WALK COUNTS AND THE DEPTH OF GRAPH NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Message-passing neural networks distinguish vertices exactly as far as color refinement does, layer for layer, so on graphs whose stable coloring has m classes they can need $m - 1$ layers. Global filters, such as spectral, diffusion and Katz filters, let every layer see the whole graph, and how much of this depth they remove has been open. We prove that they halve it: with any filters that are polynomials in the adjacency matrix, arbitrary update functions and unbounded width, exactly $\lfloor m/2 \rfloor$ layers are needed in the worst case. The layers that remain are spent on naming, since a linear filter can separate color classes without producing their indicator vectors and only a nonlinearity supplies them. The depth that full expressive power costs thus counts nonlinear layers, however far each layer sees.

1 INTRODUCTION

Color refinement, the one-dimensional Weisfeiler–Leman algorithm (1-WL), marks the exact limit of what message-passing neural networks can distinguish (Morris et al., 2019; Xu et al., 2019). It separates two vertices as soon as they have different numbers of neighbors in some class of the current coloring, and after t layers a message-passing network separates at most the vertices that t rounds of refinement separate, while suitable networks separate all of them (Morris et al., 2019, Theorems 1 and 2). A network therefore reaches its full expressive power only when its depth matches the number of rounds refinement needs, a quantity with a long history of its own: linear lower bounds (Fürer, 2001) were sharpened by Kiefer and McKay, who showed that the trivial bound of $n - 1$ rounds on graphs with n vertices is attained for infinitely many n (Kiefer & McKay, 2020), the graphs attaining it are being classified (Kiefer & de Mel, 2026), and on graphs whose stable coloring has m classes the worst case is $m - 1$, attained on paths.

The standing explanation for so long a worst case is locality. A layer informs a vertex about its neighbors only, so after t layers it has seen a ball of radius t, and the extremal graphs run on exactly this budget, a single distinction crossing the long-refinement graphs of Kiefer and McKay one step per round (Kiefer & McKay, 2020; Krebs & Verbitsky, 2015). Deep networks pay for their reach with over-smoothing (Li et al., 2018; Oono & Suzuki, 2020) and over-squashing (Alon & Yahav, 2021; Di Giovanni et al., 2023), and the standard remedy is to let every layer see far beyond its neighbors, up to the whole graph. Spectral and polynomial filters (Defferrard et al., 2016; Chien et al., 2021; He et al., 2021; Wang & Zhang, 2022), diffusion by personalized PageRank (Gasteiger et al., 2019), and decoupled architectures (Wu et al., 2019) all apply in a single layer a function of the adjacency matrix, possibly normalized by degrees, such as a power series or the attenuated resolvent of Katz (1953), and such filters stay within the power of 1-WL (Vasileiou et al., 2026). Their simplest instance, comparing the numbers of walks of every length that leave a vertex, is the classical walk partition (Powers & Sulaiman, 1982; Kriege, 2022).

How many layers do such networks need? Morris et al. (2024) single this out in their position paper on the theory of graph machine learning: most expressivity results do not quantify, or only loosely bound, the size, depth included, that a network needs to match 1-WL (their Challenge II.3), and they ask for lower bounds on the cost of a given expressive power (Challenge II.5) and for problems that only deep networks solve (Challenge IV.4). For networks with global filters the question takes a sharp form:

> *Do global filters remove the depth that message passing needs, or does part of it survive every filter?*

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

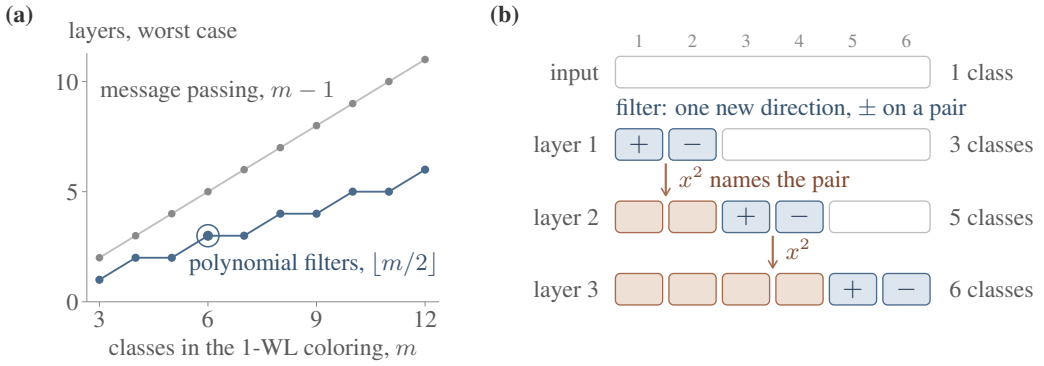

Figure 1: **(a)** Layers needed to compute the 1-WL coloring in the worst case over graphs whose coloring has m classes: $m-1$ with message passing, exactly $\lfloor m/2 \rfloor$ with filters that are polynomials in the adjacency matrix. **(b)** Why half remains, on the graph of Section 3.2 attaining $\lfloor m/2 \rfloor$ for $m = 6$ (circled in (a)), one row per layer. Each layer's filter finds one new direction, with values $\pm$ on a pair of classes (blue). Before the last layer this direction separates the pair without naming it; a nonlinearity (x^2 suffices) names it (brown), and only then can the next layer reach the next pair.

Half survives, rounded up. Over graphs whose stable coloring has m classes, message passing needs $m-1$ layers in the worst case, and a network that may apply any polynomial in the adjacency matrix in every layer, with arbitrary update functions and unbounded width, needs $\lfloor m/2 \rfloor = \lceil (m-1)/2 \rceil$ and no fewer. What survives is naming: a linear filter can separate vertices by the values it computes without producing the indicator vectors of the classes they cut out, and only a nonlinearity supplies them. Figure 1 shows both halves.

1.1 Our Results

We count layers of the most permissive kind. A layer may apply any finite set of filters from the algebra $\mathcal{A}$ generated by the adjacency matrix A, which consists of the polynomials in A and so contains every power series and every resolvent such as Katz's, and it combines each vertex's own features, the multiset of its neighbors' features and its filtered features by an arbitrary function, at any width (Definition 1). Message passing is the case without filters, and normalized filters live in the algebra $\mathcal{D}$ generated by A and the degree matrix D. Walk refinement is the refinement that saturates each step with global information: it separates two vertices as soon as, for some length l and some class C of the current coloring, they have different numbers of walks of length l ending in C, and its iteration number $\tau(G)$ is the number of steps it needs to reach the stable coloring (Definition 2).

Theorem (Informal version of Theorem 1)**.** *For every graph, the least depth of a network with filters from $\mathcal{A}$ whose features induce the 1-WL coloring is $\tau(G)$, and one fixed Katz filter per layer achieves it. Over graphs whose 1-WL coloring has $m \geq 3$ classes, the worst case is $m-1$ for message passing, exactly $\lfloor m/2 \rfloor$ with filters from $\mathcal{A}$, and between $\lfloor m/2 \rfloor - 1$ and $\lfloor (m-1)/2 \rfloor$ with normalized filters from $\mathcal{D}$.*

Global reach thus buys exactly $\lceil m/2 \rceil - 1$ of the $m-1$ worst-case layers, and no filter from $\mathcal{A}$ buys more: half of the depth was locality, and half of it never was.

The upper bound comes from a single accounting identity. A layer's gain in classes has two parts: one class for each new dimension of features its filters produce, and one for each dimension those features still lack to name the classes they separate, that is, to span their indicator vectors (Proposition 3). Before the last layer both parts are at least one: a layer that changes anything adds a dimension, and a layer that named every class it separates would be the last, the next filter closing the same space again and finding nothing new. Every layer but the last therefore gains at least two classes, and with only $m-1$ classes to gain, the depth is at most $\lfloor m/2 \rfloor$ (Proposition 4). The identity asks nothing of walks: every closure that contains its argument and does not change when repeated obeys it, which is why the same accounting governs the normalized filters of $\mathcal{D}$ (Remark C.2). For even m the accounting also prescribes the extremal graphs: each layer adds exactly one new dimension and, before the last,

falls exactly one dimension short of naming its classes. Section 3.2 builds graphs with this profile for every m, and Figure 1(b) follows one of them layer by layer.

The two costs, reach and naming, separate cleanly on explicit graphs. On paths, global filters finish in one layer where message passing needs $m - 1$ (Proposition A.2); on the graphs we construct to need $\lfloor m/2 \rfloor$ layers, message passing needs no more, and the whole depth is naming (Proposition 6).

The same accounting speaks directly to architectures (Section 3.3). A linear network, with no nonlinearity between its filters, stops at the first stage at any depth, and so does a decoupled architecture that propagates once and then applies a vertex-wise predictor (Corollary 1). Degree normalization multiplies features by the degree vector $A\mathbf{1}$, itself a computed feature, so normalized filters name the degree classes for free, and this is worth at most one layer in the worst case. A global readout, a virtual node, or global attention without positional encodings reads the same multiset at every vertex and adds nothing to message passing.

Naming has a price of its own, and it is multiplicative. The subspaces that contain the all-ones vector and are closed under entrywise products are exactly the spaces of vectors constant on the classes of a partition (Lemma 1), so indicators that filtering does not produce are reached only by multiplying. Counting rounds of entrywise products, with filtering free in between, gives the multiplicative depth; on the graphs we construct it is the number of layers minus one, so no constant number of alternations between filtering and multiplication computes the 1-WL coloring (Theorem 2).

1.2 Related Work

Expressivity and depth of message passing. Message-passing networks distinguish at most what color refinement does (Morris et al., 2019; Xu et al., 2019), and suitable ones match it round for round (Morris et al., 2019), with matching logical characterizations (Barceló et al., 2020; Grohe, 2024); Morris et al. (2023) survey this line. Depth lower bounds for message passing trade against width (Loukas, 2020), and over-squashing, the loss of information from distant vertices, is governed by the topology of the graph and is not removed by depth (Alon & Yahav, 2021; Di Giovanni et al., 2023). Morris et al. (2024) name the depth that a 1-WL-equivalent architecture requires as a challenge for the field, and Theorem 1 answers it exactly for networks with global filters.

Global filters. Polynomial and spectral filters (Defferrard et al., 2016; Kipf & Welling, 2017; Chien et al., 2021; He et al., 2021), diffusion by personalized PageRank (Gasteiger et al., 2019), and decoupled architectures (Wu et al., 2019) are networks over $\mathcal{A}$ or $\mathcal{D}$ in our sense. Vasileiou et al. (2026) prove that spectral networks built from continuous filters of an aggregation shift operator are exactly as expressive as 1-WL; we count the layers this takes. Wang & Zhang (2022) show that one polynomial filter with linear weights can produce any one-dimensional prediction when the normalized Laplacian has a simple spectrum and the features contain every frequency component; constant features contain every frequency component only when the 1-WL coloring is discrete, and from them a linear network stops at the first stage (Corollary 1). Gai et al. (2025) show that spectral invariant networks need a linear number of refinement iterations to reach their maximal power, for a refinement that reads spectral invariants of vertex pairs and is strictly stronger than 1-WL.

Iteration numbers and walk partitions. At dimension two, Lichter et al. (2019) refine colors of vertex pairs by walks of all lengths and prove that linearly many rounds remain, and Geerts (2020, Remark 2.2) observes that walks of length ℓ speed up two-dimensional refinement by a factor $\lceil \log \ell \rceil$. At dimension one, the walk partition (Powers & Sulaiman, 1982), the partitions Chan & Godsil (1997) attach to a set of vertices through walk counts, and power iterated color refinement (Kersting et al., 2014) all use global propagation, the last alternating degree-normalized propagation with clustering and stabilizing within a linear number of outer iterations. Theorem 1 gives the exact worst case of the fully adaptive process, in the number m of stable classes.

2 Preliminaries and the Network Model

This section defines the networks we count and shows that each of their layers is one step of walk refinement.

Color refinement and quotients. Throughout, G is a finite simple graph on n vertices with adjacency matrix A, degree matrix D and maximum degree Δ. We write $\mathbf{1}$ for the all-ones vector, $\mathbf{1}_S$ for the indicator vector of a set S, and $e_1, e_2, \ldots$ for standard basis vectors. For a partition π of $V(G)$, the matrix K_π has the indicators of the classes of π as its columns, and $\mathbb{K}^\pi$ is the space they span, the vectors constant on every class. Given a family X of vectors indexed by $V(G)$, possibly presented as the columns of a matrix, π_X denotes the partition by row equality: u and v lie in one class exactly when $x_u = x_v$ for all $x \in X$. A matrix H of features *induces* the partition π_H, and a partition σ *refines* π when every class of σ lies inside a class of π.

Color refinement starts from the *uniform coloring*, in which all vertices are alike, and repeatedly separates two vertices as soon as they have different numbers of neighbors in some class of the current coloring. After finitely many rounds the classes stop changing, at the *stable coloring* W, the 1-WL coloring; we write m for its number of classes. The *iteration number of color refinement* is the number of rounds after which W is first reached. A round that changes anything splits a class, so this number is at most $m - 1$, and paths attain it (Proposition A.1).

A partition π is *equitable* when all vertices of a class have the same number of neighbors in each class, equivalently when $\mathbb{K}^\pi$ is invariant under A, and W is the coarsest equitable partition (Godsil & Royle, 2001; Ramana et al., 1994). The number B_{ij} of neighbors in class j of a vertex in class i of W therefore depends only on i and j, and the $m \times m$ *quotient matrix* B satisfies $A\Phi = \Phi B$ for the characteristic matrix Φ of W. Refinement acts on a nonnegative integer matrix B by the same rule, index i seeing, for each class J of the current coloring, the total $\sum_{j \in J} B_{ij}$, and we call B *reduced* when its stable partition is discrete, as the quotient of G is. Every partition coarser than W is induced by a partition of $[m]$, and all refinement below takes place in that range, where refinement by A on $V(G)$ is refinement by B on $[m]$ (Appendix A).

The network model. A layer of a graph neural network combines, at every vertex, its own features, the multiset of its neighbors' features, and the values of filters that see the whole graph. The definition admits every such layer, with arbitrary update functions and widths.

Definition 1. Let $\mathcal{M}$ be a set of real $V(G) \times V(G)$ matrices. A *network of depth L over $\mathcal{M}$* consists of finite sets $F_0, \ldots, F_{L-1} \subseteq \mathcal{M}$ of *filters* and functions $\phi_0, \ldots, \phi_{L-1}$, and it sends initial features $H^{(0)} \in \mathbb{R}^{V(G) \times d_0}$ to

$$H_v^{(t+1)} \;=\; \phi_t\Big(H_v^{(t)},\; \{\{H_w^{(t)} : w \in N(v)\}\},\; \big((MH^{(t)})_v\big)_{M \in F_t}\Big) \qquad (t < L),$$

where H_v is the row of H at v. Widths, functions and filters may depend on G. We write $\sigma_t = \pi_{H^{(t)}}$ for the partition that the features of layer t induce.

With $\mathcal{M} = \{A\}$ this is message passing in its most permissive form, covering sum, mean and maximum aggregation as well as attention weights computed from the endpoints of an edge. We compare it with two algebras of global filters. The algebra $\mathcal{A}$ generated by I and A consists of the polynomials in A; by Cayley–Hamilton it contains every convergent power series in A, such as e^{tA}, and the resolvent $(I - \zeta A)^{-1}$ of Katz (1953). The algebra $\mathcal{D}$ generated by I, A and D contains in addition degree-normalized operators such as $D^{-1}A$, $D^{-1/2}AD^{-1/2}$ and their self-loop variants (Kipf & Welling, 2017), their polynomials, and the personalized PageRank resolvent (Gasteiger et al., 2019). Filters act on the current features, so their values are combinations of counts of walks into the current classes, weighted by degrees in the case of $\mathcal{D}$, and every network over $\mathcal{A}$ or $\mathcal{D}$ stays within 1-WL. Encodings computed from individual entries of matrices such as A^l, for instance relative positional encodings or counts of closed walks, lie outside both algebras (Appendix B).

