ON SPARSE GRAPH LIMITS FOR NEURAL COMBINATORIAL OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Graph neural networks are often trained on small graphs and deployed on larger ones, raising the question of which graph information must be preserved for combinatorial optimization values to transfer across sizes. We establish size-transfer guarantees for normalized graph optimization values using finite collections of energy features. For affine objectives, including MAX-CUT, coloring defect, and independent-set density, differences in task value are controlled by discrepancies in normalized scalar energy probes on all bounded-degree graphs. For general Lipschitz objectives, such as those with balance penalties, an additional representation error appears; this error vanishes on hyperfinite graph families but can remain bounded away from zero on expanders. Thus, at fixed degree bound, state space, Lipschitz constant, and target accuracy, a finite energy representation independent of graph size suffices universally for affine tasks and asymptotically for general Lipschitz tasks on hyperfinite families. These results yield prediction guarantees under explicit conditions on encoder accuracy, readout stability, and representation coverage, together with statistical generalization bounds. Theoretically, we prove Shapley–Folkman bounds with sharp parameter dependence, including a dimension-free ℓ_2 regime, and realize the resulting energy features with a fixed-weight message-passing network using random node initialization on bounded-treewidth graphs. Experiments with exact solvers and trained models test the resulting transfer guarantees.

1 INTRODUCTION

Graph neural networks are routinely trained on small graphs and deployed on much larger ones, raising the question of when predictions can reliably transfer across sizes. Four lines of work address different aspects of this question. Work on *size transferability* bounds how much the output of a fixed model changes as the input graph grows, for spectral filters and dense and sparse graph limits (Levie et al., 2019; 2021; Ruiz et al., 2020; Keriven et al., 2020; Maskey et al., 2023; Ruiz et al., 2023; Cerviño et al., 2023; Le & Jegelka, 2023), with a parallel empirical line on size extrapolation (Yehudai et al., 2021; Bevilacqua et al., 2021). Work on *statistical generalization* bounds the gap between training and test error, whether through classical capacity measures (Garg et al., 2020; Liao et al., 2021; Morris et al., 2023; Vasileiou et al., 2025), random-graph models (Maskey et al., 2022), or graphon, attributed-graph, graphop, and manifold limit spaces (Levie, 2023; Rauchwerger et al., 2025; Amran et al., 2026; Wang et al., 2025). Related work studies universality and expressivity on large random graphs (Keriven et al., 2021; Keriven & Vaiter, 2023). A *signal-processing* line studies sampling and filtering on finite graphs, graphons, and manifolds (Chamon & Ribeiro, 2018; Ruiz et al., 2021a;b; Wang et al., 2022b; 2024b;a; Le et al., 2024; Parada-Mayorga & Ribeiro, 2024). Finally, *neural combinatorial optimization* trains networks to output relaxed solutions to hard graph problems (Bengio et al., 2021; Cappart et al., 2023; Karalias & Loukas, 2020; Tönshoff et al., 2021; Schuetz et al., 2022; Wang et al., 2022a; Karalias et al., 2022; Wang & Li, 2023; Sanokowski et al., 2024), for which small-to-large deployment is standard but solver reliability remains debated (Angelini & Ricci-Tersenghi, 2023).

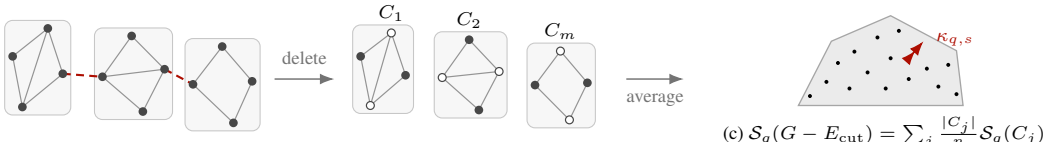
This paper connects these questions for combinatorial optimization on arbitrarily large graphs. In problems such as coloring, MAX-CUT, and maximum independent set, the objective is local but the optimum depends on a global assignment and cannot be recovered from bounded-radius neighborhoods. We ask which graph information must be retained for task values to transfer across sizes.

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.

Too coarse a representation precludes architecture-independent guarantees; too fine a representation retains information the task never uses.

Bounded-degree graph limit theory supplies a candidate representation. For fixed q (e.g. number of colors in coloring), one q-state assignment produces a quotient point: it records the fraction of vertices in each state and the normalized edge frequency between every pair of states. The *q-quotient set* $\mathcal{S}_q(G)$ collects these aggregate profiles over all assignments (see Figure 2). Hausdorff closeness of two quotient sets means that every profile attainable on either graph has a nearby counterpart on the other. Requiring this for every fixed q is *partition convergence* (Bollobás & Riordan, 2011; Borgs et al., 2017), so continuous objectives of quotient statistics have stable optima across graph sizes.

Scalar energy features offer a simpler representation by optimizing affine scores over $\mathcal{S}_q(G)$. But since these quotient sets are nonconvex in general, we can optimize instead over $\operatorname{conv} \mathcal{S}_q(G)$ and employ tools from convex optimization. However, convexification introduces mixtures of profiles that need not come from any single assignment. How close $\mathcal{S}_q(G)$ is to $\operatorname{conv} \mathcal{S}_q(G)$ is therefore a worst-case rounding, or representation, error: convexification leaves affine optima unchanged, but contributes an additional transfer error for nonlinear Lipschitz[1] tasks. Figure 1 shows the mechanism that makes this error small on hyperfinite graphs while Table 1 states the four resulting transfer regimes.



(a) G, with $|E_{\text{cut}}| \le \varepsilon n$ (b) $G - E_{\text{cut}} = \bigsqcup_j C_j, |C_j| \le K$ (c) $\mathcal{S}_q(G - E_{\text{cut}}) = \sum_j \frac{|C_j|}{n} \mathcal{S}_q(C_j)$ dots $\longrightarrow$ shaded convex hull

Figure 1: Hyperfinite decomposition and convexification. In a hyperfinite family, with $n := |V(G)|$, deleting E_{cut}, $|E_{\text{cut}}| \le \varepsilon n$, leaves components $C_1, \dots, C_m$ with $|C_j| \le K(\varepsilon)$ and K independent of n. Independent component assignments make $\mathcal{S}_q(G - E_{\text{cut}})$ a vertex-weighted Minkowski sum (dots); Shapley–Folkman bounds its gap $\kappa_{q,s}$ to the shaded convex hull.

Table 1: Task-value transfer by regime. Here D is the maximum-degree bound; $T_F(X) := \operatorname{opt}_{z \in \mathcal{S}_q(X)} F(z)$, where opt is the task's minimum or maximum; L_s is the Lipschitz constant under the ℓ^s quotient metric; $\eta_{q,s}$ is the ℓ^s-Hausdorff distance between the convexified sets; $\kappa_{q,s}(X)$ is the set-to-hull distance; and $n_X := |V(X)|$. Under an (ε_X, K_X) decomposition, $\rho_1 = O_D(\varepsilon_X + q^2 K_X/n_X)$ and $\rho_2 = O_D(\varepsilon_X + \min\{qK_X/n_X, \sqrt{K_X/n_X}\})$. Note that the regime $\sqrt{K_X/n_X}$ is independent of q

Task class	Graph assumption	Bound on $\lvert T_F(G) - T_F(H)\rvert$
Affine: MAX-CUT, coloring defect, independent-set density	Any D-bounded G, H	$L_s \eta_{q,s}(G, H)$
General L_s-Lipschitz, including finite balance penalties	Any D-bounded G, H	$L_s[\eta_{q,s} + \kappa_{q,s}(G) + \kappa_{q,s}(H)]$
Same general class, $s = 1$	Both graphs admit such decompositions	$L_1[\eta_{q,1} + \rho_1(G) + \rho_1(H)]$
Same general class, $s = 2$	Same	$L_2[\eta_{q,2} + \rho_2(G) + \rho_2(H)]$

The q- and K/n-dependences of the ℓ^1 rate, and both branches of the ℓ^2 rate, are separately sharp; the optimal ε coefficient remains open (Appendix C.7). For a concrete example, appendix A works through the affine/nonlinear distinction and these bounds for coloring.

Contributions. For finite simple graphs of maximum degree at most a fixed D, we obtain:

1. **Transfer of task values.** Differences in a finite set of normalized energy features control differences in affine task values on all graphs. For general Lipschitz objectives, an additional representation error appears, quantified in Table 1 (Theorem 11).

[1]Lipschitzness is defined by equipping the partition convergence topology with an ℓ^s metric, for $s \in \{1, 2\}$

2. **When energy features are enough.** We show using Shapley–Folkman arguments that this additional error vanishes on hyperfinite graph sequences (Theorem 5 and Corollary 6), at essentially sharp rates (Remark 7). Outside this regime, a construction based on expanders gives an $\Omega(1)$ separation (Proposition 8).
3. **Consequences for learning.** These results yield prediction and generalization guarantees that do not depend on graph size (Theorem 11, Proposition 27). On bounded-treewidth graphs, the required energy features can be computed by a fixed-weight message-passing network with random node initialization, while high-girth constructions show that logarithmic depth is necessary in general (Theorems 15–16).

1.1 Related work

We cover only the closest related works here; Appendix B gives the full graph-limit genealogy and theorem-level distinctions.

Quotient convergence. Our (p, W) coordinates and Hausdorff convergence of all $\mathcal{S}_q(G_n)$ are exactly the partition convergence of Borgs et al. (2017). Dense and uniformly upper-regular theories connect quotients to constrained energies (Borgs et al., 2012; 2018), but their normalization or regularity excludes bounded degree. At the vertex scale, unconstrained energies fail in general; Corollary 6 identifies the hyperfinite regime in which they suffice.

Convexification. Shapley–Folkman bounds for separable nonconvex programs are classical (Starr, 1969; Aubin & Ekeland, 1976; Udell & Boyd, 2016). We specialize them to quotient sets, prove matching lower bounds, and identify the associated Lagrangian gap (Remark 7 and Proposition 9).

Graph learning and neural optimization. Size generalization asks when models trained on small graphs predict on larger ones (Yehudai et al., 2021); we ask which task information must be retained. Neural relaxations produce fractional assignments, whereas overlap-gap barriers concern their optimization (Karalias & Loukas, 2020; Schuetz et al., 2022; Gamarnik & Sudan, 2017). Our transfer and rounding bounds isolate the preceding representation question from that optimization question.

2 Quotient sets and partition convergence

To study transfer across graph sizes, we first need a representation that characterizes exactly the information relevant to the task. We summarize each graph by the statistics produced by all possible assignments, collected in a *quotient set*, and compare these summaries across graph sizes using *partition convergence*.

Fix a nonnegative integer D. Throughout, G is a finite, simple, nonempty graph with vertex set $V := V(G)$, edge set $E := E(G)$, maximum degree $\Delta(G) \le D$, and $n := |V(G)|$ vertices; $[m] := \{1, \dots, m\}$ for every positive integer m, q is the state-space size, and $\vec{E}(G) := \{(u, v) : uv \in E(G)\}$ is the oriented edge set.

Definition 1 (q-quotient and quotient set)**.** *For an arbitrary assignment $c : V(G) \to [q]$, define its node weights $p = (p_i)_{i=1}^q$ and oriented edge weights $W = (W_{ij})_{i,j=1}^q$ by*

$$p_i(G, c) := \frac{1}{n}\#\{v : c(v) = i\}, \qquad W_{ij}(G, c) := \frac{1}{n}\#\{(u, v) \in \vec{E}(G) : c(u) = i,\ c(v) = j\}. \tag{1}$$

The weighted q-vertex graph $G/c \equiv (p(G, c), W(G, c))$ is the q-quotient *induced by c. The* quotient set *of G is the collection of all quotients induced by some assignment*

$$\mathcal{S}_q(G) := \{G/c \mid c : V(G) \to [q]\}. \tag{2}$$

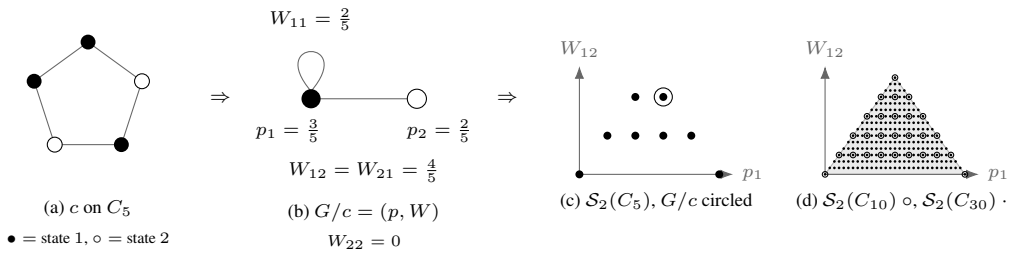
Figure 2(a)–(c) traces one assignment into its quotient point and then into the full quotient set. Vertex normalization makes the quotient well defined for edgeless graphs and linear under vertex-weighted disjoint unions. Moreover, each quotient satisfies $W = W^\top$, $\sum_{i,j} W_{ij} = 2|E(G)|/n \le D$, and $\sum_j W_{ij} \le D p_i$ for all i. With $\Delta_q := \{p \in [0, 1]^q : \sum_i p_i = 1\}$ the simplex, quotients live in the

metric space

$$\mathsf{Y}_{D,q} := \big\{(p,W) : p \in \Delta_q,\ W \in [0,D]^{q\times q},\ W = W^\top,\ \textstyle\sum_j W_{ij} \le D p_i\big\},$$
$$d_q\big((p,W),(p',W')\big) := \|p-p'\|_1 + \|W-W'\|_1. \tag{3}$$

Here and below, the ℓ^1 norm of a matrix is the entrywise norm.

Thus each quotient set $\mathcal{S}_q(G)$ is a nonempty compact subset of $\mathsf{Y}_{D,q}$, and quotient sets are compared in the Hausdorff metric $d_{\mathrm{H}}(A,B) := \max\{\sup_{a\in A}\inf_{b\in B} d_q(a,b), \sup_{b\in B}\inf_{a\in A} d_q(a,b)\}$ (permutation invariance, worked examples: Appendices C.1 and C.6).



(a) c on C_5 ● = state 1, ○ = state 2 (b) $G/c = (p, W)$ (c) $\mathcal{S}_2(C_5)$, G/c circled (d) $\mathcal{S}_2(C_{10})$ ∘, $\mathcal{S}_2(C_{30})$ ·

Figure 2: The fundamental objects on a worked example (exact rational coordinates). (a) An assignment $c = (1,1,2,1,2)$ on C_5. (b) Its q-quotient $G/c = (p,W)$ (Definition 1). (c) The quotient set $\mathcal{S}_2(C_5)$ in the (p_1, W_{12})-projection, G/c circled. (d) Partition convergence: ∘ marks $\mathcal{S}_2(C_{10})$, · marks $\mathcal{S}_2(C_{30})$; the denser grid approaches the shaded limiting region (exact formula: Appendix C.1).

Definition 2 (Partition convergence)**.** *A sequence (G_n) of D-bounded graphs is* partition convergent *if $\mathcal{S}_q(G_n)$ converges in Hausdorff distance for every positive integer q.*

To package all state-space sizes, let $\mathcal{K}(Y)$ denote the hyperspace of nonempty compact subsets of a compact metric space Y, and let $\mathfrak{X}_D^{\mathrm{part}}$ be the closure of $\{(\mathcal{S}_q(G))_{q\ge1} : \Delta(G) \le D\}$ in the product $\prod_{q\ge1}\mathcal{K}(\mathsf{Y}_{D,q})$ of Hausdorff hyperspaces. This *finite-realizable limit space* is compact and metrizable (Theorem 21, Appendix C.6), so every sequence of finite D-bounded graphs has a partition-convergent subsequence (Figure 2(d)). Thus the limit state is available without a structural assumption on the graph sequence. In Appendix C.2 we give comparisons between quotients and the richer local–global and action topologies.

Intuitively speaking, we can afford to only retain information on aggregate node and edge weights because the tasks that we are interested in are *locally-scored*. The next section identifies which tasks need the full quotient set and which can be recovered from scalar energies alone.

3 Continuous quotient problems and ground-state energies

We first express finite-state pairwise energies in the quotient coordinates, then identify exactly what scalar optimization (e.g. min coloring defect) preserves.

A homogeneous q-state pairwise spin model is specified by a unary cost $a \in \mathbb{R}^q$ and a symmetric interaction matrix $B \in \mathbb{R}^{q\times q}$. Angle brackets denote Euclidean pairings; in particular, we write $\langle a,p\rangle := \sum_i a_i p_i$ and $\langle B,W\rangle := \sum_{i,j} B_{ij}W_{ij}$. For every assignment c, its energy per vertex factors as

$$\frac{1}{n}\left(\sum_{v\in V(G)} a_{c(v)} + \sum_{uv\in E(G)} B_{c(u)c(v)}\right) = \langle a, p(G,c)\rangle + \frac{1}{2}\langle B, W(G,c)\rangle. \tag{4}$$

Minimizing this quantity is the model's ground-state energy.

Example: Potts antiferromagnet and proper q-colorings For $a = 0$ and the $q\times q$ identity matrix $B = I_q$, (4) is the density of monochromatic edges—the Hamiltonian of the q-state Potts antiferromagnet—whose ground-state energy is 0 exactly when $\chi(G) \le q$, where $\chi(G)$ is the chromatic number (Fortuin & Kasteleyn, 1972; Sokal, 2005). Appendix C.7 expands the example. In

general, the next theorem says the quotient supports every continuous post-processing, and that its topology is the coarsest one that does so. A proof can be found in Appendix E.2.1.

Theorem 3 (Completeness of continuous functions over quotient weights)**.** *Fix q, and let K_n and K be nonempty compact subsets of the host $\mathsf{Y}_{D,q}$ with $d_{\mathrm{H}}(K_n, K) \to 0$. Then for every continuous map $F : \mathsf{Y}_{D,q} \to \mathbb{R}^k$,*

$$d_{\mathrm{H}}\big(F(K_n), F(K)\big) \longrightarrow 0. \tag{5}$$

In particular, for real-valued F, $\min_{z\in K_n} F(z) \to \min_{z\in K} F(z)$ and $\max_{z\in K_n} F(z) \to \max_{z\in K} F(z)$. Conversely, $K_n \to K$ in Hausdorff distance if and only if $\min_{K_n} F \to \min_K F$ for every real-valued continuous F on $\mathsf{Y}_{D,q}$: the Hausdorff topology is the coarsest one on quotient sets that preserves every continuous query of quotient statistics.

In plainer terms, convergence of quotient sets transfers every task value expressible as a continuous optimum, which is the property a generalization theory for learned graph invariants must certify.

3.1 The nonconvex geometry of quotient sets

When the output is a single scalar affine optimum, the nonconvex geometry is redundant: the *convexified quotient set* $\mathcal{C}_q(G) := \operatorname{conv} \mathcal{S}_q(G)$ leaves every support function $h_{\mathcal{S}_q(G)}(\theta) := \max_{z\in\mathcal{S}_q(G)} \langle \theta, z\rangle$ unchanged, and the next proposition shows this loss is exact. We write $\kappa_q(G) := d_{\mathrm{H}}(\mathcal{S}_q(G), \mathcal{C}_q(G))$ for the resulting convexification gap.

Proposition 4 (Scalar ground-state information)**.** *The optimum values of all homogeneous additive q-state unary–pairwise energies determine exactly $\mathcal{C}_q(G)$. For a sequence, convergence of all support functions is equivalent to Hausdorff convergence of these convexified quotient sets.*

The full family of unconstrained scalar energies thus determines the hull $\mathcal{C}_q(G)$ and no finer information about $\mathcal{S}_q(G)$. Consequently every *affine* task transfers on every graph at a rate controlled by scalar energy discrepancy (e.g. Corollary 12). The rest of this section concerns tasks that read the full geometry. One instance when this distinction vanishes asymptotically is for hyperfinite graphs.

Hyperfinite graphs. The hull nevertheless becomes exact asymptotically when a graph decomposes into many small pieces. A sequence is *growing* if $|V(G_n)| \to \infty$. A growing bounded-degree sequence is *hyperfinite* (Schramm, 2008; Newman & Sohler, 2013; Hatami et al., 2014) if for every $\varepsilon > 0$, deleting at most $\varepsilon|V(G_n)|$ edges leaves components of order at most some $K(\varepsilon)$ independent of n. Paths, cycles, bounded-degree planar, and fixed-minor-free families are standard examples (Newman & Sohler, 2013), and the next theorem quantifies the cost of convexification under such a decomposition.

Theorem 5 (Hyperfinite convexification in ℓ^1)**.** *Suppose that deleting at most εn edges from a D-bounded graph G leaves components of order at most K. For every fixed q,*

$$d_{\mathrm{H}}(\mathcal{S}_q(G), \operatorname{conv} \mathcal{S}_q(G)) \le 4\varepsilon + \frac{2(1+D)r_q K}{n}, \qquad r_q := (q-1) + \frac{q(q+1)}{2} - 1. \tag{6}$$

Consequently, if (G_n) is a growing hyperfinite sequence of D-bounded graphs, then the left-hand side tends to zero for every fixed q.

Thus the hull is an asymptotically lossless summary. Whereas component histograms already determine hyperfinite graphs up to ε edits (Hassidim et al., 2009; Newman & Sohler, 2013), the theorem gives an explicit K/n rate from unconstrained energies alone. Its proof combines the component quotient sets by a weighted Minkowski average and applies Shapley–Folkman (see Appendix E.2.3). A concrete $O_q(n^{-1/2})$ rate for paths and cycles and the theorem's qualitative relatives are in Appendix C.7.

Corollary 6 (Scalar energies suffice on hyperfinite sequences)**.** *For a growing hyperfinite D-bounded sequence (G_n), the following are equivalent: (1) (G_n) is partition convergent; (2) for every finite q, every homogeneous q-state ground-state energy converges.*

Thus scalar energies suffice on hyperfinite families. The sharpness remark and expander counterexample below delineate this conclusion.

