A linear query lower bound for graph-isomorphism testing at maximum degree three

Scott Schubr
Luminary Oddities

October 23, 2026

Abstract

For every sufficiently large even N, we exhibit existentially a known connected simple 3-regular N-vertex target whose randomized two-sided isomorphism testing requires $\Omega(N)$ incidence queries at literal proximity 0.01, even with connected cubic inputs promised. This gives a negative answer to Goldreich's Open Problem 4.1 in its stated distance normalization. We also retain maximum-degree-three fixed targets at every sufficiently large integer order, reaching the minimum maximum-degree threshold since degree two admits size-independent query bounds. We use Dror's sparse-code and adaptive indistinguishability framework. A no-rung $16n$-vertex parity encoding has the arbitrary-bijection estimate $\mathrm{dist}_H(b, \mathrm{Im}\, A) \le (2/3)\,\mathrm{dist}_{\mathrm{iso}}(G_b, G_0)$. An adaptive coloring of imported target middle vertices, a joint defective-row budget, and median assignment rounding handle arbitrary mixing of graph roles. Bounded equivariant surgeries cover every even order with additive distance loss at most 322. An upper bound $4/429 < 0.01$ for the older capped family explains why that completion method alone cannot reach the numerical endpoint.

1 Introduction and statement

In the bounded-degree graph model, a tester asks for a specified neighbor of a named vertex of an unknown graph. Reading the entire incidence description costs linearly many queries. The fixed-target isomorphism problem asks whether the unknown graph is a relabeling of a fully known target or is far from every such relabeling.

Dror [1] proved a linear two-sided query lower bound using targets of maximum degree thirteen. His Remark 1.3 asks for the smallest maximum degree supporting this phenomenon. Goldreich and Tauber [2] show that every property of degree-two graphs has a query bound depending only on proximity. We reach the remaining endpoint.

Theorem 1. *There are constants $\varepsilon_0, c > 0$ and graphs H_N for all sufficiently large integers N such that H_N is simple, undirected, has N vertices and maximum degree at most three, and every randomized two-sided ε_0-tester for isomorphism to the known H_N makes at least cN incidence queries. Target preprocessing and computation are unrestricted. One may take $\varepsilon_0 = 10^{-6}$.*

The following strengthening removes both connectedness and regularity qualifications, at an explicit larger fixed proximity.

Theorem 2. *For every sufficiently large even N, there is a known connected simple 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 and 3-regular. Target preprocessing and computation are unrestricted.*

Odd orders are necessarily excluded for cubic graphs. Both target constructions are existential. Theorem 2 answers Goldreich's Open Problem 4.1 [3] negatively at its literal proximity 0.01, with the same unordered-edge normalization $\varepsilon(3N/2)$ and a two-sided error guarantee. The square-root bounds for almost all regular targets [4] are compatible with exceptional hard targets.

Contribution and ancestry. The contribution is the complete subcubic encoding and distance transfer under arbitrary *uncolored* bijections, together with constant-query simulation. The sparse matrix, code-membership lower bound and reduction architecture are due to Dror [1]. Parity gadgets have CFI ancestry [5]; code-distance reductions to colored graph matching already appear in Arvind et al. [6]. Partial recovery of components under edge defects occurs in O'Donnell et al. [7]; intrinsic recognition of uncolored CFI components is studied by Chen, Flum and Liu [8]. These general principles are classical. The older code-distance reduction assumes color-preserving maps; bounded color-class size is not a degree bound. Standard cubic CFI nonisomorphism also does not itself give testing-far instances: one twist changes only constantly many edges. The encoding here provides the required conjunction of degree, symmetry, robustness, size and access. Occurrence-cycle reductions to cubic graphs, with local oracle and distance transfers, already appear in Adler–Köhler–Peng [10, Section 5]. Their construction tests a property with many permitted isomorphism classes. Makarychev–Manokaran–Sviridenko [11] give broad robust approximate- isomorphism gaps under randomized Unique Games assumptions. Their large-cloud construction and almost completeness do not directly give an unconditional, exact-completeness, fixed-target cubic incidence reduction. The contribution here is the quantitative conjunction proved below, rather than a new general gap-amplification principle. The subdivided-K_4 completion cap is standard; for example, Goetze, Jungeblut and Ueckerdt [9] use it to complete subcubic graphs. The connected-cubic result uses a no-rung encoding with two nine-cycles per variable. Short-cycle recognition is inherited as a general principle; the quantitative contribution is adaptive target-middle coloring, joint charging of defective rows and imported middles, and median assignment rounding under arbitrary bijections. Connected $(3, 9)$ sparse-system selection and bounded paired surgeries preserve the oracle and all even orders. The capped-family upper bound identifies an obstruction specific to that completion method, rather than a limitation of cubic graphs in general.

