Low-degree Security of the Planted Random Subgraph Problem

Andrej Bogdanov*, Chris Jones **, Alon Rosen***, and Ilias Zadik†

Abstract. The planted random subgraph detection conjecture of Abram et al. (TCC 2023) asserts the pseudorandomness of a pair of graphs (H, G), where G is an Erdős-Rényi random graph on n vertices, and H is a random induced subgraph of G on k vertices. Assuming the hardness of distinguishing these two distributions (with two leaked vertices), Abram et al. construct communication-efficient, computationally secure (1) 2-party private simultaneous messages (PSM) and (2) secret sharing for forbidden graph structures.

We prove the low-degree hardness of detecting planted random subgraphs all the way up to $k \leq n^{1-\Omega(1)}$. This improves over Abram et al.'s analysis for $k \leq n^{1/2-\Omega(1)}$. The hardness extends to r-uniform hypergraphs for constant r.

Our analysis is tight in the distinguisher's degree, its advantage, and in the number of leaked vertices. Extending the constructions of Abram et al, we apply the conjecture towards (1) communication-optimal multiparty PSM protocols for random functions and (2) bit secret sharing with share size $(1+\epsilon)\log n$ for any $\epsilon > 0$ in which arbitrary minimal coalitions of up to r parties can reconstruct and secrecy holds against all unqualified subsets of up to $\ell = o(\epsilon \log n)^{1/(r-1)}$ parties.

1 Introduction

In the planted clique model [Jer92, Kuc95] one observes the union of an Erdős-Rényi random graph $G_0 \sim G(n, 1/2)$ and a randomly placed $k = k_n$-clique H, i.e., the graph $G = G_0 \cup H$. The goal of the planted clique detection task is to distinguish between observing G from the planted clique model and G which is simply an instance of $G(n, 1/2)$. The planted clique conjecture states that the planted clique instance remains pseudorandom whenever $k \leq n^{1/2-\Omega(1)}$ up to $n^{-\Omega(1)}$ distinguishing advantage. Conversely, multiple polynomial-time algorithms can distinguish with high probability whenever $k = \Omega(\sqrt{n})$. Research on the planted clique conjecture has gone hand-in-hand with key developments in average-case complexity theory over the last decades, including spectral and tensor algorithms [AKS98, FK08], lower bound techniques for restricted classes including the sum-of-squares hierarchy [BHK+19], low-degree

* University of Ottawa. Email: abogdano@uottawa.ca
** Bocconi University. Email: chris.jones@unibocconi.it
*** Bocconi University. Email: alon.rosen@unibocconi.it
† Yale University. Email: ilias.zadik@yale.edu

polynomial methods [Hop18], statistical query methods [FGR+17] and MCMC methods [Jer92, GZ19, CMZ23], and the development of new average-case reductions [BB20, HS24].

At this point, the conjectured hardness of the planted clique problem around $k \approx \sqrt{n}$ stands as a central conjecture in average-case complexity. But despite its popularity, the cryptographic applications have been quite limited, with one exception in the symmetric-key setting proposed by Juels and Peinado [JP97]. Recently Abram et al. [ABI+23] revisited the planted clique problem and showed how it can be useful in the context of secret sharing and secure computation. The authors specifically show that (slight variants of) the planted clique conjecture can be used to construct a computationally secure scheme whose share size is much smaller than the best existing information-theoretically secure scheme.

In order to obtain further improvements to the share size, Abram et al. proposed a new intriguing conjecture similar to planted clique. They start by defining the following general model (also introduced in [Hul22]).

Definition 1. (Planted (induced) subgraph model[1]) *Fix H to be an arbitrary unlabeled subgraph on k vertices. Then G is chosen to be a random n-vertex graph where a copy of H is placed on k vertices chosen uniformly at random (as an induced subgraph on the k vertices), and all edges without both endpoints on the k vertices appear with probability 1/2.*

When H is the k-clique, the planted subgraph model becomes exactly the planted clique model. The clique is the most structured graph possible and it is natural to wonder:

could the problem be significantly harder if a different graph H is planted?

Abram et al. suggest studying the planted random subgraph model in which H is an instance of $G(k, 1/2)$. An equivalent definition is the following.

Definition 2. (Planted random subgraph model) *One observes a pair (H, G), where G is a random n-vertex graph and H is a random k-subgraph of G with the vertex labels removed.*

Abram et al. make the following interesting conjecture.

Conjecture 1. (Planted Random Subgraph conjecture [ABI+23]) *The planted random subgraph problem is hard up to advantage $n^{-\Omega(1)}$ provided $k \leq n^{1-\Omega(1)}$, with high probability over $H \sim G(k, 1/2)$ as n grows to infinity.*

This stands in contrast to the case that H is a k-clique where a computational phase transition is expected to take place at the smaller value $k \approx n^{1/2}$.

Abram et al. confirm the planted random subgraph conjecture in the low-degree analysis framework (to be described below) but only up to the "planted

[1] A similar yet different model where one observes the *union* of a copy of H with an instance of $G(n, 1/2)$ has also been recently analyzed in the statistical inference literature [Hul22, MNWS+23, YZZ24]. For this work, we solely focus on the "induced" variant, where H appears as an induced subgraph of G.

clique threshold" $k \leq n^{1/2-\Omega(1)}$ (a result also independently proven by Huleihel [Hul22]). Their work leaves open the regime $n^{1/2-\Omega(1)} \leq k \leq n^{1-\Omega(1)}$, and in particular the question of whether there is a larger window of hardness for planted random subgraph than for planted clique.

Our main contribution is the confirmation of Conjecture 1 in the low-degree framework. We prove that the planted random subgraph problem remains hard for low-degree distinguishers of degree at most $o((\log n/\log\log n)^2)$ in the *full range* $k \leq n^{1-\Omega(1)}$. The degree is best possible up to $\log\log n$ factors, and the analysis extends also to the case of hypergraphs. See Section 2 for the precise theorem statement.

1.1 Secret sharing and leakage

For their intended cryptographic applications Abram et al. rely on a strengthening of the planted random subgraph conjecture which also allows for leaked additional information about the embedding of H in G. It is easiest to motivate these stronger conjectures through their intended application.

A *(partial) access structure* for k parties is a pair of set systems R, S over $\{1, \ldots, k\}$, where R is upward-closed, S is downward-closed, and R, S are disjoint. A bit secret sharing scheme consists of a randomized sharing algorithm that maps the secret bit $s \in \{0, 1\}$ into k shares so that sets in R can reconstruct s from their shares with probability one, while sets in S cannot distinguish $s = 0$ or $s = 1$.

In a forbidden graph access structure, R is the edge-set of a graph and S is the union of its complement $\{\{u, v\} \notin R \colon u \neq v \in [k]\}$ and the set $[k]$ of vertices. Abram et al. propose the following secret sharing scheme for any such structure:

Construction 1. Forbidden graph secret sharing:

1. The dealer samples a random n-vertex graph G and remembers a secret k-vertex subgraph H of it randomly embedded via $\phi \colon V(H) \to V(G)$.
2. The dealer publishes the pair (H_s, G), where H_s is a k-vertex graph with adjacency matrix

$$H_s(u, v) = \begin{cases} H(u, v) \oplus s, & \text{if } \{u, v\} \in R \\ \text{a random bit}, & \text{otherwise.} \end{cases} \tag{1}$$

3. The share of party v is the value $\phi(v) \in [n]$.

If $\{u, v\} \in R$, the parties reconstruct by calculating

$$H_s(u, v) \oplus G(\phi(u), \phi(v)) = H(u, v) \oplus G(\phi(u), \phi(v)) \oplus s = s. \tag{2}$$

Secrecy requires that the joint distribution $(H_s, G, \phi(u), \phi(v))$ of the public information and the shares is indistinguishable between $s = 0$ and $s = 1$ provided $\{u, v\} \in S$. In the absence of the "leakage" $(\phi(u), \phi(v))$ this is a consequence of the planted random subgraph conjecture (Conjecture 1).

