THE NUMBER OF 5-INNER EQUIVALENCE CLASSES OF SIMPLE GRAPHS ON AT MOST FIVE VERTICES

KOREA SUPERINTELLIGENCE LABS

Abstract. Eilers, Restorff, Ruiz and Sørensen close their geometric classification of graph C^*-algebras with an atlas of small graphs. A *simple graph* there is a finite directed graph with loops allowed and no multiple edges, and two such graphs with at most M vertices are *M-inner equivalent* when a chain of five elementary steps — isomorphism, a legal row addition, a legal column addition, deletion of a regular source, a legal collapse — joins them without ever leaving the simple graphs on at most M vertices. Their table records $2, 8, 35, 218$ classes for $M = 1, 2, 3, 4$ and a question mark at $M = 5$, the accompanying sentence being that at $M = 5$ they "have not attempted a complete analysis". We prove that the entry is 2172. The proof is a component certificate for the move graph on the 295 128 isomorphism classes of simple graphs with at most five vertices: a spanning forest bounds the number of classes above by 2172, a move-invariant colouring bounds it below by the same number, and the colouring is a complete invariant, so no transversal is shorter and no family of pairwise inequivalent graphs is longer. The same machinery, run at $M \leq 4$, returns the published $2, 8, 35, 218$ exactly, which is the evidence that the five clauses and their legality conditions have been transcribed correctly; and the two legality conditions that an implementation is most likely to drop — the path condition on the two additions, and the requirement that a collapsed vertex support no loop — are each shown to be load-bearing by an explicit pair of graphs that dropping it would wrongly merge. The theorems below are machine-checked in Lean 4, with no `sorry` and with the compiled evaluator as the only trusted computational step; the few statements that lie outside the formal development are labelled where they occur. Two structural by-products, computed outside the formal development and labelled as such, are that each of the two vertex-deleting moves is individually redundant at every $M \leq 5$ while the two together are not, and that every class contains a graph on exactly M vertices.

1. Introduction

The objects. Throughout, a *graph* is a finite directed graph $E = (E^0, E^1, r, s)$, and a *simple graph* is one with no multiple edges — loops are allowed. Equivalently, following Eilers–Restorff–Ruiz–Sørensen [1], a simple graph is one whose adjacency matrix A_E, with (u, v) entry the number of edges from u to v, is a $\{0, 1\}$-matrix; in terms of $\mathsf{B}_E := \mathsf{A}_E - I$ this says that the diagonal entries of B_E lie in $\{-1, 0\}$ and all other entries in $\{0, 1\}$. The number of simple graphs on n unlabelled vertices is the number of binary relations on n unlabelled points, sequence A000595 of the OEIS [7]: $2, 10, 104, 3044, 291\,968$ for $n = 1, \ldots, 5$.

On these graphs [1] defines a family of equivalence relations $\equiv_{I,M}$, one for each bound M on the number of vertices, generated by five elementary steps: isomorphism, a legal row addition, a legal column addition, the deletion of a regular source, and a legal collapse. Section 2 states all five, and the legality conditions, exactly as in the source. The point of the relation is that it is a *computable* lower approximation to stable isomorphism of the associated graph C^*-algebras: $E \equiv_{I,M} F$ implies that E and F are move equivalent, hence that $C^*(E) \otimes \mathbb{K} \cong C^*(F) \otimes \mathbb{K}$ and that the Leavitt path algebras $L_{\mathsf{k}}(E)$ and $L_{\mathsf{k}}(F)$ are Morita equivalent for any field k (see §2.4). Against it [1] places a computable upper approximation, *outer equivalence* $\equiv_O$, defined through a K-theoretic invariant; $\equiv_O$-classes are unions of $\equiv_{I,M}$-classes, and wherever the two counts agree the classification of the corresponding C^*-algebras up to stable isomorphism is settled.

The table, and the cell that was left open. The final section of [1], §7.3 "Atlas of Graph C^*-algebras of Small Graphs", tabulates the two counts for the simple graphs with at most M vertices, $M \leq 5$ [1, Table 1]:

M	1	2	3	4	5
nonisomorphic graphs	2	10	104	3044	291 968
$\equiv_{I,M}$-classes	2	8	35	218	?
$\equiv_O$-classes	2	8	35	199	1310

The text immediately preceding the table reads, verbatim — and identically in the published version and in the arXiv v2: "At $M = 4$, it then only takes a few minutes of computing time to partition these sets of graphs into $\equiv_{I,4}$- and $\equiv_O$-classes, obtaining the numbers listed in [the table above]. At $M = 5$ we have not attempted a complete analysis, as it takes hours even to compute all the K-temperatures and divide the graphs into $\equiv_O$-classes." So the missing entry is the number of $\equiv_{I,5}$-classes, and what made the analysis expensive is the *other* row of the table, the one requiring K-temperatures. The $\equiv_{I,5}$ count is a pure graph computation, and that is the cell we fill.

What is proved.

Theorem 1.1 (Theorem 3.3, Corollary 3.5 and Theorem 3.6)**.** *There are exactly* 2172 *classes of* 5*-inner equivalence among the simple graphs with at least one and at most five vertices. More precisely, there is a list of* 2172 *such graphs meeting every class exactly once; no list meeting every class at least once is shorter than* 2172*; and no repetition-free list of pairwise* $\equiv_{I,5}$*-inequivalent such graphs is longer than* 2172*.*

The two-sidedness is not decoration. A transversal by itself bounds the number of classes only from above, and a colouring by itself only from below; "exactly 2172" is the conjunction, and the bridge between the two is that the colouring we construct is a *complete* invariant (Theorem 3.4): two graphs on at most five vertices receive the same colour if and only if they are $\equiv_{I,5}$-equivalent. That is also what lets us certify *in*equivalence of concrete pairs, which is what the controls of Section 5 need.

