Resolving Goldreich's Open Problem 4.1: Linear Lower Bounds for Cubic Graph Isomorphism Testing

Scott Schubr
Luminary Oddities

October 23, 2026

Abstract

We resolve Goldreich's Open Problem 4.1 negatively: testing isomorphism to a fixed cubic graph can require linearly many queries at the specified proximity 0.01. For every sufficiently large even N, there exists a known connected simple bridgeless 3-regular N-vertex target for which every randomized two-sided tester requires $\Omega(N)$ incidence queries, even when inputs are promised connected, bridgeless and cubic. Reading the whole incidence representation gives a matching $O(N)$ upper bound, so the worst-case query complexity is $\Theta(N)$. Distance is measured by unordered edge changes with far threshold $0.01(3N/2)$; computation and target preprocessing are unrestricted. The proof develops a cubic encoding within Dror's sparse-code framework. Its central estimate is $\operatorname{dist}_H(b, \operatorname{Im} A) \leq (2/3)\operatorname{dist}_{\mathrm{iso}}(G_b, G_0)$ for a $16n$-vertex encoding under arbitrary uncolored bijections, including equal isolate padding. Adaptive target-middle colors, a joint row-defect budget and median rounding establish the estimate. Bounded equivariant surgeries produce connected cubic targets at every sufficiently large even order. The same encoding covers every sufficiently large integer order at proximity 0.01 with at most one isolate using the surgeries. This also attains the minimum possible maximum-degree threshold for linear fixed-target isomorphism-testing hardness.

1 Resolving Goldreich's Open Problem 4.1

1.1 The question and its significance

Graph isomorphism identifies graphs that have the same adjacency structure after renaming their vertices. In the bounded-degree testing model, one asks whether a small number of local neighbor queries can distinguish an input isomorphic to a fully known target from one far from every relabeling of that target. A cubic graph has exactly three neighbors at every vertex. The question is therefore about how much of a sparse graph a randomized tester must inspect, even when its reference graph is known in advance.

Goldreich's Open Problem 4.1 [3] asks whether isomorphism to every fixed 3-regular graph can be tested with sublinear queries at proximity 0.01 in the bounded-degree model. The target is specified in full, the unknown input is accessed through incidence queries, and the tester may have two-sided error. Distance counts unordered edge changes, with far threshold $0.01(3N/2)$. We answer this question negatively.

1.2 The negative answer

Theorem 1 (Resolution of Open Problem 4.1)**.** *For every sufficiently large even N, there is a known connected simple bridgeless 3-regular N-vertex graph J_N such that every randomized two-sided $(1/100)$-tester for isomorphism to J_N requires $\Omega(N)$ incidence queries. The lower bound holds even if inputs are promised connected, simple, bridgeless and 3-regular. Target preprocessing and computation are unrestricted.*

One hard family suffices to refute the universal proposal: the theorem provides such a family even within connected simple bridgeless cubic graphs. Reading all $3N$ incidence slots gives the matching $O(N)$ upper bound with unrestricted computation. Hence the worst-case query complexity is $\Theta(N)$ at the specified proximity. Odd orders are necessarily excluded for cubic graphs. The target family is selected existentially. Almost-all-target bounds [4] remain compatible with an exceptional hard family.

1.3 The degree threshold and all-order consequence

The same proof also handles every sufficiently large integer order:

Corollary 2. *For every sufficiently large integer N, there is a known simple N-vertex target of maximum degree three for which randomized two-sided $(1/100)$-testing of isomorphism requires $\Omega(N)$ incidence queries. The target and hard inputs may each consist of a connected cubic graph and at most 47 isolated vertices. Alternatively, they may be chosen with at most one isolate, with no isolate for even N.*

This reaches the minimum possible maximum-degree threshold: degree two admits query bounds depending only on proximity [2]. The conclusions concern randomized classical incidence-query complexity, with unrestricted computation and target preprocessing.

1.4 Proof mechanism and relation to prior work

Method and contribution. Dror [1] proves a linear lower bound using maximum degree thirteen; his Remark 1.3 asks for the smallest degree supporting it. We use his sparse-code and adaptive-transcript framework. The quantitative step is a cubic parity encoding with two nine-cycles per variable, together with distance transfer under arbitrary uncolored bijections. An adaptive coloring of target middle vertices, a joint charge for defective rows and imported middles, and median rounding yield the bound $\operatorname{dist}_H(b, \operatorname{Im} A) \le (2/3)\operatorname{dist}_{\mathrm{iso}}(G_b, G_0)$. For a code of rate at most $1/3$, the resulting graph-distance gap exceeds the literal 0.01 threshold. Bounded paired surgeries preserve connected cubicity and one-bit query simulation at every even order. The isolate extension then gives Corollary 2 without a parallel encoding.

