EXACT LOCALITY LIMITS OF MESSAGE PASSING FOR GRAPH-ENERGY OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

How many message-passing layers are required to minimize a graph energy? For sourced Dirichlet energies on a path of diameter D, we determine the exact minimax depth over two opposite endpoint-source orientations. Every K-local algorithm has worst-orientation energy gap at least $\frac{a^2}{2}(D-2K-2)_+$, where $a>0$ is the source amplitude, even with arbitrary nonlinear computation, unbounded messages, source-independent labels and randomization. Source flooding and a source-adapted polynomial graph filter both attain an exact minimizer in $\lfloor (D-1)/2 \rfloor$ rounds, so this is the exact complexity for every target gap below $a^2/2$. On arbitrary weighted graphs, the general bound charges the optimal energy on edges hidden from the sources, and the exact two-orientation local minimax is a least-squares value over the source balls. For K-local message-passing networks with scalar-potential outputs and locally supplied sources, this gives a necessary depth independent of width or training; it is a limit of local communication, not of neural networks in general, and global-attention models, which leave the local class, solve the same task with one or two blocks in our path controls. Trained GCN, GIN and GraphSAGE depth sweeps (10–20 seeds, training-budget ablation), including the Les Misérables and karate-club topologies compared against the exact minimax curve, separate this causal barrier from architecture- and optimization-induced error above it.

1 INTRODUCTION

Dirichlet energies are among the most basic objectives in graph signal processing (GSP): they underpin graph smoothing, interpolation, electrical networks, semi-supervised learning, and diffusion-based graph neural networks (GNNs) (Shuman et al., 2013; Gu et al., 2020; Chamberlain et al., 2021; Fu et al., 2022). Their optimization is straightforward globally, as a centralized linear solver recovers a minimizer after a gauge is fixed, but local optimization is different: information inserted at one vertex cannot affect a distant vertex until messages traverse the intervening edges.

This paper asks a narrow question: can this propagation obstruction be stated as a sharp optimization lower bound inside an actual graph-energy family? The distinction matters. A coordinate chain used in centralized oracle complexity can be consistent with a graph without being a distributed graph problem: a centralized oracle returns the full gradient, and zero-respecting restrictions concern which coordinates an algorithm chooses to activate, whereas a local algorithm stores data at vertices, communicates along edges, and may compute arbitrarily on everything that has arrived. The two models support distinct claims.

Why is this a representation-depth result. A depth-K message-passing network with a nodewise readout is K-local: its output at a vertex relies only on the radius-K input neighborhood. Therefore, a communication lower bound applies regardless of width, nonlinearity, or training. Current work studies expressivity and information compression (Loukas, 2020; Alon & Yahav, 2021; Di Giovanni et al., 2023; Black et al., 2023), and locality-based impossibility is established for graph properties (Garg et al., 2020); here the target is an optimization-error guarantee on both source orientations, with an exact threshold attained by a particular local solver and polynomial filter. The obstruction is information availability under arbitrary local computation, even with unbounded messages; it does not rely on the convergence behavior of a particular iterative method.

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.

We resolve the local question for the sourced Dirichlet energy on a path. The signed endpoint sources create a potential reduction of magnitude a across every edge at the optimum. Reversing both sources yields a second, locally indistinguishable instance. Before either endpoint is observable, an interior edge must take the same value on both orientations, although the two optima require opposite drops. This yields an additive energy cost per hidden edge and a lower bound that shrinks exactly as two wavefronts move inward.

Contributions and scope. The results sit at three levels of generality, and we keep them apart. *(i) Any weighted graph.* Every K-local algorithm pays the optimal energy of the edges hidden from the sources (Theorem 3.1), and the exact two-orientation local minimax value is a least-squares problem over vectors supported on the source balls (Theorem 3.3). *(ii) Closed forms.* On paths the value is $\frac{a^2}{2}(D-2K-2)_+$ and the exact depth is $\lfloor (D-1)/2 \rfloor$, attained by source flooding and by a polynomial filter; the same depth holds on product strips and independent-source spiders, while an asymmetric P_5 shows that hidden-edge accounting alone can be loose. *(iii) Experiments.* Exact minimax curves on eleven topologies; trained GCN, GIN and GraphSAGE on paths up to P_{31}, a grid and the Les Misérables and karate-club topologies, compared with the exact curve V_K, with 10–20 seeds, a training-budget ablation and global-attention controls. The barrier is an information-propagation limit of *local* message passing, not a universal limit of graph neural networks: architectures with global attention lie outside the model and, in our controls, pass the test at depth one or two.

These are not the first lower bounds for distributed Laplacian solving: general-purpose solvers and topology-sensitive analyses are substantially broader (Forster et al., 2022; Anagnostides et al., 2023; Gillani & Bagchi, 2019). Our contribution is an exact energy-gap minimax calculation for a transparent source family, complementing rate bounds for distributed first-order optimization (Scaman et al., 2017) and GNN depth–width bounds (Loukas, 2020). Related work is discussed in Appendix B.

2 Problem and Local Model

2.1 Sourced Dirichlet energy

Let $P_m = (\{1,\dots,m\}, \{\{i, i+1\} : 1 \le i < m\})$ be the unweighted path, let $D = m-1$ be its diameter, and let $L_m = L_{P_m}$ be its combinatorial Laplacian, abbreviated L_m from here on. For amplitude $a > 0$ and source orientation $\sigma \in \{-1, +1\}$, define

$$\begin{aligned} F_\sigma(x) &= \frac{1}{2}\sum_{i=1}^{D}(x_i - x_{i+1})^2 - a\sigma(x_1 - x_m) \\ &= \frac{1}{2}x^\top L_m x - (b^\sigma)^\top x. \end{aligned} \tag{1}$$

where $b^\sigma = a\sigma(e_1 - e_m)$. Since $(b^\sigma)^\top \mathbf{1} = 0$, the energy is bounded below despite the one-dimensional nullspace of L_m. We use the mean-zero gauge $\mathbf{1}^\top x = 0$ only to name a unique minimizer; the energy gap is gauge-invariant.

Proposition 2.1 (Exact geometry)**.** *For every $m \ge 2$, $a > 0$, and $\sigma \in \{-1, +1\}$, the mean-zero minimizer is*

$$x^\star_{\sigma,i} = a\sigma\left(\frac{D}{2} - (i-1)\right), \qquad 1 \le i \le m. \tag{2}$$

For every $x \in \mathbb{R}^m$,

$$F_\sigma(x) - F_{\sigma,\star} = \frac{1}{2}\sum_{i=1}^{D}\left(x_i - x_{i+1} - a\sigma\right)^2, \tag{3}$$

so $F_\sigma(0) - F_{\sigma,\star} = a^2 D/2$. Moreover, F_σ is globally L_m^{sm}-smooth with

$$\begin{aligned} L_m^{\mathrm{sm}} &= 2 + 2\cos(\pi/m) < 4, \\ \lambda_2(L_m) &= 2 - 2\cos(\pi/m), \end{aligned} \tag{4}$$

and satisfies the global PŁ inequality

$$F_\sigma(x) - F_{\sigma,\star} \le \frac{1}{2\lambda_2(L_m)}\|\nabla F_\sigma(x)\|_2^2. \tag{5}$$

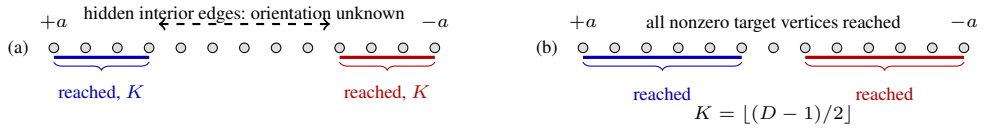


Figure 1: Two source wavefronts create a half-diameter locality barrier. (a) For $K < \lfloor (D-1)/2 \rfloor$, a band of interior edges is outside both radius-K balls; on these edges every K-local algorithm outputs the same value under $\sigma = +1$ and $\sigma = -1$, forcing the energy penalty of Theorem 3.1. (b) At $K = \lfloor (D-1)/2 \rfloor$ the source balls cover every nonzero-target vertex; the sole unreached center for even D has target zero, so SOURCEFLOOD (Section 3.3) attains the exact minimizer.

The edge differences in (3) are the graph gradient. In GSP terms, $x^\top L_m x$ weights graph-Fourier coefficients by Laplacian eigenvalues, while b^σ injects a zero-mean, broadband endpoint signal (Shuman et al., 2013). Equation (2) is the potential generated by a current of magnitude a from one endpoint to the other; D is also their effective resistance on the unweighted path.

2.2 LOCAL ALGORITHMS

The input is distributed as follows. Vertex i initially knows its incident ports, an arbitrary fixed label ℓ_i, the source value b_i^σ, and the source-independent parameter D. Labels may contain unique identifiers, coordinates, or any other side information, but are the same for $\sigma = +1$ and $\sigma = -1$. In one synchronous round, adjacent vertices exchange arbitrary messages and update arbitrary local state. There is no restriction on computation, state dimension, or message size. In short: arbitrary local computation, nonlinear updates, unbounded messages, randomization, fixed labels, and knowledge of D are all *allowed*; a source-dependent global broadcast or preprocessing step, global attention, and asynchronous or unbounded-delay communication are *not*; the excluded items are explained in Section 4.

Definition 2.2 (K-local output)**.** *A deterministic algorithm is K-local if its output at vertex i after K rounds is a deterministic function of the rooted radius-K labeled neighborhood*

$$\big(P_m[i,K],\, i,\, (\ell_j, b_j^\sigma)_{j \in P_m[i,K]},\, D\big).$$

We write its joint output as $x_\sigma^{(K)} \in \mathbb{R}^m$.

This neighborhood formulation is equivalent to K synchronous rounds in the usual LOCAL model (Linial, 1992). Allowing D strengthens the algorithm and makes the matching upper bound transparent. What is excluded is precisely a source-dependent global broadcast: a node cannot know the endpoint sign before a source lies in its communication ball.

For accuracy $\varepsilon \geq 0$, define the two-instance minimax complexity

$$K_\varepsilon^{\mathrm{loc}}(m,a) := \min\big\{K : \exists\ K\text{-local } \mathcal{A},\ \max_{\sigma=\pm 1}[F_\sigma(x_\sigma^{(K)}) - F_{\sigma,\star}] \leq \varepsilon\big\}. \tag{6}$$

The randomized counterpart, denoted $K_\varepsilon^{\mathrm{rand}}(m,a)$, permits arbitrary private or shared randomness independent of σ and places an expectation around the energy gap in (6).