Remark 7 (Sharpness and norm dependence)**.** *The two finite-size terms in Theorem 5 are separately sharp: matchings force $\Theta(\min\{1, q^2/n\})$, and unions of K-vertex expanders force $\Omega(K/n)$ already for $q = 2$. Under the Euclidean quotient metric used in Section 5, the convexification-gap bound improves to $O_D(\min\{qK/n, \sqrt{K/n}\})$, with both branches optimal up to constants. Interestingly, the second branch operates independently of q. Thus ℓ^1 and ℓ^2 have the same fixed-q topology but different uniform finite-q rates. Exact statements and proofs are in Appendices C.7 and E.2.5.*

Proposition 8 (Balanced-cut separation; informal)**.** *Outside hyperfinite families, a natural constraint can survive in the full quotient set while disappearing from every scalar energy (cf. Borgs et al., 2017, Example 4). Write $h(G) := \min_{0<|S|\le|V(G)|/2} |E(S, V(G) \setminus S)|/|S|$ for edge expansion. If $h(G_n) \ge h > 0$ and $H_n := G_n \sqcup G_n$, their convexified q-quotient sets are identical for every q, so all scalar energies agree. Yet the balanced-cut penalty $\Phi(G) := \min_{(p,W)\in\mathcal{S}_2(G)}(|p_1 - \frac{1}{2}| + W_{12})$ satisfies $\Phi(H_n) = 0$ and $\Phi(G_n) \ge \min\{1/2, h/2\}$. Hence energy features can lose a constrained task such as balanced partitioning (Nazi et al., 2019; Gatti et al., 2022), even though affine tasks still transfer (the smallest pair is in Appendix C.10).*

Proposition 8 identifies exactly what scalar energies lose: the scalar-versus-constrained gap is a Lagrangian duality gap, bounded by Theorem 5.

Proposition 9 (Lagrangian bridge; informal ((formal statement in Appendix C.7)))**.** *On every finite D-bounded graph, the gap between a Lipschitz balance-penalized optimum and its best unconstrained Lagrangian approximation lies between zero and a Lipschitz multiple of $\kappa_q(G)$.*

Each Lagrangian approximation is an unconstrained energy shifted by a constant, so constrained optima are recoverable from implementable measurements up to the convexification error and no further (Appendices C.7 and C.10; proof in Appendix E.2.9). Scalar energies are thus complete on hyperfinite sequences, with matching rates in K/n. Outside that regime the failure is an explicit, certified duality gap. The next section instantiates the calculus for representative repulsive tasks.

4 Coloring and Compact Spin Spaces

We now express representative repulsive objectives in quotient weights. Coloring and cuts penalize similar adjacent states, independent set combines an edge penalty with a vertex reward, and compact spins provide a continuous relaxation.

Coloring. Put $u_q(p, W) := \frac{1}{2}\sum_{i=1}^{q} W_{ii}$. For a graph G define its minimum normalized q-coloring defect by

$$e_q(G) := \frac{1}{n} \min_{c:V(G)\to[q]} \#\{uv \in E(G) : c(u) = c(v)\} = \min_{(p,W)\in\mathcal{S}_q(G)} u_q(p, W). \tag{7}$$

Equivalently, $e_q(G)$ is the minimum number of edge deletions per vertex making G q-colorable and Theorem 3 extends it continuously to $\mathfrak{X}_D^{\mathrm{part}}$.

Target-graph colorings and cuts. Let H be a fixed nonempty loopless graph with vertex set $[h]$. The normalized number of violated edges in an H-coloring is

$$\mathrm{defect}_H(G) := \frac{1}{n} \min_{\phi:V(G)\to[h]} \#\{uv \in E(G) : \phi(u)\phi(v) \notin E(H)\} = \min_{(p,W)\in\mathcal{S}_h(G)} \frac{1}{2} \sum_{ij\notin E(H)} W_{ij}. \tag{8}$$

Because this objective is affine in W, it is continuous under partition convergence. The complementary maximization is MAX H-COL (Färnqvist et al., 2008), and taking $H = K_q$, the complete graph on q vertices, gives MAX-q-CUT, so that $\mathrm{MaxCut}(G)/n = \max_{(p,W)\in\mathcal{S}_2(G)} W_{12}$.

Independent sets. If state 1 means "selected," rewarding state-1 mass and penalizing state-1 edges yields the exact independence ratio, where $\alpha(G)$ denotes the maximum independent-set size, $\alpha(G)/n = \max_{(p,W)\in\mathcal{S}_2(G)}(p_1 - \frac{1}{2}W_{11})$ since removing one endpoint per selected internal edge proves equality. Hence both the independence ratio and normalized minimum vertex-cover size are continuous.

Compact-spin ground-state energies. Finite state spaces also approximate compact spin spaces, uniformly over graph size thanks to the degree bound.

Proposition 10 (Compact-spin ground-state transfer)**.** *Let K be a nonempty compact metric spin space and let $h : K^2 \to \mathbb{R}$ be a continuous symmetric interaction. Then $\mathsf{E}^-_{K,h}(G) := \frac{1}{n}\min_{z\in K^{V(G)}} \sum_{uv\in E(G)} h(z_u, z_v)$ extends continuously to $\mathfrak{X}^{\mathrm{part}}_D$. The analogous maximum extends continuously as well.*

The finite-state calculus therefore also covers continuous compact-spin energies (proof in Appendix E.3.2; the unit-sphere example behind the Lovász ϑ function is in Appendix C.8), and each example above is a normalized optimum of a Lipschitz function of the quotient—the shape of task the next section transfers across graph orders.

5 Task-value transfer across graph sizes

The preceding examples are vertex-normalized, so their values compare across graph orders; we now make the comparison quantitative. Write $d_{q,1} := d_q$ and $d_{q,2}((p,W),(p',W')) := (\|p-p'\|_2^2 + \|W-W'\|_{\mathrm{F}}^2)^{1/2}$, where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm, and let $d_{\mathrm{H},s}$ be the Hausdorff metric induced by $d_{q,s}$. For two D-bounded graphs G, G', put, for $s \in \{1,2\}$,

$$\begin{aligned} \delta_{q,s}(G,G') &:= d_{\mathrm{H},s}(\mathcal{S}_q(G), \mathcal{S}_q(G')), \\ \eta_{q,s}(G,G') &:= d_{\mathrm{H},s}(\operatorname{conv}\mathcal{S}_q(G), \operatorname{conv}\mathcal{S}_q(G')), \\ \kappa_{q,s}(G) &:= d_{\mathrm{H},s}(\mathcal{S}_q(G), \operatorname{conv}\mathcal{S}_q(G)). \end{aligned} \tag{9}$$

If $\theta = (a, B)$, where $a \in \mathbb{R}^q$ and B is symmetric, write $\langle B, W\rangle := \sum_{i,j} B_{ij}W_{ij}$ and $\|B\|_{\max} := \max_{i,j}|B_{ij}|$, and define

$$\ell_\theta(p,W) := \langle a, p\rangle + \frac{1}{2}\langle B, W\rangle, \qquad \|\theta\|_* := \max\{\|a\|_\infty, \tfrac{1}{2}\|B\|_{\max}\}, \tag{10}$$

$$\gamma_\theta(G) := \min_{z\in\mathcal{S}_q(G)} \ell_\theta(z), \qquad \eta_q(G,G') := \sup_{\|\theta\|_*\le 1} |\gamma_\theta(G) - \gamma_\theta(G')|. \tag{11}$$

Support-function duality gives $\eta_q = \eta_{q,1}$. We similarly view $\eta_{q,2}$ through all scalar energies in the Euclidean dual unit ball, and abbreviate $\delta_q := \delta_{q,1}$ and $\kappa_q := \kappa_{q,1}$. Thus η compares scalar energies, δ compares attainable-statistic sets, and κ measures the information lost by scalar optimization.

Theorem 11 (Task-value size transfer)**.** *For all finite D-bounded graphs G, G', every q, and $s \in \{1,2\}$,*

$$\eta_{q,s}(G,G') \le \delta_{q,s}(G,G') \le \eta_{q,s}(G,G') + \kappa_{q,s}(G) + \kappa_{q,s}(G'). \tag{12}$$

Consequently, if $F : \mathsf{Y}_{D,q} \to \mathbb{R}$ is L_s-Lipschitz under $d_{q,s}$, then

$$|\min_{z\in\mathcal{S}_q(G)} F(z) - \min_{z'\in\mathcal{S}_q(G')} F(z')| \le L_s\big(\eta_{q,s}(G,G') + \kappa_{q,s}(G) + \kappa_{q,s}(G')\big). \tag{13}$$

The same holds for maxima. If F is affine, the minimum and maximum bounds hold with only $L_s\eta_{q,s}(G,G')$ on the right. More explicitly, suppose deleting at most $\varepsilon_G|V(G)|$ and $\varepsilon_{G'}|V(G')|$ edges leaves components of orders at most K_G and $K_{G'}$, respectively. For $s = 1$, the rightmost quantity in (12) is at most

$$\eta_q(G,G') + 4(\varepsilon_G + \varepsilon_{G'}) + 2(1+D)r_q\big(K_G/|V(G)| + K_{G'}/|V(G')|\big). \tag{14}$$

For $s = 2$, the ℓ^2 analogue is the corresponding row of Table 1. Its exact constant appears in (48).

Some simplifications are possible under additional constraints: for a task reading only k quotient coordinates, r_q in (14) improves to k (Proposition 22, Appendix C.7). For affine tasks the κ terms drop out of the transfer bound. Applying this to the examples in Section 4 we have explicit constants on *every* pair of graphs:

Corollary 12 (Transfer of coloring, cut, and independence values)**.** *Let G, G' be finite D-bounded graphs, and let J be a fixed nonempty loopless graph on h vertices. Then, with $\eta_q \le \delta_q$ by (12),*

$$|e_q(G) - e_q(G')| \le \tfrac{1}{2}\eta_q(G,G'), \quad |\operatorname{defect}_J(G) - \operatorname{defect}_J(G')| \le \tfrac{1}{2}\eta_h(G,G'), \tag{15}$$

$$\left|\frac{\operatorname{MaxCut}(G)}{|V(G)|} - \frac{\operatorname{MaxCut}(G')}{|V(G')|}\right| \le \tfrac{1}{2}\eta_2(G,G'), \qquad \left|\frac{\alpha(G)}{|V(G)|} - \frac{\alpha(G')}{|V(G')|}\right| \le \eta_2(G,G'). \tag{16}$$

In particular, if G is q-colorable, then deleting at most $\frac{1}{2}\eta_q(G, G')|V(G')|$ edges makes G' q-colorable. The compact-spin analogue is Corollary 31; proofs are in Appendices E.4.1 and E.4.2.

Rounding in neural combinatorial optimization. The expected quotient of any randomized assignment—e.g. the product measure produced by a relaxation-based loss (Karalias & Loukas, 2020; Schuetz et al., 2022)—lies in $\operatorname{conv} \mathcal{S}_q(G)$. For an *affine* loss the probabilistic method rounds it without error on every graph. Theorem 5 extends this for non-affine rounding Lipschitz objectives.

Corollary 13 (Rounding fractional quotient targets; informal)**.** *Under the decomposition hypothesis of Theorem 5, every fractional target $z^* \in \operatorname{conv} \mathcal{S}_q(G)$ is realized by a single assignment up to d_q-error $4\varepsilon + 2(1+D)r_qK/n$. Hence every Lipschitz task value follows suit, and on hyperfinite classes randomized and single-assignment optima coincide asymptotically (formal: Corollary 19, Appendix C.4).*

5.1 A QUOTIENT-AWARE GRAPH-LEARNING GUARANTEE

The next corollary spells out what an encoder and readout must supply for training at bounded order and testing at any order. Fix $J \geq 1$ and a finite list of tasks $T_j(G) = \operatorname{opt}_{z \in \mathcal{S}_{q_j}(G)} F_j(z)$, $j \in [J]$, each F_j being L_j-Lipschitz and each optimum a minimum or maximum. For tuples $R = (K_1, \dots, K_J)$, $R' = (K'_1, \dots, K'_J)$ of compact sets, write $R(G) := (\mathcal{S}_{q_1}(G), \dots, \mathcal{S}_{q_J}(G))$, $d_R(R, R') := \max_{j \in [J]} d_{\mathrm{H}}(K_j, K'_j)$, and $L_* := \max_{j \in [J]} L_j$.

Corollary 14 (Deterministic size transfer)**.** *For every $r > 0$ there is a finite training dataset $\mathcal{B}_r$ of finite D-bounded graphs, of orders bounded by a number $n_{\max}(D, \mathbf{q}, r)$ independent of the test-graph order ($\mathbf{q} = (q_1, \dots, q_J)$), such that every finite D-bounded graph G satisfies $d_R(R(G), R(B)) \leq r$ for some $B \in \mathcal{B}_r$.*

Suppose an encoder produces $\widehat{R}(G) \in \prod_{j=1}^{J} \mathcal{K}(\mathsf{Y}_{D,q_j})$, with $d_R(\widehat{R}(G), R(G)) \leq \zeta$ for every graph, and a single readout $f : \prod_{j=1}^{J} \mathcal{K}(\mathsf{Y}_{D,q_j}) \to \mathbb{R}^J$—with the same parameters at every graph size—is Λ-Lipschitz for some $\Lambda \geq 0$, meaning that $\|f(x) - f(y)\|_\infty \leq \Lambda d_R(x, y)$. If $\max_{B \in \mathcal{B}_r} \|f(\widehat{R}(B)) - (T_1(B), \dots, T_J(B))\|_\infty \leq \tau$, then on every finite D-bounded graph G,

$$\|f(\widehat{R}(G)) - (T_1(G), \dots, T_J(G))\|_\infty \leq \tau + (\Lambda + L_*)r + 2\Lambda\zeta. \tag{17}$$

Training error on a finite, bounded-order dataset thus propagates to every D-bounded test graph, with slack from the cover radius and encoder accuracy (proof: Appendix E.4.1). It remains to exhibit natural encoders that satisfy the above hypothesis.

Training dataset construction. Compactness alone makes the training set existential. Furthermore, an $(\varepsilon, K(\varepsilon))$-decomposition makes the construction finite, with $n_{\max}$ as in (39) (Proposition 18). In particular, for the class $\mathcal{T}_{D,t}$ of D-bounded graphs with treewidth at most t, Appendix C.3 gives the construction in full.

Raw-graph realization. We use a message-passing neural network (MPNN) with random node initialization (RNI) (Abboud et al., 2021; Sato et al., 2021) to break anonymous symmetry. However, it can also supply the consistent random ranks used by a local partition oracle. The next theorem compiles that observation into a single network that receives only the original graph.

Theorem 15 (Raw-graph RNI probe encoder)**.** *Fix a degree bound $D \geq 2$, a treewidth bound $h \geq 1$, a state number q, a finite probe bank $\Theta \subseteq \{\theta : \|\theta\|_* \leq 1\}$, and $\alpha \in (0, 1/2)$ and $\xi, \delta \in (0, 1)$. There is a fixed-weight RNI-MPNN $\Psi = (\Psi_\theta)_{\theta \in \Theta}$, run directly on the raw graph with vertex-mean readout, such that every graph G of maximum degree at most D and treewidth at most h satisfies*

$$\Pr\left[\max_{\theta \in \Theta} |\Psi_\theta(G) - \gamma_\theta(G)| \leq 2\alpha + \xi\right] \geq 1 - \delta. \tag{18}$$

Its parameters are independent of $|V(G)|$. More explicitly, one may take an internal component scale and query budget

$$K_\alpha = O\left(\frac{D^5 h^{O(h)} \log(D/\alpha)}{\alpha^3}\right), \qquad Q_\alpha = O\left(DK_\alpha^{4h+7}\right), \tag{19}$$

depth $O(Q_\alpha + K_\alpha)$, and finite width depending only on the fixed parameters above. Note that this is a fixed-weight representation theorem. An optimization-level guarantee is still open.

The proof compiles the partition oracle of Edelman et al. (2011). Beyond bounded treewidth, the next result exhibits a general obstruction on expanders, whose treewidth grows.

Theorem 16 (RNI depth lower bound for coloring defect)**.** *There are fixed constants $D \geq 3$ and $c_0, c_1 > 0$ and infinitely many component orders K with pairs $(A_{K,L}, H_{K,L})$, for every $L \geq 1$, of D-regular graphs on the same number of vertices whose components have order at most $2K$, such that*

$$e_2(A_{K,L}) = 0, \qquad e_2(H_{K,L}) \geq c_0. \tag{20}$$

Let one t-layer local MPNN use iid node initialization, final node states in $[-M, M]^w$, vertex-mean pooling, and a readout that is Λ-Lipschitz for $\|\cdot\|_2$, with finite (w, M, Λ) independent of L. If its expected absolute error on both graph families is at most some $\eta < c_0/2$ for every L, then $t \geq c_1 \log K$. The statement excludes global attention, virtual nodes, positional IDs, and size-dependent state or readout bounds.

Their proofs appear in Appendices E.4.6 and E.4.7; the compilation statement is in Appendix C.4.

Empirical results. We evaluate exact energy features and trained neural baselines on a dataset with maximum degree 3 and treewidth 2. The energy probes are computed by family-specific dynamic programs, so these experiments test the energy representation rather than implement the RNI-MPNN encoder of Theorem 15. The dataset contains 72 training graphs with $n \in [8, 30]$ and 66 test graphs with $n \in [120, 360]$. Below, we report test mean absolute error (MAE); learned-model entries are rounded means over seeds $0, 1, 2$:

representation or model	e_2	MaxCut $/n$	α/n
GPS–RWSE	.0046	.0085	.0039
spectral–motif ridge	.0092	.0092	.0060
45 exact generic energies + ridge	.0005	.0005	.0002

The table separates representation quality from end-to-end learning. The 45 exact probes are computed by family-specific dynamic programs and exclude the energies that directly define the targets. Their low MAE therefore shows that generic energy coordinates retain the relevant task information, not that we have learned these coordinates from raw graphs. Among the learned models, GPS–RWSE performs best and improves on the spectral–motif ridge (a handcrafted baseline using degree, triangle, and normalized-adjacency moment features) on all three tasks, although the improvement for MAX-CUT is modest. GPS–RWSE is trained only on graphs with $n \leq 30$ and evaluated on graphs with $n \in [120, 360]$, so it demonstrates size generalization to instances 4–12 times larger than those seen during training.

Statistical generalization. The preceding guarantees are deterministic. For a fixed bank of P exact probes, Proposition 25 gives, uniformly over ℓ^1-bounded linear readouts from m iid D-bounded graphs labeled in $[0, B]$, an absolute-risk gap at confidence $1 - \alpha$ of $O((W_1(1 + D) + B)\sqrt{(\log P + \log(1/\alpha))/m})$. Thus the rate is $m^{-1/2}$, logarithmic in P, and independent of graph order. At fixed D, B, Λ, P, α, Proposition 27 gives $\widetilde{O}(m^{-1/(P+2)})$ for arbitrary $[0, B]$-valued, Λ-Lipschitz readouts, equivalently $\widetilde{O}(\epsilon^{-(P+2)})$ samples for gap ϵ: the covering cost of a P-dimensional cube, but still no graph-order dependence. The full-state analogue, Theorem 26, pays $2^{(3(1+D)/\rho)^{q+q^2}}$ for one state size q. Encoder and approximation errors enter separately through the transfer bounds.

6 Discussion

In this work, we show that the quotient set is the smallest representation that captures all finite-state pairwise tasks. On (ε, K)-decomposable graphs, scalar energy features approximate this representation at the rates of Theorem 5, and these rates are essentially tight for constrained tasks. We further give representation-level guarantees: our encoder realizes the required information through message passing, while the depth lower bound identifies a limitation of bounded-state local architectures. Affine objectives continue to transfer on non-hyperfinite random-regular graphs, although whether gradient-based optimization reliably finds the corresponding solutions remains open. More broadly, an important directions for further investigation are whether these results extend to richer graph limits such as graphops, and whether bounded-state energy hierarchies are strict and their probes can be learned efficiently.

Reproducibility statement

The assumptions and normalization conventions used by the theoretical results are stated with the definitions and theorem statements in Sections 2–5; formal versions of the compressed size-transfer statements appear in Appendix C.4, and complete proofs of all deferred claims are organized theorem-by-theorem in Appendix E. Appendices C.11, C.12, and C.13 document the procedurally generated graph families, data splits, probe construction, exact-routine safeguards, model architectures, baselines, random seeds, optimization settings, and evaluation protocols for experiments E1–E9. We will release the code.

AI use statement

In this work, we used generative AI tools to generate synthetic datasets; help develop theoretical models or conceptual frameworks; formulate mathematical claims; provide critical ingredients for proving mathematical claims; assist in writing proofs; propose or refine hypotheses; design or provide feedback on research methodology and experiments; implement methods; and interpret results. We did not use generative AI tools to clean or reformat datasets or support qualitative or thematic data analysis, and translation was not applicable to this work. Additionally, we used generative AI tools to create or modify scientific figures; suggest experimental parameters; create and edit software code and other research artifacts; brainstorm; search for and identify relevant literature; summarize and analyze existing literature; identify research topics or gaps; and draft, structure, and edit the paper, including improving readability and formatting references. We have reviewed all AI-assisted work. In particular, all AI-assisted theorem statements and proofs were checked line by line by the authors, and known results or techniques suggested by AI tools were traced to their original sources and cited; AI-generated synthetic datasets were inspected by the authors to confirm they have the intended properties; AI-assisted code and figures were reviewed by the authors; and all AI-drafted text was revised by the authors to ensure it accurately reflects our contributions. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

References

Ralph Abboud, İsmail İlkan Ceylan, Martin Grohe, and Thomas Lukasiewicz. The surprising power of graph neural networks with random node initialization. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2112–2118, 2021. doi: 10.24963/ijcai.2021/291.

Ofek Amran, Tom Gilat, and Ron Levie. A graphop analysis of graph neural networks on sparse graphs: Generalization and universal approximation. In *Proceedings of the 43rd International Conference on Machine Learning (ICML)*, 2026. URL https://arxiv.org/abs/2602.08785.

Maria Chiara Angelini and Federico Ricci-Tersenghi. Modern graph neural networks do worse than classical greedy algorithms in solving combinatorial optimization problems like maximum independent set. *Nature Machine Intelligence*, 5:29–31, 2023. doi: 10.1038/s42256-022-00589-y.

Jean-Pierre Aubin and Ivar Ekeland. Estimates of the duality gap in nonconvex optimization. *Mathematics of Operations Research*, 1(3):225–245, 1976. doi: 10.1287/moor.1.3.225.

Ágnes Backhausz and Balázs Szegedy. Action convergence of operators and graphs. *Canadian Journal of Mathematics*, 74(1):72–121, 2022. doi: 10.4153/S0008414X2000070X. URL https://doi.org/10.4153/S0008414X2000070X.

Brenda S. Baker. Approximation algorithms for NP-complete problems on planar graphs. *Journal of the ACM*, 41(1):153–180, 1994. doi: 10.1145/174644.174650.

Yoshua Bengio, Andrea Lodi, and Antoine Prouvost. Machine learning for combinatorial optimization: a methodological tour d'horizon. *European Journal of Operational Research*, 290(2):405–421, 2021. doi: 10.1016/j.ejor.2020.07.063.

Itai Benjamini and Oded Schramm. Recurrence of distributional limits of finite planar graphs. *Electronic Journal of Probability*, 6(23):1–13, 2001. doi: 10.1214/EJP.v6-96.

Beatrice Bevilacqua, Yangze Zhou, and Bruno Ribeiro. Size-invariant graph representations for graph classification extrapolations. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, 2021. URL https://arxiv.org/abs/2103.05045.

