hexagon:2611.01037v1

Local Graph Limits Do Not Determine Normalized Adjacency Rank over $\mathbb{F}_2$

Yuichi Yoshida
National Institute of Informatics
yyoshida@nii.ac.jp

Research draft

Abstract

We construct two sequences of finite simple bipartite graphs of maximum degree three with the same Benjamini–Schramm limit and different limiting normalized adjacency ranks over $\mathbb{F}_2$. This disproves the finite-field rank convergence conjecture of Abért and Szegedy in characteristic two. On these graph sequences, a fixed radius-one XOR rule applied to independent uniform vertex bits has different limiting output entropies per vertex. We also give two sequences of bounded-degree two-dimensional simplicial complexes with the same local limit and different limiting second Betti numbers per vertex over $\mathbb{F}_2$.

For some absolute $\varepsilon > 0$, estimating the rank over $\mathbb{F}_2$ of an $N \times N$ binary matrix to within εN requires $\Omega(N)$ row or column queries, even with at most three ones per row and column. For a bounded-degree two-dimensional complex on M vertices, estimating the second Betti number over $\mathbb{F}_2$ to within εM requires $\Omega(M)$ queries, each of which returns all simplices containing a specified vertex. Both bounds hold for randomized adaptive algorithms that succeed with probability at least $2/3$ on every input.

1 Introduction and main results

A local limit records, for each radius, the limiting distribution of the neighborhood of a uniformly chosen vertex. Which global quantities have limits determined by this information? One example for bounded-degree graphs is the normalized rank of the adjacency matrix over $\mathbb{R}$. The empirical spectral measure puts equal mass on each eigenvalue, counted with multiplicity, so its mass at zero is one minus the normalized rank. Under local convergence, this mass converges [3, Theorem 4]. Abért asked whether normalized adjacency rank converges over finite fields [1, Question 12]. Mészáros records this as a conjecture of Abért and Szegedy [9, Introduction, after Theorem 1.9].

We answer the question in characteristic two. Even for simple bipartite graphs of maximum degree three, two sequences can have the same local limit while their normalized adjacency ranks over $\mathbb{F}_2$ approach different values.

Local convergence. For a finite graph G, let $\nu_R(G)$ be the distribution of the unlabelled rooted radius-R neighborhood of a uniformly chosen vertex. A sequence with $|V(G_j)| \to \infty$ is Benjamini–Schramm convergent if $\nu_R(G_j)$ converges for every fixed R. Its limit is a probability distribution on rooted graphs, specified by these neighborhood distributions. All graph sequences below have a common degree bound; disconnected graphs are allowed. For a field $\mathbb{K}$, write

$$r_{\mathbb{K}}(G) = \frac{\operatorname{rank}_{\mathbb{K}} \operatorname{Adj}(G)}{|V(G)|}.$$

Unless stated otherwise, matrix ranks in the proofs are over $\mathbb{F}_2$.

Theorem 1.1 (Failure of normalized rank convergence)**.** *There are two sequences G_j^0, G_j^1 of finite simple bipartite graphs of maximum degree three, a common Benjamini–Schramm limit μ, and constants $a_0 < a_1$ such that*

$$G_j^0 \longrightarrow \mu, \qquad G_j^1 \longrightarrow \mu, \qquad r_{\mathbb{F}_2}(G_j^i) \longrightarrow a_i \quad (i = 0, 1).$$

We may take $|V(G_j^0)| = |V(G_j^1)|$ for every j. Consequently, there is a Benjamini–Schramm convergent graph sequence whose normalized adjacency rank over $\mathbb{F}_2$ does not converge. The same conclusions hold over every field of characteristic two.

This disproves the Abért–Szegedy conjecture in characteristic two, already for unweighted simple graphs. In fact, the same local limit can occur with any limiting normalized rank in a nondegenerate interval; see corollary 6.2.

1.1 How the construction works

View a square binary matrix A as a bipartite graph with one vertex per row, one per column, and an edge for each nonzero entry. If A has dimension N, its graph has $2N$ vertices and adjacency rank $2 \operatorname{rank} A$, so its normalized rank is $\operatorname{rank} A/N$. A query for the neighbors of a vertex reads one row or column of A. Exploring a ball of radius R thus takes a number of queries that depends only on R and the degree bound. We seek two distributions on sparse matrices such that exploring a ball of any fixed radius gives the same distribution of queries and answers in both cases once the matrices are sufficiently large. At the same time, the rank of every matrix drawn from the second distribution should exceed that of every matrix drawn from the first by a positive fraction of the matrix dimension.

To see how information can be hidden from coordinate queries, consider a uniformly random binary vector of even parity and one of odd parity. The two distributions agree on every proper subset of the n coordinates. Simply putting these bits on the diagonal of a matrix would give little control over rank. Flipping one bit changes parity and changes the rank by just one. We seek two cosets with the same projections onto small sets of coordinates. We will associate a sparse matrix with each vector so that its rank differs by order n depending on which coset contains the vector. We use the construction of Li, Li, Han, and Liu [8] for both purposes. Its distance bound gives subspaces $B \subseteq Z_3 \subseteq \mathbb{F}_2^n$ with the following property. For any shift $b_* \in Z_3$, a uniform vector in B and a uniform vector in $b_* + B$ have the same distribution on every set of fewer than δn coordinates, for a constant $\delta > 0$. Fix the random choices of an adaptive algorithm. Each possible transcript (the sequence of queries and answers) occurs precisely when the queried coordinates take the specified values. An algorithm that reads fewer than δn coordinates therefore has the same transcript distribution when the vector is sampled from either coset.

The same construction gives sparse matrices H_1, H_2 with kernels Z_1, Z_2 and a sparse trilinear form $T(x, y, b)$. On $Z_1 \times Z_2$, the form $T(\cdot, \cdot, b)$ vanishes for every $b \in B$. There are vectors $u_i \in Z_1$, $v_i \in Z_2$, and $z_i \in Z_3$ for $1 \le i \le m$, where $m \ge \rho n$ for a constant $\rho > 0$, such that $T(u_i, v_j, z_k)$ equals one exactly when $i = j = k$. Choose $b_* = \sum_{k=1}^m z_k$. Then

$$(T(u_i, v_j, b_*))_{i,j=1}^m = I_m.$$

Adding an element of B to b_* leaves the restricted form unchanged, so it has rank at least m throughout $b_* + B$ and rank zero throughout B.

The restriction may have a dense matrix in bases of Z_1, Z_2. Write $T(x, y, b) = x^{\mathsf{T}} D_b y$. In each H_a, retain a maximal linearly independent set of its rows. We encode the restriction with the sparse block matrix

$$M_b = \begin{pmatrix} H_2 & 0 \\ D_b & H_1^{\mathsf{T}} \end{pmatrix}.$$

In $M_b(y,z)=0$, the first block row requires $y \in Z_2$. The second requires $D_b y \in \operatorname{im} H_1^{\mathsf{T}} = Z_1^{\perp}$, which says precisely that $T(x,y,b)=0$ for every $x \in Z_1$. For each such y, the vector z is unique. It follows that $\operatorname{rank} M_b$ equals the rank of the restriction plus $\operatorname{rank} H_1 + \operatorname{rank} H_2$. Each row or column of M_b can be read using only a constant number of coordinates of b.