3 LOCAL LOWER BOUND AND EXACT PATH COMPLEXITY

3.1 HIDDEN OPTIMAL ENERGY ON A GENERAL GRAPH

The path calculation is a special case of a graph-wide statement. Let $G = (V, E, w)$ be any connected graph with positive edge weights $w_e > 0$, orient its edges arbitrarily, write B for its signed incidence matrix, $W = \mathrm{diag}((w_e)_{e \in E})$, and $L_G = B^\top W B$. Let $b \in \mathbb{R}^V \setminus \{\mathbf{0}\}$ be any balanced source, $\mathbf{1}^\top b = 0$, with source set $S = \mathrm{supp}(b)$. Throughout, $\mathrm{dist}_G(u,v)$ is the *unweighted hop distance* in (V,E) (the number of edges on a shortest path, ignoring w): one edge is traversed per round whatever its weight, so this is the quantity that governs locality. Define

$$F_\sigma^G(x) = \tfrac{1}{2} x^\top L_G x - \sigma b^\top x. \tag{7}$$

Let

$$\phi = L_G^{\dagger} b, \qquad g = B\phi, \tag{8}$$

so the mean-zero optima are $\sigma\phi$ and g_e is the oriented optimal drop on edge e. Extend Definition 2.2 verbatim to G, with edge weights included in the fixed labels. Define the source-hidden edges

$$\mathcal{H}_K = \big\{\{u,v\} \in E : \ \mathrm{dist}_G(z,r) > K \text{ for every } z \in \{u,v\} \text{ and every } r \in S\big\}. \tag{9}$$

Theorem 3.1 (Hidden-energy lower bound)**.** *Every deterministic K-local algorithm satisfies*

$$\max_{\sigma=\pm 1}[F_\sigma^G(x_\sigma^{(K)}) - F_{\sigma,\star}^G] \geq \frac{1}{2}\sum_{e\in\mathcal{H}_K} w_e g_e^2. \tag{10}$$

The same bound holds in expectation for randomized local algorithms whose private or shared random bits are independent of σ.

Proof. Completing the quadratic around the mean-zero minimizer gives

$$F_\sigma^G(x) - F_{\sigma,\star}^G = \tfrac{1}{2}\|Bx - \sigma g\|_W^2.$$

For a hidden oriented edge $e = (u,v)$, both endpoint local views exclude all nonzero source labels. Hence their outputs, and thus the edge difference $d_e = x_u^{(K)} - x_v^{(K)}$, are the same on the two orientations. The loss of this edge averaged over a uniform orientation is

$$\tfrac{1}{4}w_e[(d_e - g_e)^2 + (d_e + g_e)^2] = \tfrac{1}{2}w_e(d_e^2 + g_e^2) \geq \tfrac{1}{2}w_e g_e^2.$$

Summing over $\mathcal{H}_K$ shows that the right-hand side of (10) lower-bounds the *uniform σ-average* of the gap, which is stronger than the stated maximum. For a randomized algorithm, condition on all random bits, apply the deterministic average-form argument, and then average over those bits; this is why the randomized claim needs no further work. □

At $K = 0$, the sum over all edges except those incident to S is the portion of the optimal Dirichlet energy initially invisible to the sources. The full optimal energy is $\sum_e w_e g_e^2 = b^\top L_G^{\dagger} b$; for $b = a(e_s - e_t)$ this equals $a^2 R_{\mathrm{eff}}(s,t)$. Thus (10) connects directly to effective resistance while automatically discounting hidden edges on which the target potential is nearly constant.

Remark 3.2 (What is essential)**.** *Theorem 3.1 permits any balanced, localized b, extends to n independently signed dipoles (Section A.2) and to separable even convex edge energies via a Bregman bound (Lemma F.2); Appendix C treats weighted-tree and complete-graph extremes beyond diameter.*

Theorem 3.1 is a relaxation. For the two-orientation family the exact minimax gap on *every* graph is a least-squares value. Write $\mathrm{dist}_G(v,S) = \min_{r\in S}\mathrm{dist}_G(v,r)$ and let $Z_K = \{z \in \mathbb{R}^V : \ z_v = 0 \text{ whenever } \mathrm{dist}_G(v,S) > K\}$ be the vectors supported on the union of the radius-K source balls.

Theorem 3.3 (Exact local minimax on any graph)**.** *Let $V_K = \min_{z\in Z_K} \frac{1}{2}\|Bz - g\|_W^2$. Every deterministic K-local algorithm, and every randomized one in expected gap, satisfies $\max_{\sigma=\pm1}[F_\sigma^G(x_\sigma^{(K)}) - F_{\sigma,\star}^G] \geq V_K$, and some K-local algorithm using source-independent graph-specific labels permitted by Definition 2.2 attains equality. Moreover $V_K \geq \frac{1}{2}\sum_{e\in\mathcal{H}_K} w_e g_e^2$, and V_K is non-increasing in K.*

The attainment statement is information-theoretic and non-uniform: the graph-specific labels may encode precomputed source-independent information. It does not provide a generic distributed solver.

Proof. Lower bound. Let x_σ be a deterministic K-local output and set $z = \frac{1}{2}(x_+ - x_-)$. If $\mathrm{dist}_G(v,S) > K$ the views of v under $\sigma = \pm 1$ coincide, so $x_{+,v} = x_{-,v}$ and $z_v = 0$; thus $z \in Z_K$. With $p = Bx_+ - g$ and $q = -Bx_- - g$ we have $Bz - g = \frac{1}{2}(p+q)$, and by convexity of $\|\cdot\|_W^2$,

$$V_K \leq \tfrac{1}{2}\big\|\tfrac{1}{2}(p+q)\big\|_W^2 \leq \tfrac{1}{4}(\|p\|_W^2 + \|q\|_W^2) = \tfrac{1}{2}\big(\mathrm{gap}_+ + \mathrm{gap}_-\big) \leq \max_\sigma \mathrm{gap}_\sigma,$$

using $\mathrm{gap}_\sigma = \frac{1}{2}\|Bx_\sigma - \sigma g\|_W^2$. For a randomized algorithm apply this to each realization of the coins, which are shared by the two orientations, and average. *Attainment.* Let z^K minimize $\frac{1}{2}\|Bz - g\|_W^2$ over Z_K and store z_v^K in the label of v (labels are source-independent). Vertex v outputs σz_v^K if a

SOURCEFLOOD at vertex i.
Initialize: If $b_i^\sigma \neq 0$, store $(s, r, \delta) = (b_i^\sigma, 0, D)$; otherwise store nothing.
For rounds $1, \ldots, K$**:** Send the stored tuple to every neighbor. On first receipt of (s, r, δ), store $(s, r + 1, \delta)$. Ties may be broken arbitrarily.
Output: If (s, r, δ) is stored, return $x_i = s(\delta/2 - r)$; otherwise return $x_i = 0$.

Figure 2: A local exact solver for the two-orientation path family. If both sources arrive simultaneously, either pair gives the same value.

source lies in its radius-K ball, where $\sigma = b_r^\sigma / b_r^+$ is read off any visible source r, and 0 otherwise. Then $x_\sigma = \sigma z^K \in Z_K$ and both orientations incur $\frac{1}{2}\|Bz^K - g\|_W^2 = V_K$. *Relaxation.* For $z \in Z_K$ every edge of $\mathcal{H}_K$ has $(Bz)_e = 0$, so its term in $\frac{1}{2}\|Bz - g\|_W^2$ equals $\frac{1}{2} w_e g_e^2$; the remaining terms are nonnegative. Monotonicity holds because $Z_K \subseteq Z_{K+1}$. □

The attaining algorithm stores its answer in the labels, so it certifies that V_K cannot be improved within the model; it is not a practical solver (SOURCEFLOOD below realizes the same value with generic labels on paths, strips and spiders). Theorem 3.3 settles what the hidden-energy bound left open: the exact minimax of a two-orientation family is one small linear solve on any graph, and Theorem 3.1 is the part of V_K coming from edges with both endpoints outside the source balls. Section 5.1 evaluates V_K on grids, small-world, scale-free, geometric, tree and real networks.

3.2 PATH SPECIALIZATION BY SOURCE INDISTINGUISHABILITY

An edge $\{i, i+1\}$ is source-hidden after K rounds when both endpoint K-balls exclude vertices 1 and m. The set of such central edges is

$$\mathcal{E}_K^{\mathrm{cen}} = \{\{i, i+1\} : K + 2 \leq i \leq m - K - 2\},$$
$$|\mathcal{E}_K^{\mathrm{cen}}| = (D - 2K - 2)_+. \tag{11}$$

Theorem 3.4 (Orientation lower bound)**.** *For every deterministic K-local algorithm in Definition 2.2,*

$$\max_{\sigma=\pm 1} \left[F_\sigma(x_\sigma^{(K)}) - F_{\sigma,\star}\right] \geq \frac{a^2}{2}(D - 2K - 2)_+. \tag{12}$$

Proof. For P_m, Equation (2) gives $|g_e| = a$ on every edge, and $\mathcal{H}_K = \mathcal{E}_K^{\mathrm{cen}}$. Substitute the edge count (11) into Theorem 3.1. □

The randomized counterpart is Corollary D.1 (Appendix D). The argument is local indistinguishability, not a support argument: an algorithm may output nonzero values everywhere at round zero and use identifiers, attention over neighbors or nonlinearities, but on a source-hidden edge these are the same for both orientations, and one common drop cannot match $+a$ and $-a$.

3.3 A MATCHING TOPOLOGY-AWARE ALGORITHM

The lower bound suggests the right algorithm: propagate source information from both ends. Each message contains a signed source value and its hop distance.

Theorem 3.5 (Matching upper bound)**.** *For*

$$K^\star = \left\lfloor \frac{D-1}{2} \right\rfloor, \tag{13}$$

SOURCEFLOOD *returns the mean-zero minimizer $x_\sigma^\star$ for both orientations. Consequently, for every $0 \leq \varepsilon < a^2/2$,*

$$\boxed{K_\varepsilon^{\mathrm{loc}}(m, a) = \left\lfloor \frac{D-1}{2} \right\rfloor.} \tag{14}$$

The same equality holds for $K_\varepsilon^{\mathrm{rand}}(m, a)$ under expected gap.