Béla Bollobás. The isoperimetric number of random regular graphs. *European Journal of Combinatorics*, 9(3):241–244, 1988. doi: 10.1016/S0195-6698(88)80014-3.

Béla Bollobás and Oliver Riordan. Sparse graphs: Metrics and random models. *Random Structures & Algorithms*, 39(1):1–38, 2011. doi: 10.1002/rsa.20334.

Christian Borgs, Jennifer T. Chayes, László Lovász, Vera T. Sós, and Katalin Vesztergombi. Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Annals of Mathematics*, 176(1):151–219, 2012.

Christian Borgs, Jennifer T. Chayes, Jeff Kahn, and László Lovász. Left and right convergence of graphs with bounded degree. *Random Structures & Algorithms*, 42(1):1–28, 2013.

Christian Borgs, Jennifer Chayes, and David Gamarnik. Convergent sequences of sparse graphs: A large deviations approach. *Random Structures & Algorithms*, 51(1):52–89, 2017. doi: 10.1002/rsa.20694.

Christian Borgs, Jennifer T. Chayes, Henry Cohn, and Yufei Zhao. An L^p theory of sparse graph convergence II: LD convergence, quotients, and right convergence. *The Annals of Probability*, 46 (1):337–396, 2018. doi: 10.1214/17-AOP1187.

Christian Borgs, Jennifer T. Chayes, Henry Cohn, and Yufei Zhao. An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Transactions of the American Mathematical Society*, 372(5):3019–3062, 2019.

Quentin Cappart, Didier Chételat, Elias B. Khalil, Andrea Lodi, Christopher Morris, and Petar Veličković. Combinatorial optimization and reasoning with graph neural networks. *Journal of Machine Learning Research*, 24(130):1–61, 2023. URL https://arxiv.org/abs/2102.09544.

Juan Cerviño, Luana Ruiz, and Alejandro Ribeiro. Learning by transference: training graph neural networks on growing graphs. *IEEE Transactions on Signal Processing*, 71:233–247, 2023.

Luiz F. O. Chamon and Alejandro Ribeiro. Greedy sampling of graph signals. *IEEE Transactions on Signal Processing*, 66(1):34–47, 2018. doi: 10.1109/TSP.2017.2755586.

Martin Doležal and Wiesław Kubiś. Categorical approach to graph limits. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas*, 120(2):52, 2026. doi: 10.1007/s13398-026-01846-0. URL https://doi.org/10.1007/s13398-026-01846-0.

Alan Edelman, Avinatan Hassidim, Huy N. Nguyen, and Krzysztof Onak. An efficient partitioning oracle for bounded-treewidth graphs. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 6845 of *Lecture Notes in Computer Science*, pp. 530–541. Springer, 2011. doi: 10.1007/978-3-642-22935-0_45.

Tommy Färnqvist, Peter Jonsson, and Johan Thapper. Approximability distance in the space of H-colourability problems, 2008. URL https://arxiv.org/abs/0802.0423.

Cornelius M. Fortuin and Pieter W. Kasteleyn. On the random-cluster model: I. Introduction and relation to other models. *Physica*, 57(4):536–564, 1972.

Alan Frieze and Mark Jerrum. Improved approximation algorithms for MAX k-CUT and MAX BISECTION. *Algorithmica*, 18(1):67–81, 1997.

David Gamarnik. Barriers for the performance of graph neural networks (GNN) in discrete random structures. *Proceedings of the National Academy of Sciences*, 120(46):e2314092120, 2023.

David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. *Annals of Probability*, 45(4):2353–2376, 2017.

Vikas K. Garg, Stefanie Jegelka, and Tommi Jaakkola. Generalization and representational limits of graph neural networks. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, 2020. URL https://arxiv.org/abs/2002.06157.

Alice Gatti, Zhixiong Hu, Tess Smidt, Esmond G. Ng, and Pieter Ghysels. Graph partitioning and sparse matrix ordering using reinforcement learning and graph neural networks. *Journal of Machine Learning Research*, 23(303):1–28, 2022. URL https://arxiv.org/abs/2104.03546.

Avinatan Hassidim, Jonathan A. Kelner, Huy N. Nguyen, and Krzysztof Onak. Local graph partitions for approximation and testing. In *50th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 22–31, 2009. doi: 10.1109/FOCS.2009.77.

Hamed Hatami, László Lovász, and Balázs Szegedy. Limits of locally-globally convergent graph sequences. *Geometric and Functional Analysis*, 24(1):269–296, 2014. doi: 10.1007/s00039-014-0258-7.

Nikolaos Karalias and Andreas Loukas. Erdős goes neural: an unsupervised learning framework for combinatorial optimization on graphs. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6659–6672, 2020. URL https://arxiv.org/abs/2006.10643.

Nikolaos Karalias, Joshua W. Robinson, Andreas Loukas, and Stefanie Jegelka. Neural set function extensions: learning with discrete functions in high dimensions. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://arxiv.org/abs/2208.04055.

Nicolas Keriven and Samuel Vaiter. What functions can graph neural networks compute on random graphs? The role of positional encoding. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://arxiv.org/abs/2305.14814.

Nicolas Keriven, Alberto Bietti, and Samuel Vaiter. Convergence and stability of graph convolutional networks on large random graphs. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://arxiv.org/abs/2006.01868.

Nicolas Keriven, Alberto Bietti, and Samuel Vaiter. On the universality of graph neural networks on large random graphs. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://arxiv.org/abs/2105.13099.

Dávid Kunszenti-Kovács, László Lovász, and Balázs Szegedy. Measures on the square as sparse graph limits. *Journal of Combinatorial Theory, Series B*, 138:1–40, 2019. doi: 10.1016/j.jctb.2019.01.004. URL https://doi.org/10.1016/j.jctb.2019.01.004.

Thien Le and Stefanie Jegelka. Limits, approximation and size transferability for GNNs on sparse graphs via graphops. In *Advances in Neural Information Processing Systems*, volume 36, pp. 41305–41342, 2023. URL https://arxiv.org/abs/2306.04495.

Thien Le, Luana Ruiz, and Stefanie Jegelka. A Poincaré inequality and consistency results for signal sampling on large graphs. In *International Conference on Learning Representations*, 2024. URL https://arxiv.org/abs/2311.10610.

Ron Levie. A graphon-signal analysis of graph neural networks. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://arxiv.org/abs/2305.15987.

Ron Levie, Elvin Isufi, and Gitta Kutyniok. On the transferability of spectral graph filters. In *13th International Conference on Sampling Theory and Applications (SampTA)*, pp. 1–5, 2019. URL https://arxiv.org/abs/1901.10524.

Ron Levie, Wei Huang, Lorenzo Bucci, Michael M. Bronstein, and Gitta Kutyniok. Transferability of spectral graph convolutional neural networks. *Journal of Machine Learning Research*, 22(272):1–59, 2021. URL https://arxiv.org/abs/1907.12972.

Renjie Liao, Raquel Urtasun, and Richard Zemel. A PAC-Bayesian approach to generalization bounds for graph neural networks. In *International Conference on Learning Representations*, 2021. URL https://arxiv.org/abs/2012.07690.

Richard J. Lipton and Robert Endre Tarjan. Applications of a planar separator theorem. *SIAM Journal on Computing*, 9(3):615–627, 1980. doi: 10.1137/0209046. URL https://doi.org/10.1137/0209046.

Andreas Loukas. What graph neural networks cannot learn: depth vs width. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=B1l2bp4YwS.

László Lovász. On the Shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25 (1):1–7, 1979.

László Lovász. Hyperfinite graphings and combinatorial optimization, 2017. URL https://arxiv.org/abs/1709.03179.

Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3): 261–277, 1988. doi: 10.1007/BF02126799.

Sohir Maskey, Ron Levie, Yunseok Lee, and Gitta Kutyniok. Generalization analysis of message passing neural networks on large random graphs. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://arxiv.org/abs/2202.00645.

Sohir Maskey, Ron Levie, and Gitta Kutyniok. Transferability of graph neural networks: an extended graphon approach. *Applied and Computational Harmonic Analysis*, 63:48–83, 2023.

Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. MIT Press, 2nd edition, 2018.

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 *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 4602–4609, 2019.

Christopher Morris, Floris Geerts, Jan Tönshoff, and Martin Grohe. WL meet VC. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, 2023. URL https://arxiv.org/abs/2301.11039.

Azade Nazi, Will Hang, Anna Goldie, Sujith Ravi, and Azalia Mirhoseini. GAP: generalizable approximate graph partitioning framework. arXiv:1903.00614, 2019. URL https://arxiv.org/abs/1903.00614.

Ilan Newman and Christian Sohler. Every property of hyperfinite graphs is testable. *SIAM Journal on Computing*, 42(3):1095–1112, 2013. doi: 10.1137/120890946. URL https://doi.org/10.1137/120890946.

Alejandro Parada-Mayorga and Alejandro Ribeiro. Sampling and uniqueness sets in graphon signal processing. arXiv:2401.06279, 2024. URL https://arxiv.org/abs/2401.06279.

Renfrey B. Potts. Some generalized order-disorder transformations. *Mathematical Proceedings of the Cambridge Philosophical Society*, 48(1):106–109, 1952.

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*, volume 35, pp. 14501–14515, 2022. URL https://arxiv.org/abs/2205.12454.

Levi Rauchwerger, Stefanie Jegelka, and Ron Levie. Generalization, expressivity, and universality of graph neural networks on attributed graphs. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=qKgd7RaAem.

Yuriy A. Reznik. An algorithm for quantization of discrete probability distributions. In *Data Compression Conference*, pp. 333–342, 2011. doi: 10.1109/DCC.2011.40.

Luana Ruiz, Luiz F. O. Chamon, and Alejandro Ribeiro. Graphon neural networks and the transferability of graph neural networks. In *Advances in Neural Information Processing Systems*, volume 33, 2020. URL https://arxiv.org/abs/2006.03548.

Luana Ruiz, Luiz F. O. Chamon, and Alejandro Ribeiro. Graphon signal processing. *IEEE Transactions on Signal Processing*, 69:4961–4976, 2021a. doi: 10.1109/TSP.2021.3106857.

Luana Ruiz, Fernando Gama, and Alejandro Ribeiro. Graph neural networks: architectures, stability, and transferability. *Proceedings of the IEEE*, 109(5):660–682, 2021b.

Luana Ruiz, Luiz F. O. Chamon, and Alejandro Ribeiro. Transferability properties of graph neural networks. *IEEE Transactions on Signal Processing*, 71:3474–3489, 2023. doi: 10.1109/TSP.2023.3297848.

Sebastian Sanokowski, Sepp Hochreiter, and Sebastian Lehner. A diffusion model framework for unsupervised neural combinatorial optimization. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 43346–43367, 2024. URL https://arxiv.org/abs/2406.01661.

Ryoma Sato, Makoto Yamada, and Hisashi Kashima. Random features strengthen graph neural networks. In *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)*, pp. 333–341, 2021. doi: 10.1137/1.9781611976700.38.

Oded Schramm. Hyperfinite graph limits. *Electronic Research Announcements in Mathematical Sciences*, 15:17–23, 2008. doi: 10.3934/era.2008.15.17. URL https://doi.org/10.3934/era.2008.15.17.

Martin J. A. Schuetz, J. Kyle Brubaker, and Helmut G. Katzgraber. Combinatorial optimization with physics-inspired graph neural networks. *Nature Machine Intelligence*, 4:367–377, 2022. doi: 10.1038/s42256-022-00468-6.

Maurice Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958. doi: 10.2140/pjm.1958.8.171.

Alan D. Sokal. The multivariate Tutte polynomial (alias Potts model) for graphs and matroids. In Bridget S. Webb (ed.), *Surveys in Combinatorics 2005*, volume 327 of *London Mathematical Society Lecture Note Series*, pp. 173–226. Cambridge University Press, 2005. doi: 10.1017/CBO9780511734885.009. URL https://doi.org/10.1017/CBO9780511734885.009.

Ross M. Starr. Quasi-equilibria in markets with non-convex preferences. *Econometrica*, 37(1):25–38, 1969. doi: 10.2307/1909201. URL https://doi.org/10.2307/1909201.

Jan Tönshoff, Martin Ritzert, Hinrikus Wolf, and Martin Grohe. Graph neural networks for maximum constraint satisfaction. *Frontiers in Artificial Intelligence*, 3:580607, 2021. doi: 10.3389/frai.2020.580607.

Madeleine Udell and Stephen Boyd. Bounding duality gap for separable problems with linear constraints. *Computational Optimization and Applications*, 64(2):355–378, 2016. doi: 10.1007/s10589-015-9819-4.

Antonis Vasileiou, Ben Finkelshtein, Floris Geerts, Ron Levie, and Christopher Morris. Covered forest: fine-grained generalization analysis of graph neural networks. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, 2025. URL https://arxiv.org/abs/2412.07106.

Haoyu Wang and Pan Li. Unsupervised learning for combinatorial optimization needs meta learning. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2301.03116.

Haoyu Wang, Nan Wu, Hang Yang, Cong Hao, and Pan Li. Unsupervised learning for combinatorial optimization with principled objective relaxation. In *Advances in Neural Information Processing Systems*, volume 35, 2022a. URL https://arxiv.org/abs/2207.05984.

Zhiyang Wang, Luana Ruiz, and Alejandro Ribeiro. Convolutional neural networks on manifolds: from graphs and back. In *56th Asilomar Conference on Signals, Systems, and Computers*, 2022b. URL https://arxiv.org/abs/2210.00376.

Zhiyang Wang, Luana Ruiz, and Alejandro Ribeiro. Geometric graph filters and neural networks: limit properties and discriminability trade-offs. *IEEE Transactions on Signal Processing*, 72: 2244–2259, 2024a.

Zhiyang Wang, Luana Ruiz, and Alejandro Ribeiro. Stability to deformations of manifold filters and manifold neural networks. *IEEE Transactions on Signal Processing*, 72:2130–2146, 2024b.

Zhiyang Wang, Juan Cerviño, and Alejandro Ribeiro. A manifold perspective on the statistical generalization of graph neural networks. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, 2025. URL https://arxiv.org/abs/2406.05225.

Fa-Yueh Wu. The Potts model. *Reviews of Modern Physics*, 54(1):235–268, 1982.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=ryGs6iA5Km.

Gilad Yehudai, Ethan Fetaya, Eli Meirom, Gal Chechik, and Haggai Maron. From local structures to size generalization in graph neural networks. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 11975–11986, 2021. URL https://proceedings.mlr.press/v139/yehudai21a.html.

A AFFINE VS NONLINEAR TASK: A WORKED TRANSLATION TO COLORING

In this section, we detailed an application of the theory to a particular pairwise graph task: graph coloring. The quotient formalism separates two coloring questions that are easy to conflate: minimizing the number of monochromatic edges, and doing so while controlling the color-class proportions. The first objective is affine and transfers on every graph. The second is nonlinear and exposes the rounding error created by convexification. We translate the definitions and bounds for both objectives.

Proper coloring as an affine task. Fix a palette $[q]$ and an assignment $c : V(G) \to [q]$. Its quotient coordinate p_i is the fraction of vertices receiving color i, while $W_{ii}/2$ is the number of monochromatic edges of color i, normalized by $n := |V(G)|$. The factor $1/2$ removes the two orientations of each edge. Consequently,

$$u_q(p, W) := \frac{1}{2} \sum_{i=1}^{q} W_{ii}, \qquad e_q(G) = \min_{(p,W)\in\mathcal{S}_q(G)} u_q(p, W) \tag{21}$$

is the minimum monochromatic-edge density. It vanishes exactly when G is properly q-colorable and otherwise equals the minimum number of edge deletions per vertex needed to make G q-colorable.

Because u_q is affine, convexifying the quotient set does not change its minimum. The corresponding probe is $\theta_{\text{col}} = (0, I_q)$, whose dual norm in (10) is $1/2$. Thus Corollary 12 gives, for every pair of D-bounded graphs,

$$|e_q(G) - e_q(G')| \le \frac{1}{2}\eta_q(G, G'). \tag{22}$$

In particular, if G is properly q-colorable, deleting at most $\frac{1}{2}\eta_q(G, G')|V(G')|$ edges makes G' q-colorable. This conclusion needs no decomposition assumption. It concerns repair distance, not the Boolean property of being q-colorable: for example, $e_2(C_{2m+1}) = 1/(2m + 1) \to 0$, although every odd cycle has chromatic number three (Appendix D.1).

What a balance requirement adds. Fix a target color proportion $a \in \Delta_q$ and a finite penalty $\lambda \ge 0$. Define the penalized balanced-coloring defect by

$$\text{BalDef}_{q,a,\lambda}(G) := \min_{(p,W)\in\mathcal{S}_q(G)} \big(u_q(p, W) + \lambda\|p - a\|_1\big). \tag{23}$$

Unlike (21), this objective is nonlinear in the quotient. A four-vertex star makes the lost information explicit. The two proper 2-colorings of $K_{1,3}$ have proportions $(1/4, 3/4)$ and $(3/4, 1/4)$. Their

midpoint has proportion $(1/2, 1/2)$ and zero defect, so it belongs to $\operatorname{conv} \mathcal{S}_2(K_{1,3})$. No single balanced assignment is proper, however: every two–two assignment leaves at least one of the three edges monochromatic. For $a = (1/2, 1/2)$, direct enumeration therefore gives

$$\mathrm{BalDef}_{2,a,\lambda}(K_{1,3}) = \min\{\lambda/2, 1/4\}, \qquad \mathrm{BalDef}_{2,a,\lambda}(K_{1,3} \sqcup K_{1,3}) = 0. \tag{24}$$

The second graph colors its two copies in opposite orientations. Yet self-union leaves every convexified quotient set unchanged by (49), so all scalar energy features agree on the two graphs. Equation (24) is a concrete instance of the set-to-hull information measured by κ_2.

The coloring-specific transfer bound. The balanced objective reads only $q + 1$ statistics. To expose those statistics without losing constants, define the linear projection

$$\Pi_q(p, W) := (p, \tau), \qquad \tau := \sum_{i=1}^{q} W_{ii} = 2u_q(p, W). \tag{25}$$

This projection is nonexpansive from the quotient ℓ^1 metric to ℓ^1 because $\|p - p'\|_1 + |\tau - \tau'| \le \|p - p'\|_1 + \|W - W'\|_1$. Let $\kappa_q^{\mathrm{col}}(X)$ be the Hausdorff distance between $\Pi_q \mathcal{S}_q(X)$ and its convex hull, and put $M_\lambda := \max\{\lambda, 1/2\}$. The function $(p, \tau) \mapsto \tau/2 + \lambda\|p - a\|_1$ is M_λ-Lipschitz. Theorem 11 therefore specializes to

$$|\mathrm{BalDef}_{q,a,\lambda}(G) - \mathrm{BalDef}_{q,a,\lambda}(G')| \le M_\lambda\big(\eta_q(G, G') + \kappa_q^{\mathrm{col}}(G) + \kappa_q^{\mathrm{col}}(G')\big). \tag{26}$$

Suppose that deleting at most $\varepsilon_X |V(X)|$ edges from each $X \in \{G, G'\}$ leaves components of order at most K_X, and write $n_X := |V(X)|$. Proposition 22, applied to the $q + 1$ outputs of Π_q, bounds each coloring-specific gap by

$$\kappa_q^{\mathrm{col}}(X) \le 4\varepsilon_X + \frac{2(1 + D)(q + 1)K_X}{n_X}. \tag{27}$$

Substitution into (26) yields the explicit size-transfer bound

$$\begin{aligned} &|\mathrm{BalDef}_{q,a,\lambda}(G) - \mathrm{BalDef}_{q,a,\lambda}(G')| \\ &\quad \le M_\lambda \left[\eta_q(G, G') + 4(\varepsilon_G + \varepsilon_{G'}) + 2(1 + D)(q + 1)\left(\frac{K_G}{n_G} + \frac{K_{G'}}{n_{G'}}\right)\right]. \end{aligned} \tag{28}$$

Thus the task-specific projection improves the generic q^2 factor to $q + 1$. The same estimate gives a rounding statement: every fractional balance–defect profile in $\operatorname{conv} \Pi_q \mathcal{S}_q(X)$ lies within the right-hand side of (27) of one deterministic coloring profile, and its objective increases by at most M_λ times that distance.

For comparison, the full-space Euclidean bound in (48) gives the dimension-free branch in Table 1. A valid Euclidean Lipschitz constant for (23) is $L_2 := \sqrt{q(\lambda^2 + 1/4)}$. Define the Euclidean representation remainder of a decomposable graph X by

$$\beta_{q,2}(X) := 4\varepsilon_X + \sqrt{1 + D^2}\,\min\left\{\sqrt{r_q}\frac{K_X}{n_X}, \sqrt{\frac{K_X}{n_X}}\right\} \tag{29}$$

and obtain

$$|\mathrm{BalDef}_{q,a,\lambda}(G) - \mathrm{BalDef}_{q,a,\lambda}(G')| \le L_2\big(\eta_{q,2}(G, G') + \beta_{q,2}(G) + \beta_{q,2}(G')\big). \tag{30}$$

On a growing hyperfinite family, these remainders vanish after first choosing the deletion fraction small and then taking the graph order large. Scalar energies therefore preserve ordinary coloring defect on every graph, while they preserve this balance-sensitive coloring objective asymptotically when the quotient set approaches its convex hull. Exact balance constraints are not covered by the finite-penalty statement.

B Extended related work

Section 1.1 compresses the delimitation of prior art to its quantified claims; this appendix gives the full exposition, one limit notion at a time, with example- and theorem-level deltas.

Partition convergence. For simple D-bounded graphs, our node and oriented-edge weights (p, W) are exactly the (x, X) coordinates of Borgs et al. (2017), and our Hausdorff convergence of all $\mathcal{S}_q(G_n)$ is their *partition convergence*, up to an equivalent finite-dimensional norm. Thus the topology and its compact hyperspace packaging are prior work, not a new limit notion. Their Examples 4 and 5 also separate partition and Benjamini–Schramm convergence (Benjamini & Schramm, 2001) in both directions: alternating one expander copy with two preserves a local limit but destroys partition convergence, while alternating unions of 4- and 6-cycles is partition convergent but not locally convergent.