Walk refinement. Global filters refine in stages. For a matrix algebra $\mathcal{M}$ containing I and A and a partition π, the filters of $\mathcal{M}$ applied to $\mathbb{K}^\pi$ span $\mathcal{M}\mathbb{K}^\pi$, and we write $\rho_{\mathcal{M}}(\pi) = \pi_{\mathcal{M}\mathbb{K}^\pi}$ for the partition this space induces. For $\mathcal{M} = \mathcal{A}$ the space is

$$\mathcal{C}_A(\pi) \;=\; \sum_{l \geq 0} A^l\, \mathbb{K}^\pi \;=\; \sum_{l=0}^{n-1} A^l\, \mathbb{K}^\pi,$$

the second equality by Cayley–Hamilton.

Definition 2. *Walk refinement* sends a partition π of $V(G)$ to $\rho(\pi) = \pi_{\mathcal{C}_A(\pi)}$, so that two vertices remain together exactly when, for every length $l \geq 0$ and every class C of π, they have equally

many walks of length l ending in C. Write $\pi^{(t)} = \rho^t(\pi^{(0)})$ for the iterates from the uniform coloring $\pi^{(0)}$. The *iteration number* $\tau(G)$ is the least t with $\pi^{(t)} = W$. Replacing $\mathcal{A}$ by $\mathcal{D}$ defines $\rho_{\mathcal{D}}$ and its iteration number $\tau_{\mathcal{D}}(G)$.

Walk refinement is monotone and refining, its fixed points are the equitable partitions, and its iterates stay coarser than W and reach it; the same holds for $\rho_{\mathcal{D}}$ (Appendix A). The first iterate $\pi^{(1)}$ is the walk partition of Powers & Sulaiman (1982), and the later ones feed back the classes the process itself created. The name follows Lichter et al. (2019), whose walk refinement is the two-dimensional analogue, acting on colors of vertex pairs.

A layer computes nothing that the next stage of refinement does not already see.

Proposition 1. *Let $\mathcal{M}$ be a matrix algebra containing I and A, and consider a network over a subset of $\mathcal{M}$. Then $\rho_{\mathcal{M}}(\sigma_t)$ refines σ_{t+1} for every $t < L$, and $\rho^t_{\mathcal{M}}(\sigma_0)$ refines σ_t for every $t \le L$.*

Proof. Every column of $H^{(t)}$ lies in $\mathbb{K}^{\sigma_t}$, so every column of $MH^{(t)}$ with $M \in \mathcal{M}$ lies in $\mathcal{M}\mathbb{K}^{\sigma_t}$. The neighbor multiset at v is determined by the numbers $|N(v) \cap C|$ over the classes C of σ_t, which are the entries at v of the columns of AK_{σ_t}, and the own features are determined by the columns of K_{σ_t}; both families lie in $\mathcal{M}\mathbb{K}^{\sigma_t}$ as well. So if $\rho_{\mathcal{M}}(\sigma_t)$ merges u and v, all three arguments of ϕ_t agree at u and v, and $H^{(t+1)}_u = H^{(t+1)}_v$. The second claim follows by induction, since $\rho_{\mathcal{M}}$ is monotone: a coarser partition π has a smaller $\mathbb{K}^{\pi}$, hence a smaller $\mathcal{M}\mathbb{K}^{\pi}$, which induces a coarser partition. □

Conversely, one filter carries out a whole stage. Call an attenuation ζ *faithful* for G when $I - \zeta A$ is invertible and $\pi_{(I-\zeta A)^{-1}K_\pi} = \rho(\pi)$ for every partition π of $V(G)$.

Proposition 2. *The attenuation $\zeta_G = 1/(\Delta^n + \Delta + 1)$ is faithful for G, and so is every ζ outside a finite set. For a faithful ζ, the layer with the single filter $(I - \zeta A)^{-1}$ and an update that sends distinct arguments to distinct standard basis vectors maps the features K_π to one-hot features of $\rho(\pi)$.*

The explicit value carries out every stage exactly by one linear solve, for all partitions at once: a vertex starts at most Δ^l walks of length l, so the first length at which two vertices differ dominates the attenuated tail. Generic faithfulness says that almost every attenuation does the same; the proofs are in Appendix B. The classical refinements that alternate exact linear computation over A with recoloring by equality, from the algorithm of Morgan (1965) to the partitions of Chan & Godsil (1997), are networks over $\mathcal{A}$ whose update tests equality, and walk refinement is optimal among them (Proposition B.4).

3 The Exact Depth

This section proves the main theorem. Its upper bound is an accounting identity (Section 3.1), its lower bound is a family of graphs that meets the identity with no slack (Section 3.2), and Section 3.3 draws the consequences for architectures.

Theorem 1. *Let G be a graph whose stable coloring W has m classes, and let the initial features be constant and nonzero.*

(a) The least depth of a network over $\mathcal{A}$ whose features induce W is $\tau(G)$, and one Katz filter per layer attains it, with the attenuation $\zeta_G = 1/(\Delta^n + \Delta + 1)$ or any other faithful one. Over $\{A\}$ the least depth is the iteration number of color refinement, and over $\mathcal{D}$ it is $\tau_{\mathcal{D}}(G)$, attained by layers whose filters linearly span $\mathcal{D}$.

(b) Over graphs whose stable coloring has $m \ge 3$ classes, the maximum of the least depth is $m - 1$ over $\{A\}$ and exactly $\lfloor m/2 \rfloor$ over $\mathcal{A}$; over $\mathcal{D}$ it lies between $\lfloor m/2 \rfloor - 1$ and $\lfloor (m-1)/2 \rfloor$, which coincide for even m.

In plain terms, global filters halve the worst-case depth of message passing, and no choice of filters from $\mathcal{A}$, update functions or width does better. Per graph they never cost anything: $\{A\} \subseteq \mathcal{A} \subseteq \mathcal{D}$ gives $\tau_{\mathcal{D}}(G) \le \tau(G) \le$ the iteration number of color refinement. Part (a) follows from Propositions 1 and 2 (Appendix B); in part (b), the ceiling over $\mathcal{A}$ is Proposition 4 below, extended to $\mathcal{D}$ in

Proposition C.2, and the matching graphs are paths (Proposition A.1) and the paired tail of Section 3.2. The paired tail thus supplies the kind of task that Challenge IV.4 of Morris et al. (2024) asks for: computing its 1-WL coloring is a node-level problem that no network with filters from $\mathcal{A}$ solves in fewer than $\lfloor m/2 \rfloor$ layers.

3.1 The Deficiency and the Upper Bound

Throughout this subsection and the next we work on a reduced quotient B of order m, where

$$\mathcal{C}_B(\pi) \;=\; \sum_{l \geq 0} B^l \, \mathbb{K}^{\pi}, \qquad \rho(\pi) \;=\; \pi_{\mathcal{C}_B(\pi)},$$

and the stable coloring is the discrete partition of $[m]$; the chain and the iteration number on B are the graph's (Appendix A). A stage that moves at all buys one class: if $\mathcal{C}_B(\pi)$ were no larger than $\mathbb{K}^{\pi}$, the chain would already have halted. The bound turns on the second class. A stage separates indices without necessarily containing the indicators of the classes it creates; the nonlinearity of the layer supplies those indicators, and only then does a further filter have anything new to act on. The invariant below measures how far a space is from containing the indicators of its classes.

Definition 3. The *deficiency* of a subspace $S \subseteq \mathbb{Q}^m$ is $\delta(S) = |\pi_S| - \dim S$.

Every vector of S is constant on the classes of π_S, so $S \subseteq \mathbb{K}^{\pi_S}$ and $\delta(S) \geq 0$, with $\delta(S) = 0$ exactly when S contains the indicators of its own classes. Thus $\delta(S)$ is the number of dimensions that S lacks to name its own classes, that is, to contain their indicators. The identity below is the whole engine of the upper bound, and it uses nothing about walks: a closure that contains its argument and does not change when repeated is all it asks for.

Proposition 3. *Let* cl *assign to each subspace of* $\mathbb{Q}^m$ *a subspace with* $U \subseteq \mathrm{cl}(U)$ *and* $\mathrm{cl}(\mathrm{cl}(U)) = \mathrm{cl}(U)$ *for every* U*, and put* $R(\pi) = \pi_{\mathrm{cl}(\mathbb{K}^{\pi})}$*. For a partition* π *of* $[m]$ *with* $S = \mathrm{cl}(\mathbb{K}^{\pi})$*,*

$$|R(\pi)| - |\pi| \;=\; \delta(S) \;+\; \big(\dim S - |\pi|\big),$$

both summands being nonnegative. The second is at least 1 *when* $R(\pi) \neq \pi$*, and the first is at least* 1 *when* $R(R(\pi)) \neq R(\pi)$*.*

Proof. Write $\pi' = R(\pi) = \pi_S$. The identity is $\delta(S) = |\pi'| - \dim S$ rearranged, and $\delta(S) \geq 0$ was noted above. From $\mathbb{K}^{\pi} \subseteq S$ we get $\dim S \geq |\pi|$, with equality only if $S = \mathbb{K}^{\pi}$ and hence $\pi' = \pi$. For the last claim suppose $\delta(S) = 0$, so that $S = \mathbb{K}^{\pi'}$. Then $\mathrm{cl}(\mathbb{K}^{\pi'}) = \mathrm{cl}(\mathrm{cl}(\mathbb{K}^{\pi})) = S = \mathbb{K}^{\pi'}$, and $R(\pi') = \pi'$. □

Walk refinement is the case $\mathrm{cl} = \mathcal{C}_B$, where $R = \rho$, and so is refinement by any matrix algebra containing B and the identity, the degree algebra included (Remark C.2). One further input is needed before the identity can be telescoped: on a reduced quotient the only equitable partition is the discrete one, because every equitable partition refines the stable one, so the chain moves at every step until it reaches the discrete partition. Along the chain write

$$r_t \;=\; \dim \mathcal{C}_B(\pi^{(t)}) - |\pi^{(t)}|, \qquad \delta_t \;=\; \delta\big(\mathcal{C}_B(\pi^{(t)})\big),$$

so that a stage gains $|\pi^{(t+1)}| - |\pi^{(t)}| = \delta_t + r_t$ classes: one for each new dimension, and one for each dimension the stage lacks to name the classes it separates.

Proposition 4. *Let* B *be a reduced quotient of order* $m \geq 3$*. Then* $\tau \leq \lfloor m/2 \rfloor$*. If* m *is even and equality holds, then* $r_t = \delta_t = 1$ *for every* $t < \tau - 1$*, while* $r_{\tau-1} = 1$ *and* $\delta_{\tau-1} = 0$*. If* m *is odd and equality holds, the same is true except that exactly one of* $r_0, \ldots, r_{\tau-1}, \delta_0, \ldots, \delta_{\tau-1}$ *exceeds the stated value by one.*

Proof. For $t < \tau$ the partition $\pi^{(t)}$ is not discrete, so ρ moves it and $r_t \geq 1$; for $t < \tau - 1$ it moves $\pi^{(t+1)}$ as well, so Proposition 3 gives $\delta_t \geq 1$. The chain runs from one class to m, whence

$$m - 1 \;=\; \sum_{t<\tau} (\delta_t + r_t) \;\geq\; 2(\tau - 1) + 1,$$

that is $\tau \leq m/2$. The slack in this estimate is $m - 2\tau$. For even m at equality it vanishes, forcing every listed quantity to its minimum; for odd m at equality it is 1, so exactly one of them exceeds its minimum, by one. □

A partition is equitable exactly when $\mathcal{C}_B(\pi) = \mathbb{K}^\pi$, so for even m the normal form says that an extremal chain runs through partitions each of which is one dimension short of equitable. One dimension less and refinement halts there; one more and the chain reaches all m classes too soon.

3.2 The Paired Tail

The chain below peels one pair of indices off a shrinking tail at every step. Fix $m \geq 3$ and put $q = \lfloor m/2 \rfloor$. For $0 \leq j \leq q$ let $R_j = \{2j+1, \dots, m\}$, and let P_j be the partition of $[m]$ into the singletons $\{1\}, \dots, \{2j\}$ together with R_j, the last class omitted when it is empty; so P_0 is the indiscrete partition, $|P_j| = 2j+1$ for $j < q$, and P_q is discrete. Set

$$a_0 = \mathbf{1}, \qquad \begin{aligned} b_j &= e_{2j+1} - e_{2j+2} \quad (0 \leq j \leq q-1), \\ a_j &= (m-2j)\,(e_{2j-1} + e_{2j}) - 2 \cdot \mathbf{1}_{R_j} \quad (1 \leq j \leq q), \end{aligned}$$

and define

$$B = m\,\mathbf{1}\mathbf{1}^\mathsf{T} + \sum_{j=0}^{q-1} \left(a_j b_j^\mathsf{T} + b_j a_j^\mathsf{T} \right). \tag{1}$$

Every term is symmetric with integer entries, and the j-th couples the pair $\{2j+1, 2j+2\}$ to what precedes it. At $m = 6$, the case drawn in Figure 1(b), this gives

$$B = \begin{pmatrix} 8 & 6 & 11 & 3 & 7 & 7 \\ 6 & 4 & 9 & 1 & 5 & 5 \\ 11 & 9 & 2 & 6 & 6 & 2 \\ 3 & 1 & 6 & 10 & 10 & 6 \\ 7 & 5 & 6 & 10 & 2 & 6 \\ 7 & 5 & 2 & 6 & 6 & 10 \end{pmatrix},$$

whose row sums $(42, 30, 36, 36, 36, 36)$ already display the first stage, and whose chain is 123456, $1|2|3456$, $1|2|3|4|56$, $1|2|3|4|5|6$.

Proposition 5. *For every $m \geq 3$ the matrix Equation (1) is a symmetric reduced quotient of order m with entries in $\{0, \dots, 2m\}$. Its chain is $P_0, P_1, \dots, P_q$, so $\tau = \lfloor m/2 \rfloor$, and*

$$\mathcal{C}_B(P_j) = \mathbb{K}^{P_j} \oplus \langle b_j \rangle \qquad (0 \leq j \leq q-1).$$

The proof (Appendix C) rests on three facts. The spaces $E_j = \langle a_j, b_j \rangle$ for $j < q$, together with $E_q = \langle a_q \rangle$, are pairwise orthogonal, invariant under B, and decompose $\mathbb{Q}^m$. Inside them $Bb_j = 2a_j$, which is what makes the chain move. Consequently $\mathcal{C}_B(P_j)$ exceeds $\mathbb{K}^{P_j}$ by the single direction b_j. At every stage before the last, the values of b_j separate $2j+1$ and $2j+2$ from each other and from the rest of the tail while the space contains no indicator of the three classes this creates; a nonlinearity names them, and B applied to the newly named classes produces the next direction b_{j+1}, one pair further down. At $m = 6$ the three stages each add one dimension and fall one, one and zero dimensions short of naming their classes, for gains of $2 + 2 + 1 = m - 1$; for odd m the slack unit of the normal form is spent at the last stage, which gains two classes.

A quotient becomes a graph by blowing each index up into a cycle of vertices, which preserves the quotient matrix and leaves the classes unsplit: the paired tail is realized by a connected simple graph on $2m^2 + 2m$ vertices with the same chain (Proposition C.1), and bounded-entry versions exist on 24, 32 and 60 vertices at $m = 6, 8, 10$ (Remark C.1). With Proposition 4, this settles the value $\lfloor m/2 \rfloor$ in Theorem 1(b) for every $m \geq 3$. All bounds are stated in the number m of stable classes, which is at most n; in n alone, the iteration number of color refinement reaches $n - 1$ for infinitely many n (Kiefer & McKay, 2020).

On these graphs message passing is no slower, because a stage that adds a single dimension is already a round of color refinement.