Proof. If vertex i receives the left source, then $s = a\sigma$ and $r = i - 1$, so the output is exactly (2). If it receives the right source, then $s = -a\sigma$ and $r = m - i$, which gives the same value. If D is odd, every vertex is within $\lfloor D/2 \rfloor = K^\star$ hops of an endpoint. If D is even, $K^\star = D/2 - 1$ reaches every vertex except the unique center; its exact value in (2) is zero, which is the default output. Thus the upper bound is exact.

For the lower bound, any integer $K < K^\star$ satisfies $(D - 2K - 2)_+ \geq 1$. Theorem 3.4 then gives worst-case gap at least $a^2/2 > \varepsilon$. Hence no such deterministic algorithm attains the target. Corollary D.1 gives the identical conclusion for expected randomized performance. □

Why half the diameter? Knowing either signed endpoint source and its distance determines the affine solution, so two wavefronts cover the path in half its diameter; an edge is constrained once a source reaches either endpoint, which explains the two units of slack in (12). Appendix D discusses knowing D and the coarser-accuracy tradeoff.

Extensions beyond one bit and trees. The same exact depth holds on Cartesian-product strips with equipotential fibers (Proposition A.1) and on spiders carrying n independently signed dipoles (Theorem A.2). The respective accuracy thresholds are $a^2/(2q)$ for a fiber of size q and $na^2/2$ for n legs. Appendix A gives the constructions and assumptions. The general hidden-energy bound need not be tight:

3.4 The general bound is not always attained

Reflection symmetry is not cosmetic: without it, Theorem 3.1 can be strictly loose.

Proposition 3.6 (Asymmetric sources defeat tightness)**.** *Let $G = P_5$ with unit weights and asymmetric balanced source $b = a(e_1 - e_4)$, so that vertex 5 is a dangling non-source vertex. At $K = 0$ the hidden-energy bound* (10) *equals $a^2/2$, yet every 0-local algorithm, deterministic or randomized in expected gap, satisfies*

$$\max_{\sigma=\pm 1} [F_\sigma^G(x_\sigma^{(0)}) - F_{\sigma,\star}^G] \;\geq\; \frac{3a^2}{4},$$

and this value is attained, so the exact minimax gap is $3a^2/4$.

The proof (Appendix F.7) identifies an additional boundary conflict: source vertex 4 must accommodate two source-free neighbors whose required drops differ. Hidden-edge accounting misses this cost; in the language of Theorem 3.3 the value is $V_0 = 3a^2/4$ against the hidden-energy bound $a^2/2$, and the conflict is the difference.

4 Relation to Optimization and GNN Models

4.1 Generic local baselines: gradient descent and Chebyshev acceleration

Gradient descent with step $\eta \leq 1/L_m^{\mathrm{sm}}$ takes the local form

$$x_i^{(k+1)} = x_i^{(k)} - \eta \left(\deg(i) x_i^{(k)} - \sum_{j \sim i} x_j^{(k)} - b_i^\sigma \right). \tag{15}$$

Thus each iteration is one local round, and by (5) its slow mode contracts on the $1/\lambda_2(L_m) = \Theta(D^2)$ scale. But stationary gradient descent is a weak representative of the generic local methods admitted by Definition 2.2: a degree-K polynomial graph filter is a linear K-local method of the form $x^{(K)} = p(L_m)\, b^\sigma$ with $\deg p \leq K$, the classical shift-invariant filter class of graph signal processing (Shuman et al., 2013; Hammond et al., 2011; Defferrard et al., 2016) (degree bounds the communication radius). General linear local maps need not be polynomials in L_m. A uniform spectral residual bound is achieved round by round via classical Chebyshev semi-iteration (Golub & Varga, 1961) with coefficients depending only on m.

Proposition 4.1 (Chebyshev semi-iteration baseline)**.** *Let $m \geq 3$, $0 < \varepsilon < a^2 D/2$, and $\kappa_m = L_m^{\mathrm{sm}}/\lambda_2(L_m) = \cot^2(\pi/(2m))$. Chebyshev semi-iteration on* (1) *with coefficients depending only on*

m *uses one local round per iteration and satisfies, for both orientations,*

$$F_\sigma(x_\sigma^{(K)}) - F_{\sigma,\star} \le 2a^2 D \left(\frac{\sqrt{\kappa_m} - 1}{\sqrt{\kappa_m} + 1} \right)^{2K}. \tag{16}$$

Hence $\max_\sigma \mathrm{gap} \le \varepsilon$ *within* $\lceil (\sqrt{\kappa_m}/4) \ln(2a^2 D/\varepsilon) \rceil \le \lceil (m/(2\pi)) \ln(2a^2 D/\varepsilon) \rceil$ *rounds.*

The proof is the textbook Chebyshev bound on the mean-zero subspace (Appendix F.8), so a topology-agnostic local method reaches any fixed accuracy in $O(D \log(a^2 D/\varepsilon))$ rounds; yet Theorem 3.5 still applies: for $\varepsilon < a^2/2$ no local method beats $\lfloor (D-1)/2 \rfloor$ rounds. Linear methods can, however, reach that threshold exactly:

Proposition 4.2 (Linear filters attain the exact complexity)**.** *Let* $v_k(i) = \cos\big((2i-1)k\pi/(2m)\big)$ *and* $\lambda_k = 2 - 2\cos(k\pi/m)$, $k = 0, \dots, m-1$, *be the eigenpairs of* L_m. *Then* $\langle v_k, b^\sigma \rangle = 0$ *for every even* k *and* $\langle v_k, b^\sigma \rangle \neq 0$ *for every odd* k. *Consequently the interpolating polynomial* p_m *of* $\lambda \mapsto 1/\lambda$ *at the* $\lceil D/2 \rceil$ *odd eigenvalues has degree* $\lceil D/2 \rceil - 1 = \lfloor (D-1)/2 \rfloor = K^\star$, *coefficients depending only on* m, *and satisfies* $p_m(L_m)\, b^\sigma = x_\sigma^\star$ *for both orientations: a* $K^\star$*-local linear method. Conversely, no polynomial filter of degree less than* $K^\star$ *is exact for either orientation.*

The proof is in Appendix F.9. Reflection antisymmetry of b^σ excites only the odd modes, halving the naive degree $m-2$ to exactly the causal threshold of Theorem 3.4: the exact minimax complexity is attained inside the classical linear graph-filter class, and SOURCEFLOOD buys robustness instead (Appendix D).

4.2 MESSAGE-PASSING INTERPRETATION

A depth-K message-passing GNN is K-local when its only source-dependent input is the vertex signal b^σ, so Theorem 3.4 applies to any deterministic architecture with a scalar-potential readout, whatever its width or nonlinear aggregation: an input-output approximation bound, not a training bound. Global positional encodings depending only on P_m and identifiers are permitted as fixed labels; preprocessing that already encodes σ is not, since it performs the communication being counted, and global attention falls outside the local model (Loukas, 2020; Alon & Yahav, 2021).

What this does and does not say about GNNs. The result isolates information that has not arrived at all, rather than information compressed into a narrow representation; width cannot repair a missing causal path. It concerns *local* message passing: architectures whose reach is decoupled from graph geometry, such as Graph Transformers with global attention (Rampášek et al., 2022), are outside the model, and Section 5.2 confirms on the path controls that they solve the same task with one or two blocks. Conversely, sufficient causal depth does not ensure that an optimizer finds the exact solver.

4.3 CENTRALIZED ZERO-CHAIN BOUNDS ARE DIFFERENT

A centralized oracle reveals the full gradient in one call. Zero-respecting oracle bounds constrain coordinate activation, whereas our model constrains communication and allows arbitrary local computation. Graph-compatible zero chains therefore do not by themselves imply our communication bound (Carmon et al., 2020); further context is in Appendix B.

5 EXPERIMENTS

5.1 DETERMINISTIC VALIDATION

The theory is exact and needs no empirical assumption. We nevertheless run seven deterministic checks on the exact, gauge-invariant gap (Appendix G): the identities (3), exactness of SOURCEFLOOD at $K^\star$ for $3 \le m \le 39$, comparison with gradient descent and Chebyshev semi-iteration, the weighted bound of Theorem 3.1, Propositions 3.6, 4.1 and 4.2, Theorem A.2, and a sweep of the best degree-K polynomial filter for $K \le 180$, which reaches machine zero exactly at $K^\star = 30$. These isolate approximation error; the trained-network experiments additionally introduce architecture and optimization effects.

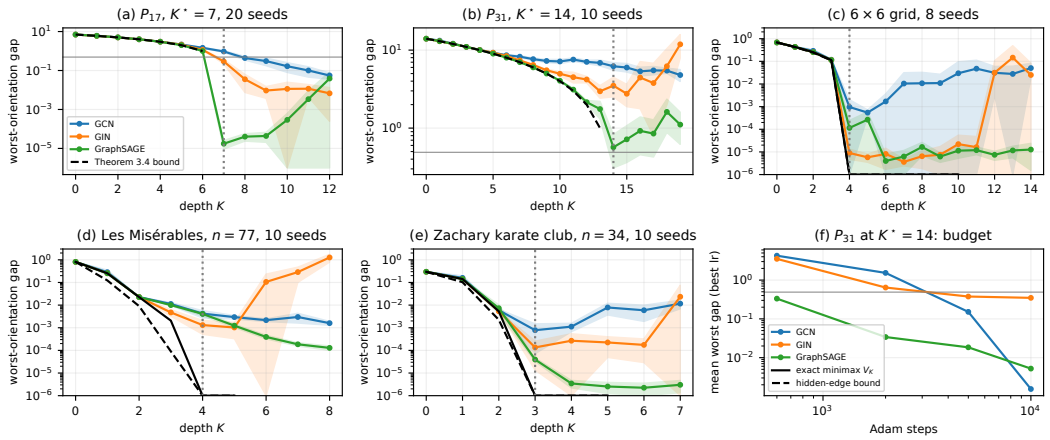


Figure 3: Trained local networks: mean worst-orientation gap, 95% confidence bands over seeds. (a,b) Paths, with the hidden-energy bound of Theorem 3.4 (dashed, zero from $K^\star$ on) and target 0.49. (c–e) Sourced Dirichlet energy on a 6×6 grid and on the topologies of Les Misérables and Zachary's karate club (unit weights, farthest-pair sources): the solid black curve is the exact minimax value V_K of Theorem 3.3, the dashed curve the hidden-edge bound, the dotted line the smallest K with $V_K = 0$. Below it the networks sit on or above V_K (every run is $\geq V_K$); once $V_K = 0$, any remaining gap is architecture- and/or optimization-induced rather than causally necessary. (f) Training budget for the P_{31} runs at $K^\star$ (best learning rate per budget). Values below 10^{-6} are clipped.