Theorem 1.2 (Theorem 4.1 and Proposition 4.2)**.** *The same construction, run with* $M = 1, 2, 3, 4$*, returns* $2, 8, 35, 218$ *classes, and counts* $2, 12, 116, 3160$ *and* 295 128 *isomorphism classes of simple graphs with at most* $1, 2, 3, 4$ *and* 5 *vertices respectively.*

Theorem 1.2 is the reason to believe Theorem 1.1, and we ran it first. The risk in this problem is not arithmetic but transcription: [1] states the two additions as operations on $\mathsf{B}_E^\bullet$ and defers their legality to a proposition imported from a second paper [2], so a reconstruction at the level of $\{0, 1\}$ adjacency matrices has four independent opportunities to go wrong. Four published numbers reproduced from the reconstructed definitions, together with the published row of graph counts — which is also OEIS A000595 — is the evidence that it did not. Section 5 adds the complementary evidence: each of the four moves is applicable somewhere, the relation is not the total relation, and each of the two legality conditions that a careless implementation would drop — the path condition on the two additions, and the requirement that a collapsed vertex support no loop — provably merges inequivalent graphs when dropped.

What the computation is. The relation $\equiv_{I,5}$ is generated, so membership in it is not decidable as stated; a certificate is needed rather than a decision. Ours is the standard component certificate (Section 3), carried on one representative of each isomorphism class:

- a *spanning forest*: a parent function on the 295 128 representatives with a strictly decreasing rank and, at each node, a named move realising the edge to its parent. Induction on the rank shows every graph is $\equiv_{I,5}$-equivalent to one of the 2172 roots.
- a *move-invariant colouring*: a function from representatives to $\{0, \dots, 2171\}$ checked to be constant along every legal move out of every representative. Hence the roots are pairwise inequivalent.

Both halves are finite checks, and they are the only steps in the paper that rest on exhaustive computation. Section 8 says exactly which finite searches were run and how large each is; the largest single one visits all 33 620 498 labelled adjacency matrices on $1, \ldots, 5$ vertices and confirms that each is a recorded relabelling of its class representative. What makes that affordable at all is a lemma proved rather than computed: each of the four moves commutes with relabelling of the vertices, so the invariance of the colouring needs checking only at the 295 128 representatives and not at all 33 620 498 labelled matrices.

Provenance. The cell is printed as "?" in the latest version of [1] on the arXiv (only v1 and v2 exist; the abstract page was checked on 3 September 2026), and we have not found the number elsewhere. What was searched, on 2–3 September 2026: the 44 distinct works citing [1] that OpenAlex (43 records) and Semantic Scholar (22 records) return, all titles and abstracts read and 11 full texts downloaded and searched for the numerals of the table and for the phrases "inner equivalen" and "atlas" — with no hit, including in a 146-page later paper by two of the four authors; the arXiv listing and the personal pages of the first author, together with about 1450 archived captures of the latter; and the OEIS, which holds the graph-count row $2, 10, 104, 3044, 291\,968$ as A000595 but returns nothing for $2, 8, 35, 218$ or for $2, 8, 35, 199, 1310$.

One gap in that search should be stated plainly, since we could not close it completely. A master's thesis of T. Medici, *On the classification of unital graph C^*-algebras. A thesis in experimental mathematics* (University of Copenhagen, 2017, advised by the first author of [1]), cites [1] and is exactly the kind of document that could carry a number of this sort. Its only known copy is a URL under `math.ku.dk` that returns 404 and has no capture in the Internet Archive (both re-checked on the date of writing), and Google Scholar and the other aggregators that might hold another copy refused automated access. **We have not read that thesis.** What we could recover is its abstract and three of its tables, indexed by Semantic Scholar from the copy that has since disappeared. The abstract begins: "We address the classification up to stable isomorphism of graph C^*-algebras corresponding to small graphs with up to three vertices with either zero, one or infinitely many edges between vertices, also including the graphs containing sinks and infinite emitters"; and it ends "Though a full classification is not obtained, we present a result, where on the one hand a certain set of the original graph C^*-algebras are fully classified computationally, and on the other hand the data of the remaining graph C^*-algebras are displayed and analysed upon." The three indexed tables (4.1 and 4.2 on page 56, 4.3 on page 63) tabulate class counts and running times, and every column of every one of them is headed either 2×2 or 3×3. So the object of the thesis stops two vertices short of $M = 5$, and is a different family of graphs besides — edge multiplicities in $\{0, 1, \infty\}$ rather than the $\{0, 1\}$ of a simple graph, with infinite emitters admitted — and nothing in what we can see of it counts the graphs this paper counts. That is much better evidence than the circumstantial argument from the title alone, but it is still not a reading of the full text, so we claim priority for the *proof* of the value 2172, which is what this paper contains, and not necessarily for the value.