To handle the leakage, we consider the following generalization. Two parties $\{u, v\} \in S$ know the location of their edge $H(u, v) = G(\phi(u), \phi(v))$ in G, which could potentially be useful to search for the "local structure of H" around their edge. The new conjecture posits that if u and v have this additional information, they still cannot distinguish whether H is planted. With an eye towards stronger security we state it below for a general ℓ.

Conjecture 2. (Planted random subgraph conjecture with ℓ-vertex leakage) *With high probability over $H \sim G(k, 1/2)$, the following two distributions are $n^{-\Omega(1)}$-indistinguishable in polynomial time for all subsets $L = \{u_1, \ldots, u_\ell\} \subseteq V(H)$ of size ℓ:*

1. *(planted) $(H, G, \phi(u_1), \ldots, \phi(u_\ell))$ where we choose uniformly at random an injective function $\phi : [k] \to [n]$ and embed H into G on the image of ϕ. The remaining edges of G are sampled randomly.*
2. *(model) $(H, G, \phi(u_1), \ldots, \phi(u_\ell))$ where we choose uniformly at random an injective function $\phi : L \to [n]$ and embed the subgraph of H on L into G on the image of ϕ. The remaining edges of G are sampled randomly.*

Assuming this conjecture with $\ell = 2$, given $(\phi(u), \phi(v))$ for $\{u, v\} \in S$, we claim that both (H_0, G) and (H_1, G) are pseudorandom and hence indistinguishable: As $\{u, v\} \in S$, the (u, v)-th bits of H_0 and H_1 in (H_s, G) are independent of all the others and cannot be used to distinguish. Once the (u, v)-th bits of H_0 and H_1 are removed, both (H_0, G) and (H_1, G) become identically distributed to the planted (H, G) with its (u, v)-th bit removed. By the conjecture, this model is indistinguishable from a uniformly random string.

The share size in this scheme is $(1 + o(1)) \log k$. In contrast, the most compact known forbidden graph scheme with perfect security has shares of size $\exp \tilde{\Theta}(\sqrt{\log k})$ [LVW17,ABF+19]. Statistical security requires shares of size $\log k - O(1)$ when R is the complete graph [ABI+23]. It is not known if computational security is subject to the same limitation.

Under the ℓ-vertex leakage assumption the secrecy holds not only against pairs of parties that are not an edge in R, but also against all independent sets up to size ℓ, i.e.,

$$S = \{I : I \text{ is an independent set of } R \text{ and } |I| \leq \ell\}.$$

By passing to r-hypergraphs instead of graphs, we naturally extend the construction to R which is an arbitrary subset of at most r parties, with security against all size-ℓ independent sets of R (see Construction 3 below). The most compact known perfectly secure forbidden r-hypergraph scheme has share size $\exp \tilde{\Theta}(\sqrt{r \log k})$ [LVW17] whereas our share size is still $(1 + o(1)) \log k$.

It would be interesting to obtain a provable separation in share size between the computationally secure Construction 3 and the best possible perfectly secure construction for some access structure. In Section 4.1 we explain why this is challenging using available methods.

1.2 Private simultaneous messages (PSM)

In a PSM, Alice and Bob are given inputs x, y to a public function $F\colon [k]^2 \to \{0,1\}$. They calculate messages $\phi(x), \phi(y)$ which are securely forwarded to Carol. Carol needs to output the value $F(x,y)$ without learning any information about x and y beyond this value.

Abram et al. propose the following PSM protocol. In a setup phase, F viewed as a bipartite graph is randomly embedded into an otherwise random host graph G via ϕ. The graph G is given to Carol and the embedding ϕ is given to Alice and Bob. Carol outputs $G(\phi(x), \phi(y))$ which must equal $F(x,y)$.

Abram et al. argue that this protocol is "secure" for a $(1-o(1))$-fraction of functions F under Conjecture 2 with leakage $\ell = 2$. Their security definition appears to additionally assume that the choice of inputs (x,y) is independent of the function F. In contrast, our security definition in Section 4.2 allows for Alice and Bob to choose their inputs jointly from some distribution that depends on the description of F. This is more natural for potential cryptographic applications; Alice and Bob should not be expected to commit to their input before they know which function they are computing. We extend our low-degree analysis to support this stronger notion of security.

Messages in this protocol are of length $\log n = (1+\epsilon)\log k$. In contrast, perfect security is known to require combined message length $|\phi(x)| + |\phi(y)| \geq (3-o(1))\log k$ [FKN94, AHMS18] (but it is not known if statistical security is subject to the same bound).

The r-hypergraph variant of the conjecture with leakage $\ell = r$ gives PSM security for r-party protocols also with message size $\log n = (1+\epsilon)\log k$ (Section 4.2). Even without a security requirement the message size must be at least $(1-o(1))\log k$ for the protocol to be correct on most inputs.

1.3 Low-degree lower bounds

We provide evidence for these conjectures in the form of lower bounds against the low-degree polynomial computational model (see e.g., [KWB19] and references therein). In this model, fixing a parameter $D = D_n$, the distinguishing algorithm is allowed to compute an arbitrary degree-D polynomial function of the bits of the input over the field $\mathbb{R}$. The algorithm succeeds if the value of the polynomial is noticeably different between the random and planted models. Degree-D polynomials serve as a proxy for $n^{O(D)}$ time computation since a degree-D polynomial in $\mathrm{poly}(n)$ input bits can be evaluated by brute force in time $n^{O(D)}$ (ignoring numerical issues).

Surprisingly, for noise-robust[2] hypothesis testing problems it has been conjectured that whenever all degree-D polynomials with $D = O(\log n)$ fail (formally, no polynomial strongly separates the two distributions [COGHK+22, Section 7]), then no polynomial-time distinguisher succeeds. This is now known as the

[2] Noise-robustness means that the planted structure is resilient to small random perturbations [Hop18, HW21].

"low-degree conjecture" of Hopkins [Hop18]. Based on this heuristic, a provable failure of $O(\log n)$-degree polynomials to strongly separate the two distributions provides a state-of-the-art prediction of the hard and easy regimes for the problem of interest.

It should be noted that there exists a certain weakness in existing low-degree hardness evidence for the planted clique problem, which also applies to our lower bound for the planted random subgraph problem (and that of [ABI+23]). Both planted clique and planted random subgraph technically do not satisfy the noise-robust assumption of the low-degree conjecture because the planted isomorphic copy of H in the graph G is not robust to small perturbations of G (if 0.01 fraction of the edges of G are randomly flipped then the copy of H will be destroyed). Noise-robustness is an important assumption; in fact, in a handful of carefully chosen noise-free problems, low-degree methods are provably weaker than other brittle polynomial-time methods such as Gaussian elimination or lattice-basis reduction techniques [ZSWB22]. That being said, the existing techniques do not appear applicable to graph settings such as planted clique or the planted random subgraph model.

2 Our result

Let H be an r-uniform hypergraph over vertex set $[k]$ chosen uniformly at random (i.e., each r-hyperedge between the vertices of $[k]$ is included independently with probability half). Let $L \subseteq V(H)$ of size ℓ. Let $\mathcal{P}_{H,L}$ and $\mathcal{Q}_{H,L}$ be the following distributions over r-uniform hypergraphs G with vertex set $[n]$, where $n \geq k \geq \ell$:

1. In the *planted distribution* $\mathcal{P}_{H,L}$, an injective map $\phi\colon [k] \to [n]$ is chosen uniformly at random among all injective maps conditioned on $\phi(u) = u$ for $u \in L$. The hyperedges of G are

$$G(u_1, \ldots, u_r) = \begin{cases} H(\phi^{-1}(u_1), \ldots, \phi^{-1}(u_r)), & \text{if } \phi^{-1}(u_1), \ldots, \phi^{-1}(u_r) \text{ exist} \\ \text{a random bit}, & \text{otherwise.} \end{cases}$$

2. In the *null distribution* $\mathcal{Q}_{H,L}$, the hyperedges of G are

$$G(u_1, \ldots, u_r) = \begin{cases} H(u_1, \ldots, u_r), & \text{if } u_1, \ldots, u_r \in L \\ \text{a random bit}, & \text{otherwise.} \end{cases}$$

Uniform r-hypergraphs on n vertices are represented by their adjacency maps $\binom{[n]}{r} \to \{\pm 1\}$, with -1 and 1 representing the presence and absence of a hyperedge, respectively.