To reach degree three, we give each occurrence of a variable its own copy, equate successive copies, and split long equations into equations for partial sums. We allocate the copies for every position that could be nonzero, so the row and column labels are the same for every b. Each original solution extends uniquely to the new variables, so the nullity is preserved. Since the number of added columns is independent of b, the difference in ranks is preserved as well. Padding with zeros gives square matrices A_b of dimension $\Theta(n)$. A row or column still depends on only a constant number of bits of b.

Finally, for each coset take the disjoint union of the graphs of A_b over all b in that coset. All summands have the same order, so choosing a uniform root samples a uniform b from the coset and then a uniform vertex of its graph. For any fixed radius and all sufficiently large n, exploring the ball reads fewer than δn coordinates of b. The two unions therefore have exactly the same rooted ball distribution at that radius. Their normalized ranks are the averages over their respective summands, so the rank gap persists. Passing to a subsequence gives the two limits in theorem 1.1.

1.2 Output entropy of the XOR rule

Put independent uniform bits X_v on the vertices and apply the same rule on every graph:

$$Y_v = \bigoplus_{u \sim v} X_u. \tag{1}$$

The empty XOR is zero, and the rule has radius one. Color each vertex by the pair (X_v, Y_v). The distribution of rooted colored neighborhoods averages over both the independent input bits and an independently chosen uniform root; the underlying graph is deterministic. Convergence of these distributions means convergence at every fixed radius. Write $H(Y)$ for the joint Shannon entropy, in bits, of the entire output vector.

Corollary 1.2. *For the graph sequences in theorem 1.1, the distributions of rooted neighborhoods in the colored graphs (G_j^i, X, Y) have the same limit, but their output entropies per vertex satisfy*

$$\frac{H(Y \text{ on } G_j^i)}{|V(G_j^i)|} \longrightarrow a_i \quad (i = 0, 1).$$

Thus normalized output entropy need not converge along a locally convergent sequence, even for the single rule equation (1).

Abért and Szegedy conjectured that, for a fixed finite automaton applied to an i.i.d. vertex coloring, the output entropy per vertex converges along every Benjamini–Schramm convergent graph sequence [2]. The XOR rule gives a counterexample because $H(\operatorname{Adj}(G)X) = \operatorname{rank}_{\mathbb{F}_2} \operatorname{Adj}(G)$. Indeed, $\operatorname{Adj}(G)X$ is uniform on the image of the adjacency matrix, which has $2^{\operatorname{rank}_{\mathbb{F}_2} \operatorname{Adj}(G)}$ elements.

1.3 Second Betti numbers

The same matrices yield a topological example. In section 8 we encode solutions of $A^{\mathsf{T}}x = 0$ by sets of triangles with zero boundary. A two-dimensional simplicial complex K consists of vertices, edges, and triangles, with all faces included. Over $\mathbb{F}_2$, its boundary matrix ∂_2 has one row per edge and one column per triangle, with an entry equal to one when the edge belongs to the triangle. Since there are no three-dimensional simplices, its second Betti number is

$\beta_2(K;\mathbb{F}_2) = \dim\ker\partial_2$. Equivalently, this is the dimension of the space of triangle sets in which every edge occurs an even number of times. For a simplicial complex, degree means degree in its 1-skeleton, the graph formed by its vertices and edges. A rooted radius-R neighborhood is the induced subcomplex on the radius-R vertex ball. Local convergence is defined by these rooted neighborhood distributions.

Theorem 1.3. *There are an absolute degree bound d and two sequences of finite two-dimensional simplicial complexes K_j^0, K_j^1 of maximum vertex degree at most d with the same local limit, but*

$$\frac{\beta_2(K_j^i;\mathbb{F}_2)}{|V(K_j^i)|} \longrightarrow b_i, \qquad b_0 \neq b_1.$$

In particular, the normalized second Betti number over $\mathbb{F}_2$ need not converge under local convergence. The same examples work over every field of characteristic two.

Elek proved that normalized Betti numbers with real coefficients converge under local convergence and can be estimated to any fixed additive accuracy using a number of queries independent of the size of the complex [6]. His result does not cover mod-p coefficients. Our examples show that the corresponding conclusions fail in characteristic two.

1.4 Query lower bounds

The two matrix distributions also give a linear lower bound on the number of row or column queries needed to estimate rank. We transfer this bound to second Betti numbers by answering each vertex query with at most four matrix queries. For a matrix, row and column weights mean the numbers of nonzero entries in the corresponding row and column. The algorithm knows the dimension and the bounds on these weights; a query specifies a row or column label and returns all nonzero entries with their indices. For a simplicial complex, the vertex set and degree bound are known. A *star query* at a vertex returns all simplices containing it, with their vertex labels. Computation is free, and later queries may depend on earlier answers.

Theorem 1.4 (Linear query lower bounds)**.** *There are absolute constants $\varepsilon_0, \varepsilon_1 > 0$ with the following properties.*

(i) For every sufficiently large N, estimating the rank of an $N \times N$ binary matrix of row and column weights at most three to absolute error at most $\varepsilon_0 N$ requires $\Omega(N)$ row or column queries.

(ii) For some absolute constant d, estimating $\beta_2(K;\mathbb{F}_2)$ to absolute error at most $\varepsilon_1|V(K)|$ on two-dimensional complexes of maximum vertex degree at most d requires $\Omega(|V(K)|)$ star queries.

In both statements, the estimate must meet the stated error bound with probability at least $2/3$ on every input. The bounds hold for randomized adaptive algorithms with unrestricted computation, both for worst-case expected query complexity and for worst-case query complexity.

Problem 68 of the sublinear algorithms problem list [11] asks how many queries suffice to estimate sparse matrix rank. If every row of a binary matrix has exactly two ones, the rows form the edges of a graph on the columns. With n columns and c connected components, the matrix has rank $n - c$. At bounded degree, Chazelle, Rubinfeld, and Trevisan [5] estimate c to within εn using a number of queries independent of n for each fixed $\varepsilon > 0$.

Bogdanov, Obata, and Trevisan [4] proved a linear query lower bound for testing satisfiability of bounded-degree systems $Ax = b$ over $\mathbb{F}_2$ with three variables per equation. Only the right-hand side b varies in their construction, so $\operatorname{rank} A$ is constant. Theorem 1.4(i) proves the linear lower

bound conjectured in Problem 68 for estimating rank over $\mathbb{F}_2$, even with at most three ones per row and column.

Section 2 states the result of Li, Li, Han, and Liu used to construct the two matrix distributions in sections 3 to 5. Section 6 uses these matrices to prove theorem 1.1. Sections 7 and 8 treat the output entropy of the XOR rule and second Betti numbers. Section 9 proves the query lower bounds.

2 Subspaces and a trilinear form

The construction needs two properties of the same pair of cosets in a binary space Z_3. First, a parity check that vanishes on the subspace $B \subseteq Z_3$ and takes the value one somewhere in Z_3 must involve at least δn coordinates. Thus B and its translates within Z_3 have the same projection onto any set of fewer than δn coordinates (section 3). Second, the restriction of $T(\cdot, \cdot, b)$ to the solution spaces Z_1, Z_2 must vanish on one coset and have rank at least ρn on another (section 4). The sparsity conditions below let us compute each row or column of the resulting matrix from a constant number of coordinates of b.