Proposition 6. *Let π be a partition of $V(G)$ with $\dim \mathcal{C}_A(\pi) = |\pi| + 1$. Then one round of color refinement from π returns $\rho(\pi)$. Consequently, on every graph whose stable coloring has an even number m of classes and needs $\tau = m/2$ stages, and on the realized paired tail for every $m \geq 3$, color refinement needs exactly τ rounds.*

Proof. One round of color refinement from π returns $\pi_{\mathbb{K}^\pi + A\mathbb{K}^\pi}$, since two vertices stay together exactly when they agree on the columns of K_π and AK_π, and a row-equality partition depends only on the span of the family. Now $\mathbb{K}^\pi \subseteq \mathbb{K}^\pi + A\mathbb{K}^\pi \subseteq \mathcal{C}_A(\pi)$, and the outer two differ by one dimension. If the middle space equalled $\mathbb{K}^\pi$, then $A\mathbb{K}^\pi \subseteq \mathbb{K}^\pi$ and $\mathcal{C}_A(\pi) = \mathbb{K}^\pi$, against the dimension count; so it equals $\mathcal{C}_A(\pi)$, and the two partitions coincide. For the consequence, Proposition 4 gives $r_t = 1$ at every stage when m is even, and Proposition 5 gives it on the paired tail for every m, so by induction color refinement follows the chain of walk refinement stage by stage. □

So on the paired tail, message passing and global filters from $\mathcal{A}$ need the same $\lfloor m/2 \rfloor$ layers, and all of that depth is spent on naming. Paths are the opposite extreme: their stable coloring has $m = \lceil n/2 \rceil$ classes, message passing needs $m - 1$ layers, and global filters need one (Propositions A.1 and A.2).

3.3 Consequences for Architectures

Linear and decoupled networks. Depth helps only when nonlinearities separate the filters. Call a network *linear* when every layer is $H^{(t+1)} = \sum_{M \in F_t} M H^{(t)} \Theta_M$ for weight matrices Θ_M.

Corollary 1. *Let $\mathcal{M}$ be a matrix algebra containing I and A. Every linear network over $\mathcal{M}$, of every depth, induces a partition coarser than $\rho_{\mathcal{M}}(\sigma_0)$, and when the initial features span $\mathbb{K}^{\sigma_0}$, one linear layer whose filters linearly span $\mathcal{M}$ attains it.*

Proof. The space $\mathcal{M}\mathbb{K}^{\sigma_0}$ contains the columns of $H^{(0)}$ and is invariant under $\mathcal{M}$ and under linear combination, so it contains the columns of every $H^{(t)}$. If the initial features span $\mathbb{K}^{\sigma_0}$, the filtered features of one layer whose filters span $\mathcal{M}$ span $\mathcal{M}\mathbb{K}^{\sigma_0}$ itself. □

A decoupled architecture, which propagates once and hands the result to a vertex-wise predictor, is thus a network of depth one, and however far it propagates it reaches no further than the first stage of ρ, or of $\rho_{\mathcal{D}}$ when it normalizes.

Normalized filters. Degree normalization multiplies by the degree vector, $Dx = (A\mathbf{1}) \odot x$ with $\odot$ the entrywise product. The algebra $\mathcal{D}$ therefore contains multiplication by every entrywise power of $A\mathbf{1}$, and these powers span the indicators of the degree classes, so normalized filters name the degree classes for free: on graphs with $m \geq 3$ stable classes the first stage of $\rho_{\mathcal{D}}$ adds at least two dimensions and, when it is not the last, gains at least three classes. This is worth at most one layer in the worst case, as Theorem 1(b) records, and exactly one on the paired tail with $m \geq 4$ (Proposition C.2).

Input colors, readouts and attention. Input features that one-hot encode c colors start the chain at c classes, and the same accounting bounds the depth by $\lfloor (m - c + 1)/2 \rfloor$ when $m \geq c + 2$, which colored graphs attain (Proposition C.3). A global sum or mean readout, a virtual node, or the global attention of graph transformers (Rampášek et al., 2022; Cai et al., 2023; Rosenbluth et al., 2024), used without positional encodings, feeds every vertex the same multiset of features; it cannot split a class, and combined with message passing it separates no more than message passing alone (Proposition B.3).

4 Multiplicative Depth

Naming a class costs a multiplication. We extend the entrywise product to subspaces, writing $U \odot U'$ for the span of the products $x \odot y$ with $x \in U$ and $y \in U'$. Neither of the two operations that build the stable coloring advances alone: the linear closure is idempotent, and products never separate anything, since a product of vectors that agree at two coordinates agrees there too. The lemma below says which subspaces the multiplicative closure fixes, and that only the whole space is fixed by both.

Lemma 1. *A subspace of $\mathbb{Q}^m$ that contains $\mathbf{1}$ and is closed under $\odot$ is $\mathbb{K}^\sigma$ for a partition σ of $[m]$. Consequently, for a reduced quotient B of order m, the only subspace of $\mathbb{Q}^m$ containing $\mathbf{1}$, invariant under B, and closed under $\odot$ is $\mathbb{Q}^m$ itself.*

The proof builds the indicator of each class as a product of normalized differences $(x - x_s\mathbf{1})/(x_u - x_s)$ (Appendix D). All of the difficulty is thus in the alternation: the nonlinearity of a layer converts

computed values into the indicators of the classes they cut out, which the next filter then propagates. A stage that separates classes without containing their indicators reaches them only by multiplying, and on the paired tail the missing dimension is literally a square:

$$\mathbb{K}^{P_{j+1}} \;=\; \mathcal{C}_B(P_j) \oplus \langle b_j \odot b_j \rangle \qquad (0 \le j \le q-2), \tag{2}$$

where $b_j \odot b_j = e_{2j+1} + e_{2j+2}$. Counting these multiplications, with filtering free, gives a second measure of the same depth, the number of alternations between filtering and multiplication. Extend $\mathcal{C}_A$ to subspaces by $\mathcal{C}_A(U) = \sum_{l \ge 0} A^l U$.

Definition 4. The *multiplicative ladder* of G is the chain of subspaces

$$S_0 \;=\; \mathcal{C}_A(\pi^{(0)}), \qquad S_{j+1} \;=\; \mathcal{C}_A\big(S_j \odot S_j\big),$$

and the *multiplicative depth* $d_\odot(G)$ is the least j with $\pi_{S_j} = W$.

Theorem 2. *Every graph whose stable coloring has $m \ge 3$ classes satisfies $\tau - 1 \le d_\odot \le m-3$, and the realized paired tail has $d_\odot = \tau - 1 = \lfloor m/2 \rfloor - 1$. Consequently $\lfloor m/2 \rfloor - 1 \le \max d_\odot \le m-3$ over graphs whose stable coloring has $m \ge 3$ classes.*

The lower bound comes from $S_j \subseteq \mathbb{K}^{\pi^{(j+1)}}$, proved by induction: products of vectors constant on the classes of a partition stay constant on them, and a closure of $\mathbb{K}^{\pi^{(j+1)}}$ lies in $\mathbb{K}^{\pi^{(j+2)}}$. On the paired tail, Equation (2) shows that one round of products supplies exactly the missing dimension, so the ladder follows the chain of walk refinement; the upper bound is a dimension count (Appendix D).

At $m = 6$ the ladder of the paired tail is short enough to see, and it is the computation drawn in Figure 1(b). There $b_0 = e_1 - e_2$, so $S_0 = \langle \mathbf{1}, b_0 \rangle$ separates 1, 2 and $\{3, 4, 5, 6\}$ while containing no indicator of those three classes. One multiplication supplies what is missing, $b_0 \odot b_0 = e_1 + e_2$, and $S_0 \odot S_0 = \mathbb{K}^{P_1}$. Closing under B produces S_1, whose extra direction over $\mathbb{K}^{P_1}$ is $b_1 = e_3 - e_4$; one further square and closure give $S_2 = \mathbb{Q}^6$, so $d_\odot = 2 = \tau - 1$.

So no constant number of alternations between filtering and multiplication computes the 1-WL coloring, unconditionally. The levels themselves are cheap: exact rational linear algebra lies in NC^2 (Borodin et al., 1982; Mulmuley, 1987), so each fixed level of the ladder is computable in parallel, while deciding color-refinement equivalence is complete for polynomial time (Grohe, 1999); what is sequential is the alternation. The same count prices a static budget: homomorphism counts of rooted trees characterize color-refinement equivalence (Dvořák, 2010; Dell et al., 2018), and the leaf budget $\lambda(G)$, the least k for which the counts of rooted trees with at most k leaves induce the stable coloring, is exactly $2^{\tau-1}$ on the paired tail, every stage past the first doubling it (Theorem D.1). There the three measures coincide, $\tau - 1 = d_\odot = \log_2 \lambda$, so one number is a stage count, a multiplication count and an alternation count at once. The same numbers are quantifier depths: walk refinement corresponds to a counting logic whose quantifier counts walks of every length, with worst-case depth $\lfloor m/2 \rfloor$ against $m - 1$ for the fragment that counts neighbors (Corollary E.1).

5 Conclusion

Message passing gains at least one class in every layer that changes anything, so on graphs with m stable classes it can need $m - 1$ layers. Global filters let every layer see the whole graph, and the guarantee doubles: every layer but the last gains at least two classes, and with filters from $\mathcal{A}$ the worst case is exactly $\lfloor m/2 \rfloor$, whatever the filters, update functions or width. The doubling has a single source: before the last layer a filter cannot name every class it separates, so each layer gains one class for each new dimension and one for each dimension still missing to name its classes, and neither count is zero.

For architecture design, the depth that full expressive power costs counts nonlinear layers, however far each layer sees: one fixed Katz filter per layer already achieves the optimum, linear and decoupled networks stop at the first stage, degree normalization lowers the worst case by at most one layer, and on the extremal graphs we construct, reach is free and the whole depth is naming.

Naming is also where the multiplications are. Multiplicative depth grows with the number of stable classes, so no constant number of alternations between filtering and multiplication computes the 1-WL coloring (Theorem 2). The layers that remain are the times a network has to stop and name what it has separated, and no amount of filtering inside a layer removes one of them.

AI USE STATEMENT

The authors used OpenAI's ChatGPT and Anthropic's Claude to help organize and polish the exposition of this paper, based on the authors' own drafts. The same tools assisted in surveying the related literature. The authors formulated and proved all mathematical claims in this paper, and checked, revised, and finalized every statement, proof, and reference. They take full responsibility for the content of the submission.

REPRODUCIBILITY STATEMENT

Every result in this paper is a theorem with a complete proof. Propositions 1, 3, 4 and 6 and Corollary 1 are proved in the main text. The appendix proves the rest: Appendix A, the statements on paths and the quotient correspondence; Appendix B, Proposition 2 and Theorem 1(a); Appendix C, Proposition 5, the realization and the degree algebra; Appendix D, Lemma 1 and Theorem 2; and Appendix E, the logical reading. Every construction is explicit. The paired tail is given in closed form in Equation (1), with its instance at $m = 6$ displayed in Section 3.2; Proposition C.1 turns it into a connected simple graph on $2m^2 + 2m$ vertices, and Remark C.1 lists the bounded-entry quotients in full.

REFERENCES

Uri Alon and Eran Yahav. On the bottleneck of graph neural networks and its practical implications. In *9th International Conference on Learning Representations (ICLR 2021)*, 2021. arXiv:2006.05205.

Pablo Barceló, Egor V. Kostylev, Mikael Monet, Jorge Pérez, Juan L. Reutter, and Juan Pablo Silva. The logical expressiveness of graph neural networks. In *8th International Conference on Learning Representations (ICLR 2020)*, 2020.

Allan Borodin, Joachim von zur Gathen, and John Hopcroft. Fast parallel matrix and GCD computations. *Information and Control*, 52(3):241–256, 1982. doi: 10.1016/S0019-9958(82)90766-5.

Chen Cai, Truong Son Hy, Rose Yu, and Yusu Wang. On the connection between MPNN and graph transformer. In *Proceedings of the 40th International Conference on Machine Learning (ICML 2023)*, volume 202 of *PMLR*, pp. 3408–3430, 2023. arXiv:2301.11956.

Ada Chan and Chris D. Godsil. Symmetry and eigenvectors. In Gena Hahn and Gert Sabidussi (eds.), *Graph Symmetry: Algebraic Methods and Applications*, volume 497 of *NATO Advanced Science Institutes Series C*, pp. 75–106. Kluwer Academic Publishers, 1997. doi: 10.1007/978-94-015-8937-6_3.

Eli Chien, Jianhao Peng, Pan Li, and Olgica Milenkovic. Adaptive universal generalized PageRank graph neural network. In *9th International Conference on Learning Representations (ICLR 2021)*, 2021. arXiv:2006.07988.

Sang Cho and Dung T. Huynh. The parallel complexity of coarsest set partition problems. *Information Processing Letters*, 42(2):89–94, 1992. doi: 10.1016/0020-0190(92)90095-D.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, 2016. arXiv:1606.09375.

Holger Dell, Martin Grohe, and Gaurav Rattan. Lovász meets Weisfeiler and Leman. In *45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)*, volume 107 of *LIPIcs*, pp. 40:1–40:14, 2018. doi: 10.4230/LIPIcs.ICALP.2018.40.

Francesco Di Giovanni, Lorenzo Giusti, Federico Barbero, Giulia Luise, Pietro Liò, and Michael M. Bronstein. On over-squashing in message passing neural networks: The impact of width, depth, and topology. In *Proceedings of the 40th International Conference on Machine Learning (ICML 2023)*, volume 202 of *PMLR*, pp. 7865–7885, 2023. arXiv:2302.02941.

Zdeněk Dvořák. On recognizing graphs by numbers of homomorphisms. *Journal of Graph Theory*, 64(4):330–342, 2010. doi: 10.1002/jgt.20461.

Alexander Farrugia and Irene Sciriha. On the main eigenvalues of universal adjacency matrices and U-controllable graphs. *Electronic Journal of Linear Algebra*, 30:812–826, 2015. doi: 10.13001/1081-3810.2935.

Martin Fürer. Weisfeiler–Lehman refinement requires at least a linear number of iterations. In *28th International Colloquium on Automata, Languages and Programming (ICALP 2001)*, volume 2076 of *LNCS*, pp. 322–333, 2001. doi: 10.1007/3-540-48224-5_27.

Jingchu Gai, Yiheng Du, Bohang Zhang, Haggai Maron, and Liwei Wang. Homomorphism expressivity of spectral invariant graph neural networks. In *13th International Conference on Learning Representations (ICLR 2025)*, 2025. arXiv:2503.00485.

Johannes Gasteiger, Aleksandar Bojchevski, and Stephan Günnemann. Predict then propagate: Graph neural networks meet personalized PageRank. In *7th International Conference on Learning Representations (ICLR 2019)*, 2019. arXiv:1810.05997; published under the author name Johannes Klicpera.

Floris Geerts. Walk message passing neural networks and second-order graph neural networks, 2020. arXiv:2006.09499.

Chris Godsil and Gordon Royle. *Algebraic Graph Theory*, volume 207 of *Graduate Texts in Mathematics*. Springer, New York, 2001. doi: 10.1007/978-1-4613-0163-9.

Martin Grohe. Equivalence in finite-variable logics is complete for polynomial time. *Combinatorica*, 19(4):507–532, 1999. doi: 10.1007/s004939970004.

Martin Grohe. The descriptive complexity of graph neural networks. *TheoretiCS*, 3, 2024. doi: 10.46298/theoretics.24.25. Article 25.

Willem H. Haemers and Gholam Reza Omidi. Universal adjacency matrices with two eigenvalues. *Linear Algebra and its Applications*, 435(10):2520–2529, 2011. doi: 10.1016/j.laa.2011.02.023.

Mingguo He, Zhewei Wei, Zengfeng Huang, and Hongteng Xu. BernNet: Learning arbitrary graph spectral filters via Bernstein approximation. In *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, pp. 14239–14251, 2021. arXiv:2106.10994.

Neil Immerman and Eric S. Lander. Describing graphs: A first-order approach to graph canonization. In Alan L. Selman (ed.), *Complexity Theory Retrospective*, pp. 59–81. Springer, New York, 1990. doi: 10.1007/978-1-4612-4478-3_5.

Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953. doi: 10.1007/BF02289026.