Prior work. Parity gadgets have CFI ancestry [5]; code-distance reductions to colored graph matching appear in Arvind et al. [6]. Their color-preserving maps do not supply the arbitrary uncolored distance estimate required here, and bounded color-class size is not a degree bound. Standard cubic CFI nonisomorphism alone does not give testing-far instances, since one twist changes only constantly many edges. Partial recovery under edge defects appears in O'Donnell et al. [7]; intrinsic recognition of uncolored CFI components is studied by Chen, Flum and Liu [8]. Occurrence-cycle cubic reductions, including local oracle and distance transfers, appear in Adler–Köhler–Peng [10, Section 5]; their tested property permits many isomorphism classes. Makarychev–Manokaran–

Sviridenko [11] obtain broad approximate-isomorphism gaps under randomized Unique Games assumptions, with large clouds and almost completeness. The result here combines unconditional exact completeness, a fixed target, cubic incidence access and the stated numerical proximity.

Proof organization. Section 2 gives the model and connected sparse-code input. Section 3 defines the cubic encoding and its oracle. Section 4 proves the arbitrary-bijection distance bound. Section 5 completes the literal 0.01 theorem, and Section 6 derives the all-order corollary. The appendix proves the required sparse-matrix existence statement.

The reduction contract. The proof separates three obligations: a satisfying codeword gives an exact isomorphism; one incidence answer reads at most one source bit; and every uncolored bijection yields a quantitatively close decoded word. The distance inequality is proved for the base encoding and equal isolate padding. The bounded surgeries are paid afterwards by triangle inequality, so their use does not require the modified graph to retain the base short-cycle characterization. The all-order corollary uses this same encoding and therefore needs no second geometric construction.

2 Model and connected sparse-code input

All graphs are finite, simple and undirected. For N-vertex graphs set

$$\operatorname{dist}_{\mathrm{iso}}(G,H)=\min_{\phi:V(G)\to V(H)\ \text{bijective}}|E(G)\triangle\phi^{-1}(E(H))|.$$

For degree bound d, an input is ε-far from the isomorphism class of H if this distance exceeds $\varepsilon dN/2$. An incidence oracle on $[N]\times[d]$ returns the indicated neighbor or 0 for a missing slot. Neighbor order is arbitrary; isomorphism is a property of the graph rather than of the slot names. A tester accepts every representation of an isomorphic input with probability at least $2/3$ and rejects every far input with probability at least $2/3$. We count worst-case queries.

Lemma 3. *For some $\delta>0$, every sufficiently large n divisible by three admits $A\in\mathbb{F}_2^{n\times(n/3)}$ of row weight three and column weight nine, with connected bipartite incidence support and every set of at most $\lfloor\delta n\rfloor$ rows independent.*

Its complete configuration-model proof appears in Appendix A. Fix such a matrix, put $m=n/3$ and $C=\operatorname{Im}A$, and order its row coordinates. Write (v_{i1},v_{i2},v_{i3}) for the three distinct variables of row i, with this order independent of b. All bit additions are in $\mathbb{F}_2$. Define

$$\operatorname{dist}_H(b,C)=\min_{x\in\mathbb{F}_2^m}|\{i:(Ax)_i\neq b_i\}|.$$

Order each variable's nine occurrences by positions $j\in\mathbb{Z}/9\mathbb{Z}$, independently of b.

Lemma 4 (Adaptive indistinguishability, following Dror)**.** *Algorithms making at most $\lfloor\delta n\rfloor$ coordinate queries have identical transcript laws on Ax, for uniform $x\in\mathbb{F}_2^m$, and on uniform $b\in\mathbb{F}_2^n$. Moreover, for all sufficiently large n, uniform b has $\operatorname{dist}_H(b,C)>n/6$ with probability at least $3/4$.*

Proof. Fix the algorithm's random coins and a realized decision-tree path. Every collection of at most $\lfloor\delta n\rfloor$ distinct queried rows is independent. Its coordinates of Ax are therefore uniform independent bits. A newly queried coordinate remains unbiased given the preceding answers; a repeated answer is determined. Induction on the path length proves the transcript assertion, including adaptive queries.