For a vector v, write $\mathrm{wt}(v) = |\operatorname{supp}(v)|$. Orthogonal complements are with respect to the standard binary dot product. We use the convention $\min \varnothing = +\infty$ when defining minimum weights.

Definition 2.1 (Algebraic assumptions)**.** Fix constants $K \geq 1$, integers $w, \tau \geq 1$, and $\rho, \delta > 0$. We require the following objects and conditions, with size parameter n.

(i) Subspaces $Z_a \subseteq \mathbb{F}_2^{n_a}$ for $a = 1, 2, 3$, with $n_3 = n$ and $n_1, n_2 \leq Kn$. For $a = 1, 2$, there is a binary matrix H_a with $Z_a = \ker H_a$ whose row and column weights are at most w.

(ii) A subspace $B \subseteq Z_3$ satisfying the minimum-weight bound

$$\min\{\mathrm{wt}(v) : v \in B^{\perp} \setminus Z_3^{\perp}\} \geq \delta n. \tag{2}$$

(iii) A trilinear form $T : \mathbb{F}_2^{n_1} \times \mathbb{F}_2^{n_2} \times \mathbb{F}_2^{n} \to \mathbb{F}_2$ given by

$$T(x, y, b) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n} t_{ijk} x_i y_j b_k, \qquad t_{ijk} \in \mathbb{F}_2, \tag{3}$$

such that each coordinate, in each of the three positions, occurs in at most τ nonzero monomials. Moreover,

$$T(x, y, s) = 0 \qquad (x \in Z_1,\ y \in Z_2,\ s \in B). \tag{4}$$

(iv) An integer $m \geq \rho n$ and vectors $u_1, \ldots, u_m \in Z_1$, $v_1, \ldots, v_m \in Z_2$, and $z_1, \ldots, z_m \in Z_3$ satisfying

$$T(u_i, v_j, z_k) = \begin{cases} 1, & i = j = k, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

The query algorithm may know all the objects in this definition. Only the vector b sampled from a coset is hidden.

Theorem 2.2 (From Li, Li, Han, and Liu [8])**.** *For some choice of the constants in definition 2.1, the conditions of that definition hold for an unbounded increasing sequence of values of n. The ratio of each value to the preceding one is at most a constant $\lambda > 1$.*

Arbitrarily large values of n suffice for the graph sequences. The ratio bound lets us pad a matrix to any sufficiently large dimension while keeping the rank difference proportional to that dimension (section 9). Appendix A derives theorem 2.2 from [8].

3 Cosets that are indistinguishable under adaptive queries

Fix a set S of coordinates. If the projections of B and Z_3 onto S differed, a parity check supported on S would witness the difference. This observation gives equality of the projected distributions on small sets. We then apply it to the set of coordinates queried along each possible execution of an adaptive algorithm. We write $\mathrm{Unif}(U)$ for the uniform distribution on a finite nonempty set U.

Lemma 3.1 (Equality of coordinate projections)**.** *Let $B \subseteq Z \subseteq \mathbb{F}_2^n$ be linear subspaces, and let*

$$d = \min\{\mathrm{wt}(v) : v \in B^\perp \setminus Z^\perp\}.$$

Here $d = +\infty$ if $B = Z$. For every $S \subseteq [n]$ with $|S| < d$, the coordinate projection $\pi_S : \mathbb{F}_2^n \to \mathbb{F}_2^S$ satisfies $\pi_S(B) = \pi_S(Z)$. Consequently, for every $b_* \in Z$,

$$\pi_S(\mathrm{Unif}(B)) = \pi_S(\mathrm{Unif}(b_* + B)). \tag{6}$$

Proof. If $\pi_S(B)$ were a proper subspace of $\pi_S(Z)$, there would be a linear functional on $\mathbb{F}_2^S$ vanishing on the former but not the latter. Extending its coefficient vector by zero outside S gives $v \in B^\perp \setminus Z^\perp$ of weight at most $|S|$, a contradiction. The projection of a uniform vector in B is uniform on the subspace $\pi_S(B)$, since all fibers have equal cardinality. Finally, $\pi_S(b_*) \in \pi_S(B)$, so translating by $\pi_S(b_*)$ does not change that uniform distribution. □

Lemma 3.2 (Identical transcript distributions)**.** *Under the hypotheses of lemma 3.1, every randomized adaptive algorithm that reads fewer than d coordinates of its input on every execution has identical transcript distributions under $\mathrm{Unif}(B)$ and $\mathrm{Unif}(b_* + B)$.*

Proof. First fix the internal random tape. A leaf of the resulting decision tree specifies a list of queried coordinates and their returned values. The event of reaching that leaf is precisely the event that the input agrees with those values on the queried coordinates. Repeated queries impose no additional coordinates. By equation (6), this event has the same probability under the two input distributions. This applies to every leaf, including leaves reached by early stopping. Averaging over the random tape proves the claim. Since equality holds for each tape, any output computed from the tape and the transcript also has the same law under both input distributions. □

4 Constructing matrices whose ranks differ between the cosets

By lemma 3.2, we can choose any shift $b_* \in Z_3$ and keep the cosets indistinguishable to algorithms reading fewer than δn coordinates. We now choose the shift to make the rank of $T(\cdot, \cdot, b)$ on $Z_1 \times Z_2$ differ between the cosets. We then construct a sparse matrix with the same rank difference. Each of its rows and columns can be computed from at most τ coordinates of b. Throughout this section, $Z_a = \ker H_a$ for $a = 1, 2$, as in definition 2.1.

Put $r_a = \mathrm{rank}\, H_a$, and let $G_a \in \mathbb{F}_2^{n_a \times (n_a - r_a)}$ have columns forming a basis of Z_a, for $a = 1, 2$. These possibly dense matrices are used only in the analysis. For $b \in \mathbb{F}_2^n$, define matrices $D_b \in \mathbb{F}_2^{n_1 \times n_2}$ and $K_b \in \mathbb{F}_2^{(n_1 - r_1) \times (n_2 - r_2)}$ by

$$(D_b)_{ij} = \sum_{k=1}^{n} t_{ijk} b_k, \qquad K_b = G_1^\mathsf{T} D_b G_2. \tag{7}$$

Thus $x^\mathsf{T} D_b y = T(x, y, b)$, and K_b represents the restriction of this bilinear form to $Z_1 \times Z_2$. Each row and column of D_b has at most τ potential nonzero positions. A complete row or column can be computed by reading at most τ coordinates of b.