We use $a = 1$ and target gap $0.49 < a^2/2$. For $m \in \{15, 31, 63, 127, 255\}$ the exact local round counts are $6, 14, 30, 62, 126$; gradient descent reaches the target after $114, 630, 3177, 15210, 70512$ rounds and Chebyshev semi-iteration after $10, 24, 55, 125, 279$ rounds, within its analytic caps from Proposition 4.1 and a factor 2.3 of flooding (log–log slopes 1.048, 2.218, 1.149, descriptive only; Figure 5, Appendix G.1). On a weighted path with conductances $w_i \in [0.362, 2.763]$, at $K = 15$ the theorem certifies 0.443 of the initial gap versus 0.964 retained by gradient descent.

Exact minimax beyond paths. We evaluate V_K and the hidden-edge bound on eleven topologies (Table 1, Appendix E): grids, a strip, Watts–Strogatz, Barabási–Albert, a random tree and geometric graph, Zachary's karate club and Les Misérables. The checks recover $V_K = \frac{a^2}{2}(D - 2K - 2)_+$ on P_{31}, the value $3a^2/4$ on the asymmetric P_5 and the strip law, and V_K is at least the hidden-edge bound everywhere. The boundary-conflict share of the initial gap is at most 8% except on the random tree (19%) and the asymmetric P_5; the exact-solution depth need not equal the distance (6 versus 14 on an 8×8 grid, since the potential is constant on the unreached anti-diagonal).

5.2 Trained message-passing diagnostic

This diagnostic asks three questions: do learned local networks respect the causal barrier; is a failure above the barrier architectural or an optimization artifact; and does the barrier disappear for a model outside the local class? We train GCN (Kipf & Welling, 2017), GIN (Xu et al., 2019) and mean-aggregation GraphSAGE (Hamilton et al., 2017) to output node potentials; one shared parameter set sees both orientations, only the local source input changes with σ, fixed features are normalized position, degree and diameter, and there is no global readout or source-dependent preprocessing. Width is 32; training is full-batch Adam on the mean of the two energy gaps (600 steps, rate 0.003 unless stated), evaluating the lowest-training-loss iterate on the same two instances, so this is a finite-family representation diagnostic, not a held-out benchmark. We use 20 seeds on P_9, P_{10}, P_{17}, 10 on P_{31} ($K^\star = 14$), and report 95% confidence intervals for the seed mean (protocol: Appendix G.1).

Causal limit or optimization budget? We rerun depths $K^\star - 1, \ldots, K^\star + 3$ for 10 000 steps with rates $\{3 \cdot 10^{-3}, 10^{-2}, 3 \cdot 10^{-2}\}$ and 6 seeds; the value at N steps is the best iterate within N steps, rate chosen per budget by seed mean (a disclosed search; Figure 6). On P_{17} at $K^\star = 7$, GCN's mean worst gap falls from 0.278 (600 steps) through 0.023 (2000) to 2×10^{-6} (5000), all seeds succeeding:

its 600-step failure was optimization, not a representational limit. On P_{31} GCN behaves alike (all seeds by 5000 steps) and GraphSAGE succeeds by 2000 steps. GIN plateaus at 0.35 with $4/6$ seeds at the target, two failing at every budget and rate, so there we cannot separate an optimization plateau from an architectural limit. Below $K^\star$ no budget helps.

Learned models on real topologies, against the exact curve. We pose the same task on the topologies of Les Misérables ($n = 77$) and Zachary's karate club ($n = 34$), both of diameter 5: unit weights, unit opposite sources at the farthest pair, and fixed source-independent labels (degree, diameter, two lowest nonzero Laplacian eigenvectors). We compute V_K (Theorem 3.3) before training, then train the three architectures (10 seeds, 3000 steps). The exact values are $V_K = 0.810, 0.255, 0.023, 0.0021, 0$ for $K = 0, \dots, 4$ (Les Misérables, initial gap 1.81) and $0.296, 0.142, 0.0053, 0$ for $K = 0, \dots, 3$ (karate club, initial gap 0.646): exact solvability needs depth 4 and 3 although the sources are 5 hops apart. All 510 runs lie at or above V_K, and for $K \leq 2$ GIN and GraphSAGE match it to 2×10^{-3} or better. At the exact depth the residual gap is small but positive ($\leq 0.23\%$ of the initial gap on Les Misérables, $\leq 0.13\%$ on the karate club, all architectures). Since $V_K = 0$, this residual is not causally necessary and reflects architecture and/or optimization effects. Deeper networks do not help: GIN on Les Misérables reaches 1.27 at $K = 8$, worse than depth zero, again with $V_K = 0$. On the 6×6 grid GIN's mean worst gap is $0.683, 0.433, 0.254, 0.115, 0.000$, equal to V_K (GraphSAGE within 10^{-3}, GCN slightly above), with exact depth 4 against source distance 10. Trained networks thus closely track, and in several settings attain, the value of Theorem 3.3, rather than merely the weaker hidden-edge bound.

Why this matters for graph learning. Sourced Dirichlet energies underlie label propagation and harmonic extension, graph-regularized semi-supervised inference and diffusion-based or implicit GNN layers (Shuman et al., 2013; Gu et al., 2020; Chamberlain et al., 2021; Fu et al., 2022); two opposite sources is the smallest instance. Theorem 3.3 turns depth selection into a computation done before training: the least K with $V_K \leq \varepsilon$ is necessary for every local architecture and can be far below the diameter or source distance (grid: 4 versus 10). It also separates causal from model-specific failure: below that depth no width or training can overcome the locality barrier; above it, excess over V_K is attributable to architecture and/or optimization. The budget ablation shows that this excess is purely optimization-induced for several GCN and GraphSAGE settings, while the P31 GIN result leaves the two causes unresolved. The scope is narrow: unit-weight graphs, a two-orientation family, small graphs, energy gap rather than accuracy.

Global attention leaves the model. Table 2 (Appendix G.1) trains a Transformer and a GCN-plus-attention hybrid (GPS-style (Rampášek et al., 2022)) of one or two blocks on P_{17}, P_{31}, P_{63} ($K^\star = 7, 14, 30$). All reach gap below 6×10^{-4}; the far end's output responds to the source input, so these models are not K-local for $K < m - 1$ and the theorem does not apply.

6 Limitations and Broader Impact

The closed-form characterizations cover paths, the product strips of Section A.1 and the spiders of Section A.2. Theorem 3.3 computes the exact two-orientation value on any graph, but closed forms for general trees, irregular weighted graphs and nonsymmetric source patterns, and exact values for families of many correlated orientations, remain open; randomization is covered only in expected gap. The threshold measures causal reach, not linearity (Propositions 4.1 and 4.2), so the contrast with plain gradient descent is not the generic price of topology-agnosticism.

7 Conclusion

Exact local minimax gaps of sourced Dirichlet energies are least-squares values over source balls, in closed form on paths, strips and spiders; trained local networks, including on real topologies, track them, and global attention leaves the model.

AI USE STATEMENT

During the preparation of this manuscript, the author(s) utilized AI assistance tools solely for language polishing, grammar correction, and improving stylistic clarity. All conceptual contributions, methodology, theoretical derivations, experimental designs, and data analyses were conceived, performed, and verified entirely by the human authors. The authors have thoroughly reviewed, edited, and approved all generated text, and take full responsibility for the contents of this publication.

REPRODUCIBILITY STATEMENT

Proofs accompany the main results or appear in the appendices. The supplementary scripts reproduce the deterministic checks (including the exact-minimax study on 11 graphs), the seeded GNN diagnostics, the training-budget ablation, the global-attention controls and the grid experiment, and an audit asserts that every trained run respects the theorems. Raw per-run results, configuration, and software versions are included, with predictions, histories and checkpoints for the base sweep. Appendix G specifies the protocol.

REFERENCES

Uri Alon and Eran Yahav. On the bottleneck of graph neural networks and its practical implications. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=i80OPhOCVH2.

Ioannis Anagnostides, Christoph Lenzen, Bernhard Haeupler, Goran Zuzic, and Themis Gouleakis. Almost universally optimal distributed laplacian solvers via low-congestion shortcuts. *Distributed Computing*, 36(4):475–499, 2023. doi: 10.1007/s00446-023-00454-0.

Alexandr Andoni, Robert Krauthgamer, and Yosef Pogrow. On solving linear systems in sublinear time. In *10th Innovations in Theoretical Computer Science Conference (ITCS 2019)*, volume 124 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 3:1–3:19. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. doi: 10.4230/LIPIcs.ITCS.2019.3.

Mitchell Black, Zhengchao Wan, Amir Nayyeri, and Yusu Wang. Understanding oversquashing in GNNs through the lens of effective resistance. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 2528–2547. PMLR, 2023. URL https://proceedings.mlr.press/v202/black23a.html.

Yair Carmon, John C. Duchi, Oliver Hinder, and Aaron Sidford. Lower bounds for finding stationary points i. *Mathematical Programming*, 184(1–2):71–120, 2020. doi: 10.1007/s10107-019-01406-y.

Benjamin Paul Chamberlain, James Rowbottom, Maria I. Gorinova, Michael M. Bronstein, Stefan Webb, and Emanuele Rossi. GRAND: Graph neural diffusion. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 1407–1418. PMLR, 2021. URL https://proceedings.mlr.press/v139/chamberlain21a.html.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*, volume 29, pp. 3844–3852, 2016.

Francesco Di Giovanni, Lorenzo Giusti, Federico Barbero, Giulia Luise, Pietro Liò, and Michael M. Bronstein. On over-squashing in message passing neural networks: The impact of width, depth, and topology. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 7865–7885. PMLR, 2023. URL https://proceedings.mlr.press/v202/di-giovanni23a.html.

Vijay Prakash Dwivedi, Ladislav Rampášek, Michael Galkin, Ali Parviz, Guy Wolf, Anh Tuan Luu, and Dominique Beaini. Long range graph benchmark. In *Advances in Neural Information Processing Systems, Datasets and Benchmarks Track*, volume 35, 2022.