Since $|C| \le 2^{n/3}$, the entropy bound for Hamming balls gives

$$\Pr[\mathrm{dist}_H(b, C) \le n/6] \le 2^{-n(2/3 - H_2(1/6))}.$$

Here H_2 is binary entropy. The exponent is positive, since $H_2(1/6) < 2/3$ is equivalent to $6^6 < 2^4 5^5$, or $46656 < 50000$. The probability is at most $1/4$ for all sufficiently large n. No full-column-rank assumption is needed. □

3 The connected cubic encoding

For every variable v, create vertices $z(v, a, j)$ with $a \in \mathbb{F}_2$ and $j \in \mathbb{Z}/9\mathbb{Z}$, and join each fixed-bit set in a 9-cycle. For every row i, create outer vertices $o(i, k, a)$ with $k \in [3]$ and $a \in \mathbb{F}_2$, and middle vertices $q(i, u, w)$ with $u, w \in \mathbb{F}_2$. Put

$$t(i, u, w) = (u, w, b_i + u + w),$$

and join $q(i, u, w)$ to $o(i, k, t_k(i, u, w))$ for each k. The internal row graph is a once-subdivided K_4: the four parity-b_i tuples are its branch vertices, and each outer vertex subdivides the edge between the two tuples sharing its coordinate and value.

If row i uses $v = v_{ik}$ at its occurrence position j, join $o(i, k, a)$ to $z(v, a, j)$ for both bits a. Call the resulting graph G_b. It has

$$|V(G_b)| = 18m + 10n = 16n, \qquad |E(G_b)| = 24n.$$

Every vertex has degree three. Distinct variables in a row and distinct occurrence positions ensure simplicity. Each connected row gadget ties both bit cycles of every incident variable; connected incidence support therefore makes G_b connected. Variable, outer and middle roles are construction descriptions, not colors supplied to the tester or bijection.

Lemma 5 (Completeness)**.** *For every b, x, $G_b \cong G_{b+Ax}$. In particular, if $b \in C$, then $G_b \cong G_0$.*

Proof. Use the bijection

$$\begin{aligned} z(v, a, j) &\mapsto z(v, a + x_v, j), \\ o(i, k, a) &\mapsto o(i, k, a + x_{v_{ik}}), \\ q(i, u, w) &\mapsto q(i, u + x_{v_{i1}}, w + x_{v_{i2}}). \end{aligned}$$

Its third target bit is $(b_i + (Ax)_i) + (u + x_{v_{i1}}) + (w + x_{v_{i2}}) = b_i + u + w + x_{v_{i3}}$. All cycle, row and external edges are consequently preserved. □

Lemma 6 (One-bit simulation)**.** *A fixed labeling and fixed incidence slots permit each answer about G_b to read at most one coordinate of b.*

Proof. At a variable vertex, slots one and two return predecessor and successor, and slot three its external outer neighbor, all independently of b. At an outer vertex, slot one returns its variable neighbor. The other two slots enumerate middle neighbors: for $k = 1$ fix $u = a$ and enumerate w; for $k = 2$ fix $w = a$ and enumerate u; for $k = 3$ enumerate u and set $w = b_i + a + u$. At a middle vertex, slot k returns $o(i, k, t_k(i, u, w))$. Only row bit b_i can be needed. The lists are reciprocal and have no repeated neighbors; all vertex and slot names are fixed before b is known. □

Lemma 7 (Bridgelessness)**.** *Every graph G_b is bridgeless.*

Proof. Each internal row edge lies on a cycle of the subdivided K_4, and each variable-cycle edge lies on its designated 9-cycle. Consider an external edge at an occurrence of variable v in row i. Choose a second occurrence of v, in a distinct row i'. Inside each of these two row gadgets, join the opposite-bit outer vertices for v by a simple path. In each bit cycle of v, join the two occurrence positions by a simple path. These four paths, together with the four external edges at the two occurrences, form a simple cycle containing the chosen edge: the row gadgets are distinct, the two bit cycles are disjoint, and the four paths have disjoint interiors. Thus every edge lies on a cycle, proving the claim. □

4 Distance under arbitrary uncolored bijections

Recognizing uncolored CFI components using short cycles has explicit ancestry in Chen–Flum–Liu [8, Lemma 7.12]. Here we need the following exact fact.