2 Model and 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.

The following is Dror's Lemma 2.3 [1]; Appendix A reconstructs its proof.

Lemma 3. *For some $\delta > 0$, every sufficiently large even n admits a matrix $A \in \mathbb{F}_2^{n \times (n/2)}$ of row weight three and column weight six such that every set of at most $\lfloor \delta n \rfloor$ rows is independent.*

Fix such A, write $m = n/2$ and $C = \operatorname{Im} A$, and fix independently of b an ordering (v_{i1}, v_{i2}, v_{i3}) of each row's three distinct variables. Order each variable's six occurrences by $j = 1, \ldots, 6$. All bit

additions below are in $\mathbb{F}_2$. Put

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

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 $b \in \mathbb{F}_2^n$. For every fixed $0 < \beta < 1/2$ with $H_2(\beta) < 1/2$, uniform b satisfies $\operatorname{dist}_H(b, C) > \beta n$ with probability at least $3/4$ for all large n. In particular, $\beta = 1/100$ is admissible.*

Proof. Fix random coins and a realized decision-tree path. The distinct rows queried so far together with a new queried row are independent while their number is at most $\lfloor \delta n \rfloor$. Their coordinates of Ax are uniform and independent, so the next new answer remains unbiased after conditioning on earlier answers. Repeated answers are determined. Induction proves the transcript assertion.

Since $|C| \leq 2^{n/2}$, the entropy estimate for a Hamming ball, with $0 < \beta < 1/2$, gives

$$\Pr[\operatorname{dist}_H(b, C) \leq \beta n] \leq 2^{-n/2} \sum_{j=0}^{\lfloor \beta n \rfloor} \binom{n}{j} \leq 2^{-n(1/2 - H_2(\beta))}.$$

For $\beta = 1/100$, $H_2(\beta) < 9/100$: use $\log_2 100 < 7$, $\log_2 e < 2$, and $-(1-\beta)\ln(1-\beta) \leq \beta$. For every admissible fixed β the exponent is strictly positive, so the probability is at most $1/4$ for large n. No full-column-rank assumption is needed. □

3 Subcubic encoding

We define F_b on a fixed named vertex set independent of b. For each variable v create

$$z(v, a, p), \qquad a \in \mathbb{F}_2, \quad p \in \mathbb{Z}/28\mathbb{Z}.$$

For each a, join the vertices in the cycle $p \leftrightarrow p + 1$. Add the single *rung* $\{z(v, 0, 0), z(v, 1, 0)\}$. Occurrence j uses position $4j$ on each cycle. This variable component has 56 vertices and 57 internal edges. Interchanging the two cycles preserves every occurrence position; completeness does not require arbitrary socket permutations.

For each row i create outer vertices $o(i, k, a)$, $k \in [3], a \in \mathbb{F}_2$, and middle vertices $q(i, u, w)$, $u, w \in \mathbb{F}_2$. Set

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

Join $q(i, u, w)$ to $o(i, k, t_k(i, u, w))$ for each k. If v_{ik} has occurrence number j in row i, add $\{o(i, k, a), z(v_{ik}, a, 4j)\}$ for $a \in \mathbb{F}_2$. The internal row graph is a once-subdivided K_4.

Designated cycle positions have degree three; other cycle vertices have degree two. Every row vertex has degree three in the full graph. Thus

$$|V(F_b)| = 56m + 10n = 38n, \qquad |E(F_b)| = 93n/2.$$

There are $21n$ degree-two and $17n$ degree-three vertices. Distinct variables in a row and distinct sockets for occurrences ensure simplicity. Variable and row roles are construction descriptions, not colors supplied to an isomorphism.