For each k, the matrix $(T(u_i, v_j, z_k))_{i,j=1}^m$ in equation (5) has a single one, in position (k, k). To obtain all m diagonal entries at once, set

$$b_* := \sum_{k=1}^{m} z_k, \qquad \mu_0 := \mathrm{Unif}(B), \qquad \mu_1 := \mathrm{Unif}(b_* + B). \tag{8}$$

Lemma 4.1 (Rank of K_b on the two cosets)**.** *For every $b \in B$, $K_b = 0$. For every $b \in b_* + B$, $K_b = K_{b_*}$ and $\operatorname{rank} K_{b_*} \geq m$.*

Proof. The first assertion is exactly equation (4), applied to basis vectors of Z_1 and Z_2. Linearity in the third argument then gives $K_{b_*+s} = K_{b_*}$ for $s \in B$. By equation (5),

$$T(u_i, v_j, b_*) = \sum_{k=1}^{m} T(u_i, v_j, z_k) = \mathbf{1}_{i=j}.$$

Let $P \in \mathbb{F}_2^{(n_1-r_1)\times m}$ and $Q \in \mathbb{F}_2^{(n_2-r_2)\times m}$ have as columns the coordinates of u_i and v_i in the bases G_1 and G_2, respectively. Then $P^{\mathsf{T}} K_{b_*} Q = I_m$, so $\operatorname{rank} K_{b_*} \geq m$. □

We next construct a sparse matrix whose rank is $\operatorname{rank} K_b$ plus a term independent of b. We encode $\ker K_b$ with sparse equations, avoiding the possibly dense bases G_1, G_2 in the input. For $y \in Z_2$, the condition $T(x, y, b) = 0$ for every $x \in Z_1$ is equivalent to $D_b y \in Z_1^{\perp} = \operatorname{im} H_1^{\mathsf{T}}$. To encode this condition with a unique auxiliary vector, first retain a maximal independent subset of the rows of each H_a, $a = 1, 2$, and continue to call the resulting $r_a \times n_a$ matrix H_a. This preserves its kernel and sparsity, while making H_a^{T} injective. We choose the rows using H_a alone. Now $D_b y \in Z_1^{\perp}$ holds exactly when there is a unique $z \in \mathbb{F}_2^{r_1}$ with $D_b y + H_1^{\mathsf{T}} z = 0$. Together with $H_2 y = 0$, this equation gives the matrix

$$M_b = \begin{pmatrix} H_2 & 0 \\ D_b & H_1^{\mathsf{T}} \end{pmatrix} \in \mathbb{F}_2^{(n_1+r_2)\times(n_2+r_1)}. \tag{9}$$

Lemma 4.2 (Rank of M_b)**.** *For every $b \in \mathbb{F}_2^n$,*

$$\operatorname{rank} M_b = r_1 + r_2 + \operatorname{rank} K_b. \tag{10}$$

Proof. Every $y \in Z_2$ can be written uniquely as $G_2 a$ with $a \in \mathbb{F}_2^{n_2-r_2}$. Since $K_b = G_1^{\mathsf{T}} D_b G_2$, the condition $K_b a = 0$ says exactly that $D_b y \in Z_1^{\perp}$. For each such a there is a unique z with $H_1^{\mathsf{T}} z = D_b y$. The map $a \mapsto (G_2 a, z)$ is thus a linear isomorphism from $\ker K_b$ to $\ker M_b$. Therefore

$$\dim \ker M_b = (n_2 - r_2) - \operatorname{rank} K_b.$$

Subtracting from the number $n_2 + r_1$ of columns proves the identity. □

Proposition 4.3 (Two distributions on sparse matrices)**.** *Sampling $b \sim \mu_i$ and returning M_b defines a matrix distribution for each $i = 0, 1$. These distributions satisfy the following properties:*

(i) $\operatorname{rank} M_b = r_1 + r_2$ *for $b \in B$, and* $\operatorname{rank} M_b = r_1 + r_2 + \operatorname{rank} K_{b_*}$ *for $b \in b_* + B$, where* $\operatorname{rank} K_{b_*} \geq \rho n$*;*

(ii) each row and column has at most $D := w + \tau$ positions that can be nonzero as b varies;

(iii) every full row or column query is simulated using at most τ coordinate queries to b;

(iv) every adaptive matrix algorithm using q queries with $\tau q < \delta n$ has exactly the same transcript distribution under the two distributions.

Proof. The rank formula follows from lemmas 4.1 and 4.2. Each row or column of M_b has at most w nonzero entries from H_2 or H_1^{T}, both of which are known, and at most τ potential nonzero positions from D_b. By equation (7), row i of D_b depends only on those b_k for which $t_{ijk} = 1$ for some j. At most τ nonzero monomials contain x_i, so there are at most τ such indices k. The same argument for column j counts the monomials containing y_j. Thus q matrix queries can be simulated by at most τq coordinate queries. When $\tau q < \delta n$, lemma 3.2 and equation (2) give identical distributions for the simulated transcripts. □

5 Reducing row and column weights to three

Yuster [12, Theorem 1.1] transforms a square matrix into a square matrix of the same nullity with at most three nonzero entries per row and column. Edelsbrunner and Parsa [7, Theorem 4.1] give a related reduction for computing the rank of a binary matrix. For our family M_b, we also need row and column labels independent of b and a way to read each row or column from a constant number of coordinates of b.

We modify the equations of M_b in two steps. We limit the number of occurrences of each variable by giving its occurrences separate copies and equating successive copies. We then split each long equation into equations for partial sums. Every original solution will extend uniquely to the new variables, which is enough to preserve the rank difference as long as the new dimensions are the same for every b.

For this reason we allocate copies for all positions that *can* be nonzero. Allocating them only for entries that are one could make the dimensions themselves reveal information about b before the first query. The following lemma applies to any family $M(b)$ with a known set of potential nonzero positions.

Lemma 5.1 (Reduction preserving nullity)**.** *Let $D \geq 1$ be an integer, and let $M(b) \in \mathbb{F}_2^{p \times q}$ be indexed by $b \in \mathbb{F}_2^n$. Suppose a set of positions $\mathcal{P} \subseteq [p] \times [q]$ is known to the algorithm and satisfies $\operatorname{supp} M(b) \subseteq \mathcal{P}$ for every b. The set $\mathcal{P}$ does not depend on b. Assume that $\mathcal{P}$ has at most D positions in each row and column, and every potentially nonzero coefficient can be computed from at most L coordinates of b. There is a matrix $\widehat{M}(b) \in \mathbb{F}_2^{p' \times q'}$ with $p', q' \leq (D+2)(p+q)$ such that:*

(i) every row and column has at most three nonzero entries;

(ii) $\dim \ker \widehat{M}(b) = \dim \ker M(b)$*, and hence*

$$\operatorname{rank} \widehat{M}(b) = q' - q + \operatorname{rank} M(b); \tag{11}$$

(iii) the dimensions and indexing do not depend on b;

(iv) each full row or column query to $\widehat{M}(b)$ is answered by reading at most L coordinates of b.

Adding zero rows or columns to make the matrix square of dimension $\max\{p', q'\}$ preserves its rank and the query bound in (iv).

Proof. We define $\widehat{M}(b)$ as the coefficient matrix of a new system: its columns are the variables below, and its rows are their constraints. For each original variable x_j in $M(b)x = 0$, create copies $x_{j,1}, \ldots, x_{j,D}$ and impose

$$x_{j,t} + x_{j,t+1} = 0 \qquad (1 \leq t < D). \tag{12}$$

Assign distinct copies to the potential occurrences of x_j in $\mathcal{P}$. Unused copies are retained. This assignment is known to the algorithm and does not depend on b.

List the positions of $\mathcal{P}$ in each row i in increasing column order. For the tth position in this list, let $a_{i,t}(b)$ be its coefficient and let $x(i,t)$ be the copy assigned to it. To accumulate the row sum, introduce variables $y_{i,0}, \dots, y_{i,D}$ and impose

$$y_{i,0} = 0, \qquad y_{i,D} = 0, \qquad y_{i,t-1} + y_{i,t} + a_{i,t}(b)x(i,t) = 0 \quad (1 \le t \le D). \tag{13}$$

If the list for row i has fewer than t positions, omit the last term in the tth equation.

