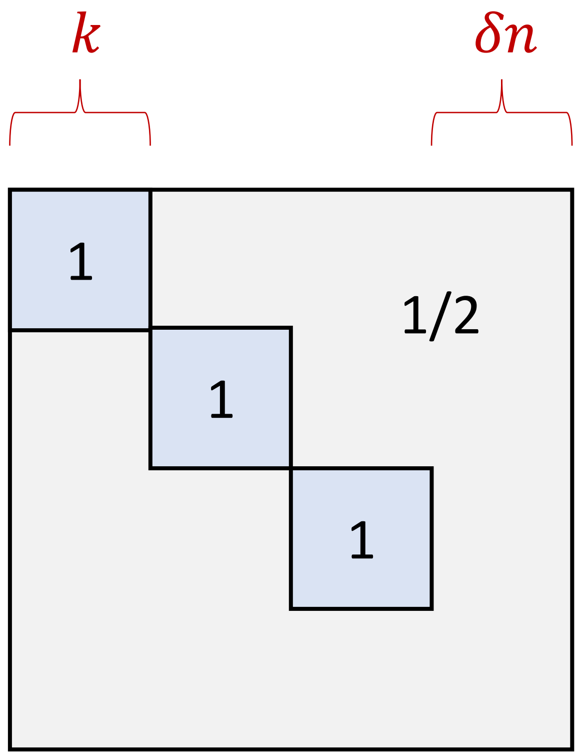


II. Recovery

Hardness of recovering a planted q-coloring

Warm-Up: Partial Coloring

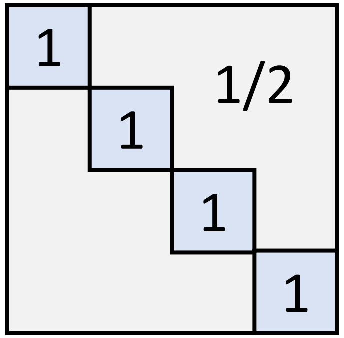
- Cliques of size k with δ fraction of vertices un-colored
  - $\delta = \Theta(1)$ or even $\delta = n^{-o(1)}$
- Exact recovery is easy when $k \gg \sqrt{n}$
- Exact recovery is hard when $k \ll \sqrt{n}$
  - Why: even if all cliques except one are revealed, still left with a hard instance of planted clique
  - Formally: reduction from planted clique
- Adding cliques doesn't make recovery easier
- But this argument won't work for coloring ($\delta = 0$)



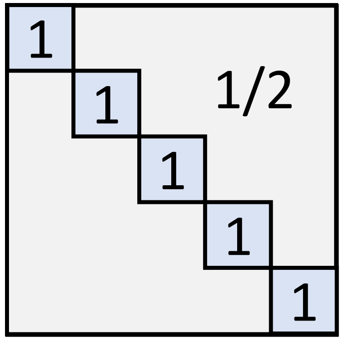
True Coloring

- **Goal**: hardness of recovery in $\mathbb{P}_q$ when $k \ll \sqrt{n}$
- Want to construct $\widetilde{\mathbb{Q}}$ such that:
  - W.h.p., $\widetilde{\mathbb{Q}}$ is not q-colorable
  - Distinguishing $\mathbb{P}_q$ vs $\widetilde{\mathbb{Q}}$ is hard (for low-degree tests)
- $\widetilde{\mathbb{Q}}$ = G(n,1/2)? Easy when $k \gg 1$ (total edge count)
- $\widetilde{\mathbb{Q}}$ = G(n,1/2+ϵ)? Easy when $k \gg n^{1/4}$ (triangle count)
- ???
- $\widetilde{\mathbb{Q}} = \mathbb{P}_{q+1}$ Not q-colorable; hard when $k \ll \sqrt{n}$