Kristian Kersting, Martin Mladenov, Roman Garnett, and Martin Grohe. Power iterated color refinement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28, pp. 1904–1910, 2014. doi: 10.1609/aaai.v28i1.8992.

Sandra Kiefer and T. Devini de Mel. A classification of long-refinement graphs for colour refinement. In *2026 ACM-SIAM Symposium on Discrete Algorithms (SODA 2026)*, pp. 388–407, 2026. arXiv:2510.20802.

Sandra Kiefer and Brendan D. McKay. The iteration number of colour refinement. In *47th International Colloquium on Automata, Languages, and Programming (ICALP 2020)*, volume 168 of *LIPIcs*, pp. 73:1–73:19, 2020. doi: 10.4230/LIPIcs.ICALP.2020.73.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations (ICLR 2017)*, 2017. arXiv:1609.02907.

Andreas Krebs and Oleg Verbitsky. Universal covers, color refinement, and two-variable counting logic: Lower bounds for the depth. In *30th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS 2015)*, pp. 689–700, 2015. doi: 10.1109/LICS.2015.69.

Nils M. Kriege. Weisfeiler and Leman go walking: Random walk kernels revisited. In *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, pp. 20119–20132, 2022. arXiv:2205.10914.

Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In *32nd AAAI Conference on Artificial Intelligence (AAAI 2018)*, 2018. doi: 10.1609/aaai.v32i1.11604. arXiv:1801.07606.

Moritz Lichter, Ilia Ponomarenko, and Pascal Schweitzer. Walk refinement, walk logic, and the iteration number of the Weisfeiler–Leman algorithm. In *34th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS 2019)*, pp. 1–13, 2019. doi: 10.1109/LICS.2019.8785694.

Andreas Loukas. What graph neural networks cannot learn: depth vs width. In *8th International Conference on Learning Representations (ICLR 2020)*, 2020. arXiv:1907.03199.

H. L. Morgan. The generation of a unique machine description for chemical structures: a technique developed at Chemical Abstracts Service. *Journal of Chemical Documentation*, 5(2):107–113, 1965. doi: 10.1021/c160017a018.

Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and Leman go neural: Higher-order graph neural networks. In *33rd AAAI Conference on Artificial Intelligence (AAAI 2019)*, pp. 4602–4609, 2019. doi: 10.1609/aaai.v33i01.33014602.

Christopher Morris, Yaron Lipman, Haggai Maron, Bastian Rieck, Nils M. Kriege, Martin Grohe, Matthias Fey, and Karsten Borgwardt. Weisfeiler and Leman go machine learning: The story so far. *Journal of Machine Learning Research*, 24(333):1–59, 2023. arXiv:2112.09992.

Christopher Morris, Fabrizio Frasca, Nadav Dym, Haggai Maron, İsmail İlkan Ceylan, Ron Levie, Derek Lim, Michael M. Bronstein, Martin Grohe, and Stefanie Jegelka. Position: Future directions in the theory of graph machine learning. In *Proceedings of the 41st International Conference on Machine Learning (ICML 2024)*, volume 235 of *PMLR*, pp. 36294–36307, 2024. arXiv:2402.02287.

Ketan Mulmuley. A fast parallel algorithm to compute the rank of a matrix over an arbitrary field. *Combinatorica*, 7(1):101–104, 1987. doi: 10.1007/BF02579205.

Kenta Oono and Taiji Suzuki. Graph neural networks exponentially lose expressive power for node classification. In *8th International Conference on Learning Representations (ICLR 2020)*, 2020. arXiv:1905.10947.

David L. Powers and Mohammad M. Sulaiman. The walk partition and colorations of a graph. *Linear Algebra and its Applications*, 48:145–159, 1982. doi: 10.1016/0024-3795(82)90104-5.

Motakuri V. Ramana, Edward R. Scheinerman, and Daniel H. Ullman. Fractional isomorphism of graphs. *Discrete Mathematics*, 132(1-3):247–265, 1994. doi: 10.1016/0012-365X(94)90241-0.

Ladislav Rampášek, Mikhail Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. Recipe for a general, powerful, scalable graph transformer. In *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*, pp. 14501–14515, 2022. arXiv:2205.12454.

Eran Rosenbluth, Jan Tönshoff, Martin Ritzert, Berke Kisin, and Martin Grohe. Distinguished in uniform: Self-attention vs. virtual nodes. In *12th International Conference on Learning Representations (ICLR 2024)*, 2024. arXiv:2405.11951.

Antonis Vasileiou, Juan Cervino, Pascal Frossard, Charilaos I. Kanatsoulis, Christopher Morris, Michael T. Schaub, Pierre Vandergheynst, Zhiyang Wang, Guy Wolf, and Ron Levie. Graph learning should move beyond restrictive views of spectral and message-passing GNNs, 2026. arXiv:2602.10031.

Xiyuan Wang and Muhan Zhang. How powerful are spectral graph neural networks. In *Proceedings of the 39th International Conference on Machine Learning (ICML 2022)*, volume 162 of *PMLR*, pp. 23341–23362, 2022. arXiv:2205.11172.

Felix Wu, Tianyi Zhang, Amauri Holanda de Souza Jr., Christopher Fifty, Tao Yu, and Kilian Q. Weinberger. Simplifying graph convolutional networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML 2019)*, volume 97 of *PMLR*, pp. 6861–6871, 2019. arXiv:1902.07153.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *7th International Conference on Learning Representations (ICLR 2019)*, 2019. arXiv:1810.00826.

A PRELIMINARIES IN FULL

This appendix supplies the quotient correspondence on which Sections 3 and 4 compute, the facts about walk refinement that make $\tau(G)$ and $\tau_{\mathcal{D}}(G)$ in Definition 2 well defined, and the two statements about paths that Theorem 1(b) and Section 1 use.

A.1 COLOR REFINEMENT AND QUOTIENTS

We keep the notation of Section 2. All linear algebra takes place over $\mathbb{Q}$; every space below is spanned by integer vectors, so the row-equality partitions it induces do not change when the scalars are extended to $\mathbb{R}$, the field of the features in Definition 1. The stable coloring W is the coarsest equitable partition of $V(G)$, so every equitable partition refines it (Godsil & Royle, 2001; Ramana et al., 1994), and the iteration number of color refinement is at most $n-1$ on graphs with n vertices, a bound attained for infinitely many n (Kiefer & McKay, 2020).

Since W is equitable, the number B_{ij} of neighbors in class j of a vertex in class i depends only on i, and the resulting $m \times m$ matrix B is the *quotient matrix* of G. Writing $\Phi \in \{0,1\}^{V(G)\times[m]}$ for the characteristic matrix of W, this reads

$$A\Phi \;=\; \Phi B.$$

A partition π of $V(G)$ coarser than W is induced by a partition σ of $[m]$, with $K_\pi = \Phi K_\sigma$. Since $A^l\Phi = \Phi B^l$, the columns $A^l K_\pi$ are the lifts of $B^l K_\sigma$, and row equality on $V(G)$ is the preimage of row equality on $[m]$. The chains of this paper stay coarser than W (Appendix A.2), so Sections 3 and 4 and Appendices C and D work on B, where the stable coloring is the discrete partition of $[m]$. The correspondence extends to the degree algebra: the degree of a vertex in class i is the row sum $(B\mathbf{1})_i$, so the degree matrix of G satisfies $D\Phi = \Phi\,\mathrm{diag}(B\mathbf{1})$, and every word in A and D applied to Φ is Φ times the same word in B and $\mathrm{diag}(B\mathbf{1})$.

A round that changes anything splits at least one class, so a refinement process that starts from the uniform coloring and reaches an m-class coloring runs for at most $m-1$ rounds. Paths attain that bound, and Theorem 1 measures walk refinement against it.

Proposition A.1. *For $n \geq 3$ the stable coloring of the path P_n has $m = \lceil n/2 \rceil$ classes, and the iteration number of color refinement on P_n is $m-1$.*

Proof. Label $V(P_n) = \{1,\dots,n\}$ along the path and let $d(v) = \min\{v-1,\, n-v\}$ be the level of v, its distance to the nearer endpoint, so that the levels are $0,1,\dots,h$ with $h = \lfloor (n-1)/2 \rfloor \geq 1$. A vertex at level $j < h$ has its neighbors at levels $j-1$ and $j+1$, an endpoint having only the one at level 1; the vertices at level h have their neighbors at levels $h-1$ or h. In particular the level partition is equitable.

We claim that after t rounds the coloring is the partition into level sets of $\min\{d(v),\, t\}$. At $t=0$ this is the uniform coloring. Assume it for some $t < h$. The levels below t are then already separated from one another and from the rest, so only the class of vertices at level t or above can split. Within it, a vertex at level t sees the color $t-1$ on one side while a vertex above level t sees the color t on both; at $t=0$ the two are told apart by their degrees. The round splits off level t and nothing further, which is the claim at $t+1$.

At $t = h$ the coloring is the level partition, which is equitable and hence stable, while at $t = h-1$ it is strictly coarser. Therefore $m = h+1 = \lceil n/2 \rceil$ and the iteration number is $h = m-1$. □

Walk refinement finishes the same graphs in a single stage: the whole Krylov space of $\mathbf{1}$ is available at once, and on a path it is the whole of $\mathbb{K}^W$.

Proposition A.2. *For $n \geq 3$ walk refinement reaches the stable coloring of P_n in one stage: $\pi^{(1)} = W$, so $\tau(P_n) = 1$. The stage adds $m-1$ dimensions and separates no class it cannot name, $\dim \mathcal{C}_A(\pi^{(0)}) = m = |W|$; in the notation of Proposition 4, $r_0 = m-1$ and $\delta_0 = 0$.*

Proof. The level partition W of Proposition A.1 is equitable and $\mathbf{1} \in \mathbb{K}^W$, so $\mathcal{C}_A(\pi^{(0)}) = \mathrm{span}\{A^l\mathbf{1} : l \geq 0\} \subseteq \mathbb{K}^W$, a space of dimension $m = \lceil n/2 \rceil$. The eigenvalues of A are $2\cos(\pi k/(n+1))$ for $k = 1,\dots,n$, pairwise distinct, with eigenvectors $s_k(j) = \sin(\pi kj/(n+1))$. Writing $\mathbf{1} = \sum_k c_k s_k$,

the space $\operatorname{span}\{A^l\mathbf{1} : l \geq 0\}$ is spanned by the s_k with $c_k \neq 0$, by a Vandermonde argument in the distinct eigenvalues, so its dimension is the number of k with $\langle \mathbf{1}, s_k \rangle \neq 0$. For even k the symmetry $s_k(n+1-j) = -s_k(j)$ gives $\langle \mathbf{1}, s_k \rangle = 0$. For odd k, summing the sine series with $\theta = \pi k/(n+1)$,

$$\langle \mathbf{1}, s_k \rangle \;=\; \sum_{j=1}^{n} \sin(j\theta) \;=\; \frac{\sin(n\theta/2)\,\sin((n+1)\theta/2)}{\sin(\theta/2)} \;=\; \cot\!\Big(\frac{\pi k}{2(n+1)}\Big) \;>\; 0,$$

since $\sin((n+1)\theta/2) = \sin(\pi k/2) = (-1)^{(k-1)/2}$ and $\sin(n\theta/2) = \sin(\pi k/2 - \theta/2) = (-1)^{(k-1)/2}\cos(\theta/2)$ for odd k, while $0 < \pi k/(2(n+1)) < \pi/2$. There are exactly $\lceil n/2 \rceil = m$ odd indices in $\{1, \ldots, n\}$, so $\mathcal{C}_A(\pi^{(0)})$ has dimension m and equals $\mathbb{K}^W$. Hence $\pi^{(1)} = \pi_{\mathbb{K}^W} = W$, and $\tau(P_n) = 1$ because W is not the uniform coloring for $n \geq 3$. The dimension count gives $r_0 = m - 1$ and $\delta_0 = |\pi^{(1)}| - \dim \mathcal{C}_A(\pi^{(0)}) = 0$. □

A.2 Walk Refinement

Let $\mathcal{M}$ be one of the algebras $\mathcal{A}$ and $\mathcal{D}$ of Section 2, and write $\mathcal{C}_{\mathcal{M}}(\pi) = \mathcal{M}\mathbb{K}^\pi$, so that $\mathcal{C}_{\mathcal{A}} = \mathcal{C}_A$ and $\rho_{\mathcal{M}}(\pi) = \pi_{\mathcal{C}_{\mathcal{M}}(\pi)}$ as in Definition 2. The map $\rho_{\mathcal{M}}$ is monotone and refining. If σ is coarser than π then $\mathbb{K}^\sigma \subseteq \mathbb{K}^\pi$, hence $\mathcal{C}_{\mathcal{M}}(\sigma) \subseteq \mathcal{C}_{\mathcal{M}}(\pi)$, and π_X is coarser for smaller X, so $\rho_{\mathcal{M}}(\sigma)$ is coarser than $\rho_{\mathcal{M}}(\pi)$; and $\mathbb{K}^\pi \subseteq \mathcal{C}_{\mathcal{M}}(\pi)$, as $I \in \mathcal{M}$, makes $\rho_{\mathcal{M}}(\pi)$ a refinement of π. The fixed points are the equitable partitions: a family always lies in the space of vectors constant on its own row-equality classes, so $\rho_{\mathcal{M}}(\pi) = \pi$ gives $\mathcal{C}_{\mathcal{M}}(\pi) \subseteq \mathbb{K}^\pi$, hence $\mathcal{C}_{\mathcal{M}}(\pi) = \mathbb{K}^\pi$ and $A\mathbb{K}^\pi \subseteq \mathbb{K}^\pi$; conversely an equitable partition has $\mathbb{K}^\pi$ invariant under A and, the degree being constant on each class, under D, so $\mathcal{M}\mathbb{K}^\pi = \mathbb{K}^\pi$. The chain $\pi^{(0)}, \rho_{\mathcal{M}}(\pi^{(0)}), \ldots$ consequently stabilizes at an equitable partition, which refines W; and since $\rho_{\mathcal{M}}$ fixes W and $\pi^{(0)}$ is coarser than W, monotonicity keeps every iterate coarser than W. The limit is W. This makes $\tau(G)$ and $\tau_{\mathcal{D}}(G)$ well defined and confirms that both chains stay in the range where the quotient correspondence applies. The same argument runs verbatim from any initial coloring $\pi^{(0)}$, with W the coarsest equitable partition refining it.

The first update $\pi^{(1)}$ of ρ is the walk partition of Powers and Sulaiman (Powers & Sulaiman, 1982), which need not be equitable and in general falls short of W; the later iterations are what close that gap. The name follows Lichter et al. (2019), whose walk refinement is the two-dimensional analogue, acting on colors of vertex pairs.

B The Network Model in Full

This appendix records what Definition 1 covers and what it excludes, proves Proposition 2, part (a) of Theorem 1 and the monotonicity $\tau_{\mathcal{D}}(G) \leq \tau(G) \leq$ the color-refinement iteration number, treats global readouts, and presents the algorithmic special case of the model in which walk refinement is optimal.

B.1 Scope of the Model

With $\mathcal{M} = \{A\}$, Definition 1 is message passing in its most permissive form (Xu et al., 2019; Morris et al., 2019), covering sum aggregation with arbitrary message and update functions, the mean and maximum aggregators, and attention weights computed from the two endpoints of an edge, each being a function of the neighbor multiset. With $\mathcal{M} \subseteq \mathcal{A}$ it adds every fixed filter that is a function of A: polynomials of any degree, the Katz resolvent of Proposition 2, the exponential e^{zA}, and the fixed point of a linear propagation, all polynomials in A by Cayley–Hamilton. The filters that architectures deploy normalize by degree and therefore live in $\mathcal{D}$, not in $\mathcal{A}$: the normalized operators $D^{-1}A$, $D^{-1/2}AD^{-1/2}$ and the Laplacians (Kipf & Welling, 2017; Wu et al., 2019; Defferrard et al., 2016), with reciprocals of zero degrees read as zero, and personalized PageRank (Gasteiger et al., 2019), which is the resolvent of a normalized operator, not of A.