Sebastian Forster, Gramoz Goranci, Yang P. Liu, Richard Peng, Xiaorui Sun, and Mingquan Ye. Minor sparsifiers and the distributed laplacian paradigm. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 989–999. IEEE, 2022. doi: 10.1109/FOCS52979.2021.00099.

Guoji Fu, Peilin Zhao, and Yatao Bian. p-laplacian based graph neural networks. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 6878–6917. PMLR, 2022. URL https://proceedings.mlr.press/v162/fu22e.html.

Vikas 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*, pp. 3419–3430. PMLR, 2020. URL https://proceedings.mlr.press/v119/garg20c.html.

Iqra Altaf Gillani and Amitabha Bagchi. A distributed laplacian solver and its applications to electrical flow and random spanning tree computation. *arXiv preprint arXiv:1905.04989*, 2019. URL https://arxiv.org/abs/1905.04989.

Gene H. Golub and Richard S. Varga. Chebyshev semi-iterative methods, successive overrelaxation iterative methods, and second order richardson iterative methods. part i. *Numerische Mathematik*, 3:147–156, 1961. doi: 10.1007/BF01386013.

Fangda Gu, Heng Chang, Wenwu Zhu, Somayeh Sojoudi, and Laurent El Ghaoui. Implicit graph neural networks. In *Advances in Neural Information Processing Systems*, volume 33, pp. 11984–11995, 2020.

William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, volume 30, pp. 1024–1034, 2017.

David K. Hammond, Pierre Vandergheynst, and Rémi Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011. doi: 10.1016/j.acha.2010.04.005.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=SJU4ayYgl.

Nathan Linial. Locality in distributed graph algorithms. *SIAM Journal on Computing*, 21(1):193–201, 1992. doi: 10.1137/0221015.

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.

Saeed Masiha, Negar Kiyavash, and Patrick Thiran. Sharp first-order lower bounds under sublevel α-polyak–łojasiewicz conditions. *arXiv preprint arXiv:2606.28278*, 2026. URL https://arxiv.org/abs/2606.28278.

Arkadi S. Nemirovsky and David B. Yudin. *Problem Complexity and Method Efficiency in Optimization*. John Wiley & Sons, Chichester, 1983.

Yurii Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*, volume 87 of *Applied Optimization*. Kluwer Academic Publishers, 2004. doi: 10.1007/978-1-4419-8853-9.

Alex Olshevsky and John N. Tsitsiklis. Convergence speed in distributed consensus and averaging. *SIAM Journal on Control and Optimization*, 48(1):33–55, 2009. doi: 10.1137/060678324.

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, 2022.

Kevin Scaman, Francis Bach, Sébastien Bubeck, Yin Tat Lee, and Laurent Massoulié. Optimal algorithms for smooth and strongly convex distributed optimization in networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 3027–3036. PMLR, 2017. URL https://proceedings.mlr.press/v70/scaman17a.html.

David I. Shuman, Sunil K. Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013. doi: 10.1109/MSP.2012.2235192.

Shreyas Sundaram and Christoforos N. Hadjicostis. Finite-time distributed consensus in graphs with time-invariant topologies. In *2007 American Control Conference*, pp. 711–716. IEEE, 2007. doi: 10.1109/ACC.2007.4282726.

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.

Pengyun Yue, Cong Fang, and Zhouchen Lin. On the lower bound of minimizing polyak–łojasiewicz functions. In *Proceedings of the Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pp. 2948–2968. PMLR, 2023. URL https://proceedings.mlr.press/v195/yue23a.html.

A Tight Extensions

A.1 A tight non-tree family

The exact phenomenon is not confined to trees. Let H be any connected unweighted graph with $q = |V(H)|$, and form the Cartesian-product strip $G = P_m \square H$. Its layers are $\{i\} \times V(H)$, and consecutive layers are joined by a matching of q edges. Put balanced fiber sources

$$b = \frac{a}{q} \sum_{h \in V(H)} \big(e_{(1,h)} - e_{(m,h)}\big). \tag{17}$$

Proposition A.1 (Exact complexity on product strips)**.** *For every connected H, the mean-zero optimum is*

$$x^\star_{\sigma,(i,h)} = \frac{a\sigma}{q}\left(\frac{D}{2} - (i-1)\right).$$

If D is available at the source fibers, then for every $0 \le \varepsilon < a^2/(2q)$ the deterministic and expected-randomized local complexities both equal $\lfloor (D-1)/2 \rfloor$. No separate knowledge of q is needed: the fiber size enters only through the locally known source values $\pm a\sigma/q$.

The proof is in Appendix F.1.

For $q \ge 2$ and $m \ge 2$, these strips contain cycles, giving a non-tree family on which the hidden-energy bound and flooding upper bound are simultaneously tight (Appendix D discusses the structural property responsible).

A.2 Pricing n independent bits: a spider family

On the path, the two instances differ in a single bit. The hidden-energy calculus is not limited to one bit: it prices independent source bits additively, and the exact law survives. For $n \ge 1$ and odd $D_0 \ge 3$, let $\mathrm{Sp}(n, D_0)$ be the unit-weight spider (star of paths) with hub c and legs $c = u_{j,0}, u_{j,1}, \dots, u_{j,D_0+1}$ for $1 \le j \le n$. Put on leg j the dipole $s_j = u_{j,1}$, $t_j = u_{j,D_0+1}$, and take

$$b^{\boldsymbol{\sigma}} = a \sum_{j=1}^{n} \sigma_j \big(e_{s_j} - e_{t_j}\big), \qquad \boldsymbol{\sigma} \in \{-1,+1\}^n. \tag{18}$$

All 2^n instances share every label; drawing $\boldsymbol{\sigma}$ uniformly makes the source distribution carry n independent bits.

Theorem A.2 (Exact n-bit law). *(i) Up to the global constant, the minimizer is $x^\star_c = 0$ and $x^\star_{u_{j,r}} = -a\sigma_j(r-1)$ for $1 \le r \le D_0 + 1$ (the formula is not evaluated at $r = 0$: that index is the shared hub $u_{j,0} = c$, whose value is 0 for every leg, consistent with $x^\star_{u_{j,1}} = 0$); the optimal drop is $a\sigma_j$ on the D_0 segment edges of leg j between s_j and t_j, and 0 on hub edges. (ii) Every K-local algorithm, deterministic or randomized with coins independent of $\boldsymbol{\sigma}$, satisfies*

$$\mathbb{E}_{\boldsymbol{\sigma}}\left[F_{\boldsymbol{\sigma}}(x^{(K)}) - F_{\boldsymbol{\sigma},\star}\right] \;\ge\; \frac{a^2}{2}\, n\,(D_0 - 2K - 2)_+$$

for $\boldsymbol{\sigma}$ uniform on $\{-1,+1\}^n$; the maximum over $\boldsymbol{\sigma}$ obeys the same bound. (iii) A flooding algorithm using source-independent position labels and D_0 is exact for all 2^n orientations after $K^\star = (D_0 - 1)/2$ rounds. Consequently, for every $0 \le \varepsilon < na^2/2$, the deterministic and expected-randomized minimax complexities over the 2^n-instance family both equal $(D_0 - 1)/2$.

The proof (Appendix F.5) conditions on σ_{-j} and applies the two-point average of Theorem 3.1 leg by leg. This refines the one-bit reading of the path result in two ways. The bound is average-case, not just worst-case: a uniformly random instance already retains $na^2/2$ per unresolved layer, so all n bits price additively at $a^2/2$ per bit per hidden edge. And the depth threshold is set by the leg length, not the spider diameter $2(D_0 + 1)$, because every vertex outside the open source segments has target potential zero — a property that placing the same n dipoles on one long path would destroy, since the plateau between consecutive blocks is then offset by $a\sum_j \sigma_j D_j$ and exact recovery would need reach to distant blocks regardless of the lower bound. The hub does not accelerate recovery: it reaches all n sources within one round, but the direct wavefront of leg j reaches every hidden edge of that leg no later than any message relayed through the hub, so the hub's early knowledge of the other bits is causally shadowed. Oddness of D_0 matters similarly: for even D_0 each segment center has nonzero target potential, and a boundary conflict as in Proposition 3.6 leaves a one-round gap.

B Related Work

Distributed optimization and Laplacian systems. Local indistinguishability is foundational in distributed computing (Linial, 1992). Network-dependent lower and upper bounds for distributed smooth optimization connect convergence to spectral properties (Scaman et al., 2017). Distributed Laplacian solvers are studied in models ranging from message passing to the CONGEST model (Gillani & Bagchi, 2019; Forster et al., 2022; Anagnostides et al., 2023), and local solvers with query access approximate single coordinates of Laplacian solutions from small neighborhoods (Andoni et al., 2019). Those works address general systems and graph classes, and their lower bounds are order-level. The closest prior statements live in consensus and averaging: convergence-time lower bounds there are spectral and order-level (Olshevsky & Tsitsiklis, 2009), while finite-time consensus shows that, for a fixed linear time-invariant iteration, the number of steps a node needs before it can recover the asymptotic value is governed by the degree of a local minimal polynomial (Sundaram & Hadjicostis, 2007). Relative to these, Theorem 3.4 covers arbitrary nonlinear, unbounded-message local algorithms rather than a fixed linear scheme; it is an exact, constant-sharp minimax law in energy-gap units with a matching algorithm; and Proposition 4.2 recovers the minimal-polynomial phenomenon inside our family — halved by source parity and certified optimal by the model-free bound. We are not aware of a prior locality law of this exact form for a Laplacian-type objective.

First-order lower bounds. Zero chains are a standard device in oracle complexity (Nemirovsky & Yudin, 1983; Nesterov, 2004; Carmon et al., 2020). Recent work gives sharp first-order lower bounds under sublevel α-PŁ geometry (Yue et al., 2023; Masiha et al., 2026). These results motivate the model comparison above, but the proof of Theorem 3.4 is independent of them.

C Topology Beyond Diameter

On a weighted tree, current flows only along the unique s–t path and $|g_e| = a/w_e$ there. Theorem 3.1 therefore charges $\frac{a^2}{2}\sum_{e\in\mathcal{H}_K\cap P_{st}} w_e^{-1}$: precisely the hidden resistance along that path, with no charge for dangling branches. At the opposite extreme, on the complete graph K_m, $\phi = (a/m)(e_s - e_t)$ and all edges with nonzero optimal drop touch a source. The hidden-energy bound is zero at $K = 0$, and the optimum is indeed locally recoverable at round zero when m is known. These two extremes

bracket the path case analyzed in the main text: on the path, every non-source edge lies on the unique geodesic between the two sources, so its target drop is nonzero. This comparison does not assert extremality among all weighted graphs.

