AICR · OPEN PROBLEM

A Character-Sum Conjecture for Planted Subgraphs

$\sum_{\pi \in \mathrm{Sym}(n)} (-1)^{\langle \pi \circ H + H, \alpha \rangle}$ *is small, for most* H *and every* α

Status. Statement: AI-written from Damiano Abram, Amos Beimel and Yuval Ishai, *Cryptography from Planted Graphs: Security with Logarithmic-Size Messages*, IACR ePrint 2023/1929. Not yet checked by a human. The source derives the exact formula that governs low-degree detection of a planted subgraph, observes that the character sum in it should be small for most graphs, and says in as many words that it leaves the rigorous study of the question to future work. No bound is proved and none is stated as a target — see Remark 1.

Category. *Average-Case Hardness.*

Conjecture (informal). *Hide a small graph* H *on* n *nodes inside a random graph on* N *nodes and ask how hard it is to notice. The source derives a formula for exactly how well a low-degree polynomial can do, and inside that formula sits a sum over all permutations of the* n *nodes: for each permutation, compare the permuted copy of* H *with* H *itself, and take the parity of their disagreement on a chosen set of edges. If those parities come out even and odd about equally often, the sum nearly cancels and* H *is hard to detect; if* H *is highly structured — a clique, an independent set — every term is* $+1$ *and the sum is as large as it can be. The conjecture is that cancellation is the typical case: for most graphs* H*, the sum is small for every choice of edge set.*

1 The setting

The source builds cryptography from planted-graph assumptions, obtaining private simultaneous messages protocols and secret-sharing schemes with public information whose messages or shares are only $O(\log n)$ bits — where the information-theoretic bounds are $\Omega(\log n)$ and, for forbidden-graph secret sharing, $2^{\widetilde{O}(\sqrt{\log n})}$. The

security of those constructions rests on the difficulty of detecting a planted n-node subgraph H inside a random N-node graph, and the source studies that difficulty against the natural class of attacks: low-degree polynomial distinguishers.

Its Theorem 4.7 yields a formula measuring how easily a degree-D polynomial detects a planted H. With no hints given to the distinguisher, the quantity is

$$\sum_{0<w(\alpha)\le D} \binom{N-V(\alpha)}{n-V(\alpha)}^2 \cdot \sum_{\pi\in \mathrm{Sym}(n)} (-1)^{\langle \pi\circ H+H,\alpha\rangle},$$

and the lower its value, the harder detection becomes. Two consequences the source draws from it are worth keeping in view: cliques are the easiest subgraph to detect, independently of N, t and D; and the argument stops working once D reaches $\log^2 n$, which is no accident, since the largest clique in a random N-node graph has at most $2\log n$ nodes with high probability.

The inner sum is the object of this statement. The source observes that the vectors $(\pi\circ H+H)_{\pi\in\mathrm{Sym}(n)}$ describe how *structured* H is — if H is a clique or an independent set all of them are zero, every term of the sum is +1, and the sum attains its maximum $|\mathrm{Sym}(n)|$ — and then says: "We conjecture that, for most graphs H, the sum $\sum_{\pi\in\mathrm{Sym}(n)}(-1)^{\langle\pi\circ H+H,\alpha\rangle}$ should be small for all choices of α, as $\langle\pi\circ H+H,\alpha\rangle$ should assume the values 0 and 1 almost equally often. We leave the rigorous study of this problem to future work."

2 Notation and parameters

- n is the number of nodes of the planted graph H; N the number of nodes of the ambient random graph, in the application.
- A graph on n nodes is encoded as a vector of $\binom{n}{2}$ bits, one per potential edge; H is such a vector. Sums of graphs are coordinatewise over $\mathbb{F}_2$.
- α is a subset of the edges of an n-node graph, encoded the same way; $w(\alpha)$ is the number of edges in it and $V(\alpha)$ the total number of nodes they touch.
- $\langle\cdot,\cdot\rangle$ is the inner product of two such vectors over $\mathbb{F}_2$, so $(-1)^{\langle\cdot,\cdot\rangle}\in\{+1,-1\}$.

- $\pi \circ H$ is the graph obtained by permuting the *nodes* of H — not the edges — according to $\pi \in \mathrm{Sym}(n)$, encoded again as a vector.
- D is the degree of the polynomial distinguisher, in the application; the sum below does not involve it.