A layer of Definition 1 reads the neighbor multiset of the current features $H_w^{(t)}$. A layer that pools filtered features nonlinearly over neighbors, for instance $\max_{w \in N(v)}(MH^{(t)})_w$, counts as two layers here, a filter layer followed by an aggregation layer, and may cover more than one stage; linear aggregation of filtered features stays inside a single layer, since $\sum_{w \in N(v)}(MH^{(t)})_w = (AMH^{(t)})_v$ and $AM \in \mathcal{A}$ whenever $M \in \mathcal{A}$.

What lies outside is measured against the algebras $\mathcal{A}$ and $\mathcal{D}$, never against Definition 1 itself, whose unrestricted filter set excludes nothing. A layer that reads a structural value attached to each *pair* of vertices, as attention with relative positional encodings does, is outside both algebras: it can compare the whole multiset of those values where a filter forms one weighted sum. On the four-regular graph on seven vertices obtained from $K_{4,3}$ by adding two disjoint edges inside the side of size four, color refinement and walk refinement merge all seven vertices, while the multisets $\{\{A^2(v,w) : w \in V(G)\}\}$ take the two values $\{\{4,3,3,3,1,1,1\}\}$ and $\{\{4,4,4,1,1,1,1\}\}$. Counts of closed walks, the diagonal entries of A^l, leave both algebras as well: on the disjoint union of a triangle and a hexagon, a regular graph on which every network over $\mathcal{A}$ or $\mathcal{D}$ from constant features keeps all nine vertices together, the diagonal of A^3 is 2 on the triangle and 0 on the hexagon. Aggregating over the vertices at a fixed distance leaves the algebras for the same reason, a distance-k matrix being in general no polynomial in A; on this particular graph it is, the distance-two matrix being $J - I - A$ with $J = \mathbf{1}\mathbf{1}^{\mathsf{T}}$. Architectures coloring tuples or subgraphs leave dimension one altogether.

B.2 Faithful Attenuations

A stage of walk refinement asks for every walk length at once, and one rational attenuation delivers all of them in a single linear solve.

Proposition B.1. *Let $\zeta_G = 1/(\Delta^n + \Delta + 1)$. For every partition π of $V(G)$,*

$$\rho(\pi) \;=\; \pi_{H_\pi}, \qquad H_\pi \;=\; (I - \zeta_G A)^{-1} K_\pi.$$

Proof. If $\Delta = 0$ both sides equal π, so assume $\Delta \geq 1$. Then $\zeta_G \Delta < 1$, the Neumann series converges, and $H_\pi = \sum_{l \geq 0} \zeta_G^l A^l K_\pi$. Fix two vertices u and v, and for a class C of π set

$$d_l \;=\; (A^l K_\pi)_{uC} - (A^l K_\pi)_{vC} \;\in\; \mathbb{Z},$$

so that the entries of H_π in column C differ by $\sum_{l \geq 0} d_l \, \zeta_G^l$. If $\rho(\pi)$ keeps u and v together, then every d_l vanishes for every class, and the two rows of H_π agree.

Suppose instead $\rho(\pi)$ separates them, and choose a class C for which the d_l do not all vanish. By Cayley–Hamilton the sequence $(d_l)_{l \geq 0}$ satisfies a linear recurrence of order n, so vanishing at $l = 0, \ldots, n-1$ would force vanishing throughout, and the first nonzero term occurs at some index $l_0 < n$. A vertex is the start of at most Δ^l walks of length l, whence $|d_l| \leq \Delta^l$, while $|d_{l_0}| \geq 1$. The tail is therefore dominated by the leading term:

$$\sum_{l > l_0} |d_l| \, \zeta_G^l \;\leq\; \zeta_G^{l_0} \cdot \frac{\Delta^{l_0+1} \zeta_G}{1 - \Delta \zeta_G} \;\leq\; \zeta_G^{l_0} \cdot \frac{\Delta^n \zeta_G}{1 - \Delta \zeta_G} \;<\; \zeta_G^{l_0} \;\leq\; |d_{l_0}| \, \zeta_G^{l_0},$$

the strict inequality being the choice $\zeta_G(\Delta^n + \Delta) < 1$, so the two rows of H_π differ. The two partitions therefore agree, and since the estimate involves only n and Δ, they do so for every π at once. □

We call an attenuation with this property *faithful* for G, as in Proposition 2. Faithfulness depends on which attenuation is chosen: on the seven-vertex graph with edges $01, 05, 06, 12, 13, 14, 23, 24$, the profile $(I - zA)^{-1}\mathbf{1}$ at $z = 1/4$ takes the value $20/7$ at both vertex 0 and vertex 3, whose degrees are 3 and 2, while ρ separates them at its first step. The explicit ζ_G is needed only for one value that works for every graph and every partition at once; almost every attenuation, and almost every polynomial filter of the right degree, is faithful.

Proposition B.2. *Fix G and a partition π, and put $r = \dim \mathcal{C}_A(\pi) - |\pi|$, at most $m - 1$ when π is coarser than W. For all but finitely many z with $\det(I - zA) \neq 0$, $\pi_{(I - zA)^{-1} K_\pi} = \rho(\pi)$; and $\pi_{\sum_{l \leq r} \theta_l A^l K_\pi} = \rho(\pi)$ for every $\theta \in \mathbb{R}^{r+1}$ outside a finite union of hyperplanes.*

Proof. Both families lie in $\mathcal{C}_A(\pi)$, so both partitions are coarser than $\rho(\pi)$ whatever the parameters. The spaces $U_j = \sum_{l \leq j} A^l \mathbb{K}^\pi$ increase, and $U_{j+1} = U_j$ makes U_j invariant under A, hence equal to $\mathcal{C}_A(\pi)$; since $\dim U_0 = |\pi|$, the chain arrives at $j = r$. So a pair u, v separated by $\rho(\pi)$ has $d_l := (A^l K_\pi)_{uC} - (A^l K_\pi)_{vC} \neq 0$ for some $l \leq r$ and some class C, the linear form $\theta \mapsto \sum_{l \leq r} \theta_l d_l$

is nonzero, and its kernel is a hyperplane; finitely many triples (u, v, C) give a finite union. For the resolvent, $z \mapsto ((I - zA)^{-1}K_\pi)_{uC} - ((I - zA)^{-1}K_\pi)_{vC}$ is a rational function whose expansion at 0 has the coefficients d_l, so it is not identically zero and has finitely many zeros. □

A random attenuation, or random filter coefficients, therefore carry out a stage of walk refinement almost surely. Since G has finitely many partitions, Proposition B.2 also shows that every z outside a finite set is faithful for G, which is the last claim of Proposition 2.

B.3 Proof of Proposition 2

Proof of Proposition 2. Let ζ be faithful for G, let π be a partition of $V(G)$, and write σ for the partition by equality of the arguments of ϕ at $H = K_\pi$. The arguments include the rows of K_π and of $(I-\zeta A)^{-1}K_\pi$, so σ refines $\pi_{(I-\zeta A)^{-1}K_\pi} = \rho(\pi)$ by faithfulness; and $\rho(\pi)$ refines σ by Proposition 1, since $(I - \zeta A)^{-1} \in \mathcal{A}$. So $\sigma = \rho(\pi)$, and ϕ returns the indicators $K_{\rho(\pi)}$. The attenuation ζ_G is faithful by Proposition B.1, and every ζ outside a finite set is faithful by Proposition B.2 applied to the finitely many partitions of $V(G)$. □

One attenuation serves every layer, faithfulness being a property of G alone.

B.4 Proof of Theorem 1, Part (a)

Proof of Theorem 1(a). Constant nonzero initial features have column space $\langle \mathbf{1} \rangle = \mathbb{K}^{\pi^{(0)}}$: they induce the uniform coloring $\sigma_0 = \pi^{(0)}$ and span its space. Let $\mathcal{M}$ be $\mathcal{A}$ or $\mathcal{D}$ and let a network of depth L over a subset of $\mathcal{M}$ have features inducing W. By Proposition 1, $\rho^L_{\mathcal{M}}(\pi^{(0)})$ refines $\sigma_L = W$, and every iterate of $\rho_{\mathcal{M}}$ is coarser than W (Appendix A.2), so $\rho^L_{\mathcal{M}}(\pi^{(0)}) = W$ and L is at least the iteration number of $\rho_{\mathcal{M}}$. For the filter set $\{A\}$ the same argument runs with one round of color refinement in place of $\rho_{\mathcal{M}}$: the arguments of ϕ_t at v are the row $H^{(t)}_v$, the multiset of the rows $H^{(t)}_w$ over $w \in N(v)$, and $(AH^{(t)})_v = \sum_C |N(v) \cap C|\, H^{(t)}_C$, the sum over the classes C of σ_t; all three are determined by the class of v and the counts $|N(v) \cap C|$, that is by the rows of K_{σ_t} and AK_{σ_t}. So σ_{t+1} is coarser than $\pi_{\mathbb{K}^{\sigma_t} + A\mathbb{K}^{\sigma_t}}$, the result of one round of color refinement from σ_t, and induction with the monotonicity of color refinement shows that t rounds refine σ_t; a network reaching W has depth at least the color-refinement iteration number.

For attainment over $\mathcal{A}$, the row-equality partition of a matrix depends only on its column space, which $H^{(0)}$ shares with $K_{\pi^{(0)}}$; so the layer of Proposition 2, with one Katz filter and any faithful attenuation, returns $K_{\rho(\pi^{(0)})}$ from $H^{(0)}$, and t such layers return $K_{\pi^{(t)}}$, which induces W at $t = \tau(G)$. Over $\{A\}$ the layer with no filter and a ϕ_t sending distinct pairs $(H^{(t)}_v, \{\{H^{(t)}_w : w \in N(v)\}\})$ to distinct standard basis vectors returns the indicators of one round of color refinement, the multiset of the neighbors' features carrying the multiset of their classes; there the spanning hypothesis is unnecessary. Over $\mathcal{D}$ take a layer whose filter set F_t is a finite set spanning $\mathcal{D}$ linearly, with ϕ_t sending distinct arguments to distinct standard basis vectors. When the columns of $H^{(t)}$ span $\mathbb{K}^{\sigma_t}$, the columns of $MH^{(t)}$ over $M \in F_t$ span $\mathcal{D}\mathbb{K}^{\sigma_t}$, so the argument partition refines $\rho_{\mathcal{D}}(\sigma_t)$, while $\rho_{\mathcal{D}}(\sigma_t)$ refines it by Proposition 1; the layer returns $K_{\rho_{\mathcal{D}}(\sigma_t)}$, and $\tau_{\mathcal{D}}(G)$ such layers reach W.

Every network over $\{A\}$ is a network over $\mathcal{A}$, and every network over $\mathcal{A}$ is a network over $\mathcal{D}$, since $\{A\} \subseteq \mathcal{A} \subseteq \mathcal{D}$. The three least depths are therefore ordered, $\tau_{\mathcal{D}}(G) \leq \tau(G) \leq$ the color-refinement iteration number, for every graph. □

Features that induce $\pi^{(0)}$ without spanning $\mathbb{K}^{\pi^{(0)}}$ cost at most one further layer, an encoding layer with $F_0 = \varnothing$ and ϕ_0 one-hot being legal in Definition 1. The extra layer is sometimes necessary: on P_5 with $H^{(0)} = 0$ walk refinement reaches W in one stage (Proposition A.2) while one layer sees only degrees, since every argument of ϕ_0 other than the size of the neighbor multiset is zero. This is why Theorem 1 asks for nonzero constant features.

B.5 Global Readouts

Proposition B.3. *Let a layer of Definition 1 be allowed to read, in addition to its arguments, the multiset $\{\{H_w^{(t)} : w \in V(G)\}\}$ of all current features, as a global sum or mean readout, a virtual node, or global self-attention without structural or positional encodings does. A layer whose update reads only $H_v^{(t)}$ and this multiset cannot split a class of σ_t. Combined with message passing, every such layer produces a partition coarser than or equal to one round of color refinement from σ_t, and the finest partition attainable is exactly that round.*

Proof. The global multiset is the same argument at every vertex, so $H_u^{(t)} = H_v^{(t)}$ implies $H_u^{(t+1)} = H_v^{(t+1)}$, and no class of σ_t splits. With the neighbor multiset present as well, the update at v is a function of the pair $(H_v^{(t)}, \{\{H_w^{(t)} : w \in N(v)\}\})$ and of an argument common to all vertices, so σ_{t+1} is coarser than the row-equality partition of that pair, which is one round of color refinement from σ_t, the pair being determined by the class of v and the counts $|N(v) \cap C|$ over the classes C of σ_t. Ignoring the global argument and encoding the pair injectively attains that round. Global attention $\sum_{w \in V(G)} \alpha(H_v^{(t)}, H_w^{(t)})\, H_w^{(t)} \Theta$ with no encodings, α and Θ its weights, a virtual node, and the sum and mean readouts are all functions of $H_v^{(t)}$ and the global multiset; a global sum readout is the filter $J = \mathbf{1}\mathbf{1}^{\mathsf{T}}$, and $J\mathbb{K}^{\pi} \subseteq \langle \mathbf{1} \rangle$. □

The bound is on what a layer can separate, not an equality for every update: a combined layer may discard what its message-passing branch distinguishes. Graph transformers that reach beyond message passing do so through the encodings, which is where Appendix B.1 places them (Rosenbluth et al., 2024).

B.6 The Linear/Equality Model

The classical procedures behind Definition 2 differ in what they propagate and agree in how they are organized. A round of color refinement forms the columns K_π and AK_π and recolors by row equality. Morgan's iteration propagates $A^k\mathbf{1}$ and recolors by the values obtained (Morgan, 1965). The partition Chan and Godsil attach to a subset S compares the entries of $A^l\mathbf{1}_S$ for all $l \geq 0$ at once (Chan & Godsil, 1997). The attenuated profile $(I - zA)^{-1}K_\pi$ of Katz (Katz, 1953) is, by Cayley–Hamilton, a polynomial in A applied to K_π. Power iterated color refinement is organized the same way and propagates with $D^{-1}A$, where D is the diagonal matrix of degrees (Kersting et al., 2014). In each case a linear computation is followed by a recoloring that reads only equalities. The model below is that pattern over the adjacency matrix, with no restriction on the linear step; it is the special case of Definition 1 over $\mathcal{A}$ in which every ϕ_t is an equality test.

Definition B.1. A *state* is a partition π of $V(G)$ together with a finite set X of vectors in $\mathbb{Q}^{V(G)}$, its *features*; the initial state consists of the uniform coloring and the single feature $\mathbf{1}$. One *iteration* carries a state to the next in two steps. A *linear stage* forms finitely many new features from the columns of K_π and the members of X by addition, scaling, and application of A, the scalars being arbitrary and allowed to depend on G. An *equality gate* then takes X' to be X together with the new features, and sets $\pi' := \pi_{X'}$.

After a gate every feature is constant on the classes of the new coloring, so $X' \subseteq \mathbb{K}^{\pi'}$ and a state carries nothing beyond its coloring into the next iteration. The features available to a linear stage at π therefore span $\mathbb{K}^{\pi}$, and the features it can form are exactly the elements of

$$\mathcal{C}_A(\pi) \;:=\; \sum_{l \geq 0} A^l\, \mathbb{K}^{\pi} \;=\; \sum_{l=0}^{n-1} A^l\, \mathbb{K}^{\pi},$$

the second equality by Cayley–Hamilton. Call a stage *saturated* when its features span $\mathcal{C}_A(\pi)$. A computation in the model is a network over $\mathcal{A}$: its linear stage applies finitely many polynomials in A to features spanning $\mathbb{K}^{\pi}$, and its gate is the ϕ_t that one-hot encodes the values at v of the features in X', a function of the filtered rows. Iterating ρ is the best a computation in the model can do.