Lemma 5 (Completeness)**.** *If $b = Ax$, then $F_0 \cong F_b$.*

Proof. The map

$$z(v, a, p) \mapsto z(v, a + x_v, p),$$
$$o(i, k, a) \mapsto o(i, k, a + x_{v_{ik}}),$$
$$q(i, u, w) \mapsto q(i, u + x_{v_{i1}}, w + x_{v_{i2}})$$

is bijective. Its third row bit is correct because $b_i = x_{v_{i1}} + x_{v_{i2}} + x_{v_{i3}}$. Every cycle edge, rung, row edge and external edge is preserved. □

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

Proof. Cycle slots one and two return predecessor and successor. Slot three returns the rung neighbor at $p = 0$, the appropriate outer neighbor at $p = 4j$, or 0 otherwise. All are independent of b. At an outer vertex, slot one is the external neighbor. The other two slots enumerate middle neighbors as follows: 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 the corresponding row bit can be needed. These are reciprocal adjacency lists without repeated neighbors, and all vertex and slot names are fixed before b is known. □

4 Distance preservation under arbitrary bijections

Proposition 7. *For every b,*

$$\mathrm{dist}_H(b, C) \leq 48 \ \mathrm{dist}_{\mathrm{iso}}(F_b, F_0).$$

This also holds after adding any equal number of isolates to both graphs.

Proof. Call a vertex of *variable type* if it has degree two, or has degree three and a degree-two neighbor. Exactly variable vertices satisfy this predicate: every designated cycle position has two degree-two neighbors, whereas no row vertex has a degree-two neighbor. The induced variable-type components are precisely the 56-vertex variable components. The induced row-type components are the 10-vertex row components. With padding, degree-zero vertices form a third type and are excluded from row type.

Fix any bijection $\phi : V(F_b) \to V(F_0)$, including padding. Pull target adjacency back to the source set and let

$$K = |E(F_b) \triangle \phi^{-1}(E(F_0))|.$$

Let D contain all endpoints of discrepant edges, including edges present only in the pulled target. Then $|D| \leq 2K$. Set $W = D \cup N_{F_b}(D)$; degree at most three gives $|W| \leq 8K$. A source variable or row component is *bad* if it meets W. Their counts B_v, B_c obey

$$B_v + B_c \leq 8K, \tag{1}$$

since these components are disjoint.

At each vertex outside D, the whole adjacency list agrees with its pulled-target list. At a vertex outside W, this holds at all source neighbors too. Its degree and every neighbor's degree agree with their image degrees. Hence intrinsic type is preserved outside W. Extra target neighbors cannot evade this observation, since the whole list at the vertex agrees.

A component avoiding W, called *good*, has exact internal adjacency, and its connected image has one intrinsic type. Its image lies in a single target component of that type. Equal component

sizes, 56 or 10, and injectivity force the image to be that whole target component. Distinct good components have distinct images.

The rung is the unique bridge of a variable component. A good variable component map therefore permutes its two complete cycles. For each good source variable v, define x_v so that its bit-a cycle maps to the target bit-$(a + x_v)$ cycle. Set x_v arbitrarily on bad variables. Independent cycle reflections and changes of occurrence positions do not change this global bit interchange.

Now consider a good source row i with all three incident variable components good. It maps onto an entire target row component. Internal degree distinguishes its middle and outer vertices, with degrees three and two. The two source outer vertices for one coordinate have external edges into the same good source variable component. Their images have external edges into the same good target variable component. Distinct source variables map to distinct target variable components, and the target row uses three distinct variables. Thus source coordinate pairs map to target coordinate pairs, permuted. External-edge preservation shows that source value a becomes $a + x_v$ in that target pair.

Each source middle tuple has parity b_i, and its image is a target middle tuple of parity zero. Coordinate permutation preserves parity, so

$$b_i + x_{v_{i1}} + x_{v_{i2}} + x_{v_{i3}} = 0.$$

This row is satisfied by x. No alignment of source and target row names is assumed. If independent cycle reflections give incompatible occurrence images, the external edges of this good row cannot all be preserved, so that attempted exception does not satisfy the hypotheses.

Every violated row is consequently bad itself or meets a bad variable. Each variable occurs six times, and (1) gives

$$\operatorname{dist}_H(b, C) \le |\{i : (Ax)_i \ne b_i\}| \le B_c + 6B_v \le 6(B_c + B_v) \le 48K.$$

Minimize over ϕ. Padding does not change the proof: outside D, degrees agree, so no good nonisolated component maps into isolates. □

5 Linear query hardness

Take $\beta = 1/100$ and $\varepsilon_0 = \beta/10000$. If $\operatorname{dist}_H(b, C) > \beta n$, Proposition 7 gives $\operatorname{dist}_{\text{iso}}(F_b, F_0) > \beta n/48$. For arbitrary sufficiently large N put

$$n = 2\lfloor N/76 \rfloor, \qquad r = N - 38n \in \{0, \dots, 75\}.$$

Define H_N as F_0 plus r isolates, and use the same padding on F_b. Since $N \le 113n$ for $n \ge 1$,

$$\varepsilon_0(3N/2) \le 339\beta n/20000 < \beta n/48.$$

Each far vector therefore gives an ε_0-far graph. Codewords give isomorphic graphs, and graph queries use at most one bit query each.