Normalized quotients and ground-state energies. The quotient calculus originates in the dense theory: Borgs et al. (2012) prove that cut-metric convergence, Hausdorff convergence of quotient sets, and *microcanonical* ground-state and free-energy convergence are equivalent for dense sequences (their Theorem 2.8), that plain (unconstrained) energies are strictly weaker (their Corollary 3.6 and Examples 6.1 and 6.3—the dense ancestors of our Corollary 23), and that energies are Lipschitz in cut distance (their Theorem 4.6, the dense ancestor of Corollary 12); their fractional quotient sets are closures of integer ones, not convex hulls, and convexity plays no role in that regime. For bounded degree, Borgs et al. (2013) relate right and left convergence through free energies of soft-core models, again without quotient geometry. Borgs et al. (2018, Definition 2.3) extend the dense equivalences to sparse graphs with unbounded average degree under uniform upper regularity, calling Hausdorff convergence of their quotient sets *quotient convergence*. For a nonempty simple graph their edge coordinate is

$$\beta_{ij} = \frac{W_{ij}}{\bar{d}(G)}, \qquad \bar{d}(G) := \frac{2|E(G)|}{|V(G)|}, \tag{31}$$

whereas we retain W_{ij}. Neither theory has content at bounded degree, for two different reasons: the dense normalization divides edge counts by n^2, so every D-bounded quotient collapses to zero edge weights and each side of the dense equivalence holds vacuously there, while a uniformly upper-regular sequence of simple graphs has unbounded average degree (Borgs et al., 2019, Proposition C.15), so D-bounded sequences sit outside the L^p class altogether. Under the vertex normalization, which keeps bounded degree meaningful, the equivalence acquires content—and in that form it is false, by the expander pair of Proposition 8; only the reverse direction is at issue, the forward direction being free of hypotheses (their Theorem 3.3, or Theorem 3 here). Two of their examples are the nearest precursors of this paper: their Example 3.3.5 shows unconstrained energies insufficient even under upper regularity, and their Example 3.3.7 (after Borgs et al., 2017) observes, qualitatively, that quotients of growing disjoint unions converge to the *convex hull* of the component quotients—Theorem 5 turns exactly this observation into a universal finite-graph upper bound for arbitrary (ε, K)-decompositions, with worst-case q- and K/n-dependences matched separately. Their Theorem 3.3 proves that normalized quotient convergence implies convergence of ordinary and microcanonical ground-state energies, and their Theorem 2.15 makes metric convergence, quotient convergence, and *microcanonical* ground-state convergence equivalent for uniformly upper-regular weighted sequences. On D-bounded sequences, partition convergence forces $\bar{d}(G_n)$ to converge (the total W-mass is $\bar{d}$); when the limit is positive the two quotient coordinates differ only by a convergent scalar, so on such sequences the topologies essentially agree. The delta claimed here is therefore not the topology but the *energy family and its quantitative reach*: (i) Corollary 6 characterizes partition convergence on hyperfinite classes by *unconstrained* unary–pairwise energies alone, a strictly weaker family than the microcanonical one—and Corollary 23 shows by an expander alternation that this weaker family cannot suffice in general, so the hyperfiniteness hypothesis is doing real work; (ii) Proposition 9 bounds the penalized microcanonical-versus-bounded-Lagrangian gap by the convexification error κ_q, turning the relation between the two energy families into a finite-graph inequality rather than a limit statement; and (iii) all of this comes with the explicit finite-n rate of Theorem 5 and the matching lower bound of Remark 7, neither of which appears in the cited works. The distinction in (i) matters for learning: a microcanonical energy is not a feature map—a learner's energy-based features are unconstrained optima—so the classical characterizations, correct as they are, are not implementable as order-independent feature vectors, and the question of when the implementable family suffices is not posed there.

Small-component decompositions and hyperfiniteness. Hyperfiniteness is standard terminology for bounded-degree graph families that can be broken into uniformly small components after deleting few edges (Schramm, 2008; Newman & Sohler, 2013; Hatami et al., 2014). Hatami et al. (2014,

Theorem 10.5 and Corollary 10.6) prove that a growing, locally convergent hyperfinite bounded-degree sequence is local–global convergent. Since $\mathcal{S}_q$ is a continuous first-moment image of the radius-one colored-neighborhood set, their theorem implies partition convergence under those hypotheses. Our Theorem 5 instead gives a finite-graph distance to $\operatorname{conv}\mathcal{S}_q$, and Corollary 6 characterizes partition convergence on a hyperfinite sequence by convergence of all scalar pairwise ground-state energies. This quantitative convexification and scalar characterization are not statements of Theorem 10.5.

Our novelty audit covered the dense series and its bounded-degree and L^p successors above (Borgs et al., 2012; 2013; 2017; 2018), intrinsic partition representations by measures and square-graphons (Kunszenti-Kovács et al., 2019; Doležal & Kubiś, 2026), hyperfinite graphings in combinatorial optimization (Lovász, 2017), the local–global and action literature cited below, the generic Shapley–Folkman theorem (Starr, 1969), the classical Shapley–Folkman duality-gap estimates for separable nonconvex programs (Aubin & Ekeland, 1976; Udell & Boyd, 2016), of which Proposition 9 is a quotient-set instantiation, and overlap-gap-property obstructions for local algorithms on sparse random regular graphs and for fixed-depth graph neural networks on particular random structures (Gamarnik & Sudan, 2017; Gamarnik, 2023), which block *optimization* in their stated random-instance regimes and are complementary to the *representation* ceiling certified here. Within that searched scope, we did not find a prior finite-graph bound on $d_{\mathrm{H}}(\mathcal{S}_q(G), \operatorname{conv}\mathcal{S}_q(G))$ from a bounded-component edge decomposition, nor the resulting scalar-energy characterization without an assumed local limit. This is a scoped absence claim, not an exhaustive claim over all decomposable optimization. On sharpness, the K/n term is matched at fixed q, the $r_q = \Theta(q^2)$ factor is matched on perfect matchings, and the Euclidean rate is matched in both of its regimes (Remark 7); hyperfiniteness is non-removable (Corollary 23). The optimal constant in the 4ε term and the optimal number of finite probes remain open. Ground-state coordinates are efficiently computable on bounded treewidth (Proposition 17) and generally hard beyond it.

Partition oracles, Baker's technique, and the convex hull. On hyperfinite classes, partition oracles (Hassidim et al., 2009; Newman & Sohler, 2013) give sublinear-query access to an (ε, K)-decomposition, and every additive graph parameter—MAX-CUT, minimum vertex cover, maximum independent set—is then estimated to additive εn from the empirical distribution of component isomorphism types; Baker's layering (Baker, 1994) and separator recursions (Lipton & Tarjan, 1980) give classical polynomial-time approximation schemes for many planar optimization problems. Qualitatively, then, it is classical that on hyperfinite graphs each task value is a local statistic of component types, and Proposition 17 and Proposition 18 re-derive that statement in quotient language. Theorem 5 says something different in two respects. First, it concerns the *representation* rather than any one task: if N_K denotes the number of connected D-bounded types of order at most K, the component-type histogram has dimension $N_K \le K2^{K(K-1)/2}$. By contrast, the convex hull of the quotient set lives in dimension $r_q = \Theta(q^2)$ and is read off, up to error $2(1+D)\rho$, by the $P = |\mathcal{N}_\rho|$ scalar energies in any finite ρ-net $\mathcal{N}_\rho$ of the dual-unit probe ball. The theorem gives the universal compression-loss upper bound $4\varepsilon + 2(1+D)r_qK/n$ for *all* 1-*Lipschitz* q-state pairwise tasks simultaneously; an L-Lipschitz task multiplies the bound by L. Separate examples match the worst-case K/n and q^2 dependences; the optimal ε coefficient remains open. This includes the constrained ones (balanced cut, penalized prescribed-balance coloring) that are not additive parameters and that the partition-oracle estimators do not address. Second, the object that suffices is the family of *unconstrained energies*, not the histogram; the classical equivalences reviewed above required microcanonical energies for the same conclusion. For bounded treewidth, Theorem 15 compiles the explicit local oracle of Edelman et al. (2011) into a fixed-weight RNI-MPNN on the raw graph; this is a representational compilation, not a faster partition oracle. The matching lower bounds of Remark 7 and the expander separation of Proposition 8 have no counterpart in the partition-oracle literature, whose guarantees are one-sided. Conversely, we claim no algorithmic improvement over partition oracles: the encoder has the same query scale, and Theorem 16 shows that $\Omega(\log K)$ local depth can be necessary for bounded-state RNI networks with Lipschitz mean readout that must work uniformly over disjoint-union repetitions.

Size transfer and graph-learning generalization. In the GNN literature, *size transferability* compares one fixed model on graphs or discretizations of different orders (Le & Jegelka, 2023); the cited graphop theorem compares regular discretizations of one common limiting operator. By contrast, *size generalization* commonly means learning on smaller graphs and predicting larger ones (Yehudai et al., 2021). These notions are distinct from statistical empirical-to-population generalization, as

studied for attributed DIDM and graphop representations by Rauchwerger et al. (2025); Amran et al. (2026). Section 5 is target-side: it bounds normalized optimum values, and then gives a deterministic generalization certificate for a separately quotient-stable predictor. It neither proves stability of an arbitrary GNN nor supplies an optimization algorithm. A further relevant line is learning-based combinatorial optimization, where relaxation losses train networks to output fractional assignments whose expected objective is then rounded (Karalias & Loukas, 2020; Schuetz et al., 2022); the normalized task values studied here are exactly the targets of such pipelines, and Corollary 13 states the corresponding target-side rounding guarantee on decomposable graphs.

Expressivity and symmetry breaking. Anonymous message passing is bounded by the 1-WL test (Xu et al., 2019; Morris et al., 2019); random node initialization and unique identifiers restore universality at bounded order (Abboud et al., 2021; Sato et al., 2021; Loukas, 2020). We use this bounded-order universality to compile a local bounded-treewidth partition oracle (Edelman et al., 2011) into a raw-graph RNI-MPNN (Theorem 15); it does not by itself imply the variable-size result. General depth–width lower bounds for message passing are known (Loukas, 2020); Theorem 16 gives a task-specific $\Omega(\log K)$ obstruction for bounded-state RNI networks with Lipschitz mean readout that must work uniformly over disjoint-union repetitions.

Accordingly, the paper does not replace these richer topologies. Its new representation statements are restricted to finite-state pairwise task values and, for the raw-graph neural realization, bounded-treewidth inputs.

C Extensions and omitted details

This appendix collects the material the main text defers: worked quotient computations, the comparison with richer limit topologies, effective encoders, per-section deferred discussion, the geometry of the convexification error, and the experimental protocol. Each subsection opens from the main-text passage that points to it.

C.1 Quotient sets in three simple examples

A single edge. Assign states 1 and 2 to the two endpoints of K_2. Then $p_1 = p_2 = 1/2$ and $W_{12} = W_{21} = 1/2$: the two off-diagonal entries record the two orientations of the same undirected edge. This example also explains the factor $1/2$ that appears in undirected edge energies.

Edgeless graphs. For the edgeless graph I_n, the quotient set is the denominator-n grid of points $(p, 0)$ and therefore

$$\mathcal{S}_q(I_n) \longrightarrow \Delta_q \times \{0\}. \tag{32}$$

Cycles. For a binary assignment on C_n, if j vertices receive state 1 in r cyclic runs, then $(p_1, W_{12}) = (j/n, 2r/n)$. Hence $\mathcal{S}_2(C_n)$ projects to $\{(j/n, 2r/n) : 1 \le r \le \min(j, n-j)\} \cup \{(0,0), (1,0)\}$ and converges to $\{(p_1, W_{12}) : W_{12} \le 2\min(p_1, 1-p_1)\}$, as shown in Figure 2(d).

Perfect matching. If M_n is a perfect matching on an even number n of vertices, then

$$\mathcal{S}_q(M_n) \longrightarrow \{(p, W) : p \in \Delta_q,\ W \ge 0,\ W = W^\top,\ W\mathbf{1} = p\}, \tag{33}$$

where $\mathbf{1}$ is the all-ones vector; rational symmetric matrices with the displayed row sums are realized by choosing colored edge-type counts. The two examples share limiting state proportions but not edge weights, so state proportions alone cannot preserve a pairwise interaction.

C.2 Topological comparison of quotient space

First-moment factor of local–global and action profiles. The quotient is deliberately coarser than a colored rooted-neighborhood law. Let O be a uniformly sampled vertex, let $L_1^{G,c}$ be the law of its colored radius-one neighborhood, and let

$$\mathcal{Q}_{q,1}(G) := \{L_1^{G,c} : c : V(G) \to [q]\} \tag{34}$$

be the radius-one local–global coordinate. A continuous linear map T_q reads from such a law

$$p_i = \Pr(c(O) = i), \qquad W_{ij} = \mathbb{E}\big[\mathbf{1}_{\{c(O)=i\}} \#\{u \sim O : c(u) = j\}\big]. \tag{35}$$

Consequently,

$$\mathcal{S}_q(G) = T_q(\mathcal{Q}_{q,1}(G)). \tag{36}$$

The same map appears inside graphop action convergence. Regard the unnormalized adjacency matrix A_G as an operator on the uniform vertex probability space and put $v_i := \mathbf{1}_{\{c=i\}}$. The action observation records the joint law of $(v_1, \dots, v_q, A_G v_1, \dots, A_G v_q)$, while the quotient keeps only

$$p_i = \mathbb{E}[v_i(O)], \qquad W_{ij} = \mathbb{E}[v_i(O)(A_G v_j)(O)]. \tag{37}$$

Equation (37) is a pointwise moment identity from a hard one-hot observation to one quotient; it does not assert continuity of the hard slice for convergent action profiles.

C.3 Exact computations on bounded treewidth and explicit training sets

Proposition 17 (Exact quotient computation on bounded treewidth)**.** *Fix $q \geq 2$, D, t, and $\rho > 0$. Let $\mathcal{T}_{D,t}$ be the class of D-bounded graphs of treewidth at most t, and let $\mathcal{N}_\rho$ be a finite ρ-net of $\{\theta : \|\theta\|_* \leq 1\}$ in the norm $\|\cdot\|_*$. Given $G \in \mathcal{T}_{D,t}$ on n vertices together with a width-t tree decomposition, the full quotient set $\mathcal{S}_q(G)$ can be listed exactly by dynamic programming. Consequently, on $\mathcal{T}_{D,t}$: (i) the encoder $\widehat{R} = R$ of Corollary 14 is realizable with $\zeta = 0$ in polynomial time for fixed (q, t, D); (ii) every probe energy $\gamma_\theta(G)$ in (11) is computable in the same time; and (iii) for all $G, G' \in \mathcal{T}_{D,t}$ with orders n, n',*

$$\delta_q(G, G') \;\leq\; \max_{\theta \in \mathcal{N}_\rho} |\gamma_\theta(G) - \gamma_\theta(G')| + 2(1+D)\rho + c_{q,D,t}\Big(n^{-1/2} + n'^{-1/2}\Big), \tag{38}$$

where $c_{q,D,t} := 8\sqrt{(1+D) r_q (t+1) D} + 2(1+D) r_q$.

Explicit training sets on decomposable classes. Corollary 14 obtains the training set $\mathcal{B}_r$ from compactness, which bounds neither its size nor the order $n_{\max}$ of its members. On classes with an explicit decomposition both become explicit, because Theorem 5 reduces every quotient set to a convex combination of the finitely many quotient sets of connected graphs of order at most K.

Proposition 18 (Explicit training sets)**.** *Let $\mathcal{C}$ be a class of D-bounded graphs such that, for every $\varepsilon > 0$, deleting at most εn edges of any $G \in \mathcal{C}$ leaves components of order at most $K(\varepsilon)$. Fix q and $r > 0$, put $\varepsilon := r/8$, $K := K(\varepsilon)$, let $\mathcal{T}_K$ be the set of isomorphism types of connected D-bounded graphs of order at most K, $N_K := |\mathcal{T}_K| \leq K\, 2^{K(K-1)/2}$, and*

$$n_0 := \left\lceil \frac{8(1+D)K\,(r_q + N_K)}{r} \right\rceil, \qquad n_{\max} := n_0 + N_K K. \tag{39}$$

Let $\mathcal{B}_r$ be the (finite) set of all disjoint unions of graphs from $\mathcal{T}_K$ with at most $n_{\max}$ vertices. Then every $G \in \mathcal{C}$ satisfies $d_{\mathrm{H}}(\mathcal{S}_q(G), \mathcal{S}_q(B)) \leq r$ for some $B \in \mathcal{B}_r$. For a task tuple with state spaces $q_1, \dots, q_s$ the same $\mathcal{B}_r$ works with r_q replaced by $\max_j r_{q_j}$, so Corollary 14 holds on $\mathcal{C}$ with the explicit $n_{\max}$ of (39).

Every member of $\mathcal{B}_r$ has treewidth at most $K - 1$, so its quotient set is computable by Proposition 17 with this fixed bound. For the specialization $\mathcal{C} = \mathcal{T}_{D,t}$, restrict $\mathcal{T}_K$ to the types of treewidth at most t; the proof is unchanged, and then $\mathcal{B}_r \subset \mathcal{T}_{D,t}$. The constant N_K is crude—for $D = 2$ the connected graphs of order at most K are paths and cycles, so $N_K \leq 2K$—and the proof (Appendix E.4.4) loses a factor N_K only in rounding the component-type proportions to a common denominator.

C.4 Formal statements for Section 5

The informal statements of Section 5 compress the following.

Corollary 19 (Rounding fractional quotient targets; formal)**.** *Suppose deleting at most εn edges from a D-bounded G leaves components of order at most K, and let $z^* \in \operatorname{conv} \mathcal{S}_q(G)$ be any fractional target. Then there is an assignment $c : V(G) \to [q]$ with*

$$d_q\big(G/c,\; z^*\big) \;\leq\; 4\varepsilon + \frac{2(1+D) r_q K}{n}, \tag{40}$$

and hence $F(G/c) \le F(z^*) + L\big(4\varepsilon + 2(1+D)r_q K/n\big)$ *for every* L*-Lipschitz* $F : \mathsf{Y}_{D,q} \to \mathbb{R}$*. On hyperfinite classes the attainable-in-expectation and attainable-by-one-assignment task values therefore coincide asymptotically.*

Proof. The distance claim is Theorem 5, applied to $z^* \in \operatorname{conv} \mathcal{S}_q(G)$; the task bound is the L-Lipschitz inequality for F. □

The raw-graph theorem uses the following compilation step. It is stated separately because bounded-order universality alone does not justify a single network on graphs of arbitrary order.

Lemma 20 (Compiling a bounded local algorithm)**.** *Fix* D, Q, P *and* $0 < B < \infty$*. Let* $\mathcal{A}$ *be a comparison-based rooted local algorithm, starting at its root, makes at most* Q *adjacency-list queries in a* D*-bounded graph and queries only vertices reached from that root. Suppose it uses two independent iid* $U[0,1]$ *ranks per vertex—one for its randomized decisions and one to break every BFS or enumeration tie, including loops over discovered vertex sets—and returns an isomorphism-equivariant vector* $a_G(v) \in [-B, B]^P$*. For every* $\xi, \zeta > 0$ *there is a finite-width RNI-MPNN of depth* $O(Q)$*, with parameters independent of graph order, whose root outputs* $\widehat{a}_G(v)$ *satisfy, under a coupling with the same ranks,*

$$\Pr\left[\left\|\frac{1}{n}\sum_v \widehat{a}_G(v) - \frac{1}{n}\sum_v a_G(v)\right\|_\infty \le \xi\right] \ge 1 - \zeta \tag{41}$$

for every finite input graph G*.*

C.5 Architecture behind the raw-graph encoder

The construction in Theorem 15 starts from iid ranks, which seed a consistent local partition oracle; at a root v, the network reconstructs its oracle part $C(v)$ and returns $(\gamma_\theta(C(v)))_{\theta\in\Theta}$. Vertex-mean pooling is not learned: for the partitioned graph H it obeys the exact identity

$$\frac{1}{n}\sum_{v\in V(G)} \gamma_\theta(C(v)) = \sum_{C\in H} \frac{|C|}{n}\gamma_\theta(C) = \gamma_\theta(H). \tag{42}$$

The deleted edges then cost at most 2α. Independent channels and a coordinatewise median amplify the oracle's constant success probability to $1-\delta$. Define the exact vector $\gamma_\Theta(X) := (\gamma_\theta(X))_{\theta\in\Theta}$. Every 1-Lipschitz head $\psi : \ell^\infty(\Theta) \to \mathbb{R}$ obeys $|\psi(\gamma_\Theta(G)) - \psi(\gamma_\Theta(G'))| \le \delta_q(G, G')$. For any fixed pair, the theorem and a union bound add the corresponding approximation errors with high probability; no uniform-over-graphs claim is made.

C.6 Deferred material for Section 2

The partition limit space. 'Hyperspaces' are spaces of subsets of another space; $\mathcal{K}(Y)$ denotes the hyperspace of nonempty compact subsets of a compact metric space Y. The *finite-graph partition state, ambient product, and finite-realizable limit space* of Section 2 are, respectively,

$$\Theta_D^{\mathrm{part}}(G) := \big(\mathcal{S}_q(G)\big)_{q\ge1}, \quad \mathfrak{A}_D^{\mathrm{part}} := \prod_{q\ge1} \mathcal{K}(\mathsf{Y}_{D,q}), \quad \mathfrak{X}_D^{\mathrm{part}} := \overline{\{\Theta_D^{\mathrm{part}}(G) : \Delta(G) \le D\}}. \tag{43}$$

For $x \in \mathfrak{X}_D^{\mathrm{part}}$, write $\mathcal{S}_q(x)$ for its qth coordinate. The metric

$$d_D^{\mathrm{part}}(x,y) := \sum_{q\ge1} 2^{-q} \min\{1, d_{\mathrm{H}}(\mathcal{S}_q(x), \mathcal{S}_q(y))\} \tag{44}$$

induces the product topology (Figure 2(d)); Appendix C.2 locates quotients inside the richer local–global (Hatami et al., 2014) and action (Backhausz & Szegedy, 2022) topologies. The packaging exists so that limits never have to be assumed (proof in Appendix E.1.1).

Theorem 21 (Compactness)**.** *The ambient space* $\mathfrak{A}_D^{\mathrm{part}}$ *and the finite-realizable space* $\mathfrak{X}_D^{\mathrm{part}}$ *are compact metrizable. Every sequence of finite* D*-bounded graphs has a partition-convergent subsequence.*

Every construction in this paper can therefore extract convergent subsequences with no structural hypothesis on the graphs.

One coordinate per task. Permutation invariance is restored at the set level: quotient sets are unions of permutation orbits (worked examples in Appendix C.1). A fixed finite-state task needs only one quotient coordinate. If $Q \geq q$ and $s : [Q] \to [q]$ is a surjection, aggregating node and edge weights over the fibers of s maps $\mathcal{S}_Q(G)$ onto $\mathcal{S}_q(G)$. The countable family is kept only to cover all finite state spaces at once. Color names add no information: each quotient set is invariant under simultaneously permuting p and the rows and columns of W, and because permutations act by ℓ^1-isometries and quotient sets are unions of permutation orbits, the Hausdorff distance computed from d_q coincides with the distance between the label-free orbit quotients.