Lemma 8. *The only simple* 9*-cycles of* G_b *are its designated variable cycles. Every simple* 6*- or* 8*-cycle lies in one row gadget.*

Proof. The internal variable graphs are 9-cycles and the internal row graphs are once-subdivided K_4's, whose cycles have lengths six or eight. An external cycle uses an even number of external edges. With exactly two, its internal pieces would lie in one row and one variable. A row uses a variable once, so these endpoints are the opposite values of one coordinate; their variable vertices belong to disconnected bit cycles. This is impossible. With $2k \geq 4$ external edges, the k internal row runs have length at least two, and the k internal variable runs length at least one. These bounds hold even on revisiting gadgets: outer vertices are nonadjacent and each socket has just one external edge. Its length is at least $5k \geq 10$. □

4.1 Row rigidity and occupancy

Fix any vertex bijection $\phi : G_b \to G_0$. Let K be its unordered-edge symmetric difference. A source edge is unmatched if its image is absent. Write $q_{\mathrm{cyc}}, q_{\mathrm{row}}, q_{\mathrm{ext}}$ for the unmatched counts by source cycle, internal-row and external roles, without assumptions on images. Equal edge cardinalities give

$$q_{\mathrm{cyc}} + q_{\mathrm{row}} + q_{\mathrm{ext}} = K/2. \tag{1}$$

A source row is intact if all twelve internal edges match. Its four overlapping 6-cycles map into one target row by Lemma 8. Their ten-vertex union fills that row, and the twelve matched edges exhaust its internal edges. This is an internal isomorphism. Internal degree distinguishes middle and outer vertices; disjoint middle neighborhoods distinguish the three opposite outer pairs. The pairs are therefore permuted, with value flips $y(i,k)$ satisfying

$$\sum_{k=1}^{3} y(i,k) = b_i. \tag{2}$$

Lemma 9. *A nonintact source row has at least two unmatched internal edges. If it has exactly two, it retains a matched* 6*- or* 8*-cycle.*

Proof. Suppose exactly one half-edge is unmatched, over underlying K_4 edge uv. Write c, d for the other branch vertices. The two triangles avoiding uv give intact 6-cycles sharing the subdivided cd path. Their images are distinct 6-cycles in one target row. Their union contains nine vertices: all four middle vertices and five outers. The shared path identifies its middle endpoints; the centers of

the other four-edge paths between those endpoints identify the other middles. Thus every source middle maps to a target middle. The omitted outer over uv has one still-matched leg. Its image must be the remaining target outer: it neighbors a target middle and the other nine vertices are occupied. Its other middle neighbor is then precisely the image of the other endpoint of uv, contradicting the single unmatched edge. Here the complementary edge of a subdivided K_4 is determined by the five already identified edges.

With exactly two unmatched half-edges, remove their at most two underlying K_4 edges. Two adjacent deletions leave a triangle; two opposite deletions leave a 4-cycle. Fewer underlying deletions also leave a triangle. The corresponding subdivided cycle has all edges matched. □

Let B count nonintact source rows and a those with exactly two defects. Their disjoint internal edge sets give

$$q_{\text{row}} \geq 2a + 3(B - a) = 3B - a. \tag{3}$$

The $n - B$ intact rows fill distinct whole target rows. Call the remaining B target rows unoccupied. Every target-row image of a source variable vertex lies in an unoccupied row, by injectivity. For each of the a two-defect rows choose the cycle from Lemma 9. Its image is in an unoccupied target row. These a target rows are distinct: source cycles from distinct rows are vertex-disjoint, while two disjoint cycles of length at least six cannot fit in one ten-vertex row. Each occupies at least three target middles. We use only this cycle occupancy, not a type-preserving map on a defective row.

4.2 Adaptive middle colors and median rounding

Proposition 10. *For every b, $\operatorname{dist}_H(b, C) \leq (2/3)\operatorname{dist}_{\text{iso}}(G_b, G_0)$. The same inequality holds after adding any equal number of isolates.*

Proof. For the arbitrary ϕ above, color target variable vertices by their bit, and target outer vertices by their value. Call this binary color χ. It agrees at both ends of every target variable-cycle or external edge. Let P consist of images of matched source cycle edges. Its maximum degree is two. At each target middle choose its color to minimize disagreements on its incident P-edges. Their other endpoints are target outers with fixed colors. With two differing neighbor bits exactly one disagreement is unavoidable; with one neighbor or two equal bits choose that bit, and with no neighbor choose zero. Let c count unavoidable disagreements.

The target parity-zero row has middle tuple 000, which contributes zero, and three other middles, each contributing at most one. An occupied target row contains no source-variable image and hence contributes zero. Each of the a cycle-anchored unoccupied rows has at least three middles occupied by source-row images, so contributes at most one. Therefore

$$c \leq 3B - 2a, \qquad c + B \leq 4B - 2a \leq \tfrac{4}{3}(3B - a) \leq \tfrac{4}{3}q_{\text{row}}.$$

This coloring may depend on ϕ; it is a decoding existence argument, not oracle advice. No matched-external-edge or full-star condition is imposed on the imported middle vertices.