Every equation has at most three variables. Each x-copy belongs to at most two equations in equation (12) and one equation in equation (13). Each y-variable belongs to exactly two equations, counting the endpoint conditions. Thus the column weights are also at most three.

The copy equations identify all copies of each original variable. Starting with $y_{i,0} = 0$, the recurrence in equation (13) determines the y-variables uniquely from the original variables. The additional condition $y_{i,D} = 0$ is precisely the original row equation. Restriction to one copy of each x_j is therefore a linear isomorphism between the two kernels. Counting equations and variables gives

$$p' = (D-1)q + (D+2)p, \tag{14}$$
$$q' = Dq + (D+1)p. \tag{15}$$

Both are at most $(D+2)(p+q)$, and rank-nullity gives equation (11).

A partial-sum equation contains at most one coefficient $a_{i,t}(b)$, and the copy $x(i,t)$ occurs in no other partial-sum equation. Thus each new row or column contains at most one coefficient that depends on b. Its position is known, and its value can be computed by reading at most L coordinates of b. Adding zero rows or columns preserves the rank and requires no further queries to b. □

Applying the lemma to M_b gives the matrices we use in both the graph construction and the query lower bound.

Corollary 5.2 (Two distributions on square matrices)**.** *Suppose the conditions of definition 2.1 hold for n. There is a family $A_b \in \mathbb{F}_2^{N_0 \times N_0}$, $b \in \mathbb{F}_2^n$, with row and column weights at most three. The dimension and row and column labels are independent of b, and $\rho n \le N_0 \le Cn$ for a constant C depending only on the constants in definition 2.1. Each full row or column query to A_b can be simulated using at most τ coordinates of b.*

Let $\mu_0 = \mathrm{Unif}(B)$ and $\mu_1 = \mathrm{Unif}(b_* + B)$, with $b_* = \sum_{k=1}^m z_k$ as in equation (8). The rank of A_b is $R_0(n)$ for every $b \in B$ and $R_1(n)$ for every $b \in b_* + B$, where $R_1(n) - R_0(n) \ge \rho n$. Every adaptive algorithm making q queries to A_b, with $\tau q < \delta n$, has the same transcript distribution under $b \sim \mu_0$ and $b \sim \mu_1$.

Proof. For equation (9), put $p = n_1 + r_2$ and $q = n_2 + r_1$. Then

$$p + q \le 2(n_1 + n_2) \le 4Kn.$$

There are at most $D = w + \tau$ potentially nonzero positions in each row and column, and each coefficient is computable from at most τ coordinates of b. Apply lemma 5.1 with $L = \tau$, obtaining a matrix with dimensions p', q'. Pad to a square of dimension $N_0 = \max\{p', q'\}$ and call the result A_b. Put

$$C = 4K(D+2). \tag{16}$$

Both p' and q' are at most

$$(D+2)(p+q) \le Cn.$$

By equation (11), the rank increases by $q' - q$ for every b. The two ranks therefore still differ by at least ρn. Since their difference is at most N_0, we also have $N_0 \ge \rho n$. The new query simulation still uses at most τ coordinates of b per query. Now apply lemma 3.2. □

6 The same local limit with different normalized ranks

The rank of A_b depends on which coset contains b. For an algorithm making fewer than $\delta n/\tau$ queries, however, the transcript has the same distribution under $b \sim \mu_0$ and $b \sim \mu_1$. To obtain a deterministic graph for each coset, we take the disjoint union of the graphs associated with its vectors.

For a square binary matrix A, let $G(A)$ have one vertex per row and one per column, with an edge between row i and column j exactly when $A_{ij} = 1$. It is a simple bipartite graph, and

$$\operatorname{Adj}(G(A)) = \begin{pmatrix} 0 & A \\ A^{\mathsf{T}} & 0 \end{pmatrix}. \tag{17}$$

Its kernel is $\ker A^{\mathsf{T}} \oplus \ker A$, so its rank is $2 \operatorname{rank} A$. If A is $N_0 \times N_0$, then $r_{\mathbb{F}_2}(G(A)) = \operatorname{rank} A/N_0$. A full adjacency-list query is a full row or column query to A.

Proof of theorem 1.1. For each n in the sequence of theorem 2.2, take the matrices A_b of corollary 5.2. Their common dimension satisfies $\rho n \le N_0 = N_0(n) \le Cn$. Their ranks are $R_0(n)$ for $b \in B$ and $R_1(n)$ for $b \in b_* + B$. Define two deterministic graphs by taking all vectors in each coset:

$$\Gamma_n^0 = \bigsqcup_{b \in B} G(A_b), \qquad \Gamma_n^1 = \bigsqcup_{b \in B} G(A_{b_*+b}). \tag{18}$$

The graphs have the same order, $2N_0|B|$, which tends to infinity. Every summand in Γ_n^i has adjacency rank $2R_i(n)$. By additivity of rank over disjoint unions,

$$r_{\mathbb{F}_2}(\Gamma_n^1) - r_{\mathbb{F}_2}(\Gamma_n^0) = \frac{R_1(n) - R_0(n)}{N_0} \ge \frac{\rho}{C}. \tag{19}$$

Now fix a radius R. Since the summands have equal order, a uniform root in Γ_n^i can be sampled by drawing $b \sim \mu_i$ and choosing a vertex independently from the common $2N_0$ row and column labels. Starting at that vertex, read the adjacency lists of the vertices in the radius-R ball, including those on its boundary, and retain the induced rooted graph. At degree three this takes at most q_R queries, where q_R depends only on R. For all sufficiently large n we have $\tau q_R < \delta n$, so corollary 5.2 gives identical distributions for the exploration transcripts under μ_0 and μ_1. Forgetting the labels gives

$$\nu_R(\Gamma_n^0) = \nu_R(\Gamma_n^1) \qquad \text{for all sufficiently large } n. \tag{20}$$

There are finitely many rooted radius-R types at a fixed degree bound, so their probability distributions form a compact simplex. A diagonal subsequence argument makes (Γ_n^0) locally convergent; equation (20) gives the same limit for (Γ_n^1). Both normalized ranks lie in $[0, 1]$, so along a further subsequence they converge to limits a_0 and a_1. By equation (19), $a_1 - a_0 \ge \rho/C > 0$. Relabel the subsequences G_j^0, G_j^1. Interleaving them preserves their common local limit but makes normalized rank nonconvergent.

Finally, a binary matrix has the same rank over $\mathbb{F}_2$ and over every extension field of characteristic two: the nonzero minors defining its rank remain nonzero, and all larger minors remain zero. This proves the last assertion. □

Remark 6.1. The disjoint unions in equation (18) can have exponentially many vertices in n. The query lower bound in section 9 concerns a single matrix A_b. Its dimension is $\Theta(n)$, so the threshold of $\delta n/\tau$ queries is linear in that dimension.