$\mathbb{P}_q$



$\widetilde{\mathbb{Q}}$

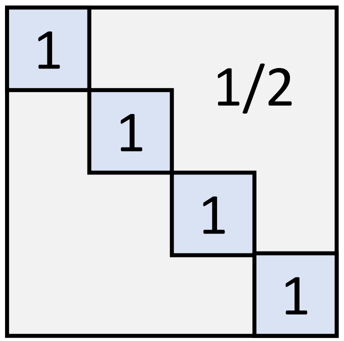


Testing q vs $q + \ell$

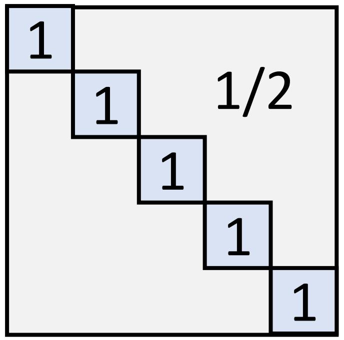
Theorem: Let $1 \leq q < q + \ell \leq n$.

- (Easy) If $q^2 \ll \ell n$ then there is a degree-1 polynomial that strongly separates $\mathbb{P}_q$ and $\mathbb{P}_{q+\ell}$.
- (Hard) If $q^2 \gg \ell n$ then no degree-O(log n) polynomial strongly separates $\mathbb{P}_q$ and $\mathbb{P}_{q+\ell}$.

$\mathbb{P}_q$



$\mathbb{P}_{q+\ell}$



Easy when $q^2 \ll \ell n$, hard when $q^2 \gg \ell n$

*Now $\gg$ hides $n^{o(1)}$

Testing q vs $q + \ell$: Proof (Lower Bound)

- To rule out strong separation between $\mathbb{P}$ and $\mathbb{Q}$, suffices to show

$$\mathrm{Adv}_{\leq D}(\mathbb{P}, \mathbb{Q}) \coloneqq \max_{f \deg D} \frac{\mathrm{E}_{\mathbb{P}}[f]}{\sqrt{\mathrm{E}_{\mathbb{Q}}[f^2]}} = O(1)$$

- Standard formula:

$$\mathrm{Adv}_{\leq D}^2(\mathbb{P}, \mathbb{Q}) = \sum_h \big(\mathrm{E}_{\mathbb{P}}[h]\big)^2$$

  where $\{h\}$ is an orthonormal basis for degree-D polynomials w.r.t. $\mathbb{Q}$
- Straightforward if $\mathbb{Q}$ has independent coordinates, e.g. G(n,1/2)
- Our proof builds on [Schramm, **W** ’20; Rush, Skerman, **W**, Yang ‘22]

Recovery: Summary

- Testing planted q-coloring versus planted-$(q+\ell)$-coloring
  - Easy for low-degree polynomials when $q^2 \ll \ell n$, hard when $q^2 \gg \ell n$
  - $\ell = 1$: hard when $q^2 \gg n$, i.e., $k \; := \frac{n}{q} \ll \sqrt{n}$
- Conjecture: no poly-time algorithm can distinguish q vs q+1 if $k \ll \sqrt{n}$
  - If true, this conjecture implies: no poly-time algorithm can <u>recover</u> a planted q-coloring when $k \ll \sqrt{n}$
  - I.e., simple algorithm (common neighbors) is optimal
  - Planted coloring is no easier than planted clique (for recovery)
- Alternative: low-degree lower bound for recovery [Schramm, **W** '20]

III. Refutation

Hardness of refuting q-colorability in G(n,1/2)

Refutation: Prior Work

- Recall: refuting q-colorability in G(n,1/2) is easy when $k := \frac{n}{q} \gg \sqrt{n}$
- Sum-of-squares (SoS) lower bounds
  - A particular SoS formulation fails when $k \ll \sqrt{n}$ [Kothari, Manohar '21]
  - Open to characterize the more canonical formulation (equality constraints)
- Our approach: formulate a new type of refutation lower bound
  - Directly based on low-degree polynomials
  - Advantages: simplicity, no choice of formulation
  - No formal relation to SoS

Low-Degree Refutation

Definition: A polynomial $f\colon\{0,1\}^{\binom{n}{2}} \to \mathbb{R}$ *strongly separates* $\mathbb{Q}$ = G(n,1/2) *from q-colorable graphs* if

(1) $f(A) \geq 1$ for *every* q-colorable graph A

(2) $\mathrm{E}_{\mathbb{Q}}[f^2] = o(1)$

- Implies refutation: output NO if $f(A) < 1$, MAYBE otherwise
  - If graph has a q-coloring, output is *always* MAYBE
  - If graph is drawn from $\mathbb{Q}$, output is NO w.h.p. (Chebyshev)