Remark 1.3 (a consequence of the published 1310; not part of the formal development). Since $\equiv_O$-classes are unions of $\equiv_{I,5}$-classes, and [1] computes 1310 outer classes at $M = 5$, our 2172 says that the two approximations are further apart at $M = 5$ than at $M = 4$: writing k for the number of outer classes that split and t for the total number of inner classes into which they split, $t - k = 2172 - 1310 = 862$, and since a split class yields at least two inner classes, $t \geq 2k$, whence $k \leq 862$. At $M = 4$ the corresponding figures are recorded in [1]: 17 outer classes split into 36 inner classes. This arithmetic uses the published 1310, which we have not verified, and it is not part of the machine-checked development.

2. Preliminaries

2.1. **Graphs, matrices, and vertices of the four kinds.** Let $E = (E^0, E^1, r, s)$ be a finite directed graph, s the source and r the range map. A vertex v is *regular* if $s^{-1}(v)$ is nonempty, and *singular* otherwise; for finite graphs the singular vertices are exactly the sinks. A vertex v is a *source* if $r^{-1}(v) = \emptyset$. A vertex *supports a loop* if there is an edge from it to itself. We write A_E for the adjacency matrix, $\mathsf{B}_E = \mathsf{A}_E - I$, and $\mathsf{B}_E^\bullet$ for B_E with the rows of the singular vertices deleted. All of this is the notation of [1].

For a simple graph on the vertex set $\{0, \ldots, n-1\}$ we identify A_E with a $\{0,1\}$-matrix A, so that: v is regular iff row v of A is nonzero; v is a source iff column v of A is zero; v supports a loop iff $A[v][v] = 1$; and there is a directed path from u to v iff the (u, v) entry of $A \vee A^2 \vee \cdots \vee A^n$ is 1, the powers and the join being Boolean. (A shortest directed walk between distinct vertices is a path and so uses at most $n-1$ edges, so the truncation at A^n loses nothing.)

2.2. **The four moves and their legality.** The two additions are the derived moves of [1, §2.5], whose legality is governed by a proposition imported from [2]. We restate both from [1], in the notation of this paper; numbering throughout is that of the published version (see the bibliography for how it differs from the arXiv v2).

Proposition 2.1 ([1, Proposition 2.12], quoting [2])**.** *Let E be a graph with finitely many vertices. Suppose $u, v \in E^0$ are distinct vertices with a path from u to v. Let $E_{u,v}$ equal the identity matrix except at the (u, v) entry, where it is 1; then $\mathsf{B}_E E_{u,v}$ is B_E with the uth column added into the vth column, and $E_{u,v}\mathsf{B}_E$ is B_E with the vth row added into the uth row. Then:*

- (i) *if u supports a loop, or there is an edge from u to v and u emits at least two edges, then $\mathsf{A}_E \sim_{ME} \mathsf{B}_E E_{u,v} + I$;*
- (ii) *if v is regular and either v supports a loop or there is an edge from u to v, then $\mathsf{A}_E \sim_{ME} E_{u,v}\mathsf{B}_E + I$.*

Here $\sim_{ME}$ is *move equivalence*, the equivalence relation on graphs generated by graph isomorphism and the four moves (O) out-splitting, (I) in-splitting, (R) reduction and (S) removal of a regular source [1, Definition 2.6]; see also [3, 4]. Following [1, §7.3] we call a row or column addition in a matrix B_E representing a simple graph *legal* if it meets the requirements of Proposition 2.1 and produces another such matrix, and we call a collapse (Move (Col), which deletes a vertex v and replaces each pair of edges $x \to v \to y$ by an edge $x \to y$) *legal* if it is applied to a regular vertex not supporting a loop and takes a simple graph to another simple graph. The published version of [1] does not itself number Move (Col); it takes the notation from [2], whose arXiv v2 numbers it Definition 2.7.

We now write out what the four moves do to the $\{0,1\}$-matrix A, since that is the form in which the computation is carried out and since the translation is the one place where a reconstruction could silently diverge from [1].

Lemma 2.2. *Let A be the adjacency matrix of a simple graph on $\{0, \ldots, n-1\}$ and $B = A - I$.*

- (ii) *Let A' be the matrix of the row addition $E_{u,v}B + I$. Then A' agrees with A off row u, and on row u*

$$A'[u][j] = A[u][j] + A[v][j] \ (j \neq v), \qquad A'[u][v] = A[u][v] + A[v][v] - 1.$$

  Hence A' is again a $\{0,1\}$-matrix if and only if $A[u][j] + A[v][j] \leq 1$ for all $j \neq v$ and $A[u][v] + A[v][v] \geq 1$; the second of these is exactly the side condition "v supports a loop or there is an edge from u to v" of Proposition 2.1(ii). In Boolean form, $A'[u][j] = A[u][j] \vee A[v][j]$ for $j \neq v$ and $A'[u][v] = A[u][v] \wedge A[v][v]$.
- (iii) *Let A' be the matrix of the column addition $BE_{u,v} + I$. Then A' agrees with A off column v, and on column v*

$$A'[i][v] = A[i][v] + A[i][u] \ (i \neq u), \qquad A'[u][v] = A[u][v] + A[u][u] - 1,$$

  so A' is a $\{0,1\}$-matrix iff $A[i][v] + A[i][u] \leq 1$ for $i \neq u$ and $A[u][v] + A[u][u] \geq 1$, the latter being implied by the side condition of Proposition 2.1(i).
- (iv) *Deleting a regular source w deletes row and column w.*

(v) *Collapsing a regular vertex v that supports no loop replaces A by the matrix on $\{0,\dots,n-1\}\setminus\{v\}$ with entries $A[x][y]+A[x][v]\cdot A[v][y]$, which is a $\{0,1\}$-matrix iff no $x,y\neq v$ have $A[x][y]=A[x][v]=A[v][y]=1$.*

Moreover, performing (ii) or (iii) in $\mathsf{B}_E^\bullet$ rather than in B_E gives the same simple graph.