Corollary 6.2 (An interval of limiting normalized ranks)**.** *For the local limit μ in theorem 1.1, every $a \in [a_0, a_1]$ is the limiting normalized adjacency rank over $\mathbb{F}_2$ of some sequence of simple bipartite graphs of maximum degree three converging to μ.*

Proof. For $\theta \in [0,1]$, take the disjoint union of p_j copies of G_j^0 and q_j copies of G_j^1, with $p_j/(p_j+q_j) \to \theta$. Since $|V(G_j^0)| = |V(G_j^1)|$, the fraction of vertices in copies of G_j^0 is $p_j/(p_j+q_j)$. Both the rooted neighborhood distributions and the normalized ranks are averaged with these same weights. The local limit remains μ and the normalized rank converges to $\theta a_0 + (1-\theta)a_1$. □

Contrast with real rank. By [3, Theorem 4], the masses at zero of the real adjacency spectral measures converge along the interleaved sequence. Each such mass equals one minus the normalized real rank. Hence

$$\lim_j r_{\mathbb{R}}(G_j^0) = \lim_j r_{\mathbb{R}}(G_j^1), \qquad \lim_j r_{\mathbb{F}_2}(G_j^0) \neq \lim_j r_{\mathbb{F}_2}(G_j^1).$$

7 Entropy and the number of solutions

For the XOR rule equation (1), the entire output vector is $Y = \mathrm{Adj}(G)X$. We compute $H(Y)$ from the rank of $\mathrm{Adj}(G)$. To find the distribution of the colors (X_v, Y_v) in a ball of radius R, it suffices to know the underlying ball of radius $R+1$.

Lemma 7.1. *Let $A \in \mathbb{F}_2^{m\times n}$ and let X be uniform on $\mathbb{F}_2^n$. Then*

$$H(AX) = \operatorname{rank} A, \qquad \log_2 |\ker A| = n - \operatorname{rank} A.$$

Proof. Every element of $\operatorname{im} A$ has the same number of preimages, $|\ker A| = 2^{n-\operatorname{rank} A}$. Thus AX is uniform on the $2^{\operatorname{rank} A}$ elements of $\operatorname{im} A$ and has entropy $\operatorname{rank} A$ bits. □

Proof of corollary 1.2. Applying lemma 7.1 to $Y = \mathrm{Adj}(G)X$ gives

$$\frac{H(Y)}{|V(G)|} = r_{\mathbb{F}_2}(G).$$

The colors (X_v, Y_v) in a radius-R ball are determined by the underlying radius-$(R+1)$ ball and the independent input bits there. The distribution of a colored rooted radius-R ball is therefore determined by $\nu_{R+1}(G)$. The two graph sequences have the same limits of $\nu_{R+1}(G)$ for every R, so their distributions of rooted colored neighborhoods also have the same limits. By theorem 1.1, the output entropies per vertex tend to a_0 and a_1, respectively. □

We can also count the assignments $x \in \mathbb{F}_2^{V(G)}$ satisfying $\sum_{u\sim v} x_u = 0$ at every vertex v. These are precisely the vectors in $\ker \mathrm{Adj}(G)$, so there are $2^{|V(G)|-\operatorname{rank}\mathrm{Adj}(G)}$ solutions. The base-two logarithm of their number, divided by $|V(G)|$, is $1 - r_{\mathbb{F}_2}(G)$ and hence tends to $1 - a_i$ along (G_j^i).

8 From matrix rank to the second Betti number

We adapt the construction of Edelsbrunner and Parsa [7, Section 3] to obtain simplicial complexes of bounded vertex degree whose vertex set does not depend on the matrix entries. Let A be an $N \times N$ binary matrix with at most three ones per row and column. Represent each row by a triangulated surface and each column by a cycle. Attach each row surface to the cycles corresponding to the ones in that row. A set of row surfaces has zero boundary exactly when an even number of the chosen surfaces meet each column cycle. These are the equations $A^{\mathsf{T}}x = 0$, so the space of such choices has dimension $N - \operatorname{rank} A$.

To transfer the query lower bound, we must also keep the vertex labels and number of vertices independent of the entries of A, and answer each star query with a bounded number of row or column queries. We achieve this by reserving vertices for both a disk and a cylinder at each of three boundary cycles of a row surface. We use one of the two pieces and leave the unused vertices isolated.

Lemma 8.1. *There are absolute integers $c_{\mathrm{top}}, d_{\mathrm{top}}$ such that every $A \in \mathbb{F}_2^{N\times N}$ of row and column weights at most three has an associated two-dimensional simplicial complex $K(A)$ with:*

(i) exactly $c_{\mathrm{top}}N$ vertices, on a label set independent of A;

(ii) maximum vertex degree at most d_{top};

(iii) $\beta_2(K(A);\mathbb{F}_2) = N - \operatorname{rank} A$;

(iv) simulation of each star query using at most four full row or column queries to A.

For square matrices A and A' of row and column weights at most three, $K(A \oplus A')$ is isomorphic to the disjoint union of $K(A)$ and $K(A')$.

Proof. Fixed pieces. Choose a triangulated sphere with three disjoint open disks removed, whose three boundary cycles have a common constant length h. Also fix a triangulated disk with an h-cycle boundary and a triangulated cylinder with two h-cycle boundaries. Choose the triangulations so that any edge or triangle whose vertices all lie on boundary cycles is already a face of those cycles. Such triangulations can be obtained by subdivision; choose them once and for all. Number the vertices of each boundary cycle from 1 to h in cyclic order, and identify vertices with the same number whenever two cycles are glued. Vertices outside the boundary cycles are distinct for different pieces. Thus gluing along whole boundary cycles produces a simplicial complex: simplices from different pieces can coincide only on their shared boundary cycles.

For each column j, create a disjoint h-cycle C_j numbered in the same way. For each row i, create a disjoint copy P_i of the three-holed sphere, with boundary cycles numbered $1, 2, 3$. At each boundary cycle, reserve distinct sets of interior vertices for a disk and a cylinder. The cylinder's other boundary will use the vertices of a column cycle. All these reserved vertices remain in the vertex set, even if unused. The total number of vertices in the cycles C_j, the surfaces P_i, and the sets reserved for disks and cylinders is $c_{\mathrm{top}}N$ for a fixed integer c_{top}.

Connecting the pieces. Read the nonzero columns in row i in increasing order, say $j_1, \dots, j_{d_i}$, where $d_i \le 3$. For $t \le d_i$, attach the cylinder joining boundary cycle t of P_i to C_{j_t}. Cap each remaining boundary cycle of P_i with the reserved disk. Vertices reserved for an unused disk or cylinder are isolated. The resulting surface S_i is connected and has d_i boundary components, one at each C_{j_t}. If $d_i = 0$, S_i is a sphere. The surfaces have disjoint interiors. At most three of them meet along any C_j, which gives a uniform bound on the vertex degree. For a block diagonal matrix $A \oplus A'$, the surfaces for each block attach only to cycles for that block. This gives the claimed disjoint union.

Counting triangle sets with zero boundary. Assign a coefficient in $\mathbb{F}_2$ to every triangle. The assignment has zero boundary if the sum of triangle coefficients incident to each edge is zero. An interior edge of S_i belongs to exactly two of its triangles, so those two coefficients must agree. Any two triangles of the connected surface S_i can be joined through such interior edges. Consequently, all triangles in S_i have a common coefficient x_i. This also holds when $d_i = 0$ and S_i is a sphere.

It remains to check the edges of each cycle C_j. Each surface S_i attached to C_j contributes one triangle at each edge of C_j. The boundary coefficient at every such edge is therefore $\sum_i A_{ij}x_i$. Thus the assignments with zero boundary are in linear bijection with

$$\{x \in \mathbb{F}_2^N : A^{\mathsf{T}}x = 0\}.$$

The unused isolated vertices contribute no triangles or constraints. Since $K(A)$ has no three-dimensional simplices, its second Betti number is the dimension of the kernel of its boundary matrix, namely $N - \operatorname{rank} A$. This proves (iii).