Proposition B.4. *Let σ_t be the coloring held after t iterations by a computation in the model of Definition B.1. Then $\pi^{(t)}$ refines σ_t, with equality when every stage is saturated. Consequently $\tau(G)$ is the least number of iterations in which a computation in the model reaches W.*

Proof. Induct on t, the case $t = 0$ being the initial state. The features present at σ_t and those formed by the next stage all lie in $\mathcal{C}_A(\sigma_t)$, and the gate compares a subfamily of that space, so σ_{t+1} is coarser than $\pi_{\mathcal{C}_A(\sigma_t)} = \rho(\sigma_t)$. The inductive hypothesis and monotonicity make $\rho(\sigma_t)$ coarser than $\rho(\pi^{(t)}) = \pi^{(t+1)}$. A saturated stage compares a spanning set of $\mathcal{C}_A(\sigma_t)$ and returns $\rho(\sigma_t)$ exactly, so equality is preserved along such a computation.

For the last claim, $\pi^{(t)}$ is strictly coarser than W for $t < \tau(G)$, hence so is σ_t, and no computation has reached W; the saturated computation reaches it at $t = \tau(G)$. □

C The Exact Depth in Full

This appendix proves Proposition 5, realizes its quotient as a connected graph, records the scope of Proposition 6 for odd m, and proves the part of Theorem 1(b) that concerns $\mathcal{D}$ together with the depth bound for colored inputs.

Throughout, B is a reduced quotient of order $m \geq 3$ (Appendix A.1), on which

$$\mathcal{C}_B(\pi) \;=\; \sum_{l \geq 0} B^l \, \mathbb{K}^{\pi}, \qquad \rho(\pi) \;=\; \pi_{\mathcal{C}_B(\pi)},$$

the stable coloring is the discrete partition of $[m]$, and the chain $\pi^{(t)}$ and the iteration number τ are the graph's. We use the notation of Section 3.2; note that $a_q = 0$ when m is even and $a_q = e_{m-2} + e_{m-1} - 2e_m$ when m is odd.

C.1 The Paired Tail

Proof of Proposition 5. Put $E_j = \langle a_j, b_j \rangle$ for $j < q$ and $E_q = \langle a_q \rangle$. The proof rests on these being pairwise orthogonal B-invariant subspaces spanning $\mathbb{Q}^m$, after which the closures can be read off directly.

Layers. Each b_j is supported on $\{2j+1, 2j+2\} \subseteq R_j$ with coordinate sum zero, and for $j \geq 1$ the vector a_j is supported on $\{2j-1, 2j\} \cup R_j = R_{j-1}$, also with coordinate sum zero since $|R_j| = m - 2j$. Now fix $j < q$. Each vector spanning $E_{j+1}, \ldots, E_q$ is supported in R_j with coordinate sum zero, while a_j is constant on R_j, equal to -2 there for $j \geq 1$ and to 1 for $j = 0$; so a_j is orthogonal to all of them, and to b_j for the same reason. Among those same vectors, a_{j+1} is constant on the support of b_j and the rest vanish on it, so b_j is orthogonal to all of them as well. The E_j are therefore pairwise orthogonal, and they are spanned by $2q$ nonzero vectors for even m and $2q+1$ for odd, which is m in both cases; so they decompose $\mathbb{Q}^m$. For invariance, note that

$$Bx \;=\; m\,\mathbf{1}\langle \mathbf{1}, x\rangle \;+\; \sum_{j<q} \big(a_j \langle b_j, x\rangle + b_j \langle a_j, x\rangle\big),$$

and for $x \in E_k$ every inner product on the right vanishes except those with a_k and b_k; since $\mathbf{1} = a_0$, the first term survives only for $k = 0$. Hence $BE_k \subseteq E_k$ for every k, with $Ba_q = 0$ in the odd case. Taking $x = b_j$, where $\langle \mathbf{1}, b_j\rangle = \langle a_j, b_j\rangle = 0$ and $\langle b_j, b_j\rangle = 2$, gives the identity

$$Bb_j \;=\; 2a_j \qquad (0 \leq j \leq q-1),$$

which is what makes the chain move.

Entries. The supports of $b_0, \ldots, b_{q-1}$ are the disjoint pairs $\{1,2\}, \{3,4\}, \ldots$, so at most one of them is nonzero at a given index; write $j(x)$ for the index of that one when it exists. Reading off Equation (1),

$$B_{rs} \;=\; m \;+\; (a_{j(s)})_r \, (b_{j(s)})_s \;+\; (b_{j(r)})_r \, (a_{j(r)})_s,$$

a correction being omitted when its j is undefined. For $j \geq 1$ the coordinates of a_j are $m - 2j$ on $\{2j-1, 2j\}$, -2 on R_j and 0 elsewhere, and $a_0 = \mathbf{1}$; so $|(a_j)_x|$ exceeds 2 only when $x \in \{2j-1, 2j\}$, that is only when $j(x) = j - 1$. The two corrections cannot both be of that kind, since that would require $j(r) = j(s) - 1$ and $j(s) = j(r) - 1$ at once. One correction is therefore at most $m - 2$ and the other at most 2, giving $|B_{rs} - m| \leq m$; for $m = 3$ the sum in Equation (1) has a single term and both corrections are at most 1. So the entries of B lie in $\{0, \ldots, 2m\}$. Its first row is sharper:

$j(1) = 0$, so the second correction is $(b_0)_1(a_0)_s = 1$, while the first is -1 for $s = 2$, is $\pm(m-2)$ for $s \in \{3, 4\}$, and vanishes otherwise. Hence

$$B_{12} = m, \qquad B_{1s} \geq 3 \quad (s \neq 1).$$

Closures. Fix $j < q$ and put $V_j = E_0 \oplus \cdots \oplus E_j$, of dimension $2j+2 = |P_j|+1$. The vectors spanning $V_j^{\perp}$, those of $E_{j+1}, \ldots, E_q$, are supported in R_j with coordinate sum zero, hence orthogonal to $e_1, \ldots, e_{2j}$ and to $\mathbf{1}_{R_j}$, which span $\mathbb{K}^{P_j}$. So $\mathbb{K}^{P_j} \subseteq V_j$, and since V_j is B-invariant, $\mathcal{C}_B(P_j) \subseteq V_j$. On the other hand $\mathbb{K}^{P_j}$ is not B-invariant: as B is symmetric it suffices that $(\mathbb{K}^{P_j})^{\perp}$ is not, and this space contains b_j while $Bb_j = 2a_j$ does not. Indeed its members vanish at $1, \ldots, 2j$, and for $j \geq 1$ the vector a_j has the coordinate $m - 2j \neq 0$ at $2j-1$, while $a_0 = \mathbf{1}$ has $\langle \mathbf{1}, \mathbf{1} \rangle = m \neq 0$. So $\mathcal{C}_B(P_j)$ properly contains $\mathbb{K}^{P_j}$, and as $\dim V_j = \dim \mathbb{K}^{P_j} + 1$ it equals V_j. Finally $b_j \in V_j$ is orthogonal to $\mathbb{K}^{P_j}$, so $V_j = \mathbb{K}^{P_j} \oplus \langle b_j \rangle$.

The chain. Walk refinement merges $s \neq t$ at P_j exactly when every vector of $\mathcal{C}_B(P_j)$ takes equal values at s and t, that is when $e_s - e_t$ is orthogonal to $\mathbb{K}^{P_j}$ and to b_j. Orthogonality to the singleton indicators $e_1, \ldots, e_{2j}$ forces $s, t \in R_j$, and orthogonality to $\mathbf{1}_{R_j}$ is then automatic. Since b_j takes the value 1 at $2j+1$, the value -1 at $2j+2$ and 0 elsewhere, $\langle e_s - e_t, b_j \rangle = 0$ forces s and t to receive the same value, hence both to lie outside $\{2j+1, 2j+2\}$. So the merged pairs are exactly those inside R_{j+1}:

$$\rho(P_j) \;=\; P_{j+1} \qquad (0 \leq j \leq q-1).$$

The chain from the indiscrete partition is therefore $P_0, \ldots, P_q$, and as P_q is discrete while P_{q-1} is not, $\tau = q$. Every iterate is coarser than the stable coloring (Appendix A.2), so at $t = q$ the stable coloring refines the discrete partition and B is reduced. □

The chain has cell counts $1, 3, \ldots, 2q-1, m$: every stage before the last contributes one new dimension and gains two classes, and the last gains one class for even m and two for odd m. The paired tail thus realizes the normal form of Proposition 4, with the odd slack unit spent at the final stage.

C.2 Realization

A quotient becomes a graph by blowing each class up into N copies arranged in a cycle, which preserves the quotient matrix and leaves the classes unsplit.

Proposition C.1. *Let B be a symmetric reduced quotient of order m, and let N be an even integer with $N \geq B_{ij}$ for $i \neq j$ and $N > B_{ii}$ for every i. Then B is the quotient matrix of a simple graph on mN vertices whose stable coloring has exactly m classes. If some row of B is positive off the diagonal and has an entry at least 2 there, that graph is connected.*

Proof. Take $[m] \times \mathbb{Z}_N$ as vertex set with classes $C_i = \{i\} \times \mathbb{Z}_N$. For $i < j$ join (i, u) to $(j, u), (j, u+1), \ldots, (j, u+B_{ij}-1)$; inside C_i take the circulant with connection set $\{\pm 1, \ldots, \pm\lfloor B_{ii}/2 \rfloor\}$, together with $N/2$ when B_{ii} is odd. The bounds on N make the listed neighbors distinct and keep 0 out of the connection sets, so the graph is simple. A vertex (i, u) has B_{ij} neighbors in C_j for $i < j$ by construction, and (j, v) has the $B_{ij} = B_{ji}$ neighbors $(i, v), \ldots, (i, v - B_{ij} + 1)$ in C_i; inside its own class it has $2\lfloor B_{ii}/2 \rfloor$ neighbors, and one more when B_{ii} is odd. The class partition is therefore equitable with quotient matrix B.

The rotation $(i, u) \mapsto (i, u+1)$ is an automorphism transitive on each class, so refinement never splits a class; and since B is reduced, refinement separates them (Appendix A.1). The stable coloring is thus the class partition.

For connectivity, let row i be positive off the diagonal with $B_{ij} \geq 2$. Then (i, u) is joined to (k, u) for every $k \neq i$, so each level $[m] \times \{u\}$ is connected, and the entry $B_{ij} \geq 2$ joins (i, u) to a vertex of C_j one level away, linking consecutive levels. □

The paired tail has entries at most $2m$, so $N = 2m + 2$ is admissible, and its first row is positive off the diagonal with $B_{12} = m \geq 3$. The resulting graph has the same chain and iteration number as B by Appendix A.1. Together with Proposition 4 this settles the exact value: for every $m \geq 3$ the

maximum of τ over graphs whose stable coloring has m classes equals $\lfloor m/2 \rfloor$, and it is attained by connected simple graphs on exactly $2m^2 + 2m$ vertices. By Theorem 1 the same number is the least depth of a network over $\mathcal{A}$ on those graphs.

Remark C.1. The theorem is stated in m, and the witnesses just built have $n = 2m^2 + 2m$ vertices, so on them τ is of order $\sqrt{n}$. The size comes from the construction: a quotient with entries bounded by a constant realizes on $O(m)$ vertices, and extremal quotients with bounded entries exist. At $m = 6$, $m = 8$ and $m = 10$ the matrices

$$\begin{pmatrix} 0 & 2 & 0 & 1 & 1 & 0 \\ 2 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 & 0 & 1 \\ 1 & 1 & 2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 & 0 & 2 & 0 & 1 \\ 0 & 0 & 1 & 0 & 2 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 4 & 0 & 2 & 0 & 2 & 6 & 4 & 2 \\ 0 & 0 & 0 & 4 & 2 & 0 & 6 & 2 & 4 & 2 \\ 4 & 0 & 2 & 1 & 0 & 4 & 2 & 5 & 0 & 4 \\ 0 & 4 & 1 & 0 & 2 & 4 & 5 & 0 & 2 & 4 \\ 2 & 2 & 0 & 2 & 3 & 2 & 0 & 2 & 3 & 6 \\ 0 & 0 & 4 & 4 & 2 & 5 & 1 & 1 & 3 & 2 \\ 2 & 6 & 2 & 5 & 0 & 1 & 3 & 2 & 1 & 2 \\ 6 & 2 & 5 & 0 & 2 & 1 & 2 & 1 & 3 & 2 \\ 4 & 4 & 0 & 2 & 3 & 3 & 1 & 3 & 4 & 0 \\ 2 & 2 & 4 & 4 & 6 & 2 & 2 & 2 & 0 & 0 \end{pmatrix}$$

are reduced quotients whose chains have cell counts $1, 3, 5, 6$, then $1, 3, 5, 7, 8$, then $1, 3, 5, 7, 9, 10$, so that $\tau = 3, 4, 5$; Proposition C.1 realizes them with $N = 4, 4, 6$ on 24, 32 and 60 vertices against the 84, 144 and 220 of the paired tail, where τ is $n/8$, $n/8$ and $n/12$. Whether such quotients exist for every even m is open.

Remark C.2. Nothing in Proposition 4 used walks. Let $\mathcal{M}$ be any algebra of $m \times m$ matrices containing B and the identity, and refine by $\pi \mapsto \pi_{\mathcal{M}\mathbb{K}^\pi}$. This closure contains its argument and is unchanged when repeated, so Proposition 3 applies verbatim; and a fixed point has $\mathcal{M}\mathbb{K}^\pi = \mathbb{K}^\pi$, hence $B\mathbb{K}^\pi \subseteq \mathbb{K}^\pi$, hence π equitable and so discrete. The bound $\lfloor m/2 \rfloor$ on the number of steps follows exactly as before, for instance for the algebra generated by B and the diagonal matrix of its row sums, the adjacency and degree pair studied through the universal adjacency matrices (Haemers & Omidi, 2011; Farrugia & Sciriha, 2015); Proposition C.2 below sharpens the bound for that algebra to $\lfloor (m-1)/2 \rfloor$.

C.3 Color Refinement on Extremal Graphs with Odd m

Proposition 6 identifies the color-refinement chain with the walk-refinement chain whenever every stage adds one dimension, which the normal form of Proposition 4 guarantees for even m at equality and Proposition 5 guarantees on the paired tail for every m. For odd m the normal form allows the slack unit to sit in some $r_t = 2$, where a stage adds two dimensions and one round of color refinement may fall short of it, and this happens: at $m = 3$ the path P_5 has $\tau = 1 = \lfloor 3/2 \rfloor$ by Proposition A.2, while color refinement needs 2 rounds by Proposition A.1. The same happens at $m = 5$: the symmetric reduced quotient of order five with rows 01104, 12210, 12012, 01122, 40220 (Appendix C.4) has $\tau = 2 = \lfloor 5/2 \rfloor$ with $r_0 = 2$, while color refinement needs 3 rounds, with cell counts $1, 2, 4, 5$. This is why Proposition 6 is stated for even m and for the paired tail rather than for every extremal graph.

C.4 The Degree Algebra

In this subsection $d = B\mathbf{1}$ is the vector of row sums of B, $D = \mathrm{diag}(d)$, and $\mathcal{D}$ is the algebra generated by I, B and D; by Appendix A.1 the chain of $\rho_{\mathcal{D}}$ on G is the lift of the chain of $\rho_{\mathcal{D}}(\pi) = \pi_{\mathcal{D}\mathbb{K}^\pi}$ on its quotient, so $\tau_{\mathcal{D}}(G)$ is computed here. Since $Dx = d \odot x$, the algebra $\mathcal{D}$ contains multiplication by every polynomial in d, and the $\mathcal{D}$-closure of $\langle \mathbf{1} \rangle$ contains the Hadamard powers $d, d \odot d, \ldots$, which span the indicators of the classes of π_d by a Vandermonde argument. Normalized filters therefore name the degree classes for free, and Lemma 1 reads this as the multiplicative step supplied by the degree matrix. That is the whole difference between the two algebras, and it is worth at most one stage in the worst case, and exactly one for even m.