Proof. Items (ii) and (iii) are the substitution $B[i][j]=A[i][j]-\delta_{ij}$ into the two matrix products, and (iv), (v) are the definitions of Move (S) and Move (Col) of [1] read on adjacency matrices. For the last sentence: the path condition of Proposition 2.1 forces u to emit an edge, so u is regular and no deleted row is row u; in (ii) the modified row is row u and the added row is row v, which is present because legality requires v regular. In (iii) every row is modified in column v, but a deleted row w is a zero row of A with $w\neq u$, so $B[w][u]=A[w][u]-\delta_{wu}=0$ and the addition leaves row w unchanged. $\square$

2.3. M-inner equivalence.

Definition 2.3 ([1, Definition 7.13]). Fix an integer M and let E and F be simple graphs with $m,n\leq M$ vertices respectively. E and F are *elementary equivalent through simple graphs of size M* if either $m=n$ and one of

(i) E is isomorphic to F,
(ii) F arises from E by performing a legal row addition in $\mathsf{B}_E^\bullet$,
(iii) F arises from E by performing a legal column addition in $\mathsf{B}_E^\bullet$,

holds, or if $m=n+1$ and

(iv) F arises from E by deleting a regular source,
(v) F arises from E by a legal collapse.

The coarsest equivalence relation containing elementary equivalence through simple graphs of size M is called *M-inner equivalence* and written $\equiv_{I,M}$.

Two features of Definition 2.3 drive everything below. First, the relation is generated, so proving that two graphs are equivalent means exhibiting a chain, and proving that they are not means exhibiting an invariant — there is no symmetry between the two directions. Second, the bound M is part of the relation and not merely of the domain: a chain joining two graphs on three vertices is allowed to pass through graphs on five, but not through graphs on six, so $\equiv_{I,M}$ genuinely depends on M and the classes at different M are not nested in any obvious way.

2.4. **What the relation means for C^*-algebras.** The atlas section of [1] places $\equiv_{I,M}$ in the following chain of implications [1, Proposition 7.14], valid for finite simple graphs with at most M vertices and for any field k:

$$E\equiv_{I,M}F \implies E\sim_{ME}F \implies \begin{cases} L_{\mathsf{k}}(E)\sim_{\text{Morita}}L_{\mathsf{k}}(F), \\ E\sim_{CE}F \implies C^*(E)\otimes\mathbb{K}\cong C^*(F)\otimes\mathbb{K}, \end{cases} \implies E\equiv_O F,$$

where $\sim_{CE}$ is Cuntz move equivalence and $\mathbb{K}$ the compacts. So a count of $\equiv_{I,M}$-classes is an upper bound for the number of stable isomorphism classes of the associated graph C^*-algebras and for the number of Morita equivalence classes of the associated Leavitt path algebras, and $\equiv_O$ supplies a lower bound; where the two counts agree, as they do for $M\leq 3$, the classification is complete. Nothing in the present paper uses this; it is why the number is worth having.

3. The count at $M=5$

3.1. **The certificate.** Fix M and let $\mathcal{V}$ be the set of isomorphism classes of simple graphs with between one and M vertices, with a chosen representative $V(t)$ for each $t\in\mathcal{V}$. Write $\operatorname{node}(g)\in\mathcal{V}$ for the class of g. A *component certificate* for $\equiv_{I,M}$ with K classes consists of

(1) a colouring $\operatorname{cl}\colon\mathcal{V}\to\{0,\dots,K-1\}$;
(2) a parent function $\operatorname{par}\colon\mathcal{V}\to\mathcal{V}$ and a rank $\operatorname{rk}\colon\mathcal{V}\to\mathbb{N}$;

(3) for each non-root t (i.e. each t with $\mathrm{par}(t) \neq t$) a *named move*: one of the four moves, its parameters, and a direction bit saying whether it runs from t to $\mathrm{par}(t)$ or the other way;
(4) a root function $\mathrm{root}\colon \{0, \ldots, K-1\} \to \mathcal{V}$;
(5) for each labelled adjacency matrix g on at most M vertices, a relabelling $\mathrm{wit}(g)$ carrying $V(\mathrm{node}(g))$ to g.

subject to the conditions: every representative is a simple graph with between 1 and M vertices; $\mathrm{wit}(g)$ is a bijection of the vertex set of g and does carry $V(\mathrm{node}(g))$ to g; every relabelling of a representative has that representative's own class; the colouring is constant along each of the four moves out of each representative; each non-root is joined to its parent by its named move, with rk strictly decreasing and cl constant; every root is the recorded root of its colour; and each colour has a root.