Answering a star query. For a vertex of P_i or an interior vertex reserved for one of its disks or cylinders, row i determines the entire star. Its sorted support specifies which pieces are used

and where each cylinder is attached. For a vertex on C_j, query column j to find the incident rows. For each such row i, the position of j in its sorted support identifies the cylinder attached to C_j. Its triangles are then known from the fixed triangulation. One column query and at most three row queries therefore recover the star. □

Proof of theorem 1.3. For each n in theorem 2.2, apply lemma 8.1 to the matrices A_b of corollary 5.2. Their common dimension is $N_0 = N_0(n)$, and rank A_b is $R_0(n)$ for $b \in B$ and $R_1(n)$ for $b \in b_* + B$. Take the two deterministic complexes

$$\mathcal{K}_n^0 = \bigsqcup_{b \in B} K(A_b), \qquad \mathcal{K}_n^1 = \bigsqcup_{b \in B} K(A_{b_*+b}).$$

Each summand in $\mathcal{K}_n^i$ has $c_{\mathrm{top}} N_0$ vertices and second Betti number $N_0 - R_i(n)$. Both unions have $c_{\mathrm{top}} N_0 |B|$ vertices, and this number tends to infinity. By additivity of the second Betti number,

$$\frac{\beta_2(\mathcal{K}_n^i; \mathbb{F}_2)}{|V(\mathcal{K}_n^i)|} = \frac{N_0 - R_i(n)}{c_{\mathrm{top}} N_0}.$$

The normalized second Betti number is larger for $i = 0$, by at least $\rho/(c_{\mathrm{top}} C)$, since $R_1(n) - R_0(n) \geq \rho n$ and $N_0 \leq Cn$.

As in the graph construction, a uniform root in $\mathcal{K}_n^i$ samples $b \sim \mu_i$ followed by a uniform vertex label independent of b. At bounded degree, exploring a ball of fixed radius takes a bounded number of star queries. Each costs at most four matrix queries, so for all sufficiently large n the exploration uses fewer than $\delta n/\tau$ matrix queries. Corollary 5.2 now gives identical laws for the rooted balls in the two unions.

There are finitely many rooted simplicial ball types at each radius for this degree bound. As in section 6, take a diagonal subsequence with a common local limit and then a subsequence along which both normalized second Betti numbers converge. They lie in $[0, 1/c_{\mathrm{top}}]$ and their gap persists. Interleaving the two sequences gives nonconvergence. The ranks of the boundary matrices, and hence the Betti numbers, are unchanged on extending scalars in characteristic two. □

For these same complexes the normalized second Betti numbers over $\mathbb{R}$ have a common limit by Elek's theorem [6]. The complexes can have three surfaces meeting along a cycle C_j, as well as isolated vertices.

9 Proof of the query lower bounds

For each n in theorem 2.2, draw b uniformly from one of the two cosets and let the algorithm query the matrix A_b of corollary 5.2. Its rank is $R_0(n)$ if $b \in B$ and $R_1(n)$ if $b \in b_* + B$, with $R_1(n) - R_0(n) \geq \rho n$. Fewer than $\delta n/\tau$ row or column queries give the same transcript distribution in both cases. To obtain a lower bound at every sufficiently large dimension N, we pad A_b with zeros while keeping the rank difference proportional to N.

The algorithm may know all the objects in definition 2.1, the shift b_*, the construction of A_b, and the two ranks $R_0(n), R_1(n)$. It can therefore choose a threshold between the ranks before making any queries.

Proof of theorem 1.4(i). Let C be the constant in corollary 5.2, and let λ bound the ratio of successive values of n in theorem 2.2. For sufficiently large N, choose the largest n in that sequence with $Cn \leq N$. Then

$$\frac{N}{C\lambda} < n \leq \frac{N}{C}.$$

Indeed, the next value in the sequence exceeds N/C and is at most λn. Choose $\varepsilon_0 = \rho/(8C\lambda)$ and pad the matrices of corollary 5.2 with zero rows and columns to dimension N. Their respective ranks $R_0(n), R_1(n)$ still satisfy

$$R_1(n) - R_0(n) \geq \rho n > 8\varepsilon_0 N.$$

If the estimate is within $\varepsilon_0 N$ of the rank, comparing it with $R_0(n) + \rho n/2$ identifies the coset containing b. Choose one of the two cosets with equal probability. An estimator that meets the error bound with probability at least $2/3$ on every input then identifies the coset with probability at least $2/3$.

Queries to the padded rows or columns return empty supports. All other queries are answered by the original matrix. Thus any algorithm limited to fewer than $\delta n/\tau$ queries has the same output distribution for either choice of coset, by corollary 5.2. Its probability of guessing the coset is therefore exactly $1/2$. Since $n > N/(C\lambda)$, this proves the $\Omega(N)$ worst-case lower bound.

For expected complexity let $q_0 = \lfloor \delta n/(2\tau) \rfloor$. Truncate the distinguisher after q_0 queries, giving any fixed answer if it has not halted. The truncated algorithm guesses the coset correctly with probability $1/2$. If the original algorithm has worst-case expected query count T, Markov's inequality bounds the truncation probability by T/q_0. Thus

$$\frac{1}{2} \geq \frac{2}{3} - \frac{T}{q_0}, \qquad T \geq \frac{q_0}{6} = \Omega(N).$$

□

Proof of theorem 1.4(ii). Let M be the target vertex count, put $N = \lfloor M/c_{\mathrm{top}} \rfloor$, and take a matrix A of dimension N. Form $K(A)$ from lemma 8.1 and pad it to M vertices with isolated vertices. Its second Betti number is still $N - \operatorname{rank} A$. Given an estimate $\widehat{\beta}$ of this Betti number, we return $N - \widehat{\beta}$ as the rank estimate. The absolute errors are equal.

Choose $\varepsilon_1 = \varepsilon_0/(2c_{\mathrm{top}})$. For $N \geq 1$, we have $M < c_{\mathrm{top}}(N+1) \leq 2c_{\mathrm{top}}N$, so an estimate of the second Betti number within $\varepsilon_1 M$ of its true value yields a rank estimate within $\varepsilon_0 N$ of the true rank. A star query on an added vertex can be answered from its label alone, and every other star query takes at most four matrix queries. The simulation therefore preserves the success probability and multiplies either the worst-case or the expected query count by at most four. Part (i) requires $\Omega(N) = \Omega(M)$ queries. □

The same matrix lower bound applies to indexed adjacency-list queries. One row or column query returns all neighbors of the corresponding vertex in $G(A)$. Sorting them by label answers a query for the neighbor at any specified index. Conversely, the degree bound lets us read the entire list with a constant number of indexed queries. By equation (17), a graph adjacency-list query on $G(A)$ reads one row or column of A, and $r_{\mathbb{F}_2}(G(A)) = \operatorname{rank} A/N$ for A of dimension N. Thus estimating normalized adjacency rank over $\mathbb{F}_2$ to absolute error ε_0 requires linearly many queries even on bipartite graphs of maximum degree three. By lemma 7.1, the same lower bound holds for estimating the output entropy per vertex of the XOR rule given adjacency-list access to the graph.