In words, the hypergraph $G \sim \mathcal{P}_{H,L}$ drawn from the planted model has the public hypergraph H embedded into a uniform choice of k vertices, and is otherwise purely random. However, the location of $L \subseteq V(H)$ is fixed and public information. The hypergraph $G \sim \mathcal{Q}_{H,L}$ drawn from the random model copies the subgraph of H on L, but it does not use the part of H outside of L; all remaining edges of the graph are chosen purely at random. Note that in both

models, the marginal distribution of G is a uniformly random hypergraph, but distinguishers know H and L.

In the case $r = 2$ of graphs, there is a slight difference between the distributions $\mathcal{P}_{H,L}, \mathcal{Q}_{H,L}$ and those described in the Introduction, namely that we have imposed the condition $\phi(u) = u$ on the leaked vertices in L. This condition is without loss of generality, and in particular, it does not affect the complexity of distinguishing $\mathcal{P}_{H,L}$ from $\mathcal{Q}_{H,L}$.

Following the low-degree framework [KWB19], we consider the degree-D-likelihood ratio $\mathcal{LR}_D(H, L)$,

$$\mathcal{LR}_D(H, L) = \sup_{\substack{p \in \mathbb{R}[G(\boldsymbol{u}) : \boldsymbol{u} \in \binom{[n]}{r}] \\ \deg p \le D}} \mathrm{Adv}_p(H, L)$$

where

$$\mathrm{Adv}_p(H, L) = \frac{\mathbf{E}_{\mathcal{P}_{H,L}}[p(G)] - \mathbf{E}_{\mathcal{Q}_{H,L}}[p(G)]}{\sqrt{\mathbf{Var}_{\mathcal{Q}_{H,L}}[p(G)]}}.$$

Here $p \in \mathbb{R}[G(\boldsymbol{u}) : \boldsymbol{u} \in \binom{[n]}{r}]$ denotes a multivariate polynomial in the quantities $G(u_1, \ldots, u_r)$ for $(u_1, \ldots, u_r) \in \binom{[n]}{r}$ with degree at most D. $\mathcal{LR}_D(H, L)$ measures the best advantage of a degree-D polynomial distinguisher that can arbitrarily preprocess H and knows L. Whenever $\mathcal{LR}_D(H, L) = o(1)$ then no D-degree polynomial can achieve strong separation between $\mathcal{P}_{H,L}$ and $\mathcal{Q}_{H,L}$ [COGHK+22, Section 7].

To gain intuition on the performance of low-degree polynomials, let us start with the simplest one, which is the bias of the edges of the hypergraph G:

$$p(G) = \sum_{1 \le u_1 < \cdots < u_r \le n} G(u_1, \ldots, u_r).$$

Assume for simplicity that $L = \emptyset$. It holds by direct expansion,

$$\begin{aligned} \mathbf{E}_{\mathcal{P}_H}[p(G)] &= \sum_{1 \le u_1 < \cdots < u_r \le k} H(u_1, \ldots, u_r) \\ \mathbf{E}_{\mathcal{Q}_H}[p(G)] &= 0 \\ \mathbf{Var}_{\mathcal{Q}_H}[p(G)] &= \binom{n}{r}. \end{aligned}$$

The likelihood ratio is

$$\mathrm{Adv}_p(H) = \Theta\left(\frac{\mathbf{E}_{\mathcal{P}_H}[p(G)]}{n^{r/2}}\right).$$

As $\mathbf{E}_{\mathcal{P}_H}[p(G)]$ is a sum of the $\binom{k}{r}$ hyperedge indicators for H, $\mathbf{E}_{\mathcal{P}_H}[p(G)]$ would have value $\pm\Theta(k^{r/2})$ for a typical choice of H, resulting in an advantage of $\Theta((k/n)^{r/2})$ (after optimizing between $p(G)$ or $-p(G)$). The advantage is $o(1)$ when $k \le n^{1-\Omega(1)}$ and therefore the distinguisher fails in this regime. Yet, when $k = \Theta(n)$ the calculation suggests the count distinguisher succeeds with $\Omega_r(1)$ probability which indeed can be confirmed by being a bit more careful in the above analysis. Our main theorem shows that other low-degree polynomials cannot substantially improve upon the edge-counting distinguisher.

Theorem 1. *Assume for some $p \in \mathbb{N}$ and constant $\epsilon > 0$, the following bounds hold on the size of H, k, the leakage number ℓ and the degree D:*

1. $k \le (n-\ell)n^{-\epsilon}/24p^2D^2 + \ell$
2. $\ell \le \min\{k,\ \epsilon^{1/(r-1)} r(\log n)^{1/(r-1)}/40\}$ *and,*
3. $D \le \epsilon^3 (\log n)^{r/(r-1)} / \left(\frac{r}{r-1}\log\log n\right)$.

Then for any $L \subseteq [k]$ with $|L| = \ell$,

$$\left(\mathbf{E}_H \mathcal{LR}_D(H,L)^{2p}\right)^{1/p} \le \frac{2^{\binom{\ell}{r-1}} n^{-\epsilon}}{1-n^{-\epsilon/2}} + \exp\left(-\Omega\left(r(\epsilon\log n)^{1+1/(r-1)}\right)\right).$$

In particular, for $p = 1$, $\ell = o((\log n)^{1/(r-1)})$, and $\epsilon = \Omega(1)$

$$\mathbf{E}_H \mathcal{LR}_D(H,L)^2 = n^{-\epsilon+o(1)}.$$

The bound is tight in the following ways:

1. **Degree:** The bound on D is optimal (for constant ϵ) up to a factor of $O(\log\log n)$. A degree-$O((r\log n)^{r/(r-1)})$ distinguisher with high advantage and time complexity $2^{O((r\log n)^{1/(r-1)})}$ exists. This is the algorithm that looks for the presence of a subgraph in G that is identical to the one induced by the first $O(r^{r/(r-1)}(\log n)^{1/(r-1)})$ vertices in H.
2. **Leakage:** When $\binom{\ell}{r-1} \ge \log(2n)$ the distinguishing advantage is constant (for any $k > \ell$). The distinguisher that looks for the existence of a vertex in G whose adjacencies in L match those of an arbitrary vertex in H outside L has constant advantage, degree $\binom{\ell}{r-1}$, and time complexity $O(n\binom{\ell}{r-1})$.
3. **Advantage:** The edge-counting distinguisher described above has advantage $(k/n)^{r/2} = n^{-\epsilon r/2}$. Our proof can show a matching lower bound in the absence of leakage. When leakage is present, assuming $\ell > r-1$, the linear distinguisher

$$\operatorname{sign}\sum_{v\notin L} G(1,\dots,r-1,v) = \operatorname{sign}\sum_{v\notin L} H(1,\dots,r-1,v)$$

has squared advantage $\Omega((k-\ell)/(n-\ell)) = \Omega(n^{-\epsilon})$ which matches the theorem statement.

2.1 Our proof

Abram et al. obtain their result as a consequence of a worst-case bound for arbitrary planted H: They prove that for all graphs H with $k \le n^{1/2-\epsilon}$ vertices,

$$\mathcal{LR}_D(H,L) \le o(1).$$

As $k = n^{1/2}$ is tight for clique their method cannot prove a better bound. In contrast, we average the likelihood ratio over the choice of H, showing that $\mathbf{E}_H[\mathcal{LR}_D(H,L)^2]$ is small all the way up to $k \le n^{1-\epsilon}$. By taking the expectation

over H, we introduce extra cancellations that are necessary to obtain the stronger bound.