If a graph tester made at most $\lfloor \delta n \rfloor$ queries, it would accept uniform Ax with probability at least $2/3$. It would accept uniform b with probability at most

$$(3/4)(1/3) + (1/4) \cdot 1 = 1/2,$$

using Lemma 4 and soundness. The identical transcript laws of Lemma 4 contradict this. As $n = \Theta(N)$, Theorem 1 follows. The full description has $3N$ slots, giving the matching trivial linear-query upper bound.

6 Connected cubic targets at proximity 0.01

We now prove Theorem 2. The parameters in this section are local: n is divisible by three, $m = n/3$, and $C = \operatorname{Im} A$ for the following matrix, rather than the $(3, 6)$ matrix used for F_b.

Lemma 8. *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 B. Fix such a matrix, order its row coordinates, and order each variable's nine occurrences by positions $j \in \mathbb{Z}/9\mathbb{Z}$, independently of b.

6.1 The no-rung encoding and exact cycles

For each variable create vertices $z(v, a, j)$, $a \in \mathbb{F}_2$, $j \in \mathbb{Z}/9\mathbb{Z}$, joined in two disjoint 9-cycles. There is no rung. Use the same ten-vertex row gadget as for F_b: $q(i, u, w)$ has outer neighbors with values $(u, w, b_i + u + w)$. Join $o(i, k, a)$ to $z(v_{ik}, a, j)$ at that occurrence. Call the graph G_b. Every vertex has degree three, and

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

Distinct row variables and distinct occurrence positions ensure simplicity. Each connected row gadget ties both bit cycles of every incident variable. Since the incidence support is connected, the whole graph is connected.

The assignment-flip formulas of Lemma 5 give, for every b, x, $G_b \cong G_{b+Ax}$. Indeed the third target value is $(b_i + (Ax)_i) + (u + x_{v_{i1}}) + (w + x_{v_{i2}}) = b_i + u + w + x_{v_{i3}}$. Variable slots return predecessor, successor and occurrence neighbor; row slots are exactly those in Lemma 6. All names and slots are fixed before b and each answer reads at most one bit.

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 9. *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 $5k \geq 10$. □

6.2 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

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

A source row is intact if all twelve internal edges match. Its four overlapping 6-cycles map into one target row by Lemma 9. 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

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

Lemma 10. *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

$$q_{\text{row}} \geq 2a + 3(B - a) = 3B - a. \tag{4}$$

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 10. 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.

6.3 Adaptive middle colors and median rounding

Proposition 11. *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

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

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 $q_{\mathrm{cyc}} + c$: 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

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

There are at most four values on each side of the median, 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

$$\sum_{v,j} |s(v, j) - t_v| \le q_{\mathrm{cyc}} + c. \tag{5}$$

For an intact row, its flips satisfy (3); 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 (5). 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 (2) 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. □

6.4 Every even order and the literal numerical gap

For even N, choose the largest positive multiple of three n with $16n \le N$. Then $r = N - 16n \in \{0, 2, \dots, 46\}$. 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 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. 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,

$$\operatorname{dist}_{\mathrm{iso}}(G_b \sqcup rK_1, G_0 \sqcup rK_1) \le \operatorname{dist}_{\mathrm{iso}}(P_b, P_0) + 322. \tag{6}$$

We use short-cycle recognition only before these surgeries.

Since $|C| \le 2^{n/3}$, the Hamming-ball estimate gives

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

The exponent is positive: $H_2(1/6) < 2/3$ is equivalent to $6^6 < 2^4 5^5$, or $46656 < 50000$. Hence uniform b has distance greater than $n/6$ with probability at least $3/4$ for large n. Proposition 11, including its isolate extension, and (6) give $\operatorname{dist}_{\mathrm{iso}}(P_b, P_0) > n/4 - 322$ 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.