C.7 Sharpness and deferred material for Section 3

Formal statements behind Remark 7 and Proposition 8. Let $M_{2m} = mK_2$ be a perfect matching, put $N_q := \binom{q+1}{2}$ and $s_* := \min\{m, \lfloor N_q/2 \rfloor\}$, and define

$$R_{q,m} := \frac{2s_*(N_q - s_*)}{N_q m}. \tag{45}$$

For every $q \geq 2$ and $m \geq 1$,

$$R_{q,m} \leq \kappa_q(M_{2m}) \leq 2R_{q,m}, \qquad \kappa_q(M_{2m}) = \Theta\big(\min\{1, q^2/m\}\big). \tag{46}$$

If X is a connected D-regular graph on K vertices with edge expansion at least $h_0 > 0$, meaning $|\partial S| \geq h_0 \min\{|S|, |V(X) \setminus S|\}$ for every $S \subseteq V(X)$, where ∂S is the set of edges with exactly one endpoint in S, and $G = mX$, then, for $n = mK$,

$$\kappa_2(G) \geq \min\{\tfrac{1}{2}, \tfrac{h_0}{2}\}\frac{K}{n}. \tag{47}$$

Finally, every D-bounded n-vertex graph for which deleting at most εn edges leaves components of order at most K satisfies

$$\kappa_{q,2}(G) \leq 4\varepsilon + \sqrt{1 + D^2}\,\min\left\{\sqrt{r_q}\frac{K}{n}, \sqrt{\frac{K}{n}}\right\}. \tag{48}$$

Moreover, $\kappa_{q,2}(M_{2m}) = \Theta(\min\{m^{-1/2}, \sqrt{N_q}/m\})$.

For the separation, let G_n be a finite graph and $H_n := G_n \sqcup G_n$. For every q,

$$\operatorname{conv} \mathcal{S}_q(H_n) = \operatorname{conv} \mathcal{S}_q(G_n). \tag{49}$$

If G_n has edge expansion at least $h > 0$, meaning $|\partial S| \geq h \min\{|S|, |V(G_n) \setminus S|\}$ for every vertex set S, where ∂S contains the edges with exactly one endpoint in S, then the continuous balanced-cut value

$$\Phi(G) := \min_{(p,W)\in\mathcal{S}_2(G)} \big(|p_1 - \tfrac{1}{2}| + W_{12}\big) \tag{50}$$

satisfies $\Phi(H_n) = 0$ and $\Phi(G_n) \geq \min\{1/2, h/2\}$.

The Shapley–Folkman mechanism. Put $r_q := (q-1) + \frac{q(q+1)}{2} - 1$. For each fixed graph, this is a common dimension bound after translating its component quotient sets: p has fixed sum one, W is symmetric, and the total mass of W is fixed on each component. In particular, $r_1 = 0$. The proof of Theorem 5 uses the following standard weighted form of the Shapley–Folkman lemma (Starr, 1969): if $A_1, \ldots, A_s \subset \mathbb{R}^m$ and $\lambda_j \geq 0$, then every point of $\sum_j \lambda_j \operatorname{conv} A_j$ can be written with at most m summands still chosen from $\operatorname{conv} A_j$ and every other summand chosen from A_j.

Proof sketch and a concrete rate. The proof writes the quotient set of the decomposed graph as a vertex-weighted Minkowski average of its component quotient sets and applies a weighted Shapley–Folkman lemma in dimension r_q; comparing the original and deleted graphs contributes 4ε across the two Hausdorff legs (Appendix E.2.3). For P_n or C_n, deleting $n/K + 1$ edges leaves paths of order K, so the bound is $4(\frac{1}{K} + \frac{1}{n}) + \frac{6r_qK}{n}$; taking $K = \lceil\sqrt{n}\rceil$ gives an explicit $O_q(n^{-1/2})$ rate. Figure 3 shows the mechanism on mK_4.

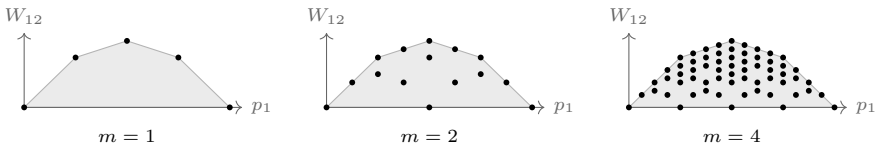


Figure 3: The (p_1, W_{12})-projection of $\mathcal{S}_2(mK_4)$ (dots, exact) inside $\operatorname{conv}\mathcal{S}_2(K_4)$ (shaded): the m-fold Minkowski average fills the hull at rate $\Theta(K/n) = \Theta(1/m)$ (Theorem 5 and Remark 7).

Why the state-space dimension is necessary. For the matching M_{2m}, an assignment is exactly a denominator-m probability type on the $N_q = \binom{q+1}{2}$ unordered color pairs. The linear map from this type to (p, W) distorts ℓ^1 distance by a factor between one and two. The exact simplex-grid covering radius is $R_{q,m}$ from (45), proving (46). This is a finite type-lattice calculation, not an asymptotic appeal to Shapley–Folkman; related quantization problems are studied by Reznik (2011).

The Euclidean proof behaves differently. Independently rounding the component selected in each convex combination makes cross terms vanish, so the squared error is controlled by $\sum_j (|C_j|/n)^2 \le K/n$. Applying this only to the at most r_q exceptional Shapley–Folkman components gives the other branch of (48). The matching type lattice attains both regimes. Thus the ℓ^1 and ℓ^2 metrics give the same fixed-q convergence but genuinely different uniform finite-q bounds.

Task-specific dimension. The dimension factor $r_q = \Theta(q^2)$ in Theorem 5 is the price of controlling *every* coordinate of the quotient at once. A task that reads only k coordinates pays for k.

Proposition 22 (Projected convexification)**.** *Let $\Pi : \mathbb{R}^q \times \mathbb{R}^{q\times q} \to \mathbb{R}^k$ be linear and nonexpansive, $\|\Pi z\|_1 \le \|z\|_1$ (for instance the restriction to any k coordinates of (p, W)). Under the hypothesis of Theorem 5,*

$$d_{\mathrm{H}}\big(\Pi\mathcal{S}_q(G), \operatorname{conv}\Pi\mathcal{S}_q(G)\big) \;\le\; 4\varepsilon + \frac{2(1+D)\,k\,K}{n}, \tag{51}$$

with d_{H} taken in the ℓ^1 norm of $\mathbb{R}^k$. Consequently, for every L-Lipschitz $F : \mathbb{R}^k \to \mathbb{R}$, $\big|\min_{z\in\mathcal{S}_q(G)} F(\Pi z) - \min_{z\in\operatorname{conv}\mathcal{S}_q(G)} F(\Pi z)\big| \le L\big(4\varepsilon + 2(1+D)kK/n\big)$, and the same holds for maxima. The factor k is sharp before saturation: for M_{2m} and every $k \le \min\{N_q - 1, 2m\}$, some restriction Π to k upper-triangular W-coordinates satisfies the following bound; here $n = 2m$ and the component scale is $K = 2$:

$$d_{\mathrm{H}}\big(\Pi\mathcal{S}_q(M_{2m}), \operatorname{conv}\Pi\mathcal{S}_q(M_{2m})\big) \ge \frac{k}{4m} = \frac{k}{4}\frac{K}{n}. \tag{52}$$

The convexification error κ_q in Theorem 11 and Proposition 9 may therefore be replaced, for any task that depends on z only through Πz, by $\kappa_\Pi(G) := d_{\mathrm{H}}(\Pi\mathcal{S}_q(G), \operatorname{conv}\Pi\mathcal{S}_q(G))$, which obeys (51). The balanced-cut value Φ of (50) reads $k = 2$ coordinates, the balance–defect tradeoffs of Appendix C.8 read $k = q + 1$, and a homogeneous energy reads $k = 1$; for an affine energy no convexification error enters its optimum, even though κ_Π itself need not vanish (Proposition 4). The proof, including (52), is in Appendix E.2.6.

Locating Theorem 5. Two qualifiers locate the theorem: Borgs et al. (2018, Example 3.3.7) observed the disjoint-union case qualitatively, and the theorem supplies the general finite-graph rate, whose q- and K/n-dependences are matched separately by Remark 7; the optimal coefficient of the ε term remains open; Hatami et al. (2014, Theorem 10.5) give a qualitative counterpart whose input is a local limit rather than a finite decomposition. One qualifier on Corollary 6: the classical equivalences (Borgs et al., 2012; 2018) use *microcanonical* energies—constrained optima not produced by our feature map—and are silent at bounded degree (see "Normalized quotients and ground-state energies" in Appendix B). In their regime they show that the unconstrained family is insufficient in general; the corollary identifies hyperfiniteness as a broad regime where the implementable family suffices, and Corollary 23 shows that the hypothesis cannot be dropped without replacement.

Growing-component instances. Taking $X = K_{D+1}$ (where $h_0 \ge 1$) gives $\frac{1}{2}K/n$ with $K = D+1$; D-regular expanders with h_0 bounded below exist for every fixed $D \ge 3$ and arbitrarily large K

(Bollobás, 1988), giving $c(h_0)K/n$ with K unbounded. The linear dependence on K/n in (6) is therefore sharp even when $q = 2$; the matching clause of Remark 7 separately makes the $r_q = \Theta(q^2)$ dependence sharp when $K = 2$.

The smallest separating pair. On expanders two graphs can be indistinguishable to every energy-based feature yet differ by a constant on a natural task; no amount of scalar-energy information extends Φ continuously, and Appendix C.10 works out the smallest such pair, K_4 versus $K_4 \sqcup K_4$, with its geometry.

The Lagrangian bridge as a duality-gap estimate. Each Lagrangian approximation in Proposition 9 is an unconstrained energy shifted by a constant, so constrained optima are recoverable from implementable measurements up to the convexification error and no further: the classical Shapley–Folkman duality-gap estimate (Aubin & Ekeland, 1976; Udell & Boyd, 2016) instantiated on quotient sets and calibrated on both sides by Theorem 5 and Proposition 8. Appendix C.10 gives the in-depth discussion and Appendices E.2.3, E.2.7, and E.2.9 the proofs.

The Potts family. Take $a = 0$ and $B = I_q$. Since $\frac{1}{2}\langle I_q, W(G,c)\rangle = \frac{1}{n}|\{uv \in E(G) : c(u) = c(v)\}|$, the energy (4) is the density of monochromatic edges—the Hamiltonian of the q-state Potts antiferromagnet. Its ground-state energy is 0 exactly when G admits a proper q-coloring, so the model detects $\chi(G) \le q$, and the zero-temperature partition function is the chromatic polynomial at q (Fortuin & Kasteleyn, 1972; Sokal, 2005). The example is the Hamiltonian $H(c) = -J\sum_{uv\in E(G)} \delta_{c(u)c(v)}$ of the q-state Potts model, where δ_{ij} is the Kronecker delta, at coupling $J = -1$ and zero external field (Potts, 1952; Wu, 1982); a nonzero field $h \in \mathbb{R}^q$ is recovered as the unary cost $a = -h$. Since the bichromatic density is $\bar{d}/2$ minus the monochromatic density, the same ground state solves MAX-q-CUT (Frieze & Jerrum, 1997); reversing the sign to $B = -I_q$ gives the ferromagnet, whose ground states are the monochromatic assignments.

Necessity of the hyperfinite hypothesis. The hyperfiniteness hypothesis of Corollary 6 cannot simply be dropped.

Corollary 23 (Scalar energies are not complete in general)**.** *There is a growing D-bounded sequence (H_i) such that every homogeneous finite-state unary–pairwise ground-state energy converges, yet (H_i) is not partition convergent.*

The proof (Appendix E.2.8) alternates a partition-convergent expander sequence with its own two-copy disjoint unions. Whether the quantifier over all q in Corollary 6 can be bounded—for example, whether the two-state energies alone suffice—remains open.

The Lagrangian bridge in full. The formal version of Proposition 9 reads as follows.

Proposition 24 (Lagrangian bridge; formal)**.** *Let G be a finite D-bounded graph, B a symmetric interaction matrix, $a \in \Delta_q$ a target balance, and $L \ge 0$ a penalty level. Write*

$$\begin{aligned} \mathsf{m}_{B,L,a}(G) &:= \min_{(p,W)\in\mathcal{S}_q(G)} \Big(\tfrac{1}{2}\langle B, W\rangle + L\|p - a\|_1\Big), \\ \mathsf{g}_{B,a}(u) &:= \min_{(p,W)\in\mathcal{S}_q(G)} \Big(\tfrac{1}{2}\langle B, W\rangle + \langle u, p - a\rangle\Big). \end{aligned} \tag{53}$$

for the penalized microcanonical value and its Lagrangian dual at multiplier $u \in \mathbb{R}^q$; each $\mathsf{g}_{B,a}(u)$ is an unconstrained unary–pairwise ground-state energy shifted by the constant $\langle u, a\rangle$. Then

$$0 \;\le\; \mathsf{m}_{B,L,a}(G) - \max_{\|u\|_\infty \le L} \mathsf{g}_{B,a}(u) \;\le\; \max\{L, \tfrac{1}{2}\|B\|_{\max}\}\, \kappa_q(G). \tag{54}$$

The content of the proposition beyond the classical duality-gap estimates (Aubin & Ekeland, 1976; Udell & Boyd, 2016) is the instantiation to quotient-set geometry: the bound is in the quotient metric, uniform over (B, a) at fixed Lipschitz level, and calibrated on both sides. On any (ε, K)-decomposable graph, Theorem 5 makes the right-hand side explicit, so unconstrained energies recover every penalized microcanonical value up to that error; this is the precise sense in which Corollary 6 strengthens microcanonical-energy characterizations on hyperfinite classes (Section 1.1). Conversely, on an expander the gap is real: with $q = 2$, B the off-diagonal indicator, $a = (\frac{1}{2}, \frac{1}{2})$,

and $L = \frac{1}{2}$, the primal value is $\Phi(G) \geq \min\{1/2, h/2\}$ by Proposition 8 while the dual is 0, so (54) certifies $\kappa_2(G) \geq \min\{1, h\}$ for every graph with edge expansion h—an $\Omega(1)$ convexification error complementing the $O(K/n)$ regime.

C.8 Deferred material for Section 4

Relaxed edge orthogonality. The finite-state calculus is no restriction: continuous spin spaces enter through finite nets at a cost controlled by the degree bound alone (Proposition 10). A concrete instance is the unit sphere with repulsive interaction $h(x, y) = \langle x, y\rangle^2$, capturing the orthogonal representations behind the Lovász ϑ function (Lovász, 1979). The relaxed edge-orthogonality energy is

$$\varepsilon_d^{\perp}(G) := \frac{1}{n} \min_{x_v \in S^{d-1}} \sum_{uv \in E(G)} \langle x_u, x_v\rangle^2. \tag{55}$$

On a finite graph it vanishes exactly when adjacent vertices can be assigned orthogonal real unit vectors in dimension d. Thus ordinary coloring and edge-orthogonal vector assignments are two zero-temperature repulsive spin models in the same quotient calculus.

C.9 Deferred material for Section 5

A finite probe net. Theorem 11 turns a family of scalar labels into a certificate for every continuous quotient task on decomposable graphs. It also has a finite version. If $\mathcal{N}_\rho$ is a finite ρ-net, in the norm $\|\cdot\|_*$, of the unit ball $\{\theta : \|\theta\|_* \leq 1\}$, then

$$\eta_q(G, G') \leq \max_{\theta \in \mathcal{N}_\rho} |\gamma_\theta(G) - \gamma_\theta(G')| + 2(1 + D)\rho. \tag{56}$$

Hence, at any fixed accuracy, a finite vector of normalized pairwise ground-state energies approximately controls every Lipschitz q-state quotient query with a prescribed Lipschitz constant, and each continuous query through its own modulus of continuity. The vector is independent of graph order, although computing its coordinates can be hard. Support-function duality identifies η_q with the Hausdorff distance between the two convexified quotient sets.

Training datasets and probe covers. In general the training dataset of Corollary 14 is existential: compactness supplies neither an efficient construction nor a useful bound on $n_{\max}$. On classes with $\kappa_{q_j} \leq \bar{\kappa}$, equation (56) turns a b-cover by finite probe vectors into a quotient cover of radius $b + 2(1 + D)\rho + 2\bar{\kappa}$; thus Theorem 5 connects the certificate to the small-component regime.

Where the representational ceiling applies. The decomposability hypothesis does real work for *constrained* tasks: relaxation-based learned combinatorial optimization is often evaluated on random-regular benchmarks (Schuetz et al., 2022). On expander instances, Proposition 9 certifies an $\Omega(1)$ gap between constrained and unconstrained optima on graph *pairs* such as $(G, G \sqcup G)$. Two qualifications keep this from being a statement about the benchmarks themselves. First, the tasks those pipelines optimize (MAX-CUT, independence) are affine, so energy-based features represent them exactly on every graph (Proposition 4) and their transfer is governed by η_q alone (Corollary 12); the obstruction there is optimization, not representation. Second, the ceiling is a worst case over pairs with equal hulls and different constrained values; a *distribution* over connected random D-regular graphs need not contain such a pair. Almost-sure LD convergence of random regular graphs is posed as a conjecture by Borgs et al. (2017), not used here. The ceiling bites when training and test laws place mass on graphs with equal hulls and different constrained values—e.g. connected and disconnected instances mixed, or families with different expansion at equal energies—and the guarantees of Section 5 for constrained tasks target planar, lattice-like, and minor-free families instead.

Generalization for linear probe readouts. Theorem 26 covers every Lipschitz predictor; the practical class used in Appendix C.11 admits a dimension-mild bound (proof in Appendix E.4.8, with the proof of Theorem 26).

Proposition 25 (Generalization for linear probe readouts)**.** *Let $B, W_1 > 0$, $\delta \in (0, 1)$, and $m \in \mathbb{N}$. Let μ be a distribution on finite D-bounded graphs with measurable labels $y(G) \in [0, B]$, draw $G_1, \ldots, G_m$ independently from μ, and write $\widehat{\mathbb{E}}_m g := m^{-1} \sum_{i=1}^m g(G_i)$. Let $\mathcal{N}$ be a set of P probes*

satisfying $\|\theta\|_* \le 1$. *For the class of* ℓ^1*-bounded linear readouts* $\mathcal{F}_{\mathrm{lin}} := \{G \mapsto \langle w, (\gamma_\theta(G))_{\theta\in\mathcal{N}}\rangle + b : \|w\|_1 \le W_1,\ |b| \le B\}$, *with probability at least* $1-\delta$,

$$\sup_{f\in\mathcal{F}_{\mathrm{lin}}} \left| \mathbb{E}\,|f-y| - \widehat{\mathbb{E}}_m|f-y| \right| \ \le\ 2W_1(1+D)\sqrt{\tfrac{2\ln(2P)}{m}} + \tfrac{2B}{\sqrt{m}} + 3\big(W_1(1+D)+2B\big)\sqrt{\tfrac{\ln(4/\delta)}{2m}}. \quad (57)$$

Generalization over the quotient state. Section 5 summarizes the following covering-number bound in words; the proof is in Appendix E.4.8.

For fixed state sizes $q_1, \dots, q_s$, write $\mathfrak{X}_s := \prod_{j=1}^s \mathcal{K}(\mathsf{Y}_{D,q_j})$, equipped with the max metric d_R of Section 5; thus $R(G) = (\mathcal{S}_{q_1}(G), \dots, \mathcal{S}_{q_s}(G)) \in \mathfrak{X}_s$.

Theorem 26 (Generalization over the quotient state)**.** *Let* $B, \Lambda > 0$, $\delta \in (0,1)$, *and* $m \in \mathbb{N}$. *Let* μ *be any distribution on finite* D*-bounded graphs with measurable labels* $y(G) \in [0,B]$, *and draw* $G_1, \dots, G_m$ *independently from* μ. *With probability at least* $1-\delta$, *uniformly over all* Λ*-Lipschitz* $f : \mathfrak{X}_s \to [0,B]$,

$$\begin{aligned} &\left| \mathbb{E}\,|f(R(G)) - y(G)| - \tfrac{1}{m}\textstyle\sum_i |f(R(G_i)) - y(G_i)| \right| \\ &\qquad \le \inf_{0<\rho\le 1+D} \left[6\Lambda\rho + B\sqrt{\tfrac{\mathsf{N}_{\mathfrak{X}}(\rho)\ln(3+\frac{B}{\Lambda\rho}) + \ln\frac{2}{\delta}}{2m}} \right]. \end{aligned} \quad (58)$$

where $\mathsf{N}_{\mathfrak{X}}(\rho) \le \prod_{j\le s} 2^{(3(1+D)/\rho)^{q_j+q_j^2}}$ *is the* ρ*-covering number of* $\mathfrak{X}_s$.

Generalization over the probe representation. The hyperspace covering number in Theorem 26 is doubly exponential in q^2. The finite probe representation of (56) lives in a cube of dimension $P = |\mathcal{N}|$, whose covering number is only polynomial in $1/\rho$ with exponent P, and by (12) the probe map $\gamma : G \mapsto (\gamma_\theta(G))_{\theta\in\mathcal{N}}$ is 1-Lipschitz from $(\mathcal{K}(\mathsf{Y}_{D,q}), d_{\mathrm{H}})$ to $(\mathbb{R}^P, \|\cdot\|_\infty)$, so every Lipschitz predictor of the probe vector is a Lipschitz predictor of the quotient state.

Proposition 27 (Generalization over the probe representation)**.** *Let* $\mathcal{N}$ *be a finite set of* P *probes with* $\|\theta\|_* \le 1$, *so that* $\gamma(G) \in [-(1+D), 1+D]^P$. *In the setting of Theorem 26, with probability at least* $1-\delta$, *uniformly over all* $f : [-(1+D), 1+D]^P \to [0,B]$ *that are* Λ*-Lipschitz for* $\|\cdot\|_\infty$,

$$\begin{aligned} &\left| \mathbb{E}\,|f(\gamma(G)) - y(G)| - \tfrac{1}{m}\textstyle\sum_i |f(\gamma(G_i)) - y(G_i)| \right| \\ &\qquad \le \inf_{0<\rho\le 1+D} \left[6\Lambda\rho + B\sqrt{\tfrac{\mathsf{N}_P(\rho)\ln(3+\frac{B}{\Lambda\rho}) + \ln\frac{2}{\delta}}{2m}} \right]. \end{aligned} \quad (59)$$

where $\mathsf{N}_P(\rho) \le (2(1+D)/\rho)^P$ *is the* ℓ^∞ ρ*-covering number of the probe cube.*

The bound is the analogue of (58) with the hyperspace replaced by the probe cube: the exponent $q + q^2$ inside a double exponential becomes the single exponent P. It is still a covering bound and is slower than Proposition 25 for linear readouts, but it is the operative uniform bound for *nonlinear* readouts of the probe vector, e.g. the heads following Theorem 15. The proof (Appendix E.4.8) is that of Theorem 26 with the state space replaced by the cube.