Proposition C.2. *Let B be a reduced quotient of order $m \geq 3$. Then $\tau_{\mathcal{D}} \leq \lfloor (m-1)/2 \rfloor$. On the paired tail Equation (1), $\tau_{\mathcal{D}} = \lfloor m/2 \rfloor - 1$ for $m \geq 4$ and $\tau_{\mathcal{D}} = 1$ for $m = 3$.*

Proof. The chain. The fixed points of $\rho_{\mathcal{D}}$ are the equitable partitions (Appendix A.2), on a reduced quotient only the discrete one, so the chain from the indiscrete partition moves at every stage until it reaches the discrete partition, which it does after finitely many stages since it refines at every move.

The ceiling. The closure $U \mapsto \mathcal{D}U$ contains its argument and is idempotent, so Proposition 3 applies: with $S = \mathcal{D}\mathbb{K}^{\pi}$ a stage gains $|\rho_{\mathcal{D}}(\pi)| - |\pi| = \delta(S) + (\dim S - |\pi|)$ classes, the second summand at least 1 at every stage and the first at least 1 at every stage but the last. So every stage before the last gains at least two classes. The first stage adds at least two dimensions: let $U = \mathcal{D}\langle \mathbf{1} \rangle$, which contains $\mathbf{1}$, d and $d \odot d = Dd$. If $\dim U = 1$ the indiscrete partition is equitable and $m = 1$. If $\dim U = 2$ then d is not constant, for otherwise D would be a scalar and $U = \mathcal{C}_B(\mathbf{1}) = \langle \mathbf{1} \rangle$; and $d \odot d \in \langle \mathbf{1}, d \rangle$ makes d satisfy a quadratic equation coordinatewise, so d takes exactly two values, $U = \langle \mathbf{1}, d \rangle = \mathbb{K}^{\pi_d}$ with $|\pi_d| = 2$, and U being B-invariant makes π_d equitable and $m = 2$. Hence $\dim U \geq 3$, and a first stage that is not the last gains $\delta(U) + (\dim U - 1) \geq 3$ classes. When $\tau_{\mathcal{D}} \geq 2$, summing the gains over the chain from one class to m gives $m - 1 \geq 3 + 2(\tau_{\mathcal{D}} - 2) + 1 = 2\tau_{\mathcal{D}}$, that is $\tau_{\mathcal{D}} \leq (m-1)/2$; when $\tau_{\mathcal{D}} = 1$ the bound holds because $m \geq 3$.

The paired tail. In the notation of Equation (1), $\langle b_j, \mathbf{1} \rangle = 0$ for all j and $\langle a_j, \mathbf{1} \rangle = 0$ for $j \geq 1$ while $a_0 = \mathbf{1}$, so

$$d \;=\; B\mathbf{1} \;=\; m^2\,\mathbf{1} + m\,b_0,$$

taking the value $m^2 + m$ at 1, $m^2 - m$ at 2 and m^2 on R_1. For $m = 3$ these are three distinct values on three indices, so $U = \mathcal{D}\langle \mathbf{1} \rangle$ contains $\mathbb{K}^{P_1} = \mathbb{Q}^3$ and $\tau_{\mathcal{D}} = 1$. For $m \geq 4$ the Hadamard powers of d span $\mathbb{K}^{P_1}$, so $U \supseteq \mathbb{K}^{P_1}$, and U being B-invariant, $U \supseteq \mathcal{C}_B(P_1) = \mathbb{K}^{P_1} \oplus \langle b_1 \rangle = V_1$ by Proposition 5. The space V_1 is B-invariant by Proposition 5 and D-invariant because d is constant on the classes of P_1 and $Db_1 = m^2 b_1$, the support of b_1 lying in R_1; it contains $\mathbf{1}$, so $U = V_1$ and the first stage lands on $\pi_{V_1} = \rho(P_1) = P_2$. For $j \geq 2$ the same two facts, d constant on the classes of P_j and $Db_j = m^2 b_j$, make $\mathcal{C}_B(P_j) = \mathbb{K}^{P_j} \oplus \langle b_j \rangle$ invariant under D, so $\mathcal{D}\mathbb{K}^{P_j} = \mathcal{C}_B(P_j)$ and the stage is the ordinary one, $\rho_{\mathcal{D}}(P_j) = P_{j+1}$. The chain is $P_0, P_2, P_3, \ldots, P_q$, of length $q - 1$. □

The first $\mathcal{D}$-stage thus adds at least two dimensions, and gains at least three classes whenever it is not the last stage; at $m = 3$ it is always the last, and gains two. On the paired tail with $m \geq 4$ the degree vector $d = m^2\mathbf{1} + mb_0$ names the first pair and nothing after it, which is why the saving there is exactly one stage, whether the degrees arrive inside the filters or as input features, the degree partition being P_1. Degrees lift to the realization of Proposition C.1 by Appendix A.1, so the realized paired tail attains $\tau_{\mathcal{D}} = \lfloor m/2 \rfloor - 1$ for every $m \geq 4$. With the ceiling this is the $\mathcal{D}$-part of Theorem 1(b): over graphs whose stable coloring has $m \geq 3$ classes, $\lfloor m/2 \rfloor - 1 \leq \max \tau_{\mathcal{D}} \leq \lfloor (m-1)/2 \rfloor$, the two ends coinciding for even m; at $m = 3$ the maximum is 1, attained by the paired tail and by P_5, and the lower end 0 is not attained.

The statement is about maxima and must not be made per graph. For odd m the two maxima can agree: the symmetric reduced quotient of order five with rows 01104, 12210, 12012, 01122, 40220 has the chain with cell counts $1, 4, 5$ under both algebras, so at $m = 5$ both maxima equal 2. On a single graph degrees can be worth more than one stage: the order-six quotient of Remark C.1 has $\tau = 3$ and finishes in one $\mathcal{D}$-stage, as do the order-eight and order-ten ones. Degrees are no substitute for A: on a graph without isolated vertices the algebra generated by $D^{-1}A$ alone closes $\langle \mathbf{1} \rangle$ at once, since $D^{-1}A\mathbf{1} = \mathbf{1}$; with isolated vertices and the zero convention, $D^{-1}A\mathbf{1}$ is the indicator of the non-isolated vertices and does separate.

C.5 Colored Inputs

The proof of Theorem 1(a) in Appendix B used only that the initial features span $\mathbb{K}^{\pi^{(0)}}$, and Appendix A.2 holds with W the coarsest equitable partition refining $\pi^{(0)}$; so for features spanning the space of any initial coloring, the least depth over $\mathcal{A}$ is the least t with $\rho^t(\pi^{(0)}) = W$.

Proposition C.3. *Let the initial features span $\mathbb{K}^{\pi^{(0)}}$ and induce c classes, and let the coarsest equitable partition W refining $\pi^{(0)}$ have $m \geq c + 2$ classes. Then the least depth of a network over $\mathcal{A}$ whose features induce W is at most $\lfloor (m - c + 1)/2 \rfloor$, and colored graphs attain it.*

Proof. The chain $\rho^t(\pi^{(0)})$ runs from c classes to m and moves at every stage before reaching W, since its fixed points are equitable and every iterate is coarser than W (Appendix A.2). By Proposition 3 with $\mathrm{cl} = \mathcal{C}_A$, each stage gains at least one class and each stage before the last at least two, so $m - c \geq 2(\tau - 1) + 1$ for the number τ of stages, that is $\tau \leq (m - c + 1)/2$. For attainment take a graph realizing the paired tail of order $m - c + 1 \geq 3$ (Proposition C.1), together with $c - 1$ isolated vertices each carrying its own color, and give the vertices of the realized tail a common color; the

initial coloring has c classes, W has m, the singleton classes never move, and the chain on the tail is that of Proposition 5, of length $\lfloor (m-c+1)/2 \rfloor$. □

The spanning hypothesis is needed: without it the bound fails, since P_5 with zero initial features has $c = 1$ and $m = 3$, so the bound reads 1, while a first layer sees only degrees and the least depth is 2 (Appendix B).

D Multiplicative Depth and the Leaf Budget

This appendix proves Lemma 1 and Theorem 2, and prices the multiplicative depth of Definition 4 against a static budget, the least number of leaves at which homomorphism counts of rooted trees already determine the stable coloring.

D.1 The Leaf Budget

For $x, y \in \mathbb{Q}^{V(G)}$, $x \odot y$ is the Hadamard product, $(x \odot y)_v = x_v y_v$, and for subspaces U, U' the space $U \odot U'$ is the span of the products $x \odot y$ with $x \in U$ and $y \in U'$; $\mathcal{C}_A$ extends to subspaces by $\mathcal{C}_A(U) = \sum_{l \geq 0} A^l U$, so that $\mathcal{C}_A(\pi) = \mathcal{C}_A(\mathbb{K}^\pi)$, as in Definition 4.

Definition D.1. For a rooted tree T let $h_T \in \mathbb{Q}^{V(G)}$ count at v the homomorphisms from T to G that send the root to v, and let $\mathcal{T}_k$ be the span of the h_T over rooted trees with at most k leaves. The *leaf budget* $\lambda(G)$ is the least k with $\pi_{\mathcal{T}_k} = W$.

Both chains increase: $\mathbf{1}$ lies in S_0 and hence in every S_j, so $S_j \subseteq S_j \odot S_j \subseteq S_{j+1}$, while $\mathcal{T}_k \subseteq \mathcal{T}_{k+1}$ by definition. Homomorphism counts from trees characterize color-refinement equivalence (Dvořák, 2010; Dell et al., 2018); the leaf budget asks how much of that family is needed.

Three rules generate the vectors h_T. A single vertex gives $\mathbf{1}$. Attaching a new root above the root of T gives Ah_T: a homomorphism of the enlarged tree that sends the new root to v is a homomorphism of T sending its root to a neighbor of v. Identifying the roots of T' and T'' gives $h_{T'} \odot h_{T''}$. Count a single vertex as one leaf; the number of leaves of any other rooted tree is then the sum over the subtrees at its root. Attaching a root preserves the count, and identifying roots gives at most the sum of the two, with equality unless one of the trees is a single vertex. Consequently $\mathcal{T}_1 = S_0$, the rooted trees with one leaf being the paths rooted at an end; each $\mathcal{T}_k$ is A-invariant; $\mathcal{T}_a \odot \mathcal{T}_b \subseteq \mathcal{T}_{a+b}$; and

$$\mathcal{T}_k \;=\; \mathcal{C}_A\Big(\sum_{a+b=k} \mathcal{T}_a \odot \mathcal{T}_b \Big) \qquad (k \geq 2), \tag{D.1}$$

the sum running over $a, b \geq 1$.

The last two facts put the right side inside $\mathcal{T}_k$; for the reverse inclusion, let T have at most k leaves and induct on the number of its vertices. If T is a single vertex then $h_T = \mathbf{1} \in \mathcal{T}_1 \odot \mathcal{T}_{k-1}$. If its root carries one subtree T' then $h_T = Ah_{T'}$ with $h_{T'}$ on the right side of Equation (D.1) by induction, and that side is A-invariant. If its root carries at least two subtrees, split them into two nonempty groups and write T as an identification of two rooted trees with a and b leaves, $a + b \leq k$; then $h_T \in \mathcal{T}_a \odot \mathcal{T}_b$, which is a summand when $a + b = k$ and is contained in $\mathcal{T}_{k-1} \subseteq \mathcal{T}_1 \odot \mathcal{T}_{k-1}$ otherwise.

Both parameters are computed on the quotient. Let Φ be the characteristic matrix of W, and define h_T, S_j and $\mathcal{T}_k$ on $\mathbb{Q}^m$ by the same rules with B in place of A, so that Equation (D.1) holds there as well. Then

$$\Phi\mathbf{1} \;=\; \mathbf{1}, \qquad A\Phi \;=\; \Phi B, \qquad \Phi(x \odot y) \;=\; \Phi x \odot \Phi y,$$

the last because Φ repeats each coordinate of its argument along a class. So Φ commutes with both closures, and the objects of Definitions 4 and D.1 on G are the images under Φ of their counterparts on B. Two vertices agree on a lifted family exactly when their classes agree on the family, so a lifted family induces W exactly when the family induces the discrete partition. Both $d_\odot$ and λ are therefore unchanged on passing to the quotient, and for the rest of the appendix B is a reduced quotient of order $m \geq 3$, in the notation of Appendix C.

D.2 Bounds on the Depth

Proof of Lemma 1. Let S be a subspace of $\mathbb{Q}^m$ containing $\mathbf{1}$ and closed under $\odot$, and $\sigma = \pi_S$, so that $S \subseteq \mathbb{K}^\sigma$; the reverse inclusion comes from writing each indicator of σ as a product. Fix a class C of σ and an index $u \in C$. For each $s \notin C$ choose $x^{(s)} \in S$ with $x^{(s)}_u \neq x^{(s)}_s$, and form

$$y \;=\; \bigodot_{s \notin C} \frac{x^{(s)} - x^{(s)}_s\,\mathbf{1}}{x^{(s)}_u - x^{(s)}_s},$$

an empty product being $\mathbf{1}$. Every factor lies in S, hence so does y, and $y_u = 1$ while $y_s = 0$ for every $s \notin C$. Being an element of $\mathbb{K}^\sigma$, the vector y is constant on C, so $y = \mathbf{1}_C$. Thus S contains the indicators of σ and equals $\mathbb{K}^\sigma$.

If S is moreover B-invariant then so is $\mathbb{K}^\sigma$, making σ equitable and hence discrete, and $S = \mathbb{Q}^m$. □

Since the ladder increases, it stabilizes at a subspace containing $\mathbf{1}$, invariant under B, and closed under $\odot$, hence at $\mathbb{Q}^m$; so $d_\odot$ is finite. The same argument runs on G: a subspace of $\mathbb{Q}^{V(G)}$ containing $\mathbf{1}$ and closed under $\odot$ is $\mathbb{K}^\sigma$ for a partition σ of $V(G)$, and A-invariance makes σ equitable, hence a refinement of W. So $\mathbb{K}^W$ is the smallest subspace of $\mathbb{Q}^{V(G)}$ containing $\mathbf{1}$, invariant under A, and closed under $\odot$.

A long walk-refinement chain forces a long ladder, and the ladder is bounded in m.

Lemma D.1. *Let B be a reduced quotient of order $m \geq 3$. Then $\tau - 1 \leq d_\odot \leq m - 3$.*

Proof. For the lower bound we show $S_j \subseteq \mathbb{K}^{\pi^{(j+1)}}$ by induction on j. At $j = 0$ this is the definition of $\pi^{(1)} = \pi_{S_0}$. If it holds at j then $S_j \odot S_j \subseteq \mathbb{K}^{\pi^{(j+1)}}$, that space being closed under $\odot$, and therefore

$$S_{j+1} \;=\; \mathcal{C}_B(S_j \odot S_j) \;\subseteq\; \mathcal{C}_B(\pi^{(j+1)}) \;\subseteq\; \mathbb{K}^{\pi^{(j+2)}}.$$

At $j = d_\odot$ the partition π_{S_j} is discrete and $\pi^{(j+1)}$ refines it, so $\pi^{(j+1)}$ is discrete and $\tau \leq d_\odot + 1$.

For the upper bound, suppose $d_\odot \geq m - 2$, so that π_{S_j} is not discrete for any $j \leq m - 3$. If $S_{j+1} = S_j$ then S_j contains $\mathbf{1}$, is B-invariant and is closed under $\odot$, so $S_j = \mathbb{Q}^m$ by Lemma 1 and π_{S_j} is discrete; along $S_0, \ldots, S_{m-3}$ the dimensions therefore increase strictly. They start at $\dim S_0 \geq 2$, since otherwise $S_0 = \langle \mathbf{1} \rangle$ would be B-invariant and would make the indiscrete partition equitable, hence discrete, against $m \geq 3$. Hence $\dim S_{m-3} \geq m - 1$. But a B-invariant subspace S with $\dim S \geq m - 1$ has π_S discrete: otherwise $|\pi_S| \leq m - 1 \leq \dim S \leq |\pi_S|$, the last step by $S \subseteq \mathbb{K}^{\pi_S}$, forcing $S = \mathbb{K}^{\pi_S}$ and making π_S equitable without being discrete. So $\pi_{S_{m-3}}$ is discrete, contrary to assumption. □

The two filtrations bracket each other, one level of the ladder against a doubling of the budget.