Lemma 3.1. *A component certificate with K colours yields a list of K graphs — the roots — such that every simple graph with at most M vertices is $\equiv_{I,M}$-equivalent to a member of the list, and such that no two members of the list are $\equiv_{I,M}$-equivalent.*

Proof sketch. Upper bound: given g, the witness condition makes g isomorphic to $V(\mathrm{node}(g))$, hence $\equiv_{I,M}$-equivalent to it by clause (i) of Definition 2.3; then induction on rk along the parent function, using at each step the named move (in the direction the bit records, and using symmetry of $\equiv_{I,M}$ when it runs backwards), reaches the root of the colour of $\mathrm{node}(g)$. Lower bound: the colouring is invariant under each generating step — under isomorphism because node is, under the four moves because it is checked to be so at every representative — so it is invariant under the generated equivalence relation, and two roots with different colours cannot be equivalent. □

The four move checks in the certificate are made only at the $|\mathcal{V}|$ representatives. That they nevertheless control every labelled matrix is the content of the equivariance lemma, which is proved rather than computed:

For a bijection q of $\{0, \ldots, n-1\}$ and a simple graph g on that vertex set, write g^q for the graph with $g^q[i][j] = g[q(i)][q(j)]$.

Lemma 3.2 (equivariance)**.** *Let q be a bijection of the vertex set of a simple graph g and let u, v be vertices. Then a row addition at (u, v) is legal in g^q if and only if a row addition at $(q(u), q(v))$ is legal in g, and in that case $\mathrm{rowAdd}(g^q, u, v) = \big(\mathrm{rowAdd}(g, q(u), q(v))\big)^q$; likewise for column additions, and likewise for the deletion of a regular source and for a legal collapse, with q replaced on the target by the bijection it induces on the smaller vertex set.*

Without Lemma 3.2 the invariance of the colouring would have to be checked at all 33 620 498 labelled matrices with at most five vertices rather than at 295 128 representatives, which is the difference between a computation that finishes and one that does not.

3.2. **The main theorem.**

Theorem 3.3. *There is a list R of 2172 simple graphs, each with between one and five vertices, such that every simple graph with at most five vertices is $\equiv_{I,5}$-equivalent to a member of R, and such that any two members of R that are $\equiv_{I,5}$-equivalent are equal.*

Proof, and what was computed. Lemma 3.1 applied to an explicit component certificate for $M = 5$, whose tables were produced by a breadth-first search of the move graph. The certificate is validated by twelve exhaustive checks, and these — together with three readings of the tables, that $M = 5$, that $K = 2172$, and that the recorded vertex bound is at most 5 — are the only steps of the proof that rest on computation. In order of cost the twelve are:

- the witness check, over all $2 + 2^4 + 2^9 + 2^{16} + 2^{25} = 33\,620\,498$ labelled adjacency matrices on $1, \ldots, 5$ vertices: each has a recorded class index below 295 128, and the recorded relabelling is a bijection carrying that class's representative to it;

- the canonicity check, over the 295 128 representatives and all relabellings of each ($1, 2, 6, 24$ or 120 of them): every relabelling of a representative is assigned that representative's own class;
- the four move checks, over each representative and each of the at most 5×5 parameter choices: the colouring is constant along every legal row addition, legal column addition, deletion of a regular source, and legal collapse;
- the reachability check (the tabulated path relation of each representative is the Boolean closure computed from its adjacency matrix), the forest check (parent in range, colour preserved, rank strictly decreasing, named move valid), and four bookkeeping checks — every representative is a simple graph with between one and five vertices, the colouring lands below K, each root is the recorded root of its colour, and each colour has a root that is a root of the forest — each over the 295 128 representatives or the 2172 colours.

That the tabulated lists of relabellings of $\{0, \dots, n-1\}$ are the complete lists for every $n \leq 5$ is *not* among the checks: it is proved, by case analysis on n, and this is where the third table reading is used. Everything else — Lemmas 3.1 and 3.2, and the passage from the checks to the conclusion — is ordinary reasoning, carried out in full and machine-checked; see Section 8. □

Theorem 3.4. *Let* $\mathrm{cl}_5(g)$ *denote the colour that the* $M = 5$ *certificate assigns to the class of* g*. For simple graphs* g, h *with at most five vertices,*

$$g \equiv_{I,5} h \quad \Longleftrightarrow \quad \mathrm{cl}_5(g) = \mathrm{cl}_5(h).$$

That is, cl_5 *is a complete invariant of* $\equiv_{I,5}$*.*

Proof sketch. "$\Rightarrow$" is the invariance half of Lemma 3.1. For "$\Leftarrow$", the forest half of Lemma 3.1 shows that each g is $\equiv_{I,5}$-equivalent to $V(\mathrm{root}(\mathrm{cl}_5(g)))$; equal colours therefore give equivalence to the same root. No new computation is involved: this is a consequence of the same certificate. □

Since cl_5 is a computable function of the adjacency matrix — a table lookup — Theorem 3.4 makes $\equiv_{I,5}$ decidable on the simple graphs with at most five vertices. We record that as a consequence rather than as part of the theorem: it is not itself part of the formal development.

Corollary 3.5. *The number of* $\equiv_{I,5}$*-classes of simple graphs with at most five vertices is exactly* 2172*.*

Proof. Theorem 3.3 gives at most 2172 classes. For the reverse inequality, each colour $i < 2172$ has a root $V(\mathrm{root}(i))$, a simple graph with at most five vertices and colour i, so all 2172 colours occur; by Theorem 3.4 graphs of different colours are inequivalent. □

The two directions of Corollary 3.5 are worth stating in the form that refutes a smaller and a larger answer separately.