D Extended Discussion

Corollary D.1 (Randomized local algorithms)**.** *For every randomized K-local algorithm whose coins are independent of the source orientation,*

$$\max_{\sigma=\pm 1} \mathbb{E}\big[F_\sigma(x_\sigma^{(K)}) - F_{\sigma,\star}\big] \geq \frac{a^2}{2}(D - 2K - 2)_+.$$

Proof. This is the randomized part of Theorem 3.1 specialized as in Theorem 3.4. □

Role of knowing D. D is a source-independent instance parameter, defined separately for each fixed m, and could equivalently be hard-coded; it need only be known at the two source vertices, which attach it to the flooded tuple. Some shared gauge-alignment is necessary here, since fronts anchored independently at the two endpoints differ by the unknown offset $a\sigma D$; a uniform algorithm over unknown path sizes would need to also specify termination, and flooding for D rounds is a simple exact upper bound there instead. No non-source node infers σ from topology: a reached node learns it from the signed tuple, while the sole unreached fiber at $K^\star$, when present, is correct only because reflection symmetry makes its target potential zero.

What the energy formulation prices. Since labels may contain coordinates and D is known, the only source-dependent unknown on the path is the single bit σ, so Theorem 3.4 is, at bottom, the exact price of disseminating one bit across graph distance in energy-gap units. We regard this as the point, not a defect: it converts bit-dissemination distance into an additive per-edge cost, a weighted-graph generalization charging each edge its own optimal energy $w_e g_e^2$ (Theorem 3.1), and accuracy thresholds ($a^2/2$ here, $a^2/(2q)$ on the strips of Proposition A.1) below which that bit is unconditionally necessary — and it is not intrinsically one-bit, since Theorem A.2 applies it additively to n independent bits.

Accuracy dependence. The exact minimax law (14) concerns the high-accuracy regime $\varepsilon < a^2/2$, where even one unresolved edge is too costly; at coarser accuracy, (12) gives the continuous tradeoff $K \geq (D - 2 - 2\varepsilon/a^2)/2$ whenever positive, though not always tightly.

Structural property of the tight non-tree strips. For $q \geq 2$ and $m \geq 2$, the product strips of Section A.1 contain cycles. Their relevant structural property is a reflection-symmetric series of equipotential fibers: the optimal energy is carried by identical longitudinal cuts, and two source wavefronts cover every nonzero-potential fiber in half the longitudinal diameter.

The generic tree mechanism behind Proposition 3.6. On trees the mechanism made explicit by the P_5 example is generic, and it requires that the attachment vertex be *reached* while part of the branch is not. If the attachment vertex were itself outside every source ball, it and its branch could both output the same σ-independent value at no cost beyond the hidden-edge term. The extra cost arises when a vertex within K hops of a source, whose target potential depends on σ, has a pendant branch that reaches beyond the balls: the branch must hold an orientation-independent level, and the reached vertex must compromise between its σ-dependent drops and that level. The P_5 reduction extends with the conflict damped by the series resistance of the branch path. Theorem 3.3 prices the conflict exactly: it is $V_K - \frac{1}{2}\sum_{e\in\mathcal{H}_K} w_e g_e^2$.

Non-local architectures. Graph Transformers and hybrids with global attention (Rampášek et al., 2022) let every vertex read every other vertex's input in one block, so the radius-K view of Definition 2.2 no longer bounds what an output depends on; the indistinguishability argument of Theorem 3.1 fails because hidden edges no longer have source-blind endpoints. The barrier is therefore a statement about the communication pattern, and a fair comparison with such models must count their quadratic (or sparse-attention) cost rather than depth alone. Our controls (Table 2) confirm the boundary empirically but do not measure that cost.

Robustness versus depth of SOURCEFLOOD. SOURCEFLOOD sends $O(\log D)$ bits plus the source-value representation and is numerically trivial, whereas the exact interpolating filter p_m of Proposition 4.2 is tailored to the endpoint-dipole family and its monomial coefficients grow rapidly with m; what the two matching solutions trade is the numerical implementation of two depth-optimal constructions.

E EXACT MINIMAX ON OTHER TOPOLOGIES

Table 1 lists, for each graph, the number of vertices, the hop distance $d(s,t)$ between the two unit sources ($b = e_s - e_t$, $a = 1$), the initial gap $F(0) - F_\star$, the smallest K with $V_K = 0$ and with hidden-edge bound 0, and the largest boundary-conflict share $\max_K (V_K - \frac{1}{2}\sum_{e\in\mathcal{H}_K} w_e g_e^2)/(F(0) - F_\star)$. Sources are opposite corners (grids), the far ends (strip) or the farthest pair (others); the karate-club and Les Misérables rows use the unweighted topology (the bundled graphs carry co-occurrence weights, which are dropped). V_K is computed by dense least squares over the ball-supported columns, and the run asserts $V_K \geq$ hidden-edge bound at every K on every graph.

Table 1: Exact local minimax V_K (Theorem 3.3) versus the hidden-edge bound (Theorem 3.1).

graph	n	$d(s,t)$	$F(0)-F_\star$	$K(V=0)$	$K(H=0)$	max conflict
path P31	31	30	15.000	14	14	0.0%
grid 8x8 (corners)	64	14	1.364	6	6	1.0%
strip 4x16 (far ends)	64	18	2.368	9	8	7.9%
Watts-Strogatz n=200 (D=13)	200	13	1.336	11	11	3.9%
Barabási-Albert n=200 (D=6)	200	6	0.579	5	5	2.5%
random tree n=120 (D=30)	120	30	15.000	20	14	18.9%
random geometric n=300 (D=17)	300	17	0.569	12	12	2.4%
Zachary karate club (D=5)	34	5	0.646	3	3	6.1%
Les Misérables (D=5)	77	5	1.810	4	4	7.3%
P5 asymmetric (Prop. nontight)	5	3	1.500	1	1	16.7%
weighted grid 8x8	64	14	1.281	7	6	2.3%

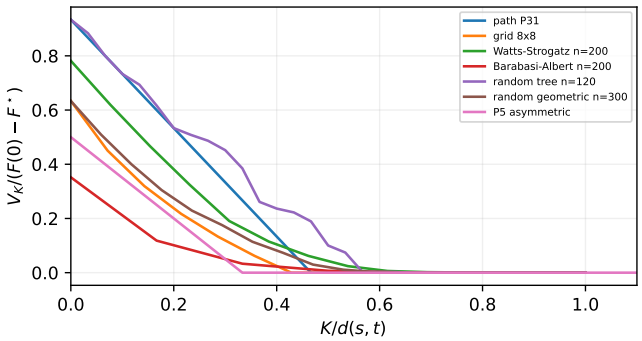


Figure 4: Normalized exact local minimax $V_K/(F(0) - F_\star)$ against $K/d(s,t)$. The path is linear and vanishes at half the distance; other topologies drop faster or slower depending on where the target potential is constant.

F DETAILED PROOFS

F.1 PROOF OF PROPOSITION A.1

Proof. The displayed signal is constant within each H-fiber, so all H-edge drops vanish. Every longitudinal edge has drop $a\sigma/q$, and applying L_G yields exactly the fiber source (17). For $K < K^\star$, at least one central layer cut is source-hidden. Its q longitudinal edges contribute $\frac{1}{2}q(a/q)^2 = a^2/(2q)$ to Theorem 3.1. Conversely, all vertices in a source fiber initialize SOURCEFLOOD with signed value

$a\sigma/q$ (or its negative); the same output rule is exact at $K^\star$. The randomized claim follows from Theorem 3.1. □

F.2 Proof of Proposition 2.1

Proof. Write the oriented edge differences as $z_i = x_i - x_{i+1}$. Since $x_1 - x_m = \sum_{i=1}^{D} z_i$, completing the square in (1) gives

$$F_\sigma(x) = \frac{1}{2}\sum_{i=1}^{D}(z_i - a\sigma)^2 - \frac{a^2 D}{2}.$$

Every vector of path-edge differences is integrable, so the minimum is attained exactly when $z_i = a\sigma$ for all i. The solutions are $x_i = c - a\sigma(i-1)$; imposing $\mathbf{1}^\top x = 0$ gives $c = a\sigma D/2$, proving (2), (3), and the initial gap.

The path-Laplacian eigenvalues, indexed from zero, are

$$\lambda_k = 2 - 2\cos(k\pi/m), \qquad k = 0, \dots, m-1.$$

Hence the Hessian norm is $\lambda_{m-1} = 2 + 2\cos(\pi/m) < 4$, and the smallest nonzero eigenvalue is $\mu := \lambda_1 = 2 - 2\cos(\pi/m)$; the main text writes $\lambda_2(L_m)$ for μ, using one-based ordering of the sorted spectrum.

Finally let $u = x - x^\star_\sigma$. Gauge components vanish from both $u^\top L_m u$ and $L_m u$. Expanding in an orthonormal eigenbasis gives

$$\begin{aligned} F_\sigma(x) - F_{\sigma,\star} &= \frac{1}{2}\sum_{k=1}^{m-1} \lambda_k \widehat{u}_k^2 \\ &\le \frac{1}{2\mu}\sum_{k=1}^{m-1} \lambda_k^2 \widehat{u}_k^2 \\ &= \frac{1}{2\mu}\|\nabla F_\sigma(x)\|_2^2. \end{aligned}$$

□

F.3 Information-flow lemma

Lemma F.1. *After K synchronous neighbor-to-neighbor rounds, the state and output of a deterministic vertex algorithm at i depend only on the radius-K rooted labeled neighborhood of i.*

Proof. The claim follows by induction. At round zero the state is a function of the vertex's own label. If every neighbor state after K rounds depends only on its radius-K neighborhood, then all messages received by i in round $K+1$ depend only on vertices at distance at most $K+1$ from i. Deterministic local computation cannot introduce dependence on any other input. This proves the claim. □

Lemma F.1 also shows that Definition 2.2 is at least as permissive as a synchronous implementation: the definition lets the output be an arbitrary function of the entire local view, even if computing that function would be expensive.