Low-Degree Refutation: Results

Theorem

- (Easy) If $k \gg \sqrt{n}$, there is a degree-O(log n) polynomial that strongly separates $\mathbb{Q}$ = G(n,1/2) from q-colorable graphs
  - Proof: spectral $f(A) = \mathrm{Tr}(A^{2m}) = \sum \lambda_i(A)^{2m} \geq \lambda_{\max}(A)^{2m}$
- (Hard) If $k \ll n^{1/3}$ then no degree-O(log n) polynomial strongly separates $\mathbb{Q}$ = G(n,1/2) from q-colorable graphs

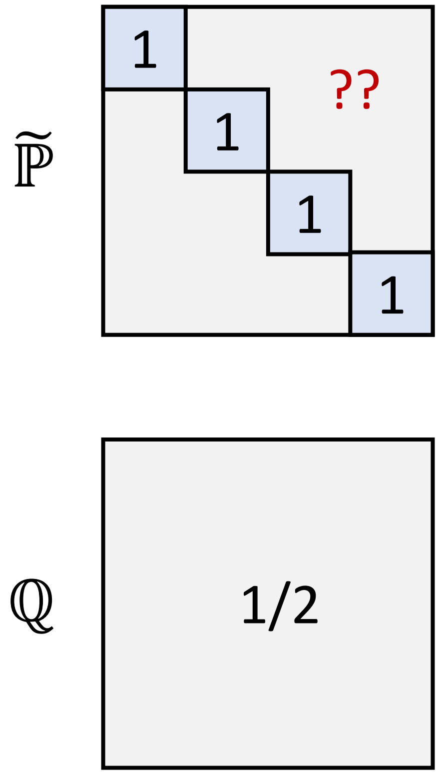
Easy when $k \gg \sqrt{n}$, hard when $k \ll n^{1/3}$, open when $\mathrm{n}^{1/3} \ll k \ll n^{1/2}$

Proof (Lower Bound)

- To show hardness of refutation in $\mathbb{Q}$ = G(n,1/2), construct $\widetilde{\mathbb{P}}$ such that:
  - W.h.p., $\widetilde{\mathbb{P}}$ is q-colorable
  - Distinguishing $\widetilde{\mathbb{P}}$ vs $\mathbb{Q}$ is hard
- **Low-degree analogue**: If $\widetilde{\mathbb{P}}$ supported on q-colorable graphs and $\mathrm{Adv}_{\leq D}\left(\widetilde{\mathbb{P}}, \mathbb{Q}\right) = O(1)$ then no degree-D polynomial strongly separates $\mathbb{Q}$ from q-colorable graphs

Proof (Lower Bound)

- **Goal**: hardness of refuting q-colorability in $\mathbb{Q}$ = G(n,1/2), for $k \ll \sqrt{n}$
- Want to construct $\widetilde{\mathbb{P}}$ such that:
  - $\widetilde{\mathbb{P}}$ supported on q-colorable graphs
  - Distinguishing $\widetilde{\mathbb{P}}$ vs $\mathbb{Q}$ is hard (for low-degree tests)
- What to do outside the cliques?
- Ber(1/2), i.e., $\widetilde{\mathbb{P}}$ = $\mathbb{P}_q$? Easy when $k \gg 1$ (total edge count)
- Ber(1/2-ϵ)? Easy when $k \gg n^{1/4}$ (triangle count)
- We can reach $k \approx n^{1/3}$: plant both cliques and ind. sets
- Open: how to go beyond this?



Refutation: Summary

- We expect it is hard to refute q-colorability in G(n,1/2) when $k \ll \sqrt{n}$
  - Refuting coloring is no easier than refuting clique
- But we only proved it (in our framework) when $k \ll n^{1/3}$
- To close the gap, suffices to construct a “quieter” planted distribution
- Maybe no such distribution exists?
  - This would imply a better refutation algorithm!
  - Quiet planting approach is “complete”
  - Proof: minimax theorem for 2-player game: distribution $\widetilde{\mathbb{P}}$ vs polynomial

Thanks!