Theorem 3.6. *Let* R *be any list of graphs such that every simple graph with at most five vertices is* $\equiv_{I,5}$*-equivalent to a member of* R*. Then* R *has at least* 2172 *members. Conversely, let* R *be any repetition-free list of simple graphs with at most five vertices, no two distinct members of which are* $\equiv_{I,5}$*-equivalent. Then* R *has at most* 2172 *members.*

Proof sketch. For the first: cl_5 maps R onto $\{0, \dots, 2171\}$, because every colour is realised by some graph g and g is equivalent to some $r \in R$, which then has the same colour by Theorem 3.4. For the second: cl_5 is injective on R by Theorem 3.4 and takes values in $\{0, \dots, 2171\}$. Again no computation beyond that of Theorem 3.3. □

4. The published entries, reproduced

The construction of Section 3 does not know that $M = 5$ is special. Run at $M \leq 4$, it returns the published row of the table of [1]. We record this because it is the control on the reconstruction of Definition 2.3 and Lemma 2.2: with any one of the five clauses or any one of the legality conditions transcribed wrongly, these four numbers would not come out.

Theorem 4.1. *For $M = 1, 2, 3, 4$ there is a list of $2, 8, 35, 218$ simple graphs respectively, each with between one and M vertices, such that every simple graph with at most M vertices is $\equiv_{I,M}$-equivalent to a member of the list and no two members of the list are $\equiv_{I,M}$-equivalent. These are the entries printed in the second row of the table of* [1] *for $M = 1, 2, 3, 4$.*

Proof, and what was computed. Four component certificates, validated by the same twelve checks as in Theorem 3.3, the largest search being over the $2 + 2^4 + 2^9 + 2^{16} = 66\,066$ labelled matrices on at most four vertices. The numbers themselves are due to [1]; what is ours is the certified reproduction. ☐

Proposition 4.2. *The certificate of Section 3 has $2, 12, 116, 3160$ and $295\,128$ nodes for $M = 1, 2, 3, 4, 5$. These are the numbers of isomorphism classes of simple graphs with at least one and at most M vertices; equivalently, the numbers on exactly n vertices are $2, 10, 104, 3044, 291\,968$ for $n = 1, \ldots, 5$.*

Proof, and what was computed. The node counts are five readings of the tables, the last of them the outcome of an enumeration that visits all 2^{25} labelled matrices on five vertices. That a node count is a count of isomorphism classes is the content of two of the certificate conditions of Section 3: the witness condition says that every labelled matrix on at most M vertices is a recorded relabelling of its node's representative, and the canonicity condition says that every relabelling of a representative is assigned that representative's own node; together they make the assignment of nodes to labelled matrices constant on relabelling orbits and injective on representatives, so the number of nodes is the number of orbits among all $\sum_{n \le M} 2^{n^2}$ labelled matrices. The identification of these numbers with published ones is not ours: the counts on exactly n vertices form the first row of the table of [1], which cites them as sequence A000595 of the OEIS [7], the number of binary relations on n unlabelled points, and the cumulative counts are the partial sums. ☐

5. The relation is not degenerate

A count of equivalence classes is only as meaningful as the relation counted, and two failure modes would be invisible in the numbers alone: a move that is never applicable, and a legality condition that has been dropped, so that the relation is coarser than intended. The following three statements close both. They are stated at $M = 4$, where the searches are cheap and where Theorem 4.1 anchors them to published data; the corresponding complete invariant cl_4 is the $M = 4$ instance of Theorem 3.4, which is what makes *in*equivalence certifiable at all.

Proposition 5.1 (non-vacuity)**.** *Each of the four moves — legal row addition, legal column addition, deletion of a regular source, legal collapse — is applicable to some simple graph on three vertices.*

Proposition 5.2 (the legality conditions bite)**.**

1. *There are a simple graph g on four vertices and distinct vertices u, v satisfying every requirement of a legal row addition* except *the existence of a path from u to v — v is regular, v supports a loop or there is an edge $u \to v$, and the result is again a simple graph — for which g and the resulting graph are* not *$\equiv_{I,4}$-equivalent.*
2. *There are a simple graph g on four vertices and a regular vertex v that* does *support a loop, at which the collapse would still produce a simple graph, for which g and that collapse are* not *$\equiv_{I,4}$-equivalent.*

Consequently an implementation that dropped the path condition of Proposition 2.1, or the requirement that a collapsed vertex support no loop, would merge genuinely inequivalent graphs.

Proposition 5.3 (non-degeneracy)**.** *$\equiv_{I,4}$ is not the total relation: the graph on four vertices with no edges and the graph on four vertices whose only edge is a loop at one vertex are not $\equiv_{I,4}$-equivalent.*

Proofs, and what was computed. Proposition 5.1 is an exhaustive search over the $2^9 = 512$ labelled matrices on three vertices for each of the four moves. Proposition 5.2 is an exhaustive search over the

$2^{16} = 65\,536$ labelled matrices on four vertices and the at most 16 parameter pairs of each, looking for a witness; both witnesses were found, and the inequivalence of each witness pair is read off cl_4 through Theorem 3.4 at $M = 4$. Proposition 5.3 is the same reading for the two named graphs: a table lookup, not a search. □

Remark 5.4 (global evidence, computed outside the formal development). An independently written program in C, sharing no code with the certificate, recomputes the class counts under modifications of the definition. Its output is reproduced below; only the first row is a theorem of this paper (Theorems 3.3 and 4.1), and the rest are measurements, reported because they show which parts of Definition 2.3 change the answer: every legality condition does, and the two vertex-deleting clauses do only jointly.