By Markov's inequality

$$\mathbf{P}_H[\mathcal{LR}_D(H,L)^2 \geq \eta] \leq \frac{\mathbf{E}_H[\mathcal{LR}_D(H,L)^2]}{\eta}\,.$$

A vanishing expectation implies concentration, namely $\mathcal{LR}_D(H,L) = o(1)$ for a $1 - o(1)$ fraction of H.

The above calculation bounds the advantage for a fixed leakage set L. In order to bound the advantage of an arbitrary set L for the cryptographic applications, we also bound the higher moments of $\mathcal{LR}_D(H,L)$. Using $p = \ell \log n$ and applying Markov's inequality with $\eta = 4n^{-\epsilon+o(1)}$

$$\begin{aligned}\mathbf{P}_H[\mathcal{LR}_D(H,L)^2 \geq \eta] &\leq \frac{\mathbf{E}_H[\mathcal{LR}_D(H,L)^{2p}]}{\eta^p} \\ &\leq \left(\frac{n^{-\epsilon+o(1)}}{\eta}\right)^p \\ &= 4^{-\ell\log n} \leq \frac{1}{n\binom{n}{\ell}}\,.\end{aligned}$$

Taking a union bound over the $\binom{k}{\ell}$ choices for L, we can deduce the stronger result that no leakage set L can attain advantage η:

$$\mathbf{P}_H\left[\max_{\substack{L\subseteq V(H)\\|L|=\ell}} \mathcal{LR}_D(H,L)^2 \geq 4n^{-\epsilon+o(1)}\right] \leq o(1)\,.$$

We summarize the final bound on the low-degree advantage for Conjecture 2 as the following corollary, which includes the parameters.

Corollary 1. *For all $p \in \mathbb{N}$ and $\eta > 0$,*

$$\begin{aligned}&\mathbf{P}_H\left[\max_{\substack{L\subseteq V(H)\\|L|=\ell}} \mathcal{LR}_D(H,L)^2 \geq \eta\right] \\ &\qquad\leq \binom{n}{\ell}\eta^{-p}\left(\frac{2^{\binom{\ell}{r-1}}n^{-\epsilon}}{1-n^{-\epsilon/2}} + \exp\left(-\Omega\left(r(\epsilon\log n)^{1+1/(r-1)}\right)\right)\right)^p\,.\end{aligned}$$

3 Proof of Theorem 1

Viewed as an $\binom{n}{r}$-dimensional vector, every G in the support of $\mathcal{Q}_{H,L}$ decomposes as (G', G_L), where G_L is the subgraph of G on L and G' is the remaining part (indexed by r-subsets that have at least one vertex in $[n] \setminus L$).

We start by claiming that without loss of generality, all polynomial distinguishers of interest are constant in the coordinates of G_L. Indeed, in both the planted $\mathcal{P}_{H,L}$ and null distributions $\mathcal{Q}_{H,L}$, the status of the hyperedges in L is

always fixed. As fixing the L-indexed inputs can only lower the degree of the distinguishing polynomial p, this assumption holds without loss of generality.

In the null G' is simply uniformly random in $\{\pm 1\}^{\binom{[n]}{r}\setminus\binom{L}{r}}$, i.e., $\mathcal{Q}_{H,L}(G', G_L) = \mathcal{Q}(G')$, where $\mathcal{Q}$ is the uniform distribution. Now, let us focus on G' for the planted $\mathcal{P}_{H,L}$. We can describe the distribution $\mathcal{P}'_{H,L}$ of G' as follows:

1. Choose a random subset S' of $k-\ell$ vertices in $[n]\setminus L$.
2. Choose a random permutation $\pi'\colon S' \to [k]\setminus L$. Extend π' to a permutation from $S'\cup L$ to $[k]$ by setting $\pi'(u) = u$ for all $u \in L$.
3. Set

$$G'(u_1,\dots,u_r) = \begin{cases} H(\pi'(u_1),\dots,\pi'(u_r)), & \text{if } u_1,\dots,u_r \in S'\cup L \\ \text{a random bit}, & \text{otherwise.} \end{cases}$$

Using the above observations we have,

$$\begin{aligned} \mathcal{LR}_D(H,L) &= \sup_{\substack{p\in\mathbb{R}[G(\boldsymbol{u}):\boldsymbol{u}\in\binom{[n]}{r}] \\ \deg p\le D}} \frac{\mathbf{E}_{\mathcal{P}_{H,L}}[p(G)] - \mathbf{E}_{\mathcal{Q}_{H,L}}[p(G)]}{\sqrt{\mathbf{Var}_{\mathcal{Q}_{H,L}}[p(G)]}} \\ &= \sup_{\substack{p\in\mathbb{R}[G'(\boldsymbol{u}):\boldsymbol{u}\in\binom{[n]}{r}\setminus\binom{[\ell]}{r}] \\ \deg p\le D}} \frac{\mathbf{E}_{\mathcal{P}'_{H,L}}[p(G')] - \mathbf{E}_{\mathcal{Q}}[p(G')]}{\sqrt{\mathbf{Var}_{\mathcal{Q}}[p(G')]}} \end{aligned}$$

Since the null distribution $\mathcal{Q}$ is a product measure, by a standard linear algebraic argument in the literature of the low-degree method (see [KWB19] or [COGHK+22, Lemma 7.2]), the optimal degree-D polynomial takes an explicit form. Using the expansion with respect to the Fourier-Walsh basis $\{\chi_\alpha(G') = \prod_{e\in\alpha} G'_e, \alpha \subseteq \binom{[n]}{r}\setminus\binom{[\ell]}{r}\}$, the explicit formula for the squared advantage is

$$\mathcal{LR}_D(H,L)^2 = \sum_{\substack{\alpha\subseteq\binom{[n]}{r}\setminus\binom{L}{r} \\ 1\le|\alpha|\le D}} \widehat{\mathcal{LR}}(\alpha|H,L)^2 \tag{3}$$

where

$$\widehat{\mathcal{LR}}(\alpha|H,L) = \mathbf{E}_{\mathcal{Q}}\frac{\mathcal{P}'_{H,L}(G')}{\mathcal{Q}(G')}\chi_\alpha(G') = \mathbf{E}_{\mathcal{P}'_{H,L}}\chi_\alpha(G').$$

Now we expand the square on the right-hand side of (3) and take the expectation over H.

$$\mathbf{E}_H\mathcal{LR}_D(H,L)^2 = \sum_{\substack{\alpha\subseteq\binom{[n]}{r}\setminus\binom{L}{r} \\ 1\le|\alpha|\le D}} \mathbf{E}\chi_\alpha(G')\chi_\alpha(G''), \tag{4}$$

where the right-hand expectation is now taken over both the choice of H and the choice of two independent "replicas" G', G'' sampled from $\mathcal{P}'_H$. The joint distribution of G' and G'' is determined by the independent choices of H, the subsets S', S'', and the permutations π', π''. Equation (4) gives a formula for

the second moment of the likelihood ratio with respect to the random variable H, which we spend the rest of this section evaluating; higher moments will be computed later.