3 The conjecture

Definition 1 (The permutation character sum). For a graph H on n nodes and an edge set α, both encoded as vectors of $\binom{n}{2}$ bits, put

$$S(H,\alpha) \;:=\; \sum_{\pi \in \mathrm{Sym}(n)} (-1)^{\langle \pi \circ H + H,\, \alpha \rangle}.$$

Note $|S(H,\alpha)| \le |\mathrm{Sym}(n)|$ always, with equality at $\alpha = 0$ and whenever $\pi \circ H = H$ for every π — in particular for H a clique or an independent set.

Conjecture 1 (Typical graphs have small character sums). *There is a function $\eta(n) = o(1)$ such that, for a $1 - o(1)$ fraction of the graphs H on n nodes,*

$$\max_{\alpha \neq 0} \left|S(H,\alpha)\right| \;\le\; \eta(n) \cdot |\mathrm{Sym}(n)|.$$

Remark 1 (The source states no target, and this one is supplied). The source says the sum "should be small" and does not quantify "small", for most H and all α. Conjecture 1 therefore supplies a target — a $o(1)$ fraction of the trivial maximum $|\mathrm{Sym}(n)|$ — and that choice is this statement's, not the source's. It is the weakest reading under which the assertion has content, since $|\mathrm{Sym}(n)|$ is attained by the structured graphs the source contrasts against. Anyone resolving it should report the bound achieved: a stronger conclusion such as $|S(H,\alpha)| \le |\mathrm{Sym}(n)| \,/\, 2^{\Omega(w(\alpha))}$, or a bound in terms of $V(\alpha)$, would be more useful to the application and should be stated as such rather than folded into this form.

Remark 2 (Why "for most H" cannot be dropped). The conjec-

ture is false for individual graphs, and visibly so: for H a clique or an independent set, $\pi \circ H = H$ for every π, so $S(H,\alpha) = |\mathrm{Sym}(n)|$ for every α. That is not a defect of the statement but the phenomenon it is about — the source's point is that those vectors measure structure, and the conjecture says structure is atypical. A resolution must therefore identify which graphs are exceptional, and the interesting failure mode is a large class of graphs with a non-trivial automorphism structure that keeps some $S(H,\alpha)$ large.

Remark 3 (What it would buy, and what it would not). The character sum is one factor of the source's detection formula; the other is the binomial factor $\binom{N-V(\alpha)}{n-V(\alpha)}^2$, summed over edge sets of weight at most D. So a bound on $S(H,\alpha)$ alone does not by itself give a hardness result: it has to be combined with that outer sum, and the source's own observation that the argument breaks once $D \geq \log^2 n$ is a bound on how far any such combination can reach. Conjecture 1 is the combinatorial half, isolated because it is a statement about graphs and permutations with no cryptography in it.

Remark 4 (A separate conjecture in the same paper). The source separately presents a candidate construction it conjectures is $(\mathrm{poly}(n), \mathrm{negl}(n))$-secure in the sense of its Definition 5.1 when $N(n) \geq n^{2+\delta}$, and says it does not know how to prove that conjecture. That is a security assumption about a specific construction rather than a combinatorial statement, and it is not this one.

4 Bibliography

[ABI23] Damiano Abram, Amos Beimel and Yuval Ishai. *Cryptography from planted graphs: security with logarithmic-size messages.* IACR ePrint 2023/1929. The source. The detection formula and the conjecture are on page 28; the summary of results and open questions is Section 1.1.5, pages 7–8.

[BE76] Béla Bollobás and Paul Erdős. *Cliques in random graph.* Mathematical Proceedings of the Cambridge Philosophical Society, 1976. The result the source cites for the largest clique in a random N-node graph having at most $2\log n$ nodes with high probability, which is why its argument stops at $D = \log^2 n$.

[KN90] Joe Kilian and Noam Nisan. Private communication, 1990. One of the sources of the information-theoretic log n share-size bound the source's constructions are measured against.

[CCX13] Ignacio Cascudo, Ronald Cramer and Chaoping Xing. *Bounds on the threshold gap in secret sharing and its applications.* IEEE Transactions on Information Theory, 2013. The other source of that log n bound.