F.4 Edge count and parity details

For $e_i = \{i, i+1\}$, both endpoint K-balls exclude the left source iff $i - 1 > K$, and exclude the right source iff $m - i - 1 > K$. Therefore $K + 2 \le i \le m - K - 2$, which contains

$$(m - K - 2) - (K + 2) + 1 = m - 2K - 3 = D - 2K - 2$$

integers when positive.

If $D = 2r$, then $K^\star = r - 1$. At that radius all vertices except $r + 1$ receive a source, and $x^\star_{\sigma,r+1} = 0$. If $D = 2r + 1$, then $K^\star = r$ and every vertex receives its unique nearest source. For all $K < K^\star$, the central-edge count is at least two in the even case and at least one in the odd case. These observations establish every boundary case used in Theorems 3.4–3.5, including $m = 2$ where $K^\star = 0$ and both nodes are sources.

F.5 Proof of Theorem A.2

Proof. (i) Direct verification of $L_G x^\star = b^{\boldsymbol{\sigma}}$: at an interior segment vertex the two incident drops $a\sigma_j$ cancel; at s_j the hub-side edge has drop 0 and the outward edge drop $a\sigma_j$, giving $a\sigma_j = b_{s_j}$; at the leaf t_j the single incident drop gives $-a\sigma_j = b_{t_j}$; at the hub all incident drops vanish. Completing the square as in Theorem 3.1 yields

$$F_{\boldsymbol{\sigma}}(x) - F_{\boldsymbol{\sigma},\star} = \tfrac{1}{2}\sum_e \big(d_e - g_e(\boldsymbol{\sigma})\big)^2, \tag{19}$$

with $g_e(\boldsymbol{\sigma}) = a\sigma_j$ on segment edges of leg j (oriented outward) and 0 on hub edges.

(ii) Fix j; condition on the coins and on σ_{-j}. The two conditioned instances $\sigma_j = \pm 1$ differ only in the source labels at s_j and t_j. For a segment edge $e = \{u_{j,r}, u_{j,r+1}\}$ with $K+2 \le r \le D_0 - K - 1$, the four tree distances $\mathrm{dist}(u_{j,r}, s_j) = r-1$, $\mathrm{dist}(u_{j,r}, t_j) = D_0 + 1 - r$, $\mathrm{dist}(u_{j,r+1}, s_j) = r$, and $\mathrm{dist}(u_{j,r+1}, t_j) = D_0 - r$ all exceed K, so both endpoint views coincide on the two conditioned instances and d_e is σ_j-independent. Averaging (19) over $\sigma_j = \pm 1$ and keeping this edge's term gives $\frac{1}{2}(d_e^2 + a^2) \ge a^2/2$. Leg j contains $(D_0 - 2K - 2)_+$ such edges, the hidden sets of distinct legs are disjoint, and every discarded term in (19) is nonnegative; averaging over σ_{-j}, summing over j, and averaging over the coins proves (ii). The maximum over $\boldsymbol{\sigma}$ dominates the average.

(iii) Each source floods the tuple (signed value, origin label, hop count). A vertex $u_{j,r}$ that receives either tuple of leg j learns σ_j (the value received from t_j is $-a\sigma_j$, and the origin is part of the tuple) and outputs $-a\sigma_j(r-1)$ from its own position label; every other vertex outputs 0. With $D_0 = 2t+1$ and $K^\star = t$, every $u_{j,r}$ with $r \ge 2$ satisfies $\min(r-1,\, D_0 + 1 - r) \le t$, and the default-output vertices ($r \le 1$ and the hub) have target exactly 0; hence the output equals $x^\star$ for every $\boldsymbol{\sigma}$ and (19) vanishes. For $K < K^\star$ we have $(D_0 - 2K - 2)_+ \ge 1$, so part (ii) bounds the $\boldsymbol{\sigma}$-average of the (expected) gap below by $na^2/2 > \varepsilon$; hence some orientation exceeds ε, which proves the complexity claim. □

F.6 Separable even edge energies

The following lemma proves the extension claimed in Remark 3.2. For a differentiable convex $\psi : \mathbb{R} \to \mathbb{R}$, write $D_\psi(u,v) = \psi(u) - \psi(v) - \psi'(v)(u-v) \ge 0$ for its Bregman divergence.

Lemma F.2 (Hidden Bregman energy)**.** *Let $G = (V, E, w)$ be connected, let each $\psi_e : \mathbb{R} \to \mathbb{R}$ be convex, differentiable, and even, let b be balanced, and set*

$$F_\sigma^G(x) = \sum_{e \in E} w_e\, \psi_e\big((Bx)_e\big) - \sigma b^\top x, \qquad \sigma \in \{-1, +1\}.$$

Suppose F_{+1}^G attains a minimizer $x^\star$ with drops $g = Bx^\star$. Then $-x^\star$ minimizes F_{-1}^G, and every K-local algorithm (deterministic, or randomized in expected gap with coins independent of σ) satisfies

$$\max_{\sigma = \pm 1} \big[F_\sigma^G(x_\sigma^{(K)}) - F_{\sigma,\star}^G\big] \;\ge\; \sum_{e \in \mathcal{H}_K} w_e\, D_{\psi_e}(0, g_e).$$

If ψ_e is uniformly convex with modulus δ_e, i.e. $D_{\psi_e}(u,v) \ge \delta_e(|u-v|)$, the e-th term is at least $w_e \delta_e(|g_e|)$. For $\psi_e(t) = t^2/2$ the statement is Theorem 3.1; for the p-Laplacian energy $\psi_e(t) = |t|^p/p$ with $p \ge 2$, each hidden edge contributes $(1 - 1/p)\, w_e |g_e|^p$.

Proof. Evenness gives $F_{-1}^G(-x) = F_{+1}^G(x)$, so $-x^\star$ minimizes F_{-1}^G, with drops $-g$ and the same optimal value. Writing $\psi'(g)$ for the entrywise derivative vector, first-order optimality of $x^\star$ reads

$B^\top W\psi'(g) = b$. Hence, for any x with $d = Bx$,

$$\begin{aligned} F^G_\sigma(x) - F^G_{\sigma,\star} &= \sum_e w_e\big[\psi_e(d_e) - \psi_e(\sigma g_e)\big] \\ &\qquad - \sigma b^\top(x - \sigma x^\star) \\ &= \sum_e w_e\big[\psi_e(d_e) - \psi_e(\sigma g_e) \\ &\qquad - \sigma\psi'_e(g_e)(d_e - \sigma g_e)\big] \\ &= \sum_e w_e\, D_{\psi_e}(d_e, \sigma g_e), \end{aligned}$$

using $\sigma\psi'_e(g_e) = \psi'_e(\sigma g_e)$ (ψ'_e is odd). Every term is nonnegative. On a hidden edge, d_e is the same for both orientations, and the two-point average is

$$\begin{aligned} &\tfrac{1}{2}\big[D_{\psi_e}(d_e, g_e) + D_{\psi_e}(d_e, -g_e)\big] \\ &\qquad = \psi_e(d_e) - \psi_e(g_e) + g_e\psi'_e(g_e), \end{aligned}$$

a convex function of d_e whose derivative $\psi'_e(d_e)$ vanishes at $d_e = 0$; its minimum value is exactly $D_{\psi_e}(0, g_e)$. Averaging the gap over σ, discarding non-hidden terms, and using maximum at least average proves the deterministic claim; conditioning on the coins gives the randomized one. For $\psi(t) = |t|^p/p$, $D_\psi(0, g) = g\psi'(g) - \psi(g) = (1 - 1/p)|g|^p$. □

F.7 Proof of Proposition 3.6

Proof. Label the edges $\{1,2\}, \{2,3\}, \{3,4\}, \{4,5\}$. A unit of current flows from vertex 1 to vertex 4, so the optimal drops are $g = (a, a, a, 0)$ and the gap identity of Theorem 3.1 reads $F^G_\sigma(x) - F^G_{\sigma,\star} = \frac{1}{2}\sum_{i=1}^4 (d_i - \sigma g_i)^2$. At $K = 0$, only edge $\{2,3\}$ has both endpoints source-free, so (10) gives $a^2/2$.

Now fix any deterministic 0-local algorithm. Vertices $2, 3, 5$ see no source, so their outputs c_2, c_3, c_5 are the same for both orientations; vertices $1, 4$ see their own signed source and may output arbitrary $x_1(\sigma), x_4(\sigma)$. Minimizing the gap over $x_1(\sigma)$ removes the first term. The third and fourth terms share $x_4(\sigma)$:

$$\min_{x_4}\big[(c_3 - x_4 - \sigma a)^2 + (x_4 - c_5)^2\big] = \tfrac{1}{2}(c_3 - c_5 - \sigma a)^2.$$

Writing $u = c_2 - c_3$ and $v = c_3 - c_5$, the best achievable per-orientation gap is therefore

$$\Gamma(\sigma) = \tfrac{1}{2}\big[(u - \sigma a)^2 + \tfrac{1}{2}(v - \sigma a)^2\big].$$

Averaging over $\sigma = \pm 1$ gives $\frac{1}{2}(u^2 + a^2) + \frac{1}{4}(v^2 + a^2) \geq \frac{3}{4}a^2$, with equality iff $u = v = 0$; hence $\max_\sigma \Gamma(\sigma) \geq 3a^2/4$ for every algorithm. Conversely, the choice $c_2 = c_3 = c_5 = 0$, $x_1(\sigma) = \sigma a$, $x_4(\sigma) = -\sigma a/2$ attains $\Gamma(\sigma) = 3a^2/4$ for both orientations. Randomized algorithms: condition on the coins and average. □