We fix $\alpha \subseteq \binom{[n]}{r} \setminus \binom{L}{r}$ and upper bound the expectation. Since we are considering the expectation of a Fourier character, it will often be zero. Let $V(\alpha)$ be the set of vertices in $[n]$ spanned by α. If $S' \cup L$ or $S'' \cup L$ does not entirely contain $V(\alpha)$ then the expectation is zero: if, say, $e \in \alpha'$ is not contained in $S' \cup L$, then $G'(e)$ is independent of all other bits appearing in $\mathbf{E}\chi_\alpha(G')\chi_\alpha(G'') = \prod_{e\in\alpha} G'(e)G''(e)$ resulting in a value of zero. Therefore

$$\begin{aligned}\mathbf{E}[\chi_\alpha(G')\chi_\alpha(G'')] &= \mathbf{E}[\chi_\alpha(G')\chi_\alpha(G'') \mid S' \cap S'' \supseteq V(\alpha) \setminus L] \cdot \mathbf{P}[S' \cap S'' \supseteq V(\alpha) \setminus L] \\ &= \mathbf{E}[\chi_\alpha(G')\chi_\alpha(G'') \mid S' \cap S'' \supseteq V(\alpha) \setminus L] \cdot \mathbf{P}[S' \supseteq V(\alpha) \setminus L]^2 \end{aligned} \tag{5}$$

by independence of S' and S''. As S' is a random k-subset of $[n] \setminus L$,

$$\begin{aligned}\mathbf{P}[S' \supseteq V(\alpha) \setminus L] &= \frac{(k-\ell)(k-\ell-1)\cdots(k-\ell-|V(\alpha)\setminus L|+1)}{(n-\ell)(n-\ell-1)\cdots(n-\ell-|V(\alpha)\setminus L|+1)} \\ &\leq \left(\frac{k-\ell}{n-\ell}\right)^{|V(\alpha)\setminus L|}. \end{aligned} \tag{6}$$

Conditioned on both S' and S'' containing $V(\alpha) \setminus L$,

$$\begin{aligned}\chi_\alpha(G')\chi_\alpha(G'') &= \prod_{(u_1,\ldots,u_r)\in\alpha} G'(u_1,\ldots,u_r)G''(u_1,\ldots,u_r) \\ &= \prod_{(u_1,\ldots,u_r)\in\alpha} H(\pi'(u_1),\ldots,\pi'(u_r))H(\pi''(u_1),\ldots,\pi''(u_r)). \end{aligned} \tag{7}$$

As H consists of i.i.d. zero mean ± 1 entries, this expression vanishes in expectation unless every hyperedge in the collection

$$(\psi(u_1),\ldots,\psi(u_r))\colon (u_1,\ldots,u_r) \in \alpha, \psi \in \{\pi', \pi''\}$$

appears exactly twice, in which case the product equals to one. This is only possible if $\pi\colon S' \to S''$ given by $\pi = (\pi'')^{-1} \circ \pi'$ restricts to an automorphism of α. In particular, π must fix the set $V(\alpha)$. As π outside L is a permutation which is chosen uniformly at random, we conclude that (7) is upper bounded by,

$$\begin{aligned}\mathbf{P}[\pi \text{ fixes } V(\alpha)] &= \frac{|V(\alpha)\setminus L|!}{(k-\ell)(k-\ell-1)\cdots(k-\ell-|V(\alpha)\setminus L|+1)} \\ &\leq \left(\frac{|V(\alpha)\setminus L|}{k-\ell}\right)^{|V(\alpha)\setminus L|}. \end{aligned} \tag{8}$$

Plugging (6) and (8) into (5) and then into (4) yields

$$\mathbf{E}\mathcal{LR}_D(H,L)^2 \leq \sum_{\substack{\alpha\subseteq\binom{[n]}{r}\setminus\binom{L}{r} \\ 1\leq|\alpha|\leq D}} \left(\frac{|V(\alpha)\setminus L|\,(k-\ell)}{(n-\ell)^2}\right)^{|V(\alpha)\setminus L|}. \tag{9}$$

This bound only depends on the hypergraph α through $|V(\alpha) \setminus L|$. For $v = 1, \ldots, rD$ let

$$N(v, D) = \left|\left\{\alpha \subseteq \tbinom{[n]}{r} \setminus \tbinom{L}{r} : |V(\alpha) \setminus L| = v,\ 1 \le |\alpha| \le D\right\}\right|. \tag{10}$$

Grouping the terms on the right-hand side by the value of $v = |V(\alpha) \setminus L|$ gives

$$\mathbf{E}\mathcal{LR}_D(H)^2 \le \sum_{v=1}^{rD} N(v, D) \cdot \left(\frac{v(k-\ell)}{(n-\ell)^2}\right)^v. \tag{11}$$

To finish the proof we will demonstrate that this sum is dominated by the leading term $v = 1$. We split this proof using the following two propositions.

In the first proposition, we bound the "low" vertex size part.

Proposition 1. *Assume that $e(k-\ell)/(n-\ell) \le n^{-\epsilon}$. Then for every $0 < \delta < \epsilon$ it holds for sufficiently large n,*

$$\sum_{v=1}^{\lfloor t \rfloor} N(v, D)\left(\frac{v(k-\ell)}{(n-\ell)^2}\right)^v \le 2^{\binom{\ell}{r-1}} \cdot \frac{n^{-\epsilon}}{1 - n^{-\epsilon+\delta}},$$

where

$$t := e^{-1}(r-1)(\delta \log n)^{1/(r-1)} - \ell \tag{12}$$

In the second proposition, we bound the "high" vertex size part.

Proposition 2. *Assume that $e(k-\ell)/(n-\ell) \le n^{-\epsilon}$. Assume also that for some $\delta > 0$ for which $0 < \delta < \epsilon$, it holds*

1. $\ell \le (r/9)(\delta \log n)^{1/(r-1)}$
   and,
2. $D \le \epsilon\delta^2 (\log n)^{r/(r-1)} / \left(\frac{r}{r-1} \log\log n\right)$.

Then for t given in (12) if also $\delta < 1/4$ it holds,

$$\sum_{v=\lfloor t \rfloor + 1}^{rD} N(v, D)\left(\frac{v(k-\ell)}{(n-\ell)^2}\right)^v \le \exp\left(-\Omega(\delta^{1/(r-1)} \epsilon r (\log n)^{r/(r-1)})\right).$$

Notice now that directly combining both the Propositions for $\delta = \epsilon/4$ directly implies Theorem 1.

3.1 Proof of Proposition 1

Proof. For fixed v, the set $V(\alpha) \setminus L$ can be chosen in $\binom{n-\ell}{v}$ ways. The subset α can then include any of the hyperedges in $V(\alpha)$ of which there are at most $\binom{v+\ell}{r}$,

except those that at completely contained in L of which there are $\binom{\ell}{r}$, leading to the bound:

$$N(v, D) \leq \binom{n-\ell}{v} \cdot 2^{\binom{v+\ell}{r}-\binom{\ell}{r}}. \tag{13}$$

Bounding $N(v, D)$ by (13) and using the standard binomial coefficient bound $\binom{a}{b} \leq (ea/b)^b$, the left hand side is at most

$$\sum_{v=1}^{t} \left(\frac{e(k-\ell)}{n-\ell} \right)^v 2^{\binom{v+\ell}{r}-\binom{\ell}{r}}$$

As $e(k-\ell)/(n-\ell) \leq n^{-\epsilon} = n^{-\epsilon+\delta} \cdot n^{\delta}$, this is bounded by

$$\sum_{v=1}^{t} n^{-(\epsilon-\delta)v} \cdot 2^{-\delta v \log_2 n + \binom{v+\ell}{r}-\binom{\ell}{r}}. \tag{14}$$

Let $f(v) = -\delta v \log_2 n + \binom{v+\ell}{r} - \binom{\ell}{r}, v \geq 1$. For all integer $v \geq 1$,

$$f(v+1) - f(v) = -\delta \log_2 n + \binom{v+\ell}{r-1} \leq -\delta \log n + \left(\frac{e(v+\ell)}{r-1} \right)^{r-1}.$$

By the definition of t, this is negative when $1 \leq v \leq t$, so $f(v)$ is maximized at $v = 1$. Therefore (14) is at most

$$\sum_{v=1}^{\lfloor t \rfloor} n^{-(\epsilon-\delta)v} \cdot 2^{-\delta \log n + \binom{\ell+1}{r}-\binom{\ell}{r}} \leq 2^{\binom{\ell}{r-1}} \cdot \frac{n^{-\epsilon}}{1 - n^{-\epsilon+\delta}}$$

using the identity $\binom{\ell+1}{r} - \binom{\ell}{r} = \binom{\ell}{r-1}$ and the geometric sum formula. □