On the generalization bounds. In Theorem 26, the hyperspace covering number is enormous, exactly as in the graphon-signal analysis it mirrors; the honest summary is that (58) is explicit and distribution-free but astronomically slow, while (57)—the class used in Appendix C.11—is dimension-mild. On classes with $\kappa_q \le \bar{\kappa}$ (Theorem 5), two graphs with the same finite ρ-net probe vector have quotient distance at most $2(1+D)\rho + 2\bar{\kappa}$. Thus any target that is Λ-Lipschitz in quotient Hausdorff distance differs across such a fiber by at most $2\Lambda((1+D)\rho + \bar{\kappa})$. This is the representation-bias term on the stated small-convexification regimes; it is separate from the κ_q-free statistical complexity bounds above.

C.10 When and why convexification is necessary

For readers who think in transferability and generalization bounds, the convexification error κ_q has a fixed address. Decompose the test error of a size-transferring predictor as

$$\text{test error} \ \le\ \underbrace{\text{training error}}_{\text{optimization}} + \underbrace{\text{generalization gap}}_{\text{Theorem 26: } \kappa_q\text{-free}} + \underbrace{\text{representation bias}}_{\text{convexification } \kappa_q \,+\, \text{probe/encoder approximation}}.$$

The reason κ_q is the unavoidable part of the third term is a ceiling on energy features. Each feature coordinate is a probe energy, hence a minimum of an affine functional over $\mathcal{S}_q(G)$; any downstream readout factors through this probe vector and therefore through the convex hull (Proposition 4). Theorem 15 approximates these coordinates on raw bounded-treewidth graphs. The hull is what energy features *are*; the nonconvex set is what tasks *need*; $\kappa_q = d_{\mathrm{H}}(\mathcal{S}_q, \operatorname{conv}\mathcal{S}_q)$ is the gap between the two. Accordingly κ_q enters the separate target-versus-representation comparison whenever the pipeline replaces the state by energy measurements—the right half of (12), the rounding radius (40), the certificate slack in (38)—but not the statistical complexity bound (57), which is explicitly κ_q-free. It also does not enter affine-task bounds stated directly in terms of the energy/hull distance η_q, such as Corollary 12; experiment E1 below distinguishes that theorem-matched distance from a looser raw-set diagnostic.

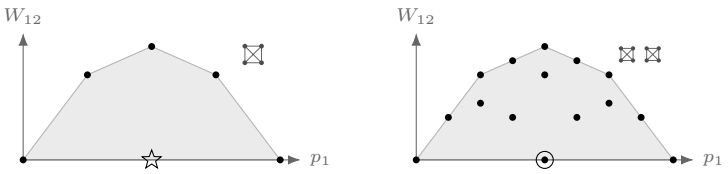
A two-graph example. Let $G = K_4$ and $G' = K_4 \sqcup K_4$. Vertex normalization makes every unary–pairwise ground-state energy of the two graphs *identical*: $\gamma_\theta(G') = \gamma_\theta(G)$ for every probe θ, by (49). So every energy feature map, of any dimension—including any network whose outputs factor through such energies—assigns both graphs the same feature vector, and every readout on it the same prediction. Figure 4 shows the two quotient sets and their common convex hull.

Yet the balanced-cut task (50), the objective of balanced graph partitioning, separates them: $\Phi(G') = 0$ (color one copy state 1, the other state 2), while $\Phi(K_4) = \frac{1}{2}$ (check $k = 0, \dots, 4$ state-1 vertices by hand: $\min\{\frac{1}{2}, \frac{1}{4} + \frac{3}{4}, 0 + 1\} = \frac{1}{2}$). Any predictor of Φ over energy features therefore errs by at least $\frac{1}{4}$ on one of the two graphs—an architecture- and training-independent impossibility of the transferability kind. Proposition 8 shows the same gap $\min\{1/2, h/2\}$ persists along growing expander sequences, so this is not a small-graph artifact, and Proposition 9 certifies it as $\kappa_2 \geq \min\{1, h\}$.

When it is not a problem. Theorem 5 and Remark 7 give the universal upper bound $4\varepsilon + O_D(q^2 K/n)$, with the q^2 and K/n dependences separately matched in the worst case. Thus, after choosing a decomposition with small ε and then sending $n \to \infty$, a finite order-independent energy representation chosen for any prescribed accuracy becomes asymptotically sufficient for the task state. A particular graph may have zero gap, and the decomposition alone does not certify an impossibility. The resulting regime map is a three-level feature hierarchy for size transfer: degree and size statistics fail already on hyperfinite instances (the matched pair C_{120} vs. $40K_3$ in Appendix C.11); probe energies are complete there but hit the K_4-versus-$K_4 \sqcup K_4$ ceiling on expanders; and past the ceiling no scalar feature map suffices—one must carry an approximation of the nonconvex set itself (Proposition 17) or solve the constrained problem per instance. Convexification is thus not an assumption of the pipeline but a sufficient regime certificate: a good decomposition proves when energy features suffice asymptotically, while a nonvanishing measured gap proves when they cannot.

C.11 Experiment details

Families: cycles C_n, paths P_n, $2 \times k$ ladders, random subcubic trees (uniform attachment to a vertex of degree < 3, seeds fixed), and disjoint triangle unions mK_3; all have maximum degree ≤ 3 and treewidth ≤ 2. Training set: 72 graphs with $8 \leq n \leq 30$; test set: 66 graphs with $120 \leq n \leq 360$. The dataset-generation seed is 11; the random-tree seeds are 1000–1009 in training and 1100–1109 in testing. Probes: 45 pairs $\theta = (a, B)$ drawn once from a standard Gaussian and normalized to $\|\theta\|_* = 1$ (seed 7); the probes that reproduce e_2, MaxCut/n, and α/n exactly ($B = I$; $B = -$offdiag; $a = (-1, 0)$, $B = \mathrm{diag}(1, 0)$) are excluded, so the readout must reconstruct the tasks from generic energies. All γ_θ values are exact: chain and cyclic transfer-matrix dynamic programs for paths and cycles, a 4-state column program for ladders, rooted dynamic programs for trees, and enumeration per triangle for mK_3. When `experiments/toy_transfer.py` is run as a script, its main guard first calls a companion self-check that compares the family solvers and closed-form task values with brute force on small instances. The individual experiment functions, and the E3–E9 scripts, do not invoke that check automatically. Readout: one ridge regression per task ($\lambda = 10^{-4}$) on standardized features. Baseline features: $1/n$, edge density $2|E|/n$, and the degree histogram $(\deg = 1, 2, 3)$—identical for C_{120} and $40K_3$. In experiment E1, the supplied routine computes the exact Hausdorff distance $\delta_2(C_n, C_m)$ between the closed-form, nonconvex quotient sets in the full metric d_q, over all 15 pairs from $n, m \in \{31, 40, 61, 80, 101, 120\}$. Using this larger raw-set distance gives the conservative maximum/median ratios .40/.27 for e_2 and MaxCut/n, and .10/.07 for α/n; the maximum is shared by four pairs, including (C_{31}, C_{40}). The bounds in (15)–(16) instead use the



(a) $\mathcal{S}_2(K_4)$: $\star = (\frac{1}{2}, 0)$ in hull only (b) $\mathcal{S}_2(K_4 \sqcup K_4)$: $(\frac{1}{2}, 0)$ attained (circled)

Figure 4: The feature blind spot, in the (p_1, W_{12})-projection with exact coordinates. (a) $\mathcal{S}_2(K_4)$ (dots) and its convex hull (shaded); the balanced zero-cut target $(p_1, W_{12}) = (\frac{1}{2}, 0)$ ($\star$) lies in the hull—as the midpoint of the two monochromatic quotients—but not in the set, and $\Phi(K_4) = \frac{1}{2}$. (b) $\mathcal{S}_2(K_4 \sqcup K_4)$ (dots): the hull is *identical* to (a), so every probe energy agrees on the two graphs, yet the set now contains $(\frac{1}{2}, 0)$ (circled; color one copy state 1, the other state 2), and $\Phi(K_4 \sqcup K_4) = 0$. Energy features see only the shaded region and cannot distinguish (a) from (b); the balanced-cut task lives in the difference between dots and shade.

convex-hull distance η_2. With that theorem-matched denominator, every tested pair containing an odd cycle has nonzero ratio $1/2$, $1/2$, and $1/8$, respectively. Total runtime: under four seconds on one CPU core.

C.12 Additional experiments

The task values and probe energies are exact; only the mesh-based Hausdorff estimate in E6 is numerical. Table 2 collects the transfer experiments.

(E3) Out-of-family transfer and a per-family reference. The split of Appendix C.11 varies only the order within five families, and on those families the task densities are close to per-family constants. Two references calibrate the reported MAE of $5 \cdot 10^{-4}$: a predictor that outputs the *training mean of the test graph's own family* (an oracle that knows the family) has MAE $(0.0072, 0.0517, 0.0122)$, ten to a hundred times larger; and with whole families held out of training—readout fitted on cycles, paths and trees, tested on ladders and mK_3; fitted on ladders, mK_3 and trees, tested on cycles and paths; fitted on the four non-tree families, tested on trees—the probe readout keeps MAE ≤ 0.003 except for the tree/e_2 and tree/MaxCut pairs $(0.011, 0.013)$, while the degree/size baseline errs by 0.10–0.36. The probe readout thus extrapolates across families it never saw, as (56) predicts for tasks that are themselves Lipschitz quotient optima.

(E4) Number of probes. Over ten independent probe draws, the in-family test MAE generally decreases with P: mean $(0.039, 0.038, 0.015)$ at $P = 3$, $(0.011, 0.011, 0.007)$ at $P = 10$, $(0.0027, 0.0028, 0.0014)$ at $P = 20$, $(0.0008, 0.0008, 0.0007)$ at $P = 45$ and $(0.0009, 0.0009, 0.0002)$ at $P = 90$; the worst draw at $P = 45$ is ≤ 0.003 on every task. The single draw reported in Appendix C.11 is therefore typical, and the tasks are recovered from roughly twenty generic energies. The ten probe seeds are 100–109, and the dataset-generation seed is 11.

(E5) The expander ceiling. For random cubic graphs G on $n \in \{8, 10, 12, 14\}$ vertices (five per order, pairing model), exact enumeration gives $\Phi(G) \in [1/6, 1/2]$ in all twenty instances. The routine uses five probes on G and forms the reported self-union features from the component-additivity identity; it does not construct $G \sqcup G$ or reoptimize the probes there. Thus $\Phi(G \sqcup G) = 0$ and $\gamma_\theta(G \sqcup G) = \gamma_\theta(G)$ for every scalar probe are analytical consequences of the self-union construction, not independent numerical observations. Together they imply that any readout of energy-based features must err by at least $\Phi(G)/2 \geq 1/12 \approx 0.083$ on one of the two graphs. The graph seeds are $1000n + s$ for $s = 0, \ldots, 4$, and the probe seed is 7.

(E6) The K/n rate. For mK_4 ($K = 4, n = 4m$), the quotient set is the m-fold Minkowski average of the five points of $\mathcal{S}_2(K_4)$, and a direct calculation gives $d_{\mathrm{H}}(\mathcal{S}_2(mK_4), \operatorname{conv} \mathcal{S}_2(mK_4)) = 7/(2m)$. Hence $m\, d_{\mathrm{H}} = 3.5$ exactly and the rate is $\Theta(K/n)$. The coarse-to-fine grid used by the routine misses the maximizer for $m \geq 8$ and prints 3.492 after rescaling, so it is only a numerical check of this rate. The exact value lies between the lower bound $\frac{1}{2}K/n$ in Remark 7 and the upper bound $2(1 + D)r_2 K/n = 24K/n$ of Theorem 5.

Table 2: Additional experiments (test MAE on 66 graphs with $n \in [120, 360]$ unless stated). E5 and E6 are reported in the text.

	e_2	MaxCut$/n$	α/n
(E1) theorem-matched task-gap/bound ratio	.50	.50	.125
(E2) C_{120} vs. $40K_3$ error: probe readout / degree baseline	0.0000 / 0.275	0.0000 / 0.275	0.0001 / 0.133
(E3) in-family: per-family training mean (oracle family)	0.0072	0.0517	0.0122
held out {ladder, mK_3}: 45 probes / baseline	0.0000 / 0.277	0.0000 / 0.277	0.0008 / 0.103
held out {cycle, path}: 45 probes / baseline	0.0001 / 0.359	0.0001 / 0.359	0.0027 / 0.192
held out {tree}: 45 probes / baseline	0.0108 / 0.249	0.0125 / 0.249	0.0015 / 0.119
(E4) MAE, mean over 10 probe draws, $P = 3$ / 10 / 20 / 45	.039/.011/.0027/.0008	.038/.011/.0028/.0008	.015/.007/.0014/.0007

C.13 Trained models

All learned-model entries are means over model seeds $\{0, 1, 2\}$; E7 and E9 use dataset-generation seed 11. In E7, GIN uses Adam with learning rate $3 \cdot 10^{-3}$ and cosine decay for 1500 epochs; GPS uses full-batch AdamW with learning rate 10^{-3}, weight decay 10^{-5}, and cosine decay for 1500 epochs. Predictions from GIN with RNI average eight fresh feature draws. In E8, cubic graphs use seeds $1000n + s$ for $s = 0, \ldots, 3$, and training uses Adam with learning rate $3 \cdot 10^{-3}$ and cosine decay for 600 epochs. In E9, training uses Adam with learning rate $3 \cdot 10^{-3}$ for 800 epochs; the cubic-graph seeds are 500–504, and each trained model uses one stochastic evaluation draw. Consequently, stochastic rows may vary in the last reported digit across software environments and reruns.

(E7) A trained GIN on the split of Appendix C.11. A four-layer GIN (width 64, sum aggregation, mean readout after every layer, MLP head, Adam, 1500 epochs, standardized targets) is trained by mean squared error on the 72 training graphs ($n \leq 30$) to predict $(e_2, \mathrm{MaxCut}/n, \alpha/n)$, with either constant node features (anonymous) or random node initialization (RNI, eight Gaussian features redrawn at every forward pass; predictions averaged over eight draws). As a stronger pre-existing neural baseline, GPS–RWSE (Rampášek et al., 2022) combines, in each of four width-64 blocks, a local GIN branch with four-head global self-attention. Write A for the adjacency matrix and D_{deg} for the diagonal degree matrix. Its deterministic node input contains the degree one-hot encoding, $1/n$, and the return probabilities $((D_{\mathrm{deg}}^{-1}A)^t_{vv})_{t=1}^{20}$; it uses layerwise mean readout, 189,575 parameters, AdamW with cosine decay, and the same targets, epoch budget, and three seeds. RWSE is global preprocessing, so this comparator is not a raw local MPNN. Alongside we fit a ridge readout on *structural* features that do separate C_{120} from $40K_3$: inverse order $1/n$, mean degree, degree histogram, triangle density, and the second to sixth spectral moments of the normalized adjacency matrix. The anonymous GIN cannot even fit the training set (MAE 0.03–0.06): all cycles and all mK_3 are 2-regular and hence 1-WL-indistinguishable, so it must output one value for C_{3m} and mK_3, whose e_2 differ by $\frac{1}{3}$—the explicit pair of Appendix C.9 observed in training. The GIN with RNI fits (0.015–0.017) and transfers to 4–12× larger graphs with test MAE $(0.018, 0.024, 0.011)$. GPS–RWSE improves this to $(0.0046, 0.0085, 0.0039)$, beating the spectral–motif ridge baseline $(0.0092, 0.0092, 0.0060)$ but remaining roughly one order of magnitude less accurate than a ridge readout of 45 exact probe energies $(0.0005, 0.0005, 0.0002)$. This is not only a width or size-channel effect: a width-32 version has 48,711 parameters, versus 54,599 for GIN, and obtains $(0.0054, 0.0112, 0.0041)$; removing $1/n$ from the width-64 model gives $(0.0074, 0.0107, 0.0058)$. Indeed, $n^{-1}\sum_v (D_{\mathrm{deg}}^{-1}A)^t_{vv} = n^{-1}\,\mathrm{Tr}((D_{\mathrm{deg}}^{-1/2}AD_{\mathrm{deg}}^{-1/2})^t)$, so the established neural baseline receives the spectral information of the hand-crafted comparator without receiving any energy probe. The result shows that the earlier GIN gap largely reflected its missing structural encoding; the remaining comparison isolates the additional information in the energy representation.

(E8) The expander ceiling for a trained model. On random cubic graphs G with $n \in \{8, 10, 12\}$ (four per order) and their two-copy unions $G \sqcup G$—both in the training set—the same GIN is trained to predict the balanced-cut value Φ of (50) and tested on four pairs with $n = 14$. The anonymous GIN outputs *identical* values on G and $G \sqcup G$ (mean absolute predicted separation 0.000) and therefore errs by the forced $\Phi(G)/2 = 0.143$ on average. This is the 1-WL fact that a mean readout of an anonymous network cannot separate a graph from its disjoint self-union on any family. The same pair underlies the energy-feature floor of Proposition 8, but the two invariances are logically distinct. The experiment illustrates the floor for this anonymous mean-readout class rather than testing that proposition. The GIN with RNI, which is not energy-based, separates the pairs partially (mean

absolute predicted separation 0.082 against the true 0.286) and lowers the test MAE to 0.103; its training error (0.133) shows it is also underfit on 24 graphs. Symmetry breaking is thus necessary but, at this size and training budget, far from sufficient to read the constrained part of the quotient set.

(E9) A relaxation-based neural MAX-CUT solver. A node-level network (four tanh message-passing layers with residual connections, width 64, RNI, sigmoid head) is trained *unsupervised* on the 72 training graphs to maximize the expected cut density $\frac{1}{n}\sum_{uv\in E(G)}[p_u(1-p_v)+p_v(1-p_u)]$ of the product measure it outputs, where p_u is the output Bernoulli marginal at vertex u; this is the standard product-Bernoulli expected-cut objective developed in Karalias & Loukas (2020) and used in Schuetz et al. (2022); the assignment is then rounded by the method of conditional expectations, which never loses against the expectation. (ReLU variants collapse to the $p \equiv \frac{1}{2}$ saddle of this loss.) Relative to the exact optimum, the expected cut density is short by 0.085 on the training graphs and by 0.098 on the test graphs ($n \in [120, 360]$), and the rounded cut by 0.049 and 0.065: the solution quality of the learned relaxation degrades by about 0.015 across a 4–$12\times$ size extrapolation, in line with the small η_2 between the families at these orders. On random cubic graphs with $n = 16$ (exact optimum by brute force), the corresponding shortfalls are 0.102 and 0.046—comparable to the hyperfinite families. This is what Theorem 11 predicts: MAX-CUT is an affine task, so expanders impose no representational ceiling on it, and the residual shortfall on all families is the local network's optimization error (a four-layer network cannot coordinate a bipartition beyond its receptive field), not a representational one.

Table 3: Trained models (three seeds, means). E7: supervised size transfer on the split of Appendix C.11. E8: the balanced-cut self-union ceiling for an anonymous mean-readout GIN. E9: an unsupervised relaxation-based MAX-CUT solver, shortfall from the exact optimum. Stochastic RNI evaluations may vary in the last reported digit across reruns.

	e_2	MaxCut$/n$	α/n
(E7) test MAE, GIN anonymous (train MAE $0.053/0.057/0.031$)	0.092	0.094	0.048
GIN + RNI (train MAE $0.015/0.017/0.014$)	0.018	0.024	0.011
GPS–RWSE (train MAE $0.0057/0.0060/0.0073$)	0.0046	0.0085	0.0039
degree/size baseline	0.0835	0.0835	0.0458
spectral–motif baseline: degrees, triangles, spectral moments (ridge)	0.0092	0.0092	0.0060
readout of 45 exact probe energies (ridge)	0.0005	0.0005	0.0002
	train MAE	test MAE	mean $\lvert\widehat{\Phi}(G)-\widehat{\Phi}(G\sqcup G)\rvert$
(E8) GIN anonymous (Φ on cubic pairs; forced error $\Phi(G)/2 = 0.143$)	0.181	0.143	0.000 (true 0.286)
GIN + RNI	0.133	0.103	0.082 (true 0.286)
	train $n \le 30$	test $n \in [120, 360]$	random cubic $n = 16$
(E9) optimum minus expected cut density of the learned relaxation	0.085	0.098	0.102
optimum minus rounded cut density (mean / worst graph)	0.049 / 0.232	0.065 / 0.154	0.046 / 0.125

D Repair cost, exactness, and richer observations

All objectives in Section 4 average a bounded cost over vertices or edges. We now quantify the resulting edit stability and show that exact colorability is excluded by the same mechanism.

D.1 Disjoint unions and negligible edits

For compact sets A, B in a vector space and $\lambda \in [0,1]$, write $\lambda A + (1-\lambda)B$ for their weighted Minkowski sum.

Proposition 28 (Unions and edit stability)**.** *For finite graphs G, H and $\lambda = |V(G)|/(|V(G)| + |V(H)|)$,*

$$\mathcal{S}_q(G \sqcup H) = \lambda \mathcal{S}_q(G) + (1-\lambda)\mathcal{S}_q(H). \tag{60}$$

Consequently, if $|V(H_n)|/|V(G_n)| \to 0$ and both sequences are D-bounded, then

$$d_D^{\mathrm{part}}\big(\Theta_D^{\mathrm{part}}(G_n \sqcup H_n), \Theta_D^{\mathrm{part}}(G_n)\big) \longrightarrow 0. \tag{61}$$

If G_n and G_n' have the same vertex set and $|E(G_n) \triangle E(G_n')| = o(|V(G_n)|)$, and both sequences are D-bounded, their partition distance also tends to zero.

The proof is given in Appendix E.5.1.

This stability is useful for normalized optimization, but it makes bounded obstructions invisible. An odd cycle is the smallest chromatic example: $\chi(C_{2m+1}) = 3$, whereas $e_2(C_{2m+1}) = 1/(2m+1) \to 0$. The limit is two-colorable after deleting $o(n)$ edges although no graph in the sequence is exactly two-colorable.

D.2 Approximate chromatic number

The odd-cycle example suggests taking the first palette size whose normalized repair cost vanishes. To specify the repair model, we call

$$\chi_{\mathrm{app}}(x) := \min\{q \le D+1 : e_q(x) = 0\} \tag{62}$$

the *edge-deletion approximate chromatic number* of a partition-limit point x. Greedy coloring gives $e_{D+1} = 0$. On a growing sequence, $e_q \to 0$ is also equivalent to deleting $o(n)$ vertices before q-coloring: one can cover the monochromatic edges by at most one chosen endpoint per bad edge, while deleting s vertices removes at most Ds incident edges.