F.8 Proof of Proposition 4.1

Proof. Set $\lambda_{\min} = \lambda_2(L_m)$, $\lambda_{\max} = L^{\mathrm{sm}}_m$, $\theta = (\lambda_{\max} + \lambda_{\min})/2$, $\delta = (\lambda_{\max} - \lambda_{\min})/2$, and $\sigma_1 = \theta/\delta$. The classical semi-iteration (Golub & Varga, 1961) is $x^{(1)} = x^{(0)} + r^{(0)}/\theta$ and, for $k \geq 1$,

$$x^{(k+1)} = x^{(k)} + \rho_{k+1}\Big[\rho_k\big(x^{(k)} - x^{(k-1)}\big) + \tfrac{2}{\delta}\, r^{(k)}\Big],$$

with $r^{(k)} = b^\sigma - L_m x^{(k)}$, $\rho_1 = 1/\sigma_1$, and $\rho_{k+1} = 1/(2\sigma_1 - \rho_k)$. Each step applies L_m once and combines locally stored vectors with scalars determined by m, so the round-K iterate is K-local under Definition 2.2. Starting from $x^{(0)} = 0$, all iterates stay in the mean-zero subspace, where the standard Chebyshev argument yields $x^{(K)} - x^\star_\sigma = r_K(L_m)(x^{(0)} - x^\star_\sigma)$ with residual polynomial $r_K(\lambda) = T_K\big((\theta - \lambda)/\delta\big)/T_K(\sigma_1)$, T_K the degree-K Chebyshev polynomial. Since the gap is $\frac{1}{2}\|x - x^\star_\sigma\|^2_{L_m}$ and $|r_K(\lambda)| \leq 1/T_K(\sigma_1)$ on $[\lambda_{\min}, \lambda_{\max}]$,

$$F_\sigma(x^{(K)}) - F_{\sigma,\star} \leq \frac{1}{T_K(\sigma_1)^2} \cdot \frac{a^2 D}{2}.$$

With $\kappa_m = \lambda_{\max}/\lambda_{\min}$ one has $\sigma_1 = (\kappa_m+1)/(\kappa_m-1)$ and $\sigma_1+\sqrt{\sigma_1^2-1} = (\sqrt{\kappa_m}+1)/(\sqrt{\kappa_m}-1)$, so $T_K(\sigma_1) \geq \frac{1}{2}\big[(\sqrt{\kappa_m}+1)/(\sqrt{\kappa_m}-1)\big]^K$, which gives (16). The method is linear and its output flips sign with σ, so both orientations have the same gap. Solving $2a^2D\rho^{2K} \leq \varepsilon$ with $\rho = (\sqrt{\kappa_m}-1)/(\sqrt{\kappa_m}+1)$ and $\ln(1/\rho) = 2\,\mathrm{artanh}(1/\sqrt{\kappa_m}) \geq 2/\sqrt{\kappa_m}$ yields the round count $\lceil(\sqrt{\kappa_m}/4)\ln(2a^2D/\varepsilon)\rceil$; finally $\kappa_m = (1+\cos(\pi/m))/(1-\cos(\pi/m)) = \cot^2(\pi/(2m))$ and $\cot(t) < 1/t$ give $\sqrt{\kappa_m} < 2m/\pi$. □

F.9 Proof of Proposition 4.2

Proof. The stated v_k are the path eigenvectors. Using $\cos\big((2m-1)k\pi/(2m)\big) = (-1)^k\cos\big(k\pi/(2m)\big)$,

$$\langle v_k, e_1 - e_m\rangle = \big(1-(-1)^k\big)\cos\big(k\pi/(2m)\big),$$

which vanishes iff k is even; for odd $k \leq m-1$ we have $0 < k\pi/(2m) < \pi/2$, so the inner product is nonzero. There are $\lceil (m-1)/2\rceil = \lceil D/2\rceil$ odd indices.

Let $x = p(L_m)b^\sigma$ with $\deg p = K$. Since $b^\sigma \perp \mathbf{1}$, so is x, and expanding the gap in the orthonormal eigenbasis,

$$F_\sigma(x) - F_{\sigma,\star} = \frac{1}{2}\sum_{k\ \mathrm{odd}} \lambda_k\big(p(\lambda_k) - 1/\lambda_k\big)^2\widehat{b}_k^2,$$

where $\widehat{b}_k \neq 0$ are the excited-mode coefficients of b^σ. Exactness is therefore equivalent to $\lambda p(\lambda) = 1$ at the $\lceil D/2\rceil$ distinct odd eigenvalues. If $K < \lceil D/2\rceil - 1$, then $q(\lambda) = \lambda p(\lambda) - 1$ is a nonzero polynomial ($q(0) = -1$) of degree at most $K+1 < \lceil D/2\rceil$ with $\lceil D/2\rceil$ distinct roots, a contradiction. Conversely, the interpolating polynomial of $1/\lambda$ at those nodes has degree $\lceil D/2\rceil - 1$ and is exact; the identity $\lceil D/2\rceil - 1 = \lfloor (D-1)/2\rfloor$ holds for every $D \geq 1$. Locality holds because $(L_m^k b^\sigma)_i$ depends only on b_j^σ with $\mathrm{dist}(i,j) \leq k$, while the coefficients of p_m are functions of m alone and hence orientation-independent side information. Linearity transfers exactness from b^{+1} to $b^{-1} = -b^{+1}$. □

G Reproduction Protocol

G.1 Trained-network protocol

The recorded environment is Python 3.12.4, PyTorch 2.9.1+cu128 (CPU execution), NumPy 2.1.0, and Matplotlib 3.9.2. We use one CPU thread, float64, deterministic PyTorch algorithms These seeds quantify initialization variability, not uncertainty over a sampled dataset. Raw results include failed fits; no seed is dropped.

Architecture and features. For vertex i, the input is $[b_i^\sigma,\ 2(i-1)/D - 1,\ \deg(i)/2,\ D/16]$. An affine map and ReLU give width-32 initial features, followed by exactly K message-passing layers and a scalar affine nodewise readout. No centering, pooling, batch normalization, residual connections, or dropout is applied. Let H be a layer's node-feature matrix, A the path adjacency, and $Q = \mathrm{diag}(\deg(i))$. GCN applies $\mathrm{ReLU}(\widetilde{Q}^{-1/2}(A+I)\widetilde{Q}^{-1/2}HW + \beta)$, with $\widetilde{Q} = Q + I$. GIN applies $\mathrm{ReLU}(\mathrm{MLP}((1+\epsilon)H + AH))$, where ϵ is trainable and initialized at zero, and the MLP has two width-32 affine maps separated by ReLU. GraphSAGE applies $\mathrm{ReLU}([H, Q^{-1}AH]W + \beta)$ without optional embedding normalization. We use direct path-neighbor sums with linear memory. Topology-dependent degree normalization is source-independent side information; it does not add source communication rounds.

Training and selection. The two inputs have $a = 1$, $\sigma \in \{-1, +1\}$ and share all fixed labels. The loss is $\frac{1}{2}\sum_\sigma[F_\sigma(x_\sigma) - F_{\sigma,\star}]$, computed directly from edge differences in (3). Adam uses its default moment parameters $(0.9, 0.999)$ and numerical epsilon 10^{-8}, zero weight decay, learning rate 0.003, and gradient-norm clipping at 10. Each run takes 600 updates, evaluating initial and post-update iterates; the lowest mean training-gap checkpoint is selected. There is no validation set, hyperparameter search, or transfer across sizes/depths. Centralized offline training is allowed: the same learned parameters are fixed for both orientations at inference, so they cannot broadcast the unknown test orientation. Shared fitting of both signs is essential to this diagnostic; separately fitting a network per sign would not test the theorem.

Table 2: Global-attention controls, outside the local model. Worst-orientation gap over 10 seeds (maximum) after 3000 steps; reach is the farthest vertex whose output changes when the left source input is perturbed. A local model of the same depth would have gap at least $\frac{1}{2}(m-3-2L) \geq 5$ for these L and m.

model	blocks L	P_{17}	P_{31}	P_{63}	reach
Transformer	1	4.7×10^{-5}	1.4×10^{-4}	2.6×10^{-4}	$m-1$
Transformer	2	2.0×10^{-5}	8.9×10^{-5}	1.2×10^{-4}	$m-1$
GCN + attention	1	1.2×10^{-5}	1.5×10^{-4}	5.7×10^{-4}	$m-1$
GCN + attention	2	1.0×10^{-9}	2.9×10^{-5}	5.4×10^{-4}	$m-1$

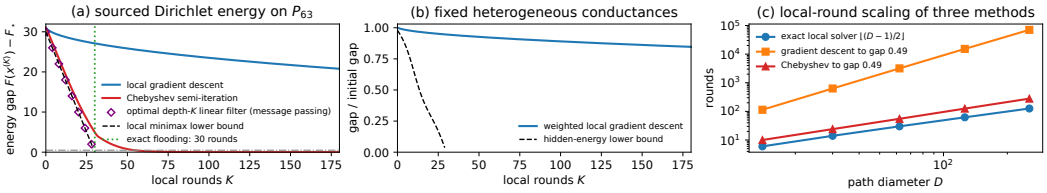


Figure 5: Deterministic validation with $a = 1$. (a) On P_{63}, the dashed black curve is the lower envelope from Theorem 3.4; flooding is exact after 30 rounds, and Chebyshev semi-iteration (Proposition 4.1) tracks the envelope before reaching the target at 55 rounds. The open diamonds are the best gap attainable by any depth-K linear filter (sampled every 4 rounds): it coincides exactly with the envelope on $[0, K^\star]$ and vanishes at $K^\star = 30$, the GNN representation-depth threshold of Section 4. (b) On a fixed heterogeneous weighted P_{63}, the normalized hidden-energy envelope automatically lower-bounds weighted local gradient descent and is far from vacuous. (c) Exact flooding scales linearly with diameter, Chebyshev near-linearly (slope 1.15), and gradient descent approximately quadratically.

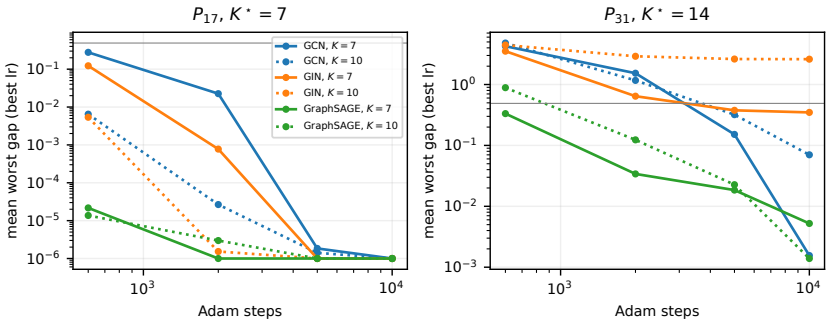


Figure 6: Training-budget ablation at $K^\star$ (solid) and $K^\star + 3$ (dotted): seed-mean worst-orientation gap against Adam steps, best learning rate per budget, 6 seeds. Horizontal line: target 0.49.