Along all source variable cycles the total color variation is at most $q_{\text{cyc}} + c$: unmatched cycle edges cost at most one, and matched ones change color only on the c selected target-middle cuts. Define

$$\alpha(v, a, j) = a + \chi(\phi(z(v, a, j))) \pmod 2, \qquad s(v, j) = \frac{\alpha(v, 0, j) + \alpha(v, 1, j)}{2}.$$

Let t_v be a median of the nine real values $s(v,j)$, and choose x_v independently with $\Pr[x_v = 1] = t_v$. For nine real cyclic values, their median t satisfies

$$\sum_j |s_j - t| \le 4(\max s_j - \min s_j) \le 2\sum_j |s_j - s_{j+1}|.$$

Pair the four sorted values below the median with the four above it; each pair contributes at most the range, giving the first inequality; cyclic variation is at least twice the range. Also

$$\sum_j |s(v,j) - s(v,j+1)| \le \tfrac{1}{2}\sum_{a,j} |\alpha(v,a,j) - \alpha(v,a,j+1)|.$$

The added bit a is constant along each cycle. Summing gives

$$\sum_{v,j} |s(v,j) - t_v| \le q_{\mathrm{cyc}} + c. \tag{4}$$

For an intact row, its flips satisfy (2); violation implies $x_{v_{ik}} \ne y(i,k)$ for some coordinate. Its probability is at most $\sum_k |t_{v_{ik}} - y(i,k)|$. For its occurrence (v,j) and bit $y = y(i,k)$,

$$|t_v - y| \le |t_v - s(v,j)| + |s(v,j) - y|, \qquad |s(v,j) - y| = \tfrac{1}{2}\sum_a |\alpha(v,a,j) - y|.$$

If the source external edge at bit a matches, then $\alpha(v,a,j) = y$. Indeed its outer image has bit $a + y$. If its other image endpoint is a target variable, external adjacency forces that bit. If it is a row vertex, it cannot belong to the already fully occupied intact target row; distinct target rows have no connecting edges, so such an edge cannot match. Thus every failed bit equality costs a distinct unmatched source external edge. Summed over intact rows, the $|s - y|$ terms cost at most $q_{\mathrm{ext}}/2$; occurrence positions are used once each in (4). The B nonintact rows cost at most B. Some assignment consequently has

$$\mathrm{dist}_H(b, C) \le q_{\mathrm{cyc}} + c + q_{\mathrm{ext}}/2 + B \le q_{\mathrm{cyc}} + \tfrac{4}{3}q_{\mathrm{row}} + q_{\mathrm{ext}}/2 \le \tfrac{4}{3}(q_{\mathrm{cyc}} + q_{\mathrm{row}} + q_{\mathrm{ext}}) = \tfrac{2}{3}K.$$

This holds for every ϕ, with no restriction on roles or occurrence images.

For equal added isolates, give target isolates color zero. Edge counts and source roles in (1) are unchanged. Intact row cycles cannot map to isolates, so the occupied-row counts remain valid. A middle whose preimage is a source isolate has no incident P-edge and contributes nothing to c. Matched source edges cannot land at an isolate; a variable endpoint mapped there has an unmatched external edge. All variation, external-error and median estimates therefore remain valid, even when ϕ mixes isolates with nonisolated vertices. □

5 Proof of the cubic theorem at proximity 0.01

For even N, choose the largest positive multiple of three n with $16n \le N$. Then $r = N - 16n \in \{0, 2, \ldots, 46\}$. Put $\eta = 0$ when $r \equiv 0 \pmod 4$, and $\eta = 2$ otherwise. If $\eta = 2$, subdivide the edge between positions zero and one on both bit cycles of one fixed variable and join the two new vertices. This adds two vertices and preserves simplicity, cubicity and connectedness. A bit flip exchanges the two new vertices. Relative to the old graph plus two isolates it costs at most seven edits: delete two edges, insert four path edges and one joining edge.

Use $(r - \eta)/4 \le 11$ paired triangle expansions (a standard cubic operation; see [9]) at original variable sites, avoiding those surgery endpoints. There are $9m = 3n$ paired positions, enough for

large n. Replace each selected vertex by a triangle, with ports fixed by predecessor, successor and external roles. Adjacent selected vertices are allowed: their edge attaches to the appropriate port at both ends. Each pair adds four vertices; each individual expansion costs at most seven edits relative to the old vertex and two isolates, by retaining one incidence, rerouting two, and adding three triangle edges. This bound also follows by sequential expansion when sites are adjacent.