Proposition 29 (Continuous defects and a lower-semicontinuous threshold)**.** *The function χ_{app} is lower semicontinuous on $\mathfrak{X}_D^{\mathrm{part}}$ and equals $\chi(G)$ on each finite graph state. For every defect-resolution scale $A > 0$,*

$$\underline{\chi}_A(x) := 1 + \sum_{q=1}^{D} \min\{1, Ae_q(x)\} \tag{63}$$

is continuous, satisfies $\underline{\chi}_A(G) \le \chi(G)$ on finite graphs, and increases pointwise to χ_{app} as $A \to \infty$.

Appendix E.5.2 contains the proof.

Approximate colorability is therefore the continuous-repair information carried by partition convergence. Exact colorability cannot be recovered by a different continuous post-processing of the same state.

Theorem 30 (Exact-colorability threshold)**.** *For every $q \le D$, the $\{0,1\}$-indicator of exact q-colorability has no continuous extension to $\mathfrak{X}_D^{\mathrm{part}}$. Moreover, suppose a continuous function U satisfies $U(\Theta_D^{\mathrm{part}}(G)) \ge \chi(G)$ for every finite D-bounded graph. If x is approached by graphs G_n with $|V(G_n)| \to \infty$, then*

$$U(x) \ge D + 1. \tag{64}$$

Both statements use the same obstruction: append a fixed clique to a growing graph. The clique changes exact feasibility but disappears from every vertex-normalized quotient.

The proof appears in Appendix E.5.3.

Thus continuity preserves a macroscopic repair cost, not a bounded witness. The same zero-set argument applies to the least dimension d with vanishing relaxed edge-orthogonality energy: it is lower semicontinuous, while a fixed edge-orthogonality obstruction can vanish under an $o(n)$-vertex perturbation.

Beyond pairwise payoffs. The quotient keeps only first edge moments: nonlinear one-hop payoffs need colored-star statistics, finite-range payoffs the corresponding local–global radius, and observed attributes typed or measure-valued quotients. None of these richer observations is part of the pairwise core model.

E Omitted proofs

Proofs are grouped to follow the sectioning of the main text; within each group they appear in order of statement.

E.1 Proofs for Section 2

The only deferred claim of the section is the compactness of the hyperspace packaging.

E.1.1 Proof of Theorem 21

Proof. Each host $\mathsf{Y}_{D,q}$ is a compact finite-dimensional set, so its Hausdorff hyperspace $\mathcal{K}(\mathsf{Y}_{D,q})$ is compact metrizable. Tychonoff's theorem makes the countable product compact, and (44) metrizes it. The finite-realizable closure is a closed subspace. □

E.2 Proofs for Section 3

The section defers the continuity theorem, the support-function characterization, the convexification theorem with its corollary, the sharp norm comparison, the separation proposition, the necessity corollary of Appendix C.7, and the Lagrangian bridge; the proofs follow in that order.

E.2.1 Proof of Theorem 3

Proof. Every continuous map on the compact host $\mathsf{Y}_{D,q}$ is uniformly continuous. If $d_{\mathrm{H}}(K_n, K) \le \delta$, each point of either set lies within δ of the other, so their images lie within the corresponding modulus of continuity. This proves (5), and the real-valued claim follows by taking the endpoints of the compact images.

For the converse, let $C(\mathsf{Y}_{D,q})$ denote the real-valued continuous functions on the host and define

$$\mathcal{U}(K) := (\min_{K} F)_{F \in C(\mathsf{Y}_{D,q})}. \tag{65}$$

Every coordinate of $\mathcal{U}$ is continuous by the first step. The map is injective: if x belongs to one of two distinct compact sets but not the other, then $F(y) = d_q(x, y)$ has minimum zero on the first set and a strictly positive minimum on the second. A continuous injection from the compact hyperspace into a Hausdorff product is a topological embedding. Therefore the initial topology induced by these minimum queries is exactly the Hausdorff topology. □

E.2.2 Proof of Proposition 4

Proof. Equation (4) ranges over every affine functional of (p, W), and such a functional has the same extrema on a set and its convex hull. Support functions uniquely determine compact convex sets. Because the host is bounded and finite dimensional, these support functions share a Lipschitz constant on the unit sphere; pointwise convergence is therefore uniform, which is equivalent to Hausdorff convergence of the convex sets. □

E.2.3 Proof of Theorem 5

Proof. Let H be the graph after the edge deletions, with components $C_1, \dots, C_m$ and weights $\lambda_j := |V(C_j)|/n$. The disjoint-union identity gives

$$\mathcal{S}_q(H) = \sum_{j=1}^{m} \lambda_j \mathcal{S}_q(C_j). \tag{66}$$

Choose one point from each $\mathcal{S}_q(C_j)$ and translate that set by its chosen point. Every translated set lies in the same linear space: the p-coordinates sum to zero, W is symmetric, and the W-coordinates sum to zero. Its dimension is at most r_q. Apply the weighted Shapley–Folkman lemma recalled in Appendix C.7 in this space. Every point of the convex hull of the sum can be represented with all but at most r_q summands chosen from the original sets. Each exceptional summand has weight at most K/n, and the diameter of $\mathsf{Y}_{D,q}$ under d_q is at most $2(1 + D)$. Therefore

$$d_{\mathrm{H}}(\mathcal{S}_q(H), \operatorname{conv} \mathcal{S}_q(H)) \le \frac{2(1 + D) r_q K}{n}. \tag{67}$$

Using the same assignment before and after the deletion shows $d_{\mathrm{H}}(\mathcal{S}_q(G), \mathcal{S}_q(H)) \le 2\varepsilon$; taking convex hulls is nonexpansive in Hausdorff distance. The triangle inequality yields (6).

For a hyperfinite sequence, choose $K = K(\varepsilon)$ first and then send $n \to \infty$. The limit superior is at most 4ε, and sending $\varepsilon \downarrow 0$ proves the final claim. □

E.2.4 Proof of Corollary 6

Proof. The forward implication follows from Theorem 3. Conversely, Proposition 4 gives Hausdorff convergence of $\operatorname{conv} \mathcal{S}_q(G_n)$ for every q, while Theorem 5 makes its distance from $\mathcal{S}_q(G_n)$ tend to zero. □

E.2.5 Proof of the sharpness claims in Remark 7

Proof. The proof has three parts: an exact type-lattice calculation for matchings in ℓ^1, a growing-component lower bound, and Euclidean rounding with its matching lower example.

Matching reduction. Index the $N = N_q$ unordered color pairs by $a \in [N]$, and let

$$\mathcal{T}_{N,m} := \{x \in \Delta_N : mx \in \mathbb{Z}^N\}$$

be the denominator-m type grid. If x_{ij} is the fraction of matching edges whose endpoint colors form $\{i, j\}$, define the symmetric matrix $B(x)$ by $B(x)_{ii} = x_{ii}$ and $B(x)_{ij} = B(x)_{ji} = x_{ij}/2$ for $i < j$, and put $L(x) := (B(x)\mathbf{1}, B(x))$. Directly from the normalization in (1),

$$\mathcal{S}_q(M_{2m}) = L(\mathcal{T}_{N,m}), \qquad \operatorname{conv} \mathcal{S}_q(M_{2m}) = L(\Delta_N). \tag{68}$$

Moreover $\|B(x) - B(y)\|_1 = \|x - y\|_1$ and $\|(B(x) - B(y))\mathbf{1}\|_1 \le \|x - y\|_1$, whence

$$\|x - y\|_1 \le d_q(Lx, Ly) \le 2\|x - y\|_1. \tag{69}$$

It remains to compute the grid radius. Given $x \in \Delta_N$, write $mx_i = a_i + f_i$ with $a_i \in \mathbb{Z}_{\ge 0}$ and $f_i \in [0, 1)$. Then $s := \sum_i f_i$ is an integer in $[0, \min\{m, N - 1\}]$. The closest denominator-m type rounds up the s largest fractional parts and rounds down the rest, at ℓ^1 cost

$$\frac{2}{m}\left(s - \sum_{i=1}^{s} f_{(i)}\right) \le \frac{2s(N - s)}{Nm}, \tag{70}$$

where $f_{(1)} \ge \cdots \ge f_{(N)}$ and the inequality uses $\sum_{i=1}^{s} f_{(i)} \ge s^2/N$. Equality is attained by taking every $f_i = s/N$ and choosing nonnegative integers a_i summing to $m - s$. Maximizing over s gives $R_{q,m}$ in (45). Equations (68)–(69) prove (46); its asymptotic form follows from $N_q = \Theta(q^2)$.

Growing components. Write A_1 for the all-state-1 quotient of $G = mX$ and A_2 for the all-state-2 quotient, and put $z^* := (1 - t)A_2 + tA_1$ with $t = 1/(2m)$. For any binary assignment, let k_j be the number of state-1 vertices in the jth copy, $\delta_j := \min\{k_j, K - k_j\}$, and cut_j its number of bichromatic edges. Then

$$p_1 = \frac{1}{m}\sum_j \frac{k_j}{K}, \qquad W_{12} = \frac{1}{m}\sum_j \frac{\mathrm{cut}_j}{K} \ge \frac{h_0}{mK}\sum_j \delta_j.$$

Rounding each k_j/K to 0 or 1 moves p_1 by at most $(mK)^{-1}\sum_j \delta_j$, while t is $1/(2m)$ from the grid $m^{-1}\mathbb{Z}$. If this rounding cost is at most $1/(4m)$, the two p-coordinates contribute at least $1/(2m)$ to d_q; otherwise the two off-diagonal W-coordinates contribute more than $h_0/(2m)$. Taking the distance from z^* to $\mathcal{S}_2(G)$ proves (47).

Euclidean upper bound. Let H be the decomposed graph, with components C_j and $\lambda_j = |C_j|/n$. For a point $z = \sum_j \lambda_j z_j$ in $\operatorname{conv} \mathcal{S}_q(H)$, independently choose $X_j \in \mathcal{S}_q(C_j)$ with $\mathbb{E}X_j = z_j$. Every quotient vector has Euclidean norm at most $\sqrt{1 + D^2}$, so independence gives

$$\mathbb{E}\left\|\sum_j \lambda_j(X_j - z_j)\right\|_2^2 \le (1 + D^2)\sum_j \lambda_j^2 \le (1 + D^2)K/n.$$

Some realization therefore lies within the square root of this bound. Alternatively, first apply Shapley–Folkman and let J denote its set of at most r_q exceptional component indices; randomizing only these gives $\sum_{j \in J} \lambda_j^2 \le r_q(K/n)^2$. Taking the better construction and paying 4ε for deletion proves (48).

For the matching lower bound, the map above satisfies $\|B(x) - B(y)\|_{\mathrm{F}} \ge 2^{-1/2}\|x - y\|_2$. If $N \ge 2m$, take x uniform on $2m$ type coordinates; its distance from $\mathcal{T}_{N,m}$ is at least $1/\sqrt{2m}$. If $N < 2m$, take all fractional parts equal to s/N for $s = \lfloor N/2 \rfloor$ as above; its Euclidean distance from the grid is $m^{-1}\sqrt{s(N - s)/N} = \Theta(\sqrt{N}/m)$. Together with the upper bound, these two cases prove $\kappa_{q,2}(M_{2m}) = \Theta(\min\{m^{-1/2}, \sqrt{N}/m\})$. □

E.2.6 Proof of Proposition 22

Proof. Let $H, C_1, \dots, C_m$ and $\lambda_j = |V(C_j)|/n$ be as in the proof of Theorem 5. Since Π is linear, applying it to (66) gives $\Pi\mathcal{S}_q(H) = \sum_j \lambda_j \Pi\mathcal{S}_q(C_j)$, and taking convex hulls (which commute with linear images and with Minkowski sums) gives $\operatorname{conv}\Pi\mathcal{S}_q(H) = \sum_j \lambda_j \operatorname{conv}\Pi\mathcal{S}_q(C_j)$. The sets $\Pi\mathcal{S}_q(C_j)$ lie in $\mathbb{R}^k$, so the weighted Shapley–Folkman lemma of Appendix C.7 applies in dimension k without any translation: every point of the hull equals $\sum_j \lambda_j a_j$ with $a_j \in \Pi\mathcal{S}_q(C_j)$ for all but at most k indices j, the exceptional a_j lying in $\operatorname{conv}\Pi\mathcal{S}_q(C_j)$. Replacing each exceptional a_j by any point of $\Pi\mathcal{S}_q(C_j)$ moves the sum by at most $\lambda_j \operatorname{diam}_{\ell^1}\big(\operatorname{conv}\Pi\mathcal{S}_q(C_j)\big) \le \lambda_j \operatorname{diam}_{\ell^1}\big(\mathcal{S}_q(C_j)\big) \le \frac{K}{n}\,2(1+D)$, using that Π is nonexpansive, that a convex hull has the diameter of the set, and that $\Upsilon_{D,q}$ has ℓ^1-diameter at most $2(1+D)$. Hence $d_{\mathrm{H}}(\Pi\mathcal{S}_q(H), \operatorname{conv}\Pi\mathcal{S}_q(H)) \le 2(1+D)kK/n$. The deletion step is as before: the same assignment on G and on H gives $d_{\mathrm{H}}(\mathcal{S}_q(G), \mathcal{S}_q(H)) \le 2\varepsilon$, nonexpansiveness of Π gives the same bound for the projected sets, and taking hulls is nonexpansive in Hausdorff distance; the triangle inequality proves (51). The Lipschitz consequence is the fact that minima (or maxima) of an L-Lipschitz function over two compact sets differ by at most L times their Hausdorff distance.

For sharpness, use the matching type map (68). Select k unordered color-pair types and one additional type. Give each selected type mass $1/(2m)$ and put the remaining mass on the additional type. This is possible when $k \le \min\{N_q - 1, 2m\}$. Project onto the corresponding upper-triangular W-coordinates. An off-diagonal coordinate is a multiple of $1/(2m)$ for every realized matching quotient but equals $1/(4m)$ at the target; a diagonal coordinate has the still larger distance $1/(2m)$ from its $1/m$ grid. Summing the k coordinate distances proves (52). □

E.2.7 Proof of Proposition 8

Proof. The disjoint-union identity gives $\mathcal{S}_q(H_n) = \frac{1}{2}\mathcal{S}_q(G_n) + \frac{1}{2}\mathcal{S}_q(G_n)$, whose convex hull equals $\operatorname{conv}\mathcal{S}_q(G_n)$. Coloring the two copies constantly with different states gives $\Phi(H_n) = 0$. For a binary assignment on G_n, put $x = p_1$ and $y = \min\{x, 1-x\}$. Its cut has size at least $hy|V(G_n)|$, hence

$$|p_1 - \tfrac{1}{2}| + W_{12} \ge \tfrac{1}{2} - y + hy \ge \min\{\tfrac{1}{2}, \tfrac{h}{2}\}.$$

□

E.2.8 Proof of Corollary 23

Proof. Let (G_i) be D-regular graphs with $|V(G_i)| \to \infty$ and edge expansion at least $h > 0$; by Theorem 21 pass to a partition-convergent subsequence, still written (G_i). Alternate $H_{2i-1} := G_i$ and $H_{2i} := G_i \sqcup G_i$. By (49), $\operatorname{conv}\mathcal{S}_q(H_{2i}) = \operatorname{conv}\mathcal{S}_q(G_i)$, so by Proposition 4 every scalar energy converges along the full sequence. But the balanced-cut query (50) is 1-Lipschitz, hence Hausdorff-continuous, and $\Phi(H_{2i}) = 0$ while $\Phi(H_{2i-1}) \ge \min\{1/2, h/2\}$, so $\mathcal{S}_2(H_i)$ does not converge. □

E.2.9 Proof of Proposition 24

Proof. Weak duality: for $\|u\|_\infty \le L$, $\langle u, p-a\rangle \le L\|p-a\|_1$ pointwise, so $\mathsf{g}_{B,a}(u) \le \mathsf{m}_{B,L,a}(G)$, proving the left inequality.

For the right inequality, write $F(p,W) := \frac{1}{2}\langle B, W\rangle + L\|p-a\|_1$ and $\varphi\big((p,W),u\big) := \frac{1}{2}\langle B, W\rangle + \langle u, p-a\rangle$ on $\operatorname{conv}\mathcal{S}_q(G) \times \{u : \|u\|_\infty \le L\}$. Both factors are compact and convex, and φ is affine in each argument, so the minimax theorem (Sion, 1958) gives

$$\min_{z \in \operatorname{conv}\mathcal{S}_q(G)} F(z) = \min_z \max_u \varphi(z,u) = \max_u \min_z \varphi(z,u) = \max_{\|u\|_\infty \le L} \mathsf{g}_{B,a}(u),$$

where the first equality uses $L\|p-a\|_1 = \max_{\|u\|_\infty \le L}\langle u, p-a\rangle$ and the last uses that an affine functional has the same minimum on $\mathcal{S}_q(G)$ and its convex hull. Finally, F is Lipschitz with constant $\max\{L, \frac{1}{2}\|B\|_{\max}\}$ for the metric d_q, because $|\frac{1}{2}\langle B, W\rangle - \frac{1}{2}\langle B, W'\rangle| \le \frac{1}{2}\|B\|_{\max}\|W - W'\|_1$ and the penalty term is L-Lipschitz in $\|p-p'\|_1$. Every minimizer of F over $\operatorname{conv}\mathcal{S}_q(G)$ lies within $\kappa_q(G)$ of $\mathcal{S}_q(G)$, so $\mathsf{m}_{B,L,a}(G) \le \min_{\operatorname{conv}\mathcal{S}_q(G)} F + \max\{L, \frac{1}{2}\|B\|_{\max}\}\,\kappa_q(G)$. □

E.3 Proofs for Section 4

The section defers the independence-ratio identity and the compact-spin transfer proposition.

E.3.1 Proof of the independence-ratio identity of Section 4

Proof. An assignment determines the selected set $S = \{v : c(v) = 1\}$. Its objective is $(|S| - |E(G[S])|)/n$. Choose one endpoint of each edge in the induced graph $G[S]$, and let C be the union of the chosen vertices. The set C covers every edge of $G[S]$ and satisfies $|C| \le |E(G[S])|$, so $S \setminus C$ is independent. Hence $|S| - |E(G[S])| \le \alpha(G)$. A maximum independent set attains equality. □

E.3.2 Proof of Proposition 10

Proof. Let $E(G) := \mathsf{E}^{-}_{K,h}(G)$. Choose a finite δ-net $Z_\delta = \{z_1, \dots, z_{q_\delta}\} \subseteq K$ and restrict all spins to this net. The resulting finite-state energy is

$$E_\delta(G) = \min_{(p,W)\in\mathcal{S}_{q_\delta}(G)} \frac{1}{2} \sum_{i,j} h(z_i, z_j) W_{ij}, \tag{71}$$

so it extends continuously by Theorem 3.

Let $\omega_h(\delta)$ be the largest change in h when both arguments move by at most δ. Restriction can only raise a minimum, while rounding a minimizing K-assignment changes each edge term by at most $\omega_h(\delta)$. Because $|E(G)|/n \le D/2$,

$$0 \le E_\delta(G) - E(G) \le \frac{D}{2}\,\omega_h(\delta). \tag{72}$$

It follows that, for two nets,

$$|E_\delta(G) - E_\eta(G)| \le \frac{D}{2}\big(\omega_h(\delta) + \omega_h(\eta)\big). \tag{73}$$

Both finite-net energies are continuous on the limit space, and finite graph states are dense, so the same inequality holds on the closure. The family (E_δ) is therefore uniformly Cauchy. Its continuous limit is independent of the nets and agrees with E on every finite graph by (72). Reversing the restriction inequality gives the maximum case. □

E.4 Proofs for Section 5

The section defers the size-transfer theorem and its corollaries, the encoder constructions with their treewidth lemma, the raw-graph RNI upper and lower bounds, and the generalization theorem; the compact-spin transfer corollary is stated as well as proved here.

E.4.1 Proofs for task-value size transfer

Proof of Theorem 11. By definition,

$$\eta_{q,s}(G, G') = d_{\mathrm{H},s}\big(\operatorname{conv}\mathcal{S}_q(G), \operatorname{conv}\mathcal{S}_q(G')\big). \tag{74}$$

For compact convex C, C' in a normed space, support-function duality gives $d_{\mathrm{H}}(C, C') = \sup_{\|\theta\|\le 1} |h_C(\theta) - h_{C'}(\theta)|$ over the dual unit ball, so $\eta_{q,s}$ is exactly the discrepancy over dual-unit scalar energies. For $s = 1$, our sets live in the subspace $\{(p, W) : W = W^\top\}$ with the ℓ^1 norm $\|p\|_1 + \|W\|_1$, whose dual norm is $\max\{\|a\|_\infty, \|B\|_{\max}\}$ over pairs $\theta = (a, B)$; restricting B to be symmetric loses no generality on this subspace, because for symmetric ΔW one has $\langle B, \Delta W\rangle = \langle \frac{1}{2}(B + B^\top), \Delta W\rangle$ and symmetrization does not increase $\|B\|_{\max}$. The probe ℓ_θ in (10) carries the factor $\frac{1}{2}$ on B, which is why $\|\theta\|_*$ uses $\frac{1}{2}\|B\|_{\max}$: the map $(a, B) \mapsto (a, \frac{1}{2}B)$ matches ℓ_θ with the dual pairing. Replacing a coefficient by its negative converts minima to support functions and identifies $\eta_{q,1}$ with η_q in (11). Convexification is nonexpansive, which proves the first inequality in (12); the triangle inequality through the two convex hulls proves the second. Minima of an L-Lipschitz function over two compact sets differ by at most L times their Hausdorff distance, proving (13). Substituting Theorem 5 gives (14), and substituting (48) gives the stated Euclidean

specialization. If F is affine, its minimum or maximum over a set equals that over its convex hull, so the two outer κ terms are unnecessary.