3.2 Proof of Proposition 2

Proof. When v is large, the bound (13) can be improved by taking into account that at most D of the hyperedges can be chosen:

$$\begin{aligned} N(v, D) &\leq \binom{n-\ell}{v} \cdot D \binom{\binom{v+\ell}{r} - \binom{\ell}{r}}{D} \\ &\leq D \left(\frac{e(n-\ell)}{v} \right)^v \cdot \left(\frac{e\binom{v+\ell}{r}}{D} \right)^D \\ &\leq \left(\frac{e(n-\ell)}{v} \right)^v \cdot \left(\frac{e(v+\ell)}{r} \right)^{rD} \cdot D \left(\frac{e}{D} \right)^D. \end{aligned}$$

Under the assumption $e(k-\ell)/(n-\ell) \le n^{-\epsilon}$ the summation of interest is at most

$$\sum_{v=t+1}^{rD} \left(\frac{e(k-\ell)}{n-\ell}\right)^v \cdot \left(\frac{e(v+\ell)}{r}\right)^{rD} \cdot D\left(\frac{e}{D}\right)^D$$
$$\le rD^2\left(\frac{e}{D}\right)^D \cdot n^{-\epsilon t}\left(\frac{e(rD+\ell)}{r}\right)^{rD}$$
$$\le rD^2\left(\frac{e}{D}\right)^D \cdot n^{-\epsilon t}\big(e(D+\ell)\big)^{rD}.$$

As $D \le \epsilon\delta^2(\log n)^{r/(r-1)}/(\frac{r}{r-1}\log\log n)$ and $\ell \le (r/9)(\delta\log n)^{1/(r-1)}$, for sufficiently large n,

$$D\log\big((D+\ell)/(\epsilon\delta^2)\big) \le \epsilon\delta^2(\log n)^{r/(r-1)},$$

Hence, for sufficiently small constant $0<\delta<1$, for sufficiently large n it holds

$$D\log\left(e(D+\ell)\right)) \le \epsilon\delta^2(\log n)^{r/(r-1)},$$

Using also the elementary inequality $D^2(e/D)^D \le 8$ we conclude that the summation of interest is at most

$$8rn^{-\epsilon t}\exp\left(\epsilon\delta^2(\log n)^{r/(r-1)}\right).$$

Plugging in the direct bound from the definition of t and the upper bound on the leaked vertices,

$$t \ge \frac{r}{7}(\delta\log n)^{1/(r-1)}$$

we conclude that the summation of interest is at most

$$8r\exp\left(-\epsilon\frac{r-1}{e}\delta^{1/(r-1)}(\log n)^{r/(r-1)} + \epsilon\delta^2(\log n)^{r/(r-1)}\right).$$

Choosing now $\delta < 1/4$ concludes the result. □

3.3 Extension to higher moments

Now we extend the calculation in Theorem 1 from $p=1$ to higher p. The $2p$-th moment of $\mathcal{LR}_D(H,L)$ is

$$\mathbf{E}_H\mathcal{LR}_D(H,L)^{2p} = \mathbf{E}_H\left(\sum_{\substack{\alpha\subseteq\binom{[n]}{r}\setminus\binom{L}{r}\\ 1\le|\alpha|\le D}} \mathbf{E}_{\substack{G'\sim\mathcal{P}'_H\\ G''\sim\mathcal{P}'_H}}\chi_\alpha(G')\chi_\alpha(G'')\right)^p$$
$$= \sum_{\substack{\alpha_1,\dots,\alpha_p\subseteq\binom{[n]}{r}\setminus\binom{L}{r}\\ 1\le|\alpha_i|\le D}} \mathbf{E}\prod_{i=1}^p \chi_{\alpha_i}(G'_i)\chi_{\alpha_i}(G''_i)$$

where the expectation is over H and also over the replicas G'_i, G''_i sampled independently from $\mathcal{P}'_H$. Each G'_i is equivalently sampled as S'_i and π'_i (and likewise G''_i as S''_i and π''_i).

Fix the Fourier characters $\alpha_1, \dots, \alpha_p$ and let $V(\alpha_i)$ be the set of vertices in $[n]$ spanned by α_i. First, the expectation is only nonzero if all of the sets S'_i and S''_i contain $V(\alpha_i) \setminus L$. By (6) this occurs with probability at most

$$\mathbf{P}\left[\forall i \in [p].\ S'_i \cap S''_i \supseteq V(\alpha_i) \setminus L\right] \le \left(\frac{k-\ell}{n-\ell}\right)^{2\sum_{i=1}^p |V(\alpha_i)\setminus L|}. \tag{15}$$

Conditioned on this event,

$$\prod_{i=1}^p \chi_{\alpha_i}(G'_i)\chi_{\alpha_i}(G''_i) = \prod_{i=1}^p \chi_{\pi'_i(\alpha_i)}(H)\chi_{\pi''_i(\alpha_i)}(H)\,.$$

When the expectation is taken over H, this is only nonzero if every hyperedge appears an even number of times among the collection of edges

$$C := (\psi_i(u_1), \dots, \psi_i(u_r)) : i \in [p],\ (u_1, \dots, u_r) \in \alpha_i,\ \psi_i \in \{\pi'_i, \pi''_i\}\,.$$

In order for this to occur, every vertex in the image of the ψ_i must be in the image of at least two ψ_i. Let us say that the collection of embeddings is a *double cover* if this occurs. Then

$$\begin{aligned}
&\mathbf{E}_{H,\pi'_i,\pi''_i} \prod_{i=1}^p \chi_{\pi'_i(\alpha_i)}(H)\chi_{\pi''_i(\alpha_i)}(H) \\
&= \mathbf{P}_{\pi'_i,\pi''_i}[C \text{ is an even collection}] \\
&\le \mathbf{P}_{\pi'_i,\pi''_i}[(\pi'_i, \pi''_i)_{i\in[p]} \text{ is a double cover}]\,.
\end{aligned} \tag{16}$$

Let $V = \sum_{i=1}^p |V(\alpha_i) \setminus L|$. We claim

$$\mathbf{P}_{\pi'_i,\pi''_i}[(\pi'_i, \pi''_i)_{i\in[p]} \text{ is a double cover}] \le \frac{(2V)^{2V}}{(k-\ell)(k-\ell-1)\cdots(k-\ell-V+1)}\,. \tag{17}$$

This is based on the following surjection a.k.a union bound. The total number of vertices mapped by all the permutations is $2V$. We take any partition of the $2V$ vertices such that every block of the partition has size at least two. There are at most $(2V)^{2V}$ such partitions. We go through the vertices in some fixed order, and for each vertex which is not the first member of its block of the partition, we obtain a factor of $\approx \frac{1}{k-\ell}$ for the probability that the vertex is mapped to the same element as the other members of its block of the partition. Since the blocks have size at least two (in order to be a double cover), we obtain at least V factors of $\approx \frac{1}{k-\ell}$ in this way. We upper bound $\approx \frac{1}{k-\ell}$ by a rising factorial to obtain the bound in (17).