The resulting P_b is simple, connected, cubic and has exactly N vertices. Bridgelessness also survives these operations. Subdividing the two chosen cycle edges preserves cycles through every old edge; deleting the new joining edge leaves a connected subdivided copy of G_b, so that edge also lies on a cycle. In a triangle expansion, each old cycle extends through the appropriate triangle ports and every new internal edge lies on the triangle. Sequential expansion covers adjacent sites. Lemma 7 therefore shows that every P_b is bridgeless.

All choices are independent of b and paired bit flips extend to full isomorphisms. New variable-site incidences require no bit query; row answers only translate the returned variable label to its fixed port. At triangle vertices two slots return triangle neighbors and the third its assigned old incidence. At a subdivision vertex the three neighbors are the old endpoints and the opposite subdivision vertex. Thus fixed names and slots still simulate each graph answer with at most one coordinate query. On each side the edit cost relative to $G_b \sqcup rK_1$ is at most $7 + 11 \cdot 14 = 161$. By the triangle inequality,

$$\mathrm{dist}_{\mathrm{iso}}(G_b \sqcup rK_1, G_0 \sqcup rK_1) \le \mathrm{dist}_{\mathrm{iso}}(P_b, P_0) + 322. \tag{5}$$

We use short-cycle recognition only before these surgeries.

By Lemma 4, uniform b has code distance greater than $n/6$ with probability at least $3/4$ for all sufficiently large n. Proposition 10, including its isolate extension, and (5) give $\mathrm{dist}_{\mathrm{iso}}(P_b, P_0) > n/4 - 322$ for every such word. The literal $1/100$ far threshold obeys

$$\frac{1}{100}\frac{3N}{2} \le \frac{6n}{25} + \frac{69}{100}, \qquad \frac{n}{4} - 322 - \left(\frac{6n}{25} + \frac{69}{100}\right) = \frac{n - 32269}{100} > 0$$

for $n > 32269$. Combine this arithmetic threshold with the asymptotic matrix and entropy thresholds.

Simulating a tester with at most $\lfloor \delta n \rfloor$ graph queries uses at most that many coordinate queries. Codewords yield isomorphic connected cubic graphs and acceptance at least $2/3$. Uniform words give far connected cubic graphs with probability at least $3/4$, so acceptance is at most $(3/4)(1/3) + 1/4 = 1/2$. Identical adaptive transcript laws contradict this. Since $n = \Theta(N)$, Theorem 1 follows at literal proximity 0.01.

6 All integer orders from the same encoding

Proof of Corollary 2. For any sufficiently large integer N, choose

$$n = 3\lfloor N/48 \rfloor, \qquad r = N - 16n \in \{0, \dots, 47\},$$

and set $Q_b = G_b \sqcup rK_1$. These graphs have N vertices and maximum degree three; their nonisolated component is connected and cubic. Codewords give isomorphic graphs by Lemma 5. Lemma 4 and the isolate extension of Proposition 10 give, with probability at least $3/4$,

$$\mathrm{dist}_{\mathrm{iso}}(Q_b, Q_0) > n/4.$$

This exceeds the required far threshold, since

$$\frac{n}{4} - \frac{3N}{200} = \frac{2n - 3r}{200} \geq \frac{2n - 141}{200} > 0$$

for $n > 70.5$, in addition to the asymptotic matrix and entropy thresholds. All slots of an isolate return zero and use no bit query. Every other answer uses at most one bit query. A tester with at most $\lfloor \delta n \rfloor$ queries would again have identical transcript laws but acceptance at least $2/3$ on uniform Ax and at most $1/2$ on uniform b. Since $n = \Theta(N)$, this proves the first assertion. The distance step uses the proved isolate extension; it does not assume that graph-isomorphism distance is invariant under isolate padding.

For the at-most-one-isolate assertion, let $h = N \bmod 2$. Apply the surgeries of Section 5 to the even order $N - h$, using the same n and even residual $r - h \leq 46$, and call the result P_b. Put $R_b = P_b \sqcup hK_1$. Relative to $G_b \sqcup rK_1$, each side still costs at most 161 edits. Proposition 10 and the triangle inequality therefore give, on the same far-word event,

$$\mathrm{dist}_{\mathrm{iso}}(R_b, R_0) > n/4 - 322.$$

The margin over $3N/200$ is at least

$$\frac{n}{4} - 322 - \left(\frac{6n}{25} + \frac{141}{200}\right) = \frac{2n - 64541}{200} > 0$$

for $n > 32270.5$, again combined with the asymptotic thresholds. Paired assignment flips extend over the surgeries and fix the isolate; its slots cost no query. The same transcript contradiction applies. For even N the graph is connected cubic; for odd N it has a connected cubic component on $N - 1$ vertices and one isolate. □

7 Finite verification and scope

The accompanying `verify_literal.py` constructs the nine-cycle encoding and checks finite consequences of the analytic argument. It tests six-, eight- and nine-cycle types in a small exhaustive word example, fixed-slot incidence answers and codeword isomorphisms. Selected arbitrary bijections include middle intrusions and isolate mixing. Exact rational calculations check unmatched-edge charges, two-defect row anchors, adaptive colors, the joint $c + B$ budget, median deviations and the final $2K/3$ inequality. All 24 even surgery residues are constructed, including adjacent expansion sites, with actual oracle, completeness and named edit comparisons.