Lemma D.2. $\log_2 \lambda \leq d_\odot \leq \lambda - 1$.

Proof. Two containments between the filtrations give the two bounds. First $S_j \subseteq \mathcal{T}_{2^j}$, by induction on j: the case $j = 0$ is $S_0 = \mathcal{T}_1$, and

$$S_{j+1} \;=\; \mathcal{C}_B(S_j \odot S_j) \;\subseteq\; \mathcal{C}_B\big(\mathcal{T}_{2^j} \odot \mathcal{T}_{2^j}\big) \;\subseteq\; \mathcal{C}_B\big(\mathcal{T}_{2^{j+1}}\big) \;=\; \mathcal{T}_{2^{j+1}}.$$

Second $\mathcal{T}_k \subseteq S_{k-1}$, by induction on k, the case $k = 1$ being $\mathcal{T}_1 = S_0$. Since $S_s \subseteq S_t$ for $s \leq t$, we have $S_s \odot S_t \subseteq S_t \odot S_t \subseteq S_{t+1}$. So if the claim holds below k then every summand of Equation (D.1) satisfies

$$\mathcal{T}_a \odot \mathcal{T}_b \;\subseteq\; S_{a-1} \odot S_{b-1} \;\subseteq\; S_{\max(a,b)} \;\subseteq\; S_{k-1},$$

because $\max(a, b) \leq k - 1$ when $a + b = k$ with $a, b \geq 1$; and S_{k-1} is B-invariant, so $\mathcal{T}_k \subseteq S_{k-1}$.

Now take $j = d_\odot$. The partition $\pi_{\mathcal{T}_{2^j}}$ refines the discrete partition π_{S_j}, so λ is defined and $\lambda \leq 2^{d_\odot}$. Taking $k = \lambda$ instead, $\pi_{S_{k-1}}$ refines the discrete partition $\pi_{\mathcal{T}_k}$, so $d_\odot \leq \lambda - 1$. □

Chaining the two lemmas gives $\tau \leq \lambda$: a lower bound on saturated stages is a lower bound on leaves.

D.3 The Three Measures Coincide

On the paired tail the ladder is the chain of closures that Proposition 5 already computed, and the dimension missing at each level is a square. The leaf filtration tracks it, one doubling to a level, so the budget there is exactly $2^{d_\odot}$.

Theorem D.1. *Let B be the paired tail Equation (1) of order $m \geq 3$ and put $q = \lfloor m/2 \rfloor$. Then*

$$S_j \;=\; \mathcal{C}_B(P_j) \quad (0 \leq j \leq q-1), \qquad \mathcal{T}_k \;=\; S_{\lfloor \log_2 k \rfloor} \quad (1 \leq k \leq 2^{q-1}),$$

and consequently $\tau - 1 = d_\odot = \log_2 \lambda = q - 1$.

Proof. The missing dimension. Fix $0 \leq j \leq q-2$, so that $|R_{j+1}| = m - 2j - 2 \geq 2$ and $|P_{j+1}| = 2j + 3$. By Proposition 5 the space $\mathcal{C}_B(P_j) = \mathbb{K}^{P_j} \oplus \langle b_j \rangle$ has dimension $2j+2$, and it lies inside $\mathbb{K}^{P_{j+1}}$ because P_{j+1} refines P_j and b_j is constant on every class of P_{j+1}. The square $b_j \odot b_j = e_{2j+1} + e_{2j+2}$ lies in $\mathbb{K}^{P_{j+1}}$ as well, and outside $\mathcal{C}_B(P_j)$: choosing $s \in R_{j+1}$, the vector $e_{2j+1} + e_{2j+2} - 2e_s$ is orthogonal to $e_1, \ldots, e_{2j}$, to $\mathbf{1}_{R_j}$ and to b_j, hence to all of $\mathcal{C}_B(P_j)$, while its inner product with $b_j \odot b_j$ is 2. Comparing dimensions gives the square identity Equation (2),

$$\mathbb{K}^{P_{j+1}} \;=\; \mathcal{C}_B(P_j) \oplus \langle b_j \odot b_j \rangle \qquad (0 \leq j \leq q-2).$$

The ladder. For the same j, the space $\mathbb{K}^{P_{j+1}}$ is closed under $\odot$ and contains $\mathcal{C}_B(P_j)$, so it contains $\mathcal{C}_B(P_j) \odot \mathcal{C}_B(P_j)$; conversely that product contains $\mathcal{C}_B(P_j)$, since $\mathbf{1} \in \mathcal{C}_B(P_j)$, and contains $b_j \odot b_j$. With Equation (2) the two are equal:

$$\mathcal{C}_B(P_j) \odot \mathcal{C}_B(P_j) \;=\; \mathbb{K}^{P_{j+1}} \qquad (0 \leq j \leq q-2). \tag{D.2}$$

So $S_j = \mathcal{C}_B(P_j)$ implies $S_{j+1} = \mathcal{C}_B(\mathbb{K}^{P_{j+1}}) = \mathcal{C}_B(P_{j+1})$ for $j \leq q-2$, and the base $S_0 = \mathcal{C}_B(\pi^{(0)}) = \mathcal{C}_B(P_0)$ carries the induction through $j = q-1$. Since $\pi_{S_j} = \rho(P_j) = P_{j+1}$, the ladder first induces the discrete partition at $j = q-1$, so $d_\odot = q-1$.

Products across levels. For $i < j \leq q-1$,

$$S_i \odot S_j \;\subseteq\; S_j. \tag{D.3}$$

Indeed $S_i \subseteq \mathbb{K}^{\rho(P_i)} = \mathbb{K}^{P_{i+1}} \subseteq \mathbb{K}^{P_j}$, and $\mathbb{K}^{P_j} \odot S_j \subseteq S_j$: a vector $w \in \mathbb{K}^{P_j}$ multiplies $\mathbb{K}^{P_j}$ into itself, and it is constant on R_j, so $w \odot b_j$ is a multiple of b_j, since b_j is supported on $\{2j+1, 2j+2\} \subseteq R_j$.

The leaf filtration. We prove $\mathcal{T}_k = S_{\lfloor \log_2 k \rfloor}$ for $1 \leq k \leq 2^{q-1}$ by induction on k, the case $k = 1$ being $\mathcal{T}_1 = S_0$. Let $k \geq 2$ and put $j = \lfloor \log_2 k \rfloor \leq q-1$. Take a summand $\mathcal{T}_a \odot \mathcal{T}_b$ of Equation (D.1) with $a \leq b$, and put $s = \lfloor \log_2 a \rfloor$ and $t = \lfloor \log_2 b \rfloor$, so that $s \leq t$ and, by induction, $\mathcal{T}_a = S_s$ and $\mathcal{T}_b = S_t$. If $s < t$ then $\mathcal{T}_a \odot \mathcal{T}_b \subseteq S_t$ by Equation (D.3), and $2^t \leq b < k < 2^{j+1}$ gives $t \leq j$. If $s = t$ then $2^{s+1} \leq a + b = k < 2^{j+1}$ gives $s + 1 \leq j$, and Equation (D.2) makes $\mathcal{T}_a \odot \mathcal{T}_b = S_s \odot S_s = \mathbb{K}^{P_{s+1}} \subseteq S_{s+1}$. Every summand therefore lies in S_j, which is B-invariant, and $\mathcal{T}_k \subseteq S_j$. For the reverse inclusion, if k is not a power of two then $\lfloor \log_2(k-1) \rfloor = j$ and $\mathcal{T}_k \supseteq \mathcal{T}_1 \odot \mathcal{T}_{k-1} \supseteq \mathcal{T}_{k-1} = S_j$; and if $k = 2^j$ then the summand with $a = b = 2^{j-1}$ is $S_{j-1} \odot S_{j-1} = \mathbb{K}^{P_j}$ by Equation (D.2), whose B-closure is S_j.

The partition $\pi_{\mathcal{T}_k} = P_{\lfloor \log_2 k \rfloor + 1}$ is discrete exactly when $\lfloor \log_2 k \rfloor = q-1$, so $\lambda = 2^{q-1}$; with $\tau = q$ from Proposition 5, the three values agree. □

Realization changes none of this, both parameters being computed on the quotient: the connected simple graphs on $2m^2 + 2m$ vertices built in Appendix C.2 have multiplicative depth $\lfloor m/2 \rfloor - 1$ and leaf budget $2^{\lfloor m/2 \rfloor - 1}$.

At $m = 6$ the ladder is short enough to see. There $q = 3$ and $b_0 = e_1 - e_2$, so $S_0 = \langle \mathbf{1}, b_0 \rangle$ separates 1, 2 and $\{3, 4, 5, 6\}$ while containing no indicator of those three classes. One multiplication supplies what is missing: $b_0 \odot b_0 = e_1 + e_2$, and $S_0 \odot S_0 = \mathbb{K}^{P_1}$. Closing under B produces S_1, whose extra direction over $\mathbb{K}^{P_1}$ is $b_1 = e_3 - e_4$; one further square and closure give $S_2 = \mathbb{Q}^6$. So $d_\odot = 2 = \tau - 1$ and $\lambda = 4 = 2^{\tau-1}$: rooted trees with three leaves reach only S_1.

Together with Lemma D.1, Theorem D.1 brackets both parameters over all graphs.

Corollary D.1. *For every $m \geq 3$, over graphs whose stable coloring has m classes,*

$$\left\lfloor \frac{m}{2} \right\rfloor - 1 \;\leq\; \max d_{\odot} \;\leq\; m-3, \qquad 2^{\lfloor m/2 \rfloor - 1} \;\leq\; \max \lambda \;\leq\; 2^{m-3}.$$

Proof. The lower bounds are Theorem D.1 read on the graphs that realize the paired tail. The upper bounds are Lemma D.1 and, for the leaf budget, $\lambda \leq 2^{d_{\odot}}$ from Lemma D.2. □

So no constant number of alternations between the linear closure and Hadamard products computes stable colorings, unconditionally. The levels themselves are cheap. Exact rational linear algebra lies in logspace-uniform NC^2 (Borodin et al., 1982; Mulmuley, 1987), and passing from one level of the ladder to the next is one round of Hadamard products followed by one closure, so for each fixed j a basis of S_j and the partition it induces are computable there. The neighboring decision problem, color-refinement equivalence of two graphs, is complete for polynomial time (Grohe, 1999), as is the relational coarsest partition problem (Cho & Huynh, 1992). What is sequential here is the alternation between the two closures.

Proof of Theorem 2. The bounds $\tau - 1 \leq d_{\odot} \leq m - 3$ for every graph with $m \geq 3$ are Lemma D.1, transferred from the quotient to the graph by the invariance of Appendix D.1. On the realized paired tail, Theorem D.1 gives $d_{\odot} = \tau - 1 = \lfloor m/2 \rfloor - 1$, and the bracket $\lfloor m/2 \rfloor - 1 \leq \max d_{\odot} \leq m - 3$ follows. □

E The Logical Reading

This appendix proves the logical reading quoted at the end of Section 4: the logic whose quantifier depth is the stage count of Definition 2, so that Theorem 1(b) reads as an exchange rate between the reach of a quantifier and the depth it needs.

The rounds of color refinement are quantifier depths in two-variable counting logic (Immerman & Lander, 1990). A stage of walk refinement compares, for every length l and every current class, the total number of walks of length l from a vertex into that class, so the matching logic is the one whose quantifier counts those walks. Quantifier depth is then the number of stages, and Section 3 reads as an exchange rate: admitting every walk length into the quantifier moves the worst-case depth from $m - 1$ to $\lfloor m/2 \rfloor$.

Definition E.1. Formulas have a single free variable and are built from $\top$ by negation, conjunction, and the *walk quantifier*. For $l \geq 0$ and $k \geq 1$, this quantifier takes a formula φ to $\exists_l^{\geq k} \varphi$, satisfied by a vertex u of G when at least k walks of length l start at u and end at a vertex satisfying φ, that is when

$$\sum_{v \models \varphi} A^l(u, v) \;\geq\; k.$$

The *depth* of a formula is the nesting depth of its walk quantifiers.

At $l = 1$ the walk quantifier counts the neighbors satisfying φ, which in the two-variable notation is $\exists^{\geq k} y\,(E(x, y) \wedge \varphi(y))$ with E the edge relation; at $l = 0$ and $k = 1$ it returns φ itself. Lichter et al. (2019) match their two-dimensional walk refinement to a logic and a pebble game of their own.

Proposition E.1. *For every $t \geq 0$, two vertices of G satisfy the same formulas of depth at most t if and only if $\pi^{(t)}$ merges them.*

Proof. Each direction comes from an induction on t.

Every formula of depth at most t is constant on the classes of $\pi^{(t)}$. At $t = 0$ the formulas are Boolean combinations of $\top$, and $\pi^{(0)}$ is the uniform coloring. Assume the claim at t. A formula of depth at most $t + 1$ is a Boolean combination of $\top$ and of formulas $\exists_l^{\geq k} \varphi$ with φ of depth at most t, so it is enough to treat one of the latter. The vertices satisfying φ form a union of classes of $\pi^{(t)}$, whence

$$\sum_{v \models \varphi} A^l(u, v) \;=\; \sum_{C} \big(A^l \mathbf{1}_C\big)(u),$$

the sum running over the classes C of $\pi^{(t)}$ on which φ holds. Each $A^l\mathbf{1}_C$ lies in $\mathcal{C}_A(\pi^{(t)})$ and is therefore constant on the classes of $\pi^{(t+1)} = \pi_{\mathcal{C}_A(\pi^{(t)})}$, and so is the displayed sum; hence $\exists_l^{\geq k}\varphi$ is constant there too.

Each class of $\pi^{(t)}$ is defined by a formula of depth at most t. At $t = 0$ take $\top$. Assume formulas χ_C of depth at most t defining the classes C of $\pi^{(t)}$. By Cayley–Hamilton the space $\mathcal{C}_A(\pi^{(t)})$ is spanned by the vectors $A^l\mathbf{1}_C$ with $l < n$, so the class of a vertex u in $\pi^{(t+1)}$ is determined by the finitely many values $(A^l\mathbf{1}_C)(u)$. Each such value is a nonnegative integer, so finitely many of the formulas $\exists_l^{\geq k}\chi_C$ determine it, all of them of depth at most $t + 1$. A conjunction of all these formulas and their negations defines the class of u in $\pi^{(t+1)}$.

Vertices that $\pi^{(t)}$ merges therefore satisfy the same formulas of depth at most t, and vertices it separates are distinguished by the formula defining the class of one of them. □

Restricting the quantifier to $l \leq 1$ replaces $\mathcal{C}_A(\pi)$ by $\mathbb{K}^\pi + A\mathbb{K}^\pi$ throughout the argument, so the same induction gives one round of color refinement per quantifier: the classical correspondence (Immerman & Lander, 1990), recovered as a fragment.

Corollary E.1. *The least depth at which the formulas of Definition E.1 distinguish every pair of vertices with different stable colors is $\tau(G)$. Over graphs whose stable coloring has $m \geq 3$ classes the maximum of that depth is $\lfloor m/2 \rfloor$, against $m - 1$ for the fragment with $l \leq 1$.*

Proof. By Proposition E.1 the formulas of depth at most t distinguish exactly the pairs of vertices that $\pi^{(t)}$ separates, and every $\pi^{(t)}$ is coarser than W (Appendix A.2); so the least depth that separates all the classes of W is the least t with $\pi^{(t)} = W$, which is $\tau(G)$. Its maximum is $\lfloor m/2 \rfloor$ by Theorem 1. In the fragment the same quantity is the iteration number of color refinement, which is at most $m - 1$ by Appendix A.1 and equals it on paths (Proposition A.1). □

The syntax has a quantifier for every length, and one linear solve suffices for a whole level of them: the attenuation ζ_G of Proposition B.1 takes π to $\rho(\pi)$, so depth t costs t solves.