10 Scope

The two graph sequences have different limiting normalized adjacency ranks over every field of characteristic two. The argument leaves open what happens in odd characteristic. The construction takes disjoint unions indexed by the vectors in each coset. It does not prescribe a regular tree as the local limit. Mészáros [10, Question 1.3] asks whether, for every prime p and every sequence of d-regular graphs G_j whose girth tends to infinity, the dimension of the kernel of $\mathrm{Adj}(G_j)$ over $\mathbb{F}_p$, divided by $|V(G_j)|$, tends to zero. Our construction also does not give a tower of connected finite covers along a prescribed chain of finite-index normal subgroups of a fixed group.

A The construction of Li, Li, Han, and Liu

We derive theorem 2.2 from the construction of Li, Li, Han, and Liu [8]. The three spaces in definition 2.1 are binary cocycle spaces in their three code blocks. The distance of the third block and a cup product on the three blocks supply the two properties used in the rank construction.

The spaces and the distance bound. Use the fixed coefficient field $\mathbb{E}$ and the $\mathbb{F}_2$-basis β of [8, Section 2.1]. Write $[x]_\beta$ for the binary coordinates of a cochain x. Let n_a be the binary blocklength of block a, and set $n = n_3$. In terms of the binary check matrices of [8, Definition 2.1], take

$$Z_a = \ker H_{Z,a} \quad (a = 1, 2, 3), \qquad H_a = H_{Z,a} \quad (a = 1, 2), \qquad B = \operatorname{im} H_{X,3}^{\mathsf{T}}.$$

Here B consists of the coboundaries in block three, so $B \subseteq Z_3$. Li, Li, Han, and Liu use the trace-dual basis for chains. In these coordinates the binary dot product is the field trace of the chain–cochain pairing [8, equation (2.6)]. The cycles of the third block thus form $B^\perp = \ker H_{X,3}$, and the boundaries form $Z_3^\perp = \operatorname{im} H_{Z,3}^{\mathsf{T}}$. The least weight of a cycle outside the boundaries is the Z-distance [8, equation (2.9)]. This is the minimum in equation (2).

The trilinear form. Write $\widetilde{u}_{a,i}$ and ξ_i for the selected cocycles and cycles in [8, equation (3.222)]. Set $\xi_{[m]} = \sum_{i=1}^{m} \xi_i$ and choose $\theta \in \mathbb{E}$ with $\operatorname{Tr}_{\mathbb{E}/\mathbb{F}_2}(\theta) = 1$. For cochains x, y, z in the respective blocks, define the binary trilinear form

$$T([x]_\beta, [y]_\beta, [z]_\beta) = \operatorname{Tr}_{\mathbb{E}/\mathbb{F}_2}(\theta \langle x \smile y \smile z, \xi_{[m]} \rangle).$$

For cocycles x, y, pairing with the cycle $\xi_{[m]}$ depends only on the cohomology class of a third cocycle z [8, Definition 2.5 and Proposition 2.10]. When $[z]_\beta \in B$, that class is zero, which gives equation (4). Now take

$$u_i = [\widetilde{u}_{1,i}]_\beta, \qquad v_i = [\widetilde{u}_{2,i}]_\beta, \qquad z_i = [\widetilde{u}_{3,i}]_\beta.$$

The cup product $\widetilde{u}_{1,i} \smile \widetilde{u}_{2,j} \smile \widetilde{u}_{3,k}$ pairs to one with ξ_t when $i = j = k = t$, and to zero otherwise [8, equations (3.223)–(3.226)]. Its pairing with $\xi_{[m]}$ is therefore one exactly when $i = j = k$. Applying $c \mapsto \operatorname{Tr}_{\mathbb{E}/\mathbb{F}_2}(\theta c)$ preserves both values and proves equation (5).

Sparsity and sizes. Section 3.4 of [8] bounds the row and column weights of the binary check matrices and the number of cubic monomials containing each binary coordinate, giving w and τ in definition 2.1. It also gives Z-distance at least δn in the third block. Their Theorem 3.11 applies to every sufficiently large even r with $R = 2^r$. At these parameters $n_a = \Theta(R^2)$ in all three blocks and $m = \Omega(R^2)$, so $n_1, n_2 \leq Kn$ and $m \geq \rho n$ for fixed K, ρ. To obtain the sequence of values of n in theorem 2.2, let r increase by a fixed even step large enough that the lower bound on the next n exceeds the upper bound on the previous one. Since R increases by a fixed factor, the ratios of successive n are bounded.

References

[1] Miklós Abért. *Some questions.* 2 November 2010, Question 12. https://users.renyi.hu/~abert/questions.pdf.

[2] Miklós Abért. *Sofic entropy and finite automata.* Lecture abstract, ICMAT, 12, 13, and 15 March 2018. https://www.icmat.es/rt/l2invariants2018/abstractabertadvanced.pdf.

[3] Miklós Abért, Andreas Thom, and Bálint Virág. *Benjamini–Schramm convergence and pointwise convergence of the spectral measure*. Preprint, Theorem 4. https://users.renyi.hu/~abert/luckapprox.pdf.

[4] Andrej Bogdanov, Kenji Obata, and Luca Trevisan. A lower bound for testing 3-colorability in bounded-degree graphs. In *Proceedings of FOCS 2002*, pages 93–102, 2002. https://lucatrevisan.github.io/pubs/bot.pdf.

[5] Bernard Chazelle, Ronitt Rubinfeld, and Luca Trevisan. Approximating the minimum spanning tree weight in sublinear time. *SIAM Journal on Computing*, 34(6):1370–1379, 2005. https://doi.org/10.1137/S0097539702403244.

[6] Gábor Elek. *Betti numbers are testable*. arXiv:0907.5302, 2009. https://arxiv.org/abs/0907.5302.

[7] Herbert Edelsbrunner and Salman Parsa. On the computational complexity of Betti numbers: reductions from matrix rank. In *Proceedings of SODA 2014*, pages 152–160, 2014. https://pub.ista.ac.at/~edels/Papers/2014-J-01-BettiRank.pdf.

[8] Yiming Li, Zimu Li, Zhengyi Han, and Zi-Wen Liu. *Parallelizable and addressable transversal non-Clifford gates on good quantum LDPC codes*. arXiv:2609.26691v3, 6 October 2026. https://arxiv.org/abs/2609.26691v3. See Definitions 2.1 and 2.5, Proposition 2.10, Theorem 3.11, equations (3.223)–(3.226), and Section 3.4.

[9] András Mészáros. *The distribution of sandpile groups of random regular graphs*. arXiv:1806.03736v4, 28 January 2020. https://arxiv.org/abs/1806.03736v4. See the introduction following Theorem 1.9.

[10] András Mészáros. *Graph convergence, determinantal processes and the sandpile group of random regular graphs*. PhD thesis, Central European University, 2021. https://www.etd.ceu.edu/2021/meszaros_andras.pdf.

[11] Open Problems in Sublinear Algorithms. *Problem 68: Approximating Rank in the Bounded-Degree Model*. Suggested by Yuichi Yoshida, Bertinoro 2014. https://sublinear.info/68.

[12] Raphael Yuster. Matrix sparsification for rank and determinant computations via nested dissection. In *Proceedings of FOCS 2008*, pages 137–145, 2008. https://math.haifa.ac.il/raphy/papers/sparsifymat.pdf.