Lemma 4's transcript proof applies unchanged to the independent rows of the new matrix, with $m = n/3$. 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 $1/2$, exactly as in Section 5. Identical adaptive transcript laws contradict this. Since $n = \Theta(N)$, Theorem 2 follows at literal proximity 0.01.

7 An obstruction to capped completion

A natural alternative is to make the original F_b cubic by attaching a constant-size cap at each degree-two site. We show that this particular completion family contains no 0.01-far inputs, regardless of how its lower distance estimate is sharpened.

Let Q_5 be K_4 on u, v, c, d with edge uv subdivided by a port p. The port has degree two and all other vertices degree three. Let Q_7 replace u by a triangle, attaching its former neighbors p, c, d at distinct triangle vertices. Both caps are simple and connected with exactly one degree-two port. The subdivided-K_4 completion cap and triangle expansion are standard constructions; see [9]. Attach one disjoint Q_5 or Q_7 by its port to every degree-two vertex of F_b. Use identical cap types at the paired sites $z(v, 0, p), z(v, 1, p)$, with every choice fixed independently of b. Write $\widehat{F}_b$ for this cubic completion and M for its order. Since there are $21n$ sites, $M \geq 38n + 5(21n) = 143n$.

Proposition 12. *For every b and every such paired cap assignment,*

$$\operatorname{dist}_{\mathrm{iso}}(\widehat{F}_b, \widehat{F}_0) \leq 4 \operatorname{dist}_H(b, C) \leq 2n, \qquad \frac{\operatorname{dist}_{\mathrm{iso}}(\widehat{F}_b, \widehat{F}_0)}{3M/2} \leq \frac{4}{429} < \frac{1}{100}.$$

Proof. The assignment-flip formulas extend to $F_b \cong F_{b+Ax}$ for every b, x, by the same third-coordinate calculation used for G_b. Paired cap types allow identical extensions over all cap vertices. Choose x minimizing the weight of $c = b + Ax$.

Map $\widehat{F}_c$ to $\widehat{F}_0$ by fixing all variable, middle and cap vertices and interchanging $o(i, 3, 0)$ with $o(i, 3, 1)$ exactly when $c_i = 1$. Every internal row edge is preserved: the third source bit $c_i + u + w$ becomes $u + w$. At each such row the two straight external edges are replaced by two crossed ones. These are four distinct unordered edges in their symmetric difference. Different rows have distinct outer vertices and occurrence sockets, so the discrepancy sets are disjoint. All other edges agree, giving exactly $4 \operatorname{wt}(c)$. This proves the first upper bound after composition with the assignment flip.

Every row of A has odd weight, so $A\mathbf{1} = \mathbf{1}$ and both 0 and $\mathbf{1}$ lie in C. Consequently $\operatorname{dist}_H(b, C) \leq \min(\operatorname{wt}(b), n - \operatorname{wt}(b)) \leq n/2$. Finally $M \geq 143n$ gives the displayed normalized bound. □

The obstruction concerns this capped input family, including paired Q_7 size adjustments. It does not exclude other cubic encodings. The no-rung construction of Section 6 avoids that size penalty and reaches literal proximity 0.01 through a different decoding argument; sharpening the capped-family lower estimate alone could not do so.

8 Finite verification

The accompanying standard-library checks test finite consequences of the analytic arguments. The original-encoding programs check exact code image distances, small $(3, 6)$ examples, fixed-slot oracle answers and selected arbitrary-map perturbations. Separate configuration enumerations test collision moments, component probabilities and expansion arithmetic. The earlier compact/capped programs document the cap repair and single-cycle anchoring approach; those historical checks are not used as verification of the changed nine-cycle geometry.

The literal-endpoint supplement `verify_literal.py` constructs the no-rung nine-cycle encoding independently of the earlier compact constructor. It checks all six-, eight- and nine-cycle types in a small exhaustive word example, actual fixed-slot oracle answers, and exact codeword isomorphisms. On selected arbitrary bijections, including middle intrusions and isolate mixing, it computes actual unmatched source edges, two-defect row anchors, the adaptive middle colors and joint $c+B$ budget. Exact rational Bernoulli expectations check the median, external-error, and final $2K/3$ inequalities. All 24 even padding residues are constructed, including adjacent expanded sites, with actual oracle, completeness and named edit comparisons. The supplement also rejects modified colors, capacity counts, subdivision rungs and bit readers through its positive exception-based gates. Reader instructions record source dependencies and normal and optimized runs.