The additional `verify_all_orders.py` checks all 48 residual sizes for both corollary constructions, isolate slots, extended codeword maps and exact rational thresholds. Its finite examples do not satisfy the asymptotic small-row-independence conclusion and are not claimed to do so. The accompanying reader instructions list the verification entry points, their software dependencies and reproducible commands.

The universal metric proof, matrix appendix, entropy estimate and adaptive transcript argument establish the theorem and its corollary; the finite checks provide reproducible tests of the stated constructions.

A Connected $(3, 9)$ systems

We prove Lemma 3 by the sparse-configuration method underlying Dror's input [1], with the degree-nine parameters made explicit. Write $(z)_k = z(z-1)\cdots(z-k+1)$. Give each of n left vertices three stubs and each of $n/3$ right vertices nine; match the $3n$ stubs uniformly.

A candidate parallel pair specifies a left vertex, two of its stubs, a right vertex, two of its stubs and one of their two pairings. There are

$$n\binom{3}{2}(n/3)\binom{9}{2}2 = 72n^2$$

candidates, each with probability $1/(3n)_2$. If X counts them, then simplicity is exactly $X = 0$, including triple edges which contribute three candidate events. For each fixed k, ordered tuples with distinct left and right vertices number $(72n^2)^k + O_k(n^{2k-1})$ and prescribe $2k$ pairings; their contribution to $\mathbb{E}(X)_k$ tends to 8^k. For a compatible exceptional tuple group its prescribed pairings into cells. Each cell contains at least two pairings; left degree three permits at most one cell at each left vertex. With q cells, $\ell = q$ left vertices, $s \le q$ right vertices and $e \ge 2q$ pairings. Equality $\ell + s - e = 0$ forces two pairings in each cell and distinct right vertices, the principal case. Every other pattern has exponent at most -1, so its $O_k(n^{\ell+s})$ realizations of probability $O_k(n^{-e})$ contribute $O_k(1/n)$. Incompatible tuples contribute zero. Thus $\mathbb{E}(X)_k \to 8^k$. For each fixed h, finite Bonferroni bounds give

$$\sum_{j=0}^{2h+1} \frac{(-1)^j \mathbb{E}(X)_j}{j!} \le \Pr[X = 0] \le \sum_{j=0}^{2h} \frac{(-1)^j \mathbb{E}(X)_j}{j!}.$$

First let $n \to \infty$ and then $h \to \infty$. Both bounds tend to $e^{-8} > 0$.

For a left set of size $s \ge 2$, put $t = \lfloor 3s/2 \rfloor$. Failure of $|N(S)| > 3s/2$ puts all its stubs in a right set of size t. For $t \le n/3$, the union bound is

$$\binom{n}{s}\binom{n/3}{t}\frac{(9t)_{3s}}{(3n)_{3s}} \le \binom{n}{s}\binom{n/3}{t}(3t/n)^{3s} \le [C\sqrt{s/n}]^s.$$

Indeed with $z = t/s \in [1, 3/2]$, elementary binomial estimates give $[e^{1+z}(3z)^{3-z}(s/n)^{2-z}]^s$. Choose $\delta < 2/9$ small enough that $C\sqrt{\delta} \le 1/2$. Splitting $2 \le s \le \delta n$ at $\sqrt{n}$ bounds the first part by a geometric tail with base $Cn^{-1/4}$ and the second by a tail of $\sum_s 2^{-s}$. Both vanish. In a simple configuration, singleton sets also have three distinct neighbors. For every nonempty such small set, its $3|S|$ edges and more than $3|S|/2$ neighbors force a neighbor incident to exactly one of its edges. Its incidence column has odd sum, so those rows cannot sum to zero over $\mathbb{F}_2$. This proves small-row independence.