If $V \le \frac{k-\ell}{2}$, then (17) can simplified to

$$\frac{(2V)^{2V}}{(k-\ell)(k-\ell-1)\cdots(k-\ell-V+1)} \le \left(\frac{8V^2}{k-\ell}\right)^V. \tag{18}$$

On the other hand, if $V \geq \frac{k-\ell}{2}$, then the right-hand side is at least 1. Combining these two possible cases, we conclude,

$$\mathbf{P}_{\pi_i',\pi_i''}[(\pi_i',\pi_i'')_{i\in[p]} \text{ is a double cover}] \leq \left(\frac{8V^2}{k-\ell}\right)^V. \tag{19}$$

Now we return to the main calculation of $\mathbf{E}_H \mathcal{LR}_D(H,L)^{2p}$. Combining (15), (19),

$$\begin{aligned}
\mathbf{E}_H \mathcal{LR}_D(H,L)^{2p} &= \sum_{\substack{\alpha_1,\dots,\alpha_p \subseteq \binom{[n]}{r}\setminus\binom{L}{r} \\ 1\leq|\alpha_i|\leq D}} \mathbf{E}\prod_{i=1}^{p}\chi_{\alpha_i}(G_i')\chi_{\alpha_i}(G_i'') \\
&\leq \sum_{\substack{\alpha_1,\dots,\alpha_p \subseteq \binom{[n]}{r}\setminus\binom{L}{r} \\ 1\leq|\alpha_i|\leq D}} \left(\frac{8V^2(k-\ell)}{(n-\ell)^2}\right)^V \\
&\leq \sum_{\substack{\alpha_1,\dots,\alpha_p \subseteq \binom{[n]}{r}\setminus\binom{L}{r} \\ 1\leq|\alpha_i|\leq D}} \left(\frac{8p^2D^2(k-\ell)}{(n-\ell)^2}\right)^{\sum_{i=1}^p |V(\alpha_i)\setminus L|} \qquad (V \leq pD) \\
&= \left(\sum_{\substack{\alpha \subseteq \binom{[n]}{r}\setminus\binom{L}{r} \\ 1\leq|\alpha|\leq D}} \left(\frac{8p^2D^2(k-\ell)}{(n-\ell)^2}\right)^{|V(\alpha)\setminus L|}\right)^p
\end{aligned}$$

The inner summation is nearly the combinatorial quantity we bounded in Equation (9) when computing $\mathbf{E}_H \mathcal{LR}_D(H,L)^2$. The only difference is the factor $8p^2D^2$ which may be larger than what we had before. This factor can be negated by scaling down $\frac{k-\ell}{n-\ell}$. Using the same counting arguments as before with the slightly stronger assumption on k, we conclude the desired moment bound.

4 Cryptographic applications

4.1 Hypergraph secret sharing

The secret sharing scheme of Abram et al. was stated for forbidden graph access structures. The construction extends to partial access structures (R,S) where R is a collection of r-subsets and S consists of all independent sets of R of size at most ℓ.

Construction 2. Forbidden hypergraph secret sharing: Syntactically replace "graph" by "r-uniform hypergraph" and (u,v) by $(u_1,\dots,u_r)$ in Construction 1.

This scheme reconstructs all $\{u_1,\dots,u_r\} \in R$ by (2).

Proposition 3. *Assume $(H,\mathcal{P}_{H,L})$ and $(H,\mathcal{Q}_{H,L})$ are (s,ϵ)-indistinguishable for all $L \subseteq V(H)$ with $|L| = \ell$. Then for every independent set $I \subseteq R$ of size at most ℓ, shares of 0 and 1 are $(s,2\epsilon)$-indistinguishable by parties in I.*

Proof. Assume parties in I can 2ϵ-distinguish shares of 0 and 1 using distinguisher D. By the triangle inequality, D ϵ-distinguishes $(H_s, G, \phi(i) : i \in I)$ from $(H, G, \phi(i) : i \in I)$ where

$$G(u_1, \ldots, u_r) = \begin{cases} H(u_1, \ldots, u_r), & \text{if } u_1, \ldots, u_r \in I \\ \text{a random bit}, & \text{otherwise.} \end{cases}$$

for at least one value of s. Let D' be the circuit that, on input $(H', G, u_i : i \in I)$, outputs $D(H' \oplus sR, G, u_i : i \in I)$. As R does not contain any hyperedges within I, by (1), $D'(\mathcal{P}_{H,I})$ is identically distributed to $D(H_s, G, \phi(i) : i \in I)$. As H is random, $D'(\mathcal{Q}_{H,I})$ is identically distributed to $D(H, G, \phi(i) : i \in I)$. Therefore D' and D have the same advantage. □

The class of access structures can be expanded to allow the reconstruction set R to consist of arbitrary sets, as long as the size of all minimal sets is at most r. This is accomplished by a reduction to size exactly r. Let $R' \subseteq [n + r - 1]$ be the r-uniform hypergraph

$$R' = \big\{A \cup \{n+1, \ldots, n+r-|A|\} : A \in R\big\}.$$

Construction 3. Apply Construction 2 to R' with the shares of parties $n + 1, \ldots, n + r - 1$ made public.

If all sets in R' can resconstruct in Construction 2 then all sets in R can reconstruct in Construction 3. As for secrecy, if Construction 2 is secure against all independent sets in R of size at most ℓ, then Construction 3 is secure against such sets of size at most $\ell - r + 1$.

Could Construction 2 give a *provable* separation between the minimum share size of information-theoretic and computational secret sharing? We argue that this is unlikely barring progress in information-theoretic secret sharing lower bounds. The share size in Construction 2 is $(1 + \Omega(1))(\log n)$. However, the share size lower bounds of [KN90,BGK20] do not exceed $\log n$ for any known n-party access structure.

In contrast, Csirmaz [Csi97] proved that there exists an n-party access structure with share size $\Omega(n/\log n)$. Using Csirmaz's method, Beimel [Bei23] constructed *total* r-hypergraph access structures that require share size $\Omega(n^{2-1/(r-1)}/r)$ for every $r \geq 3$.

We argue that Csirmaz's method cannot prove a lower bound exceeding ℓ for any (partial) access structures in which secrecy is required to hold only for sets of size up to ℓ. Csirmaz showed that a scheme with share size s implies the existence of a monotone submodular function f (the joint entropy of the shares in A) from subsets of $\{1, \ldots, n\}$ to real numbers that satisfies the additional constraints

$$f(A) + f(B) \geq f(A \cup B) + f(A \cap B) + 1 \quad \text{if } A, B \in S \text{ and } A \cup B \in R \tag{20}$$

$$f(A) \leq s \quad \text{for all } A \text{ of size } 1. \tag{21}$$

Proposition 4. *Assuming all sets in S have size at most ℓ, there exists a monotone submodular function satisfying (20) and (21) with $s = \ell$.*

As our scheme does not tolerate $\Omega(\log n)$ bits of leakage, the best share size lower bound that can be proved using Csirmaz's relaxation of secret sharing is $\ell = o(\log n)$. The proof of Proposition 4 is a natural generalization of [Csi97, Theorem 3.5] to partial access structures.

Proof (Proposition 4). The function $f(A) = \sum_{t=1}^{|A|} \max\{\ell - t + 1, 0\}$ is monotone, submodular, satisfies (20) for every $R \subseteq \overline{S}$, and (21) with $s = \ell$. □

4.2 Multiparty PSM for random functions

Given a function $F\colon [k]^r \to \{\pm 1\}$, the random hypergraph embedding of F is the r-hypergraph $\overline{F}$ on rk vertices $(x, i)\colon x \in [k], i \in [r]$ such that

$$\overline{F}((x_1, 1), \ldots, (x_r, r)) = F(x_1, \ldots, x_r).$$

All other potential hyperedges of $\overline{F}$ are sampled uniformly and independently at random.

We describe the r-partite generalization of Abram et al.'s PSM protocol. Let $\phi\colon [k] \times [r] \to [n]$ be a random injection and let G be the r-hypergraph on n vertices given by

$$G(u_1, \ldots, u_r) = \begin{cases} \overline{F}(\phi^{-1}(u_1), \ldots, \phi^{-1}(u_r)), & \text{if } \phi^{-1}(u_1), \ldots, \phi^{-1}(u_r) \text{ exist} \\ \text{a random bit}, & \text{otherwise.} \end{cases}$$

Construction 4. r-party PSM protocol for F:

In the setup phase, G is published and ϕ is privately given to the parties.
In the evaluation phase,

1. Party i is given input x_i.
2. Party i forwards $u_i = \phi(x_i, i)$ to the evaluator.
3. The evaluator outputs $G(u_1, \ldots, u_r)$.