These are finite checks, not an exhaustion of arbitrary bijections or a proof of asymptotic matrix existence. The literal proximity-0.01 conclusion rests on Proposition 11, Appendix B, the entropy bound and adaptive transcript argument. AI tools and computational checks assisted the work.

A Reconstruction of the sparse-matrix input

We reconstruct the matrix argument of Dror [1], including an elementary proof of positive simplicity probability.

Give each of n left vertices three labeled stubs and each of $n/2$ right vertices six. Match the $3n$ left stubs to right stubs uniformly. Let $(z)_k = z(z-1)\cdots(z-k+1)$.

Lemma 13. *The probability that the configuration has no parallel edges tends to e^{-5}.*

Proof. A candidate collision specifies a left vertex, two of its stubs, a right vertex, two of its stubs, and one of their two pairings. There are $n\binom{3}{2}(n/2)\binom{6}{2}\,2 = 45n^2$ candidates, each occurring with probability $1/(3n)_2$. Let X count occurring candidates. Double edges contribute one event, triple edges three; simplicity is $X = 0$.

Fix k. In the expansion of $\mathbb{E}(X)_k$ over ordered distinct candidate events, tuples with distinct left vertices and distinct right vertices number $(45n^2)^k + O_k(n^{2k-1})$ and prescribe $2k$ distinct stub pairings. Their contribution tends to 5^k.

For a remaining compatible tuple, group its distinct prescribed pairings by occupied vertex pair, or cell. Each cell contains at least two pairings. A left vertex has three stubs and hence belongs to at most one such cell. With q cells, the numbers ℓ, s of distinct left and right vertices and the number e of pairings satisfy $\ell = q$, $s \le q$, $e \ge 2q$. Equality $\ell + s - e = 0$ forces two pairings per cell and distinct right vertices. Each cell then determines one candidate; distinct tuple entries put this case in the principal class already counted. Every exceptional compatible pattern has $\ell + s - e \le -1$. There are finitely many patterns for fixed k, each with $O_k(n^{\ell+s})$ realizations of probability $1/(3n)_e = O_k(n^{-e})$. Their contribution is $O_k(n^{-1})$. Incompatible tuples have probability zero, so $\mathbb{E}(X)_k \to 5^k$ for every fixed k.

Finite Bonferroni bounds give, for fixed h,

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

Let $n \to \infty$ first and then $h \to \infty$. Both bounds converge to e^{-5}. □

For a left set S with $s = |S| \ge 2$ put $t = \lfloor 3s/2 \rfloor$. When $t \le n/2$, failure of $|N(S)| > 3s/2$ forces all $3s$ matched stubs into some t-vertex right set. The fixed-set probability is at most

$$(6t)_{3s}/(3n)_{3s} \le (2t/n)^{3s}.$$

Thus the union bound at size s is at most

$$\binom{n}{s}\binom{n/2}{t}(2t/n)^{3s} \le [C\sqrt{s/n}]^s$$

for an absolute C. To see the last estimate, write $z = t/s \in [1, 3/2]$ and apply the standard binomial estimates to obtain $[e^{1+z}(2z)^{3-z}(s/n)^{2-z}]^s$. Choose $\delta < 1/3$ with $C\sqrt{\delta} \le 1/2$. For $s \le \sqrt{n}$ these summands are bounded by $(Cn^{-1/4})^s$, whose sum from $s = 2$ is $o(1)$; for $s > \sqrt{n}$ they are at most 2^{-s} and again sum to $o(1)$.

By Lemma 13, simple configurations without a failing set of size 2 through δn have positive probability for large n. Singletons have three distinct neighbors in a simple graph. For every nonempty such S, its $3|S|$ edges and $|N(S)| > 3|S|/2$ force a neighbor incident to exactly one edge from S. The binary incidence matrix consequently has no zero sum of a nonempty set of at most $\lfloor \delta n \rfloor$ rows. This proves Lemma 3.

B Connected $(3, 9)$ systems

We prove Lemma 8 by the same sparse-configuration method, with the changed degree and constants made explicit. 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 is specified as in Lemma 13. 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 $\mathbb{E}(X)_k \to 8^k$. The finite Bonferroni bounds displayed in the preceding appendix, first letting $n \to \infty$ and then the truncation increase, give $\Pr[X = 0] \to e^{-8} > 0$.

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 8 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.