variant	$M=1$	$M=2$	$M=3$	$M=4$	$M=5$
as defined	2	8	35	218	2172
drop the path condition	2	6	14	46	336
drop the loop/edge side condition	2	5	21	131	1296
drop “v is regular” (row addition)	2	6	23	131	1217
drop both additions	2	9	72	1949	204 139
drop “v supports no loop” (collapse)	2	2	5	31	363
drop deletion of a regular source only	2	8	35	218	2172
drop legal collapse only	2	8	35	218	2172
drop both vertex-deleting moves	2	10	45	263	2435

The control at the opposite end is not in the table, because it is a theorem rather than a measurement: with no moves at all the classes are the isomorphism classes, of which there are $2, 12, 116, 3160$ and $295\,128$ by Proposition 4.2, so the machinery is not silently merging everything. (The program prints four counts per row; the entries above are its “classes among the graphs on at most M vertices” column, and in the no-move variant that column returns exactly the $2, 12, 116, 3160, 295\,128$ just quoted.) This C program also reproduces 2172 independently of the certificate, in 1.67 seconds of processor time.

6. Two structural observations

The last three rows of Remark 5.4 are of a different character from the others, and we single them out. Both statements below were obtained from the same independent C computation and lie *outside* the formal development; they are recorded as observations, not as theorems of this paper, and nothing above depends on them. Formalising either is one further certificate run with the move set restricted.

Remark 6.1 (computed, not proved here). For every $M \leq 5$: dropping clause (iv) of Definition 2.3 alone, or clause (v) alone, leaves the class counts $2, 8, 35, 218, 2172$ unchanged, whereas dropping both raises them to $2, 10, 45, 263, 2435$. So each of the two vertex-deleting moves is individually redundant given the other, at every $M \leq 5$, while the two together are not.

Remark 6.2 (computed, not proved here). For every $M \leq 5$, every $\equiv_{I,M}$-class contains a graph on exactly M vertices, and two graphs on exactly M vertices are $\equiv_{I,M}$-equivalent if and only if they are joined by a chain of legal row and column additions passing only through graphs on M vertices. In particular the count 2172 is unchanged if the domain is narrowed to the graphs on exactly five vertices and the two vertex-deleting moves are discarded.

Remark 6.2 also settles a reading question about the table of [1], whose caption does not say whether “$\equiv_{I,M}$-classes” counts the classes of the graphs with at most M vertices or the classes meeting the graphs on exactly M vertices. The two readings, and the third reading in which only the size-preserving moves are used, agree at every $M \leq 5$, so the published $2, 8, 35, 218$ does not disambiguate them and does not need to.

We have not found either observation in [1] or in the works citing it that we read.

7. Cost, and what is not attempted

The $M = 5$ certificate costs 1035.7 seconds of processor time and 6.89 gigabytes of resident memory to check, essentially all of it in the witness check over the 33 620 498 labelled matrices; the $M \leq 4$ file costs 6.25 seconds, and the controls of Section 5 another 7.5 seconds. An independent computation of the same answer in C, sharing no code with the certificate, takes 1.67 seconds and 137 megabytes. These are measurements on one machine and are quoted only to locate the frontier.

That frontier is at $M = 6$, and it is not near. There are 96 928 992 isomorphism classes of simple graphs on six vertices [7], and $2^{36} \approx 6.9 \cdot 10^{10}$ labelled matrices; the certificate as designed stores two machine words per *labelled* matrix, which is about 1.1 terabytes. Reaching $M = 6$ needs a certificate indexed by isomorphism classes rather than by labelled matrices — about 10^8 words — together with a coverage argument that does not visit every labelled matrix, which is a redesign of the witness condition rather than a larger machine.

The other open row of the table at $M = 5$ is the outer count 1310, which [1] computes and we have not verified. Doing so needs the K-temperature of [1]: the poset Γ_E of components of E, together with, at each component, either a numerical type or the K_0-group of the corresponding subquotient of $C^*(E)$. Under the standard-form hypothesis of [1] that group is the cokernel of the transpose of a block of $\mathsf{B}_E^\bullet$, so the group-theoretic part is a Smith normal form computation over $\mathbb{Z}$ on integer matrices of size at most 5 and is cheap; the poset bookkeeping is the work. With 2172 inner classes against 1310 outer classes, which outer classes split and how far is the natural next question (Remark 1.3).