The protocol is clearly functional. A reasonable notion of security with respect to random functions F should allow the parties' input choices to depend on F. An *input selector* is a randomized function I that, on input F, produces inputs $I(F) = (x_1, \ldots, x_r)$ for the r parties.

We say a protocol is (s', s, ϵ) (simulation) secure against a random function if for every input selector I there exists a size-s' simulator S for which the distributions

$$(F, G, \phi(x_1, 1), \ldots, \phi(x_r, r)) \quad \text{and} \quad (F, S(F, F(x_1, \ldots, x_r))) \tag{22}$$

are (s, ϵ)-indistinguishable, where $(x_1, \ldots, x_r)$ is the output of $I(F)$.

Proposition 5. *Assume* $(H, G, L(H))$ *with* $G \sim \mathcal{P}_{H,L(H)}$ *versus* $G \sim \mathcal{Q}_{H,L(H)}$ *are* (s, ϵ)*-indistinguishable with parameters* $|V(H)| = kr$, $|V(G)| = n$, *and* $\ell = r$. *Then Construction 4 is* $(O(\binom{n}{r}), s - O(\binom{n}{r}), \epsilon)$*-secure.*

We label the vertices of H by pairs $(x, r) \in [k] \times [r]$.

Proof. On input (F, y), the simulator S

1. chooses random $u_1, \ldots, u_r \in [n]$
2. sets $G(u_1, \ldots, u_r) = y$
3. samples all other possible hyperedges of G independently at random
4. outputs $(G, u_1, \ldots, u_r)$.

We describe a reduction R that, given a distinguisher D for (22), tells apart $(H, G, L(H))$ with $G \sim \mathcal{P}_{H,L(H)}$ versus $G \sim \mathcal{Q}_{H,L(H)}$ for some leakage function L. On input $(H, G, z_1, \ldots, z_r)$,

1. set F to be the function $F(x_1, \ldots, x_r) = H((x_1, 1), \ldots, (x_r, r))$
2. output $(F, \pi(G), \pi(z_1), \ldots, \pi(z_r))$ for a random permutation π on $[n]$ (which acts on G as a hypergraph isomorphism).

Let L be the leakage function that, on input H, runs $I(F)$ to obtain $(x_1, \ldots, x_r)$, and outputs $((x_1, 1), \ldots, (x_r, r))$.

This reduction preserves distinguishing advantage as it maps the distributions (22) into the distributions $(H, \mathcal{P}_{H,L(H)}, L(H))$ and $(H, \mathcal{Q}_{H,L(H)}, L(H))$, respectively. It can be implemented in size $O(\binom{n}{r})$, giving the desired parameters. □

References

ABF+19. Benny Applebaum, Amos Beimel, Oriol Farràs, Oded Nir, and Naty Peter. Secret-sharing schemes for general and uniform access structures. In *Advances in Cryptology – EUROCRYPT 2019: 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19–23, 2019, Proceedings, Part III*, page 441–471, 2019. 4

ABI+23. Damiano Abram, Amos Beimel, Yuval Ishai, Eyal Kushilevitz, and Varun Narayanan. Cryptography from planted graphs: security with logarithmic-size messages. In *Theory of Cryptography Conference*, pages 286–315. Springer, 2023. 2, 4, 6

AHMS18. Benny Applebaum, Thomas Holenstein, Manoj Mishra, and Ofer Shayevitz. The communication complexity of private simultaneous messages, revisited. In *EUROCRYPT (2)*, pages 261–286. Springer, 2018. 5

AKS98. Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. *Random Structures & Algorithms*, 13(3-4):457–466, 1998. 1

BB20. Matthew Brennan and Guy Bresler. Reducibility and statistical-computational gaps from secret leakage. In Jacob Abernethy and Shivani Agarwal, editors, *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 648–847. PMLR, 09–12 Jul 2020. 2

Bei23. Amos Beimel. Lower bounds for secret-sharing schemes for k-hypergraphs. In *4th Conference on Information-Theoretic Cryptography (ITC 2023)*, volume 267, pages 16:1–16:13, 2023. 17

BGK20. Andrej Bogdanov, Siyao Guo, and Ilan Komargodski. Threshold secret sharing requires a linear-size alphabet. *Theory of Computing*, 16(2):1–18, 2020. 17

BHK+19. Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K. Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019. 1

CMZ23. Zongchen Chen, Elchanan Mossel, and Ilias Zadik. Almost-linear planted cliques elude the Metropolis process. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 4504–4539. SIAM, 2023. 2

COGHK+22. Amin Coja-Oghlan, Oliver Gebhard, Max Hahn-Klimroth, Alexander S Wein, and Ilias Zadik. Statistical and computational phase transitions in group testing. In *Conference on Learning Theory*, pages 4764–4781. PMLR, 2022. 5, 7, 10

Csi97. László Csirmaz. The size of a share must be large. *Journal of Cryptology*, 10(4):223–231, 1997. 17, 18

FGR+17. Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh S. Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. *J. ACM*, 64(2), apr 2017. 2

FK08. Alan Frieze and Ravi Kannan. A new approach to the planted clique problem. In *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science*, 2008. 1

FKN94. Uri Feige, Joe Kilian, and Moni Naor. A minimal model for secure computation (extended abstract). In *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing*, STOC '94, page 554–563, New York, NY, USA, 1994. Association for Computing Machinery. 5

GZ19. David Gamarnik and Ilias Zadik. The landscape of the planted clique problem: Dense subgraphs and the overlap gap property. *arXiv preprint arXiv:1904.07174*, 2019. 2

Hop18. Samuel Hopkins. *Statistical Inference and the Sum of Squares Method*. PhD thesis, Cornell University, 2018. 2, 5, 6

HS24. Shuichi Hirahara and Nobutaka Shimizu. Planted clique conjectures are equivalent. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC 2024, page 358–366, 2024. 2

Hul22. Wasim Huleihel. Inferring hidden structures in random graphs. *IEEE Transactions on Signal and Information Processing over Networks*, 8:855–867, 2022. 2, 3

HW21. Justin Holmgren and Alexander S. Wein. Counterexamples to the Low-Degree Conjecture. In *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*, volume 185, pages 75:1–75:9, 2021. 5

Jer92. Mark Jerrum. Large cliques elude the metropolis process. *Random Struct. Algorithms*, 3:347–360, 1992. 1, 2

JP97. Ari Juels and Marcus Peinado. Hiding cliques for cryptographic security. *Designs Codes and Cryptography*, 20, 11 1997. 2

KN90. Joe Kilian and Noam Nisan. Unpublished, 1990. 17

Kuc95. Luděk Kucera. Expected complexity of graph partitioning problems. *Discrete Appl. Math.*, 57(2–3):193–212, Feb. 1995. 1

KWB19. Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *ISAAC Congress (International Society for Analysis, its Applications and Computation)*, pages 1–50. Springer, 2019. 5, 7, 10

LVW17. Tianren Liu, Vinod Vaikuntanathan, and Hoeteck Wee. Conditional disclosure of secrets via non-linear reconstruction. In *CRYPTO*, pages 758–790. Springer, 2017. 4

MNWS$^+$23. Elchanan Mossel, Jonathan Niles-Weed, Youngtak Sohn, Nike Sun, and Ilias Zadik. Sharp thresholds in inference of planted subgraphs. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 5573–5577. PMLR, 2023. 2

YZZ24. Xifan Yu, Ilias Zadik, and Peiyuan Zhang. Counting stars is constant-degree optimal for detecting any planted subgraph. *arXiv preprint arXiv:2403.17766*, 2024. 2

ZSWB22. Ilias Zadik, Min Jae Song, Alexander S Wein, and Joan Bruna. Lattice-based methods surpass sum-of-squares in clustering. In *Conference on Learning Theory*, pages 1247–1248. PMLR, 2022. 6