For the finite-probe claim, $\|p\|_1 + \|W\|_1 \le 1 + D$ on $\mathsf{Y}_{D,q}$. Approximating an arbitrary dual unit vector by $\mathcal{N}_\rho$ changes each of the two probe values by at most $(1+D)\rho$, which proves (56). □

Proof of Corollary 12. Each task is a single probe energy with an explicit dual norm: $e_q = \gamma_\theta$ for $\theta = (0, I_q)$, since $\ell_\theta = \frac{1}{2}\sum_i W_{ii} = u_q$, with $\|\theta\|_* = \frac{1}{2}$; $\mathrm{defect}_J = \gamma_\theta$ for $\theta = (0, B)$ with $B_{ij} = \mathbf{1}[ij \notin E(J)]$, again $\|\theta\|_* = \frac{1}{2}$; $\mathrm{MaxCut}\,/n = -\gamma_\theta$ for $\theta = (0, B)$ with $B_{12} = B_{21} = -1$ and zero diagonal, $\|\theta\|_* = \frac{1}{2}$; and $\alpha/n = -\gamma_\theta$ for $\theta = (a, B)$ with $a = (-1, 0)$, $B = \mathrm{diag}(1, 0)$, $\|\theta\|_* = 1$. By homogeneity of η_q in (11), $|\gamma_\theta(G) - \gamma_\theta(G')| \le \|\theta\|_* \, \eta_q(G, G')$, which gives (15)–(16); $\eta_q \le \delta_q$ is the first inequality of (12). For the last claim, $e_q(G) = 0$ and $e_q(G') \le \frac{1}{2}\eta_q(G, G')$ means some assignment of G' leaves at most $\frac{1}{2}\eta_q(G, G')|V(G')|$ monochromatic edges; deleting them makes G' q-colorable. □

Proof of Corollary 14. The finite-graph representations form a totally bounded subset of the compact finite product of hyperspaces, so they admit a finite r-net whose centers are finite graph representations. Their orders have a finite maximum $n_{\max}(D, \mathbf{q}, r)$. Choose B within r of a test graph G. The target vector moves by at most $L_* r$, while $d_R(\widehat{R}(G), \widehat{R}(B)) \le r + 2\zeta$. Inserting the prediction and target at B between those at G gives (17).

For any $\xi > 0$, choose a finite ξ-net $Z_\xi \subset K$, put $m := |Z_\xi|$, and let $\omega_h(\xi)$ be the supremum change in h when each argument moves by at most ξ. For a compact-spin target, append the finite-net coordinate $\mathcal{S}_m$ to the representation. The two rounding errors and (75) add $D\omega_h(\xi)$ to the corresponding target-transfer term; for relaxed d-dimensional edge orthogonality this is at most $4D\xi$. □

E.4.2 Compact-spin transfer

Corollary 31 (Compact-spin size transfer)**.** *For the compact-spin energy in (10), let $Z_\xi \subset K$ be a finite ξ-net of size m, and let $\omega_h(\xi)$ be the supremum change in h when each argument moves by at most ξ. Then*

$$|\mathsf{E}^-_{K,h}(G) - \mathsf{E}^-_{K,h}(G')| \le D\omega_h(\xi) + \frac{1}{2}\|h\|_\infty \delta_m(G, G'). \tag{75}$$

For $K = S^{d-1}$ and $h(x, y) = \langle x, y\rangle^2$, one may take $m \le (1 + 2/\xi)^d$ and $\omega_h(\xi) \le 4\xi$, giving $|\varepsilon^\perp_d(G) - \varepsilon^\perp_d(G')| \le 4D\xi + \frac{1}{2}\delta_m(G, G')$.

The proof combines (72) with (13) applied to the finite-net energy (71), whose integrand is $\frac{1}{2}\|h\|_\infty$-Lipschitz in W.

E.4.3 Bounded treewidth: hyperfiniteness and exact quotient computation

Lemma 32 (Explicit hyperfiniteness of bounded treewidth)**.** *Let G be D-bounded with treewidth at most t and n vertices. For every $K \ge 1$, one can delete at most $2(t+1)Dn/K$ edges so that every remaining component has at most K vertices.*

Proof. A graph of treewidth at most t on s vertices has a bag whose removal leaves components of at most $s/2$ vertices each (a standard centroid-bag argument on a tree decomposition). Deleting the at most $(t+1)D$ edges incident to that bag disconnects the bag vertices and leaves all parts of order at most $s/2$. Recurse on every part of order exceeding K. Let $T(s)$ be the maximum number of edges deleted for a piece of order s; we claim $T(s) \le (t+1)D(2s/K - 1)$ for $s > K$, which gives the lemma. Induct on s: one split costs $(t+1)D$ and yields parts $s_1, s_2, \ldots$ with $\sum_i s_i \le s$ and $s_i \le s/2$. Parts of order at most K cost nothing. If no part exceeds K the claim is immediate from $s > K$. If exactly one part exceeds K, then $s_1 \le s/2 \le s - K/2$, and $(t+1)D\big(1 + 2s_1/K - 1\big) \le (t+1)D(2s/K - 1)$. If $\ell \ge 2$ parts exceed K, then $(t+1)D\big(1 + \sum_i (2s_i/K - 1)\big) \le (t+1)D(2s/K - 1)$ using $\ell \ge 2$. □

Proof of Proposition 17. Convert the given decomposition to a nice tree decomposition with introduce, forget, join, and introduce-edge nodes, of the same width and $O((t+1)n)$ nodes; each edge is processed at its unique introduce-edge node. Scale coordinates by n: an assignment c induces the integer vector $\nu(c) = (np, nW) \in \mathbb{Z}^q \times \mathbb{Z}^{q\times q}$ with $\sum_i np_i = n$ and $\sum_{ij} nW_{ij} = 2|E(G)| \le Dn$, so the number of attainable vectors is at most $(n+1)^{q-1}(Dn+1)^{q(q+1)/2} = n^{O(q^2)}$.

For each node x with bag B_x, and each coloring $\sigma : B_x \to [q]$, let $\mathrm{DP}_x(\sigma)$ be the set of partial vectors ν realizable by colorings of the vertices introduced strictly below x (together with σ), where a vertex contributes its p-count when it is forgotten and an introduce-edge node contributes its edge's two oriented W-counts. Leaves start from the empty vector. An introduce-vertex node extends σ in q ways; an introduce-edge node adds the corresponding counts, which are determined by σ. A forget node projects σ and adds one p-count. A join node combines the two children sets of the same σ by vector sums, a Minkowski sum of two subsets of the $n^{O(q^2)}$-point grid. Each node therefore costs at most $q^{t+1} \cdot n^{O(q^2)}$ set operations, and the root returns exactly $\{\nu(c) : c\}$, hence $\mathcal{S}_q(G)$ after dividing by n. Claims (i) and (ii) follow because minima of ℓ_θ over the listed set are exact.

For (iii): apply Lemma 32 with $K := \lceil 2\sqrt{(t+1)Dn/((1+D)r_q)} \rceil$ and Theorem 5 with $\varepsilon = 2(t+1)D/K$, giving

$$\kappa_q(G) \le \frac{8(t+1)D}{K} + \frac{2(1+D)r_qK}{n} \le 8\sqrt{(1+D)r_q(t+1)D}\, n^{-1/2} + \frac{2(1+D)r_q}{n},$$

where the last term accounts for the ceiling. Since $2(1+D)r_q/n \le 2(1+D)r_q\, n^{-1/2}$, both $\kappa_q(G)$ and $\kappa_q(G')$ are at most $\big(8\sqrt{(1+D)r_q(t+1)D} + 2(1+D)r_q\big)$ times $n^{-1/2}$ and $n'^{-1/2}$ respectively, and (38) follows from (12) and (56). □

E.4.4 Proof of Proposition 18

Proof. Write $R_0 := 1 + D$; every $z \in \mathsf{Y}_{D,q}$ has $\|z\|_1 \le R_0$. Two facts are used repeatedly. (a) If $A_1, \dots, A_N \subset \mathsf{Y}_{D,q}$ and μ, μ' are probability vectors, then $d_{\mathrm{H}}\big(\sum_\tau \mu_\tau A_\tau, \sum_\tau \mu'_\tau A_\tau\big) \le R_0 \|\mu - \mu'\|_1$: map $\sum_\tau \mu_\tau a_\tau$ to $\sum_\tau \mu'_\tau a_\tau$ with the same a_τ, and symmetrically. (b) If H is a disjoint union of connected graphs of order at most K, with μ_τ the fraction of vertices lying in components of type $\tau \in \mathcal{T}_K$, then $\operatorname{conv} \mathcal{S}_q(H) = \sum_\tau \mu_\tau \operatorname{conv} \mathcal{S}_q(\tau)$ (by (66), grouping isomorphic components and using $A + A = 2A$ for convex A), and $d_{\mathrm{H}}(\mathcal{S}_q(H), \operatorname{conv} \mathcal{S}_q(H)) \le 2R_0 r_q K/|V(H)|$ by (67).

Let $G \in \mathcal{C}$ have n vertices, and let H be the graph after the (ε, K)-decomposition, so $d_{\mathrm{H}}(\mathcal{S}_q(G), \mathcal{S}_q(H)) \le 2\varepsilon = r/4$. If $n \le n_{\max}$, take $B := H \in \mathcal{B}_r$ and we are done. Otherwise let μ be the type proportions of H, and define $m'_\tau := \lceil \mu_\tau n_0/|\tau| \rceil$ and $B := \bigsqcup_\tau m'_\tau\, \tau$, of order $n_B := \sum_\tau m'_\tau|\tau| \in [n_0, n_0 + N_K K]$, so $B \in \mathcal{B}_r$. Its type proportions $\mu'_\tau = m'_\tau|\tau|/n_B$ satisfy $0 \le \mu'_\tau - \mu_\tau n_0/n_B \le |\tau|/n_B \le K/n_B$, whence $\|\mu' - \mu n_0/n_B\|_1 \le N_K K/n_B$, while $\|\mu n_0/n_B - \mu\|_1 = 1 - n_0/n_B \le N_K K/n_B$; thus $\|\mu - \mu'\|_1 \le 2N_K K/n_0$. By (b) applied to H and to B, by (a), and by the triangle inequality,

$$\begin{aligned} d_{\mathrm{H}}(\mathcal{S}_q(G), \mathcal{S}_q(B)) &\le \frac{r}{4} + \frac{2R_0 r_q K}{n} + \frac{2R_0 r_q K}{n_B} + \frac{2R_0 N_K K}{n_0} \\ &\le \frac{r}{4} + \frac{4R_0 K(r_q + N_K)}{n_0} \le \frac{r}{4} + \frac{r}{2} \le r. \end{aligned}$$

using $n > n_{\max} \ge n_0$, $n_B \ge n_0$, and the definition of n_0. For several state spaces the same B serves all q_j at once because $\mathcal{T}_K$ and μ do not depend on q; only r_q does, and it is monotone in q. The bound $N_K \le \sum_{k \le K} 2^{\binom{k}{2}} \le K 2^{K(K-1)/2}$ counts labeled graphs. □

E.4.5 Proof of Lemma 20

Proof. The proof separates the finite local computation from its neural implementation. Absorb the fixed simulation and flooding overhead into a constant-factor enlargement of Q; we keep the same symbol, so the resulting depth remains $O(Q)$. Every transcript of $\mathcal{A}$ visits at most $Q+1$ vertices in the radius-Q ball of its root. For each of the two rank coordinates, divide $[0,1]$ into M ordered bins and replace the hard bin indicators by continuous piecewise-linear soft indicators. They are exactly one-hot outside narrow transition strips around the $M-1$ thresholds and are implemented by a fixed ReLU

multilayer perceptron. Since a radius-Q ball has at most $b_D(Q) := 1 + D\sum_{j=0}^{Q-1}(D-1)^j$ vertices, a union bound makes the probability of a bin collision in either coordinate at most $b_D(Q)^2/M$. Choose M so that this is at most $\zeta\xi/(16B)$, then choose the strips with sufficiently small total measure that the probability some mark in the ball meets a strip is at most the same quantity. The expected fraction of bad roots is therefore at most $\zeta\xi/(8B)$, and Markov's inequality makes that fraction at most $\xi/(4B)$ with probability at least $1 - \zeta/2$.

Outside those roots, the soft indicators are one-hot and the two quantized ranks preserve every comparison and otherwise arbitrary enumeration order used by $\mathcal{A}$. There are only finitely many degree-D, depth-Q marked transcripts. A sum-aggregation MPNN can encode their finite multisets injectively with one-hot states, flood the transcript to the root in $O(Q)$ rounds, and implement its finite output table by a multilayer perceptron; equivalently, this is the bounded-order RNI construction of Abboud et al. (2021); Sato et al. (2021) applied only to the queried transcript. Approximate the finite table to rootwise error at most $\xi/2$ and clip the output to $[-B, B]^P$. Good roots then contribute at most $\xi/2$ to the vertex mean and bad roots at most $2B \cdot \xi/(4B) = \xi/2$. Adjusting the two probability allocations proves (41). All alphabets, widths, and weights depend on (D, Q, P, B, ξ, ζ) but not on n. □

E.4.6 Proof of Theorem 15

Proof. We first invoke the bounded-treewidth partition oracle, compile its rooted queries with Lemma 20, and then amplify the constant success probability by independent median channels.

The partition oracle of Edelman et al. (2011, Theorem 1) uses iid priority ranks for its random seed. Give each vertex a second independent iid rank and use it to order every otherwise arbitrary vertex enumeration, including loops over discovered vertex sets. Conditional on these auxiliary ranks the enumeration order is query-independent, as required in that theorem, while the cut-probability proof over the priority ranks is unchanged. The oracle gives query access to a consistent partition whose parts have order at most K_α, cuts at most αn edges with probability at least $9/10$, and uses at most Q_α adjacency queries per root, with the bounds in (19). Another $O(DK_\alpha)$ queries recover all edges induced by the returned part; this is absorbed into the stated depth.

For a root v, let $C(v)$ be its oracle part and let the local algorithm return the vector $(\gamma_\theta(C(v)))_{\theta\in\Theta}$. It is bounded by $1 + D$, comparison based, local, and equivariant. Invoke Lemma 20 with $\zeta = 1/10$ to obtain one RNI-MPNN channel with vertex-mean error at most ξ with probability at least $9/10$. If H is the graph after the oracle cut, affinity of each probe and the disjoint-union identity give

$$\frac{1}{n}\sum_{v}\gamma_\theta(C(v)) = \sum_{C\in H}\frac{|C|}{n}\gamma_\theta(C) = \gamma_\theta(H).$$

The same assignment before and after deletion shows $d_{\mathrm{H}}(\mathcal{S}_q(G), \mathcal{S}_q(H)) \le 2\alpha$, so probe duality yields $|\gamma_\theta(G) - \gamma_\theta(H)| \le 2\alpha$ simultaneously for the finite bank.

By a union bound with the oracle's $9/10$ event, one channel therefore satisfies the target inequality with probability at least $4/5$. Run $T = O(\log(1/\delta))$ independent channels and take the coordinate-wise median. A Chernoff bound makes a strict majority of channels successful with probability at least $1 - \delta$, proving (18). Parallel channels change width, not depth, and no parameter depends on graph order. □

E.4.7 Proof of Theorem 16

Proof. We separate a high-girth graph from its bipartite double cover by coloring defect, identify their depth-t local laws below the girth scale, and use replication to concentrate their mean-pooled representations.

Take a nonbipartite D-regular Lubotzky–Phillips–Sarnak Ramanujan family X_K with girth $g_K \ge c\log K$ for some constant $c > 0$ and $\lambda_{\min}(X_K) \ge -2\sqrt{D-1} = -D + \delta_D$, where $\delta_D := D - 2\sqrt{D-1} > 0$ (Lubotzky et al., 1988). Let B_K be the bipartite double cover of X_K and define $A_{K,L} := LB_K$ and $H_{K,L} := 2LX_K$. Both graphs have $2KL$ vertices and component order at most $2K$. Bipartiteness gives $e_2(A_{K,L}) = 0$. For a binary coloring of X_K, encoded by $x \in \{-1, 1\}^K$,

the number of monochromatic edges divided by K is

$$\frac{D + x^\top A_{X_K} x/K}{4} \geq \frac{\delta_D}{4} =: c_0,$$

which proves the other half of (20).

Suppose $2t + 1 < g_K$. Every radius-t unmarked rooted neighborhood in both graphs is the same D-regular tree. With iid initialization, the final node state of any t-layer local MPNN therefore has the same distribution and mean μ on both graphs. States at vertices more than $2t$ apart depend on disjoint initializations. If $\bar{h}$ is the vertex mean and the state lies in $[-M, M]^w$, the dependency-graph variance bound gives

$$\mathbb{E}\|\bar{h} - \mu\|_2^2 \leq \frac{4M^2 w\, b_D(2t)}{2KL}. \tag{76}$$

For a Λ-Lipschitz readout, the expected difference between its outputs on $A_{K,L}$ and $H_{K,L}$ is consequently at most $4\Lambda M\sqrt{w b_D(2t)/(2KL)}$. If both expected task errors are at most η, the triangle inequality forces

$$c_0 \leq 2\eta + 4\Lambda M\sqrt{\frac{w b_D(2t)}{2KL}}.$$

For fixed finite (w, M, Λ, t) and $\eta < c_0/2$, the rightmost term vanishes as $L \to \infty$, a contradiction. Hence a single network valid for every repetition count must satisfy $2t + 1 \geq g_K$, and therefore $t \geq (c/2)\log K - 1/2$; after discarding finitely many K, take $c_1 = c/4$. Global communication, deterministic positional identifiers, sum pooling, or graph-size-dependent bounds invalidate the locality or concentration step and are outside the theorem. □

E.4.8 Proofs of Theorem 26 and Propositions 25 and 27

Proof. Covering the state space. If $\mathcal{P}$ is a ρ-net of the compact metric space Y and $A \subseteq Y$ is nonempty compact, then $A_{\mathcal{P}} := \{x \in \mathcal{P} : \mathrm{dist}(x, A) \leq \rho\}$ is nonempty and $d_{\mathrm{H}}(A, A_{\mathcal{P}}) \leq \rho$; hence the subsets of $\mathcal{P}$ form a ρ-net of $\mathcal{K}(Y)$ and $\mathsf{N}_{\mathcal{K}(Y)}(\rho) \leq 2^{|\mathcal{P}|}$. Products with the max metric multiply nets, giving $\mathsf{N}_{\mathfrak{X}}(\rho) \leq \prod_j 2^{\mathsf{N}_{Y_j}(\rho)}$, where $Y_j := \mathsf{Y}_{D,q_j}$, and the volumetric bound for the ℓ^1-ball of radius $1 + D$ in dimension $q_j + q_j^2$ gives $\mathsf{N}_{Y_j}(\rho) \leq (3(1 + D)/\rho)^{q_j + q_j^2}$ for $\rho \leq 1 + D$.

Uniform bound. Fix a ρ-net $P_{\mathfrak{X}}$ of $\mathfrak{X}_s$ of size $\mathsf{N}_{\mathfrak{X}}(\rho)$ and round function values to the grid $\{0, \Lambda\rho, 2\Lambda\rho, \ldots\} \cap [0, B]$. Two Λ-Lipschitz functions sharing rounded values on $P_{\mathfrak{X}}$ differ by at most $3\Lambda\rho$ in sup norm, so the class admits a $3\Lambda\rho$-net of cardinality at most $(1 + B/(\Lambda\rho))^{\mathsf{N}_{\mathfrak{X}}(\rho)}$. For each representative, Hoeffding's inequality applies to the $[0, B]$-valued loss; a union bound over representatives and the $2 \times 3\Lambda\rho$ transfer between a function and its representative (once in the population term, once in the empirical term) give (58) with the stated logarithm since $1 + u \leq 3 + u$.

Probe representation (Proposition 27). Repeat the covering argument with the compact state space replaced by the cube $[-(1+D), 1+D]^P$ in $\|\cdot\|_\infty$: a grid of spacing 2ρ along each axis is a ρ-net with at most $\lceil (1 + D)/\rho \rceil^P \leq (2(1 + D)/\rho)^P$ points for $\rho \leq 1 + D$, and Λ-Lipschitz $[0, B]$-valued functions on the cube are covered at scale $3\Lambda\rho$ by at most $(1 + B/(\Lambda\rho))^{\mathsf{N}_P(\rho)}$ representatives exactly as before; Hoeffding, the union bound and the $2 \times 3\Lambda\rho$ transfer give (59).

Linear readouts (Proposition 25). The loss class $\{G \mapsto |f(G) - y(G)| : f \in \mathcal{F}_{\mathrm{lin}}\}$ takes values in $[0, B_*]$ with $B_* := W_1(1+D) + 2B$ and is a 1-Lipschitz image of $\mathcal{F}_{\mathrm{lin}}$. Let $\widehat{\mathfrak{R}}_m$ denote empirical Rademacher complexity. After scaling by B_*, apply the one-sided empirical Rademacher bound (Mohri et al., 2018, Theorem 3.3) to the loss class and to the class obtained by replacing each loss h with $B_* - h$, each with failure probability $\delta/2$. Their empirical complexities agree by sign symmetry, and Talagrand contraction bounds either by that of $\mathcal{F}_{\mathrm{lin}}$. With probability $1 - \delta$ the uniform gap is at most $2\widehat{\mathfrak{R}}_m(\mathcal{F}_{\mathrm{lin}}) + 3B_*\sqrt{\ln(4/\delta)/(2m)}$. Finally, $\widehat{\mathfrak{R}}_m(\mathcal{F}_{\mathrm{lin}}) \leq W_1(1+D)\sqrt{2\ln(2P)/m} + B/\sqrt{m}$: the first term is Massart's lemma for the ℓ^1-ball acting on features bounded by $1 + D$ in sup norm, and the second is the Rademacher complexity of the constant class $\{b : |b| \leq B\}$. □

E.5 Proofs for Appendix D

The repair-and-exactness appendix defers the edit-stability proposition, the approximate-chromatic-number proposition, and the exact-colorability theorem.

E.5.1 Proof of Proposition 28

Proof. An assignment on a disjoint union is exactly a pair of assignments on its components. Both p and W are their vertex-weighted averages, which proves (60). The diameter of each quotient host is at most $2(1+D)$. With $\lambda_n := |V(G_n)|/(|V(G_n)| + |V(H_n)|)$, the qth Hausdorff distance after adding H_n is at most $2(1+D)(1-\lambda_n)$. Dominated convergence in the product metric proves (61).

For graphs on the same vertex set, use the same assignment in both graphs. The node weights agree, and every edge in the symmetric difference changes the ℓ^1 mass of the oriented edge matrix by $2/|V(G_n)|$. Hence each qth Hausdorff distance is at most $2|E(G_n) \triangle E(G'_n)|/|V(G_n)|$. □

E.5.2 Proof of Proposition 29

Proof. The defects decrease with q, and $\{x : \chi_{\mathrm{app}}(x) > q\} = \{x : e_q(x) > 0\}$ is open because e_q is continuous. This proves lower semicontinuity. On a finite graph, $e_q(G) = 0$ exactly when G is q-colorable. The remaining statements follow from continuity and nonnegativity: if $\chi(G) = r$, only the $r-1$ terms with $q < r$ can contribute to (63). □

E.5.3 Proof of Theorem 30

Proof. Let I_n be the edgeless graph on n vertices. Proposition 28 gives the same partition limit for I_n and $I_n \sqcup K_{q+1}$, although only the first graph is q-colorable. Hence no continuous extension of the exact-colorability indicator exists.

For the upper-bound claim, take any growing sequence $G_n \to x$. The graphs $G_n \sqcup K_{D+1}$ have the same limit and chromatic number $D+1$. Continuity of U and its finite-graph upper-bound property imply $U(x) \geq D+1$. □