8. Verification

Theorems 3.3, 3.4 and 3.6, Corollary 3.5, Theorem 4.1 and Propositions 5.1, 5.2 and 5.3, together with Lemmas 3.1 and 3.2, are formalised and machine-checked in Lean 4 [8] against Mathlib [9]. Two statements sit slightly outside that. Lemma 2.2 is the definition of the moves in the development, so it is true by construction there and is proved on paper above. And of Proposition 4.2 it is the five node counts that are machine-checked; that a node count is a count of isomorphism classes is the argument given in its proof, which is on paper and uses two of the machine-checked certificate conditions. Remarks 1.3, 5.4, 6.1 and 6.2 and Section 7 are outside the formal development and are labelled where they occur. There is no `sorry` and no axiom declared by us anywhere in the development. The reasoning layer uses `propext`, `Quot.sound` and, through the inversion of a bijection of an initial segment and through library lemmas about lists, `Classical.choice`; the finite checks listed in the proofs of Theorems 3.3 and 4.1 and of Section 5 are carried out by Lean's compiled evaluator (`native_decide`) and each contributes one evaluation axiom of its own. The audit that matters is which evaluations each result rests on. Theorem 3.3, and with it Theorem 3.4, Corollary 3.5 and Theorem 3.6, rests on exactly the fifteen evaluations named in the proof of Theorem 3.3 and on nothing else; Proposition 4.2 rests on the five node-count readings alone; Theorem 4.1 rests, for each $M \leq 4$, on the corresponding fifteen; and each part of Propositions 5.1, 5.2 and 5.3 rests on the $M = 4$ evaluations together with one witness search or lookup of its own. That evaluator is the only trusted computational step, and an independently written C program, sharing no code with it, reproduces every class count and every node count reported above.

Two engineering points are worth recording, because they are the difference between a computation that terminates and one that does not, and because the second is a silent failure mode. First, the evaluated code lives in a module that imports nothing at all, so that it can be compiled into a shared library and the checks run compiled code rather than the interpreter; and the certificate tables, being constants, are module-initialiser code, so the $M \leq 4$ tables and the $M = 5$ tables are compiled into *two* libraries, else every small check would pay for building the $M = 5$ tables. Leaving the tables outside the compiled library cost a factor of about three on the $M \leq 4$ file alone. Second, Lean does not check that a compiled module is newer than the sources it was compiled against; after definitions were moved between modules, one stale compiled module was silently loaded by everything downstream. It was caught by comparing modification times of sources against compiled artefacts, and the entire development was then rebuilt in import order and the axiom audit re-run. Any reader reproducing this should do the same rather than trusting an incremental build.

The repository is private: repository reference to be added at submission.

References

[1] S. Eilers, G. Restorff, E. Ruiz and A. P. W. Sørensen, *Geometric Classification of Graph C^*-algebras over Finite Graphs*, Canad. J. Math. **70** (2018), no. 2, 294–353; `doi:10.4153/CJM-2017-016-7`; arXiv:1604.05439. Numbering above is that of the published version: §2.5 "Derived Moves", Definition 2.6 (move equivalence), Proposition 2.12 (row and column additions), §7.3 "Atlas of Graph C^*-algebras of Small Graphs", Definition 7.13 (M-inner equivalence), Proposition 7.14 (the chain of implications of §2.4) and Table 1. The arXiv v2 numbers §2.5, §7.3, Definition 7.13, Proposition 7.14 and Table 1 the same way; it numbers move equivalence Definition 2.19 and the proposition on row and column additions Proposition 2.26, and it gives Move (Col), which the published version leaves unnumbered, as its own Definition 2.25. The arXiv abstract page lists v1 and v2 only, and the "?" stands in both the arXiv v2 and the published version.

[2] S. Eilers, G. Restorff, E. Ruiz and A. P. W. Sørensen, *Invariance of the Cuntz splice*, Math. Ann. **369** (2017), no. 3–4, 1061–1080; `doi:10.1007/s00208-017-1570-y`; arXiv:1602.03709. This is the source of the results collected as Proposition 2.1 in [1]; the arXiv v2 of [1] cites it as the preprint arXiv:1602.03709v2, the published version as "Math. Annalen, to appear".

[3] A. P. W. Sørensen, *Geometric classification of simple graph algebras*, Ergodic Theory Dynam. Systems **33** (2013), no. 4, 1199–1220; `doi:10.1017/S0143385712000260`.

[4] T. Bates and D. Pask, *Flow equivalence of graph algebras*, Ergodic Theory Dynam. Systems **24** (2004), no. 2, 367–382; `doi:10.1017/S0143385703000348`.

[5] P. Alberca Bjerregaard, G. Aranda Pino, D. Martín Barquero, C. Martín González and M. Siles Molina, *Atlas of Leavitt path algebras of small graphs*, J. Math. Soc. Japan **66** (2014), no. 2, 581–611; `doi:10.2969/jmsj/06620581`. [1] opens its atlas section by calling it inspired by this one.

[6] B. D. McKay and A. Piperno, *Practical graph isomorphism, II*, J. Symbolic Comput. **60** (2014), 94–112; `doi:10.1016/j.jsc.2013.09.003`. Cited by [1] as the tool for producing the graph lists; the present computation enumerates the isomorphism classes itself, by filling relabelling orbits, and uses no external catalogue.

[7] OEIS Foundation Inc., *Sequence A000595: number of binary relations on n unlabeled points*, The On-Line Encyclopedia of Integer Sequences, `https://oeis.org/A000595`, terms $a(1), \ldots, a(6) = 2, 10, 104, 3044, 291\,968, 96\,928\,992$. [1] cites it, as "[Slo] A595", for the number of nonisomorphic graphs of a given size, which is the first row of its Table 1.

[8] L. de Moura and S. Ullrich, *The Lean 4 theorem prover and programming language*, in: A. Platzer and G. Sutcliffe (eds.), Automated Deduction — CADE 28, Lecture Notes in Computer Science **12699**, Springer, Cham, 2021, pp. 625–635; `doi:10.1007/978-3-030-79876-5_37`.

[9] The mathlib Community, *The Lean mathematical library*, in: Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020), ACM, 2020, pp. 367–381; `doi:10.1145/3372885.3373824`.