It remains to ensure connectedness without conditioning away positive simplicity probability. A component with s left vertices has $s/3$ right vertices by stub balance. In any disconnected configuration some component has $3 \le s \le n/2$ and $3 \mid s$. For fixed vertex sets of these sizes, all $3s$ stubs stay together with probability $1/\binom{3n}{3s}$. Thus disconnection has probability at most

$$\sum_{\substack{3 \le s \le n/2 \\ 3 \mid s}} \frac{\binom{n}{s}\binom{n/3}{s/3}}{\binom{3n}{3s}}.$$

The summand is at most $[e^{4/3}(s/n)^{5/3}]^s$. Choose $0 < \alpha < 1/2$ with $e^{4/3}\alpha^{5/3} \le 1/2$. For $s \le \sqrt{n}$ the sum is bounded by a geometric tail with base $e^{4/3}n^{-5/6}$; for $\sqrt{n} < s \le \alpha n$ it is bounded by a tail of $\sum_s 2^{-s}$. Both vanish. For $\alpha n \le s \le n/2$, natural-log entropy H gives the bound

$$(3n+1)\exp[-(5n/3)H(s/n)] \le (3n+1)\exp[-(5n/3)H(\alpha)].$$

Here $e^{MH(k/M)}/(M+1) \le \binom{M}{k} \le e^{MH(k/M)}$; the lower bound follows since k is a mode of Binomial$(M, k/M)$, whose modal probability is at least $1/(M+1)$. At most n terms remain, so this sum vanishes as well.

Both expansion and connectivity fail with probability $o(1)$. Their intersection with simplicity therefore has probability $e^{-8} - o(1) > 0$, without any independence assumption. Taking the binary incidence matrix of a successful simple configuration proves Lemma 3 for every sufficiently large multiple of three.

References

[1] A. Dror, *A Linear Lower Bound on Testing Graph Isomorphism in the Bounded-Degree Model*, ECCC TR26-268, 2026. https://eccc.weizmann.ac.il/report/2026/268/.

[2] O. Goldreich and L. Tauber, *Testing in the Bounded-Degree Graph Model with Degree Bound Two*, Computational Complexity and Local Algorithms, LNCS 15700, Springer, 2025, pp. 373–382. doi:10.1007/978-3-031-88946-2_20.

[3] O. Goldreich, *Testing Isomorphism in the Bounded-Degree Graph Model*, Computational Complexity and Local Algorithms, LNCS 15700, Springer, 2025, pp. 293–323. doi:10.1007/978-3-031-88946-2_16. Author manuscript (Open Problem 4.1): https://www.wisdom.weizmann.ac.il/~oded/COL3/bdg-iso.pdf.

[4] O. Goldreich and L. Tauber, *On Testing Isomorphism to a Fixed Graph in the Bounded-Degree Graph Model*, Computational Complexity and Local Algorithms, LNCS 15700, Springer, 2025, pp. 324–337. doi:10.1007/978-3-031-88946-2_17.

[5] J.-Y. Cai, M. Fürer and N. Immerman, *An optimal lower bound on the number of variables for graph identification*, Combinatorica 12 (1992), pp. 389–410.

[6] V. Arvind, J. Köbler, S. Kuhnert and Y. Vasudev, *Approximate Graph Isomorphism*, MFCS 2012, LNCS 7464, Springer, 2012, pp. 100–111. Full version: https://eccc.weizmann.ac.il/report/2012/078/.

[7] R. O'Donnell, J. Wright, C. Wu and Y. Zhou, *Hardness of Robust Graph Isomorphism, Lasserre Gaps, and Asymmetry of Random Graphs*, SODA 2014, pp. 1659–1677. doi:10.1137/1.9781611973402.120.

[8] Y. Chen, J. Flum and M. Liu, *Some remarks on the uncolored versions of the original CFI-graphs*, arXiv:2507.01459, 2025. https://arxiv.org/abs/2507.01459.

[9] M. Goetze, P. Jungeblut and T. Ueckerdt, *Recognition Complexity of Subgraphs of k-Connected Planar Cubic Graphs*, Algorithmica 88 (2026), article 11. doi:10.1007/s00453-025-01350-8.

[10] I. Adler, N. Köhler and P. Peng, *On Testability of First-Order Properties in Bounded-Degree Graphs and Connections to Proximity-Oblivious Testing*, SIAM Journal on Computing 53(4) (2024), pp. 825–883. doi:10.1137/23M1556253. Accepted manuscript: https://eprints.whiterose.ac.uk/id/eprint/212370/1/SubmissionSICOMP-13.pdf.

[11] K. Makarychev, R. Manokaran and M. Sviridenko, *Maximum Quadratic Assignment Problem: Reduction from Maximum Label Cover and LP-based Approximation Algorithm*, arXiv:1403.7721, 2014. https://arxiv.org/abs/1403.7721.