GLOBAL BAYESIAN OPTIMIZATION ON PARTIALLY OBSERVED GRAPHS

Anonymous authors
Paper under double-blind review

ABSTRACT

We address Bayesian optimization (BO) over graph nodes where the black-box objective is expensive to evaluate and the graph topology is partially observed. A central challenge is that BO shares information through a globally defined kernel, while incomplete graph topology makes it unavailable, limiting existing methods to local, neighborhood-based exploration. Effective global BO in this regime therefore requires inferring a coherent surrogate geometry from sparse edge observations. We reconstruct a graph-valid low-rank surrogate from revealed edges and extract its Laplacian spectrum to induce a Gaussian process kernel over all nodes. Under this kernel, the posterior is a linear model in d-dimensional node features: function evaluations learn a shared coefficient, and edge queries learn the features. For budgeted queries of adjacency entries or rows, we further derive decision-aware acquisition rules, including a knowledge-gradient rule for pair queries. We provide guarantees for surrogate identification under row and pair query patterns and for the decision error under an estimated geometry, which is charged only at the selected node and the optimum and enters the simple regret only through the identification time. On protein interaction networks, a molecular similarity graph, a Wikipedia page graph, and synthetic graphs, our method attains lower final regret than structure-free search and local graph BO on all ten objectives at matched budgets, and decision-aware pair queries improve on uniform ones at every tested budget. Code is available at https://anonymous.4open.science/r/global-graph-bo-code.

1 INTRODUCTION

Many real-world decision-making problems involve optimizing an unknown function defined over the nodes of a large graph. In such settings, each node represents a discrete action, configuration, or intervention, while querying the function value at a node is often costly, either via simulation, experimentation, or human evaluation. The goal is therefore to identify the node with the optimal function value using as few function evaluations as possible. This problem arises in diverse applications, including selecting molecules in a similarity graph (Korovina et al., 2020), prioritizing genes in a protein interaction network (Tsherniak et al., 2017; Huttlin et al., 2021), and targeting accounts in a social network (Leskovec & Mcauley, 2012). Across these domains, exhaustive evaluation over all nodes is infeasible, which necessitates sample-efficient optimization that exploits structural correlations induced by the graph topology.

Bayesian optimization (BO) (Frazier, 2018; Snoek et al., 2012) provides a principled framework for such sample-efficient optimization. It employs a probabilistic surrogate, typically a Gaussian process (GP), to guide exploration via an acquisition function. A fundamental requirement for effective BO on graphs is a globally defined kernel that propagates uncertainty and structural correlations across the entire search space. Existing methods typically construct these kernels from the graph Laplacian or spectral embeddings, implicitly assuming that the graph topology is fully known a priori (Cui et al., 2019; Ru et al., 2021; Oh et al., 2019; Xie et al., 2025).

However, in many practical settings, this assumption of full observability does not hold. Graph structures are often incomplete, costly to acquire, or revealed only incrementally through limited edge queries. We therefore study BO on graphs with partially observed topology. At each iteration, the learner performs one expensive node evaluation and, in parallel, queries a limited budget of adjacency

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.

entries or entire rows to refine its knowledge of the graph. Recent graph-specific BO methods (Wan et al., 2023) address this by incremental local exploration of the neighborhood. Without a global geometric representation, posterior uncertainty cannot propagate beyond the revealed subgraph, leading to myopic exploration and slow convergence. Existing methods for partially observed graphs explore locally, and none maintains a global posterior and acquisition rule over all nodes.

We propose a novel framework for global graph BO that enables sample-efficient optimization on large and partially observed graphs. Our core insight is that effective global BO under partial topology requires inferring a coherent surrogate geometry rather than neighborhood structure. To this end, we infer a surrogate graph from sparse edge observations by fitting a graph-valid low-rank surrogate to the revealed entries, extract its Laplacian spectral structure efficiently to obtain a global geometry, and use this geometry to define a GP surrogate and acquisition function over all nodes. The same geometry determines what each budget can learn. On the surrogate the GP posterior is a linear model in d-dimensional node features with a coefficient shared across nodes, so function evaluations learn the coefficient and edge queries learn the features. Structure acquisition therefore becomes a decision problem: we value an edge query by its expected effect on the next selected node, which yields a knowledge-gradient rule for pair queries and a simple pin-then-uniform rule for row queries.

We provide theoretical guarantees that characterize when a globally coherent surrogate graph geometry can be inferred from partial edge observations and what an estimated geometry costs the optimizer. For graphs in the surrogate class, $d + 1$ revealed rows identify the surrogate while d do not suffice in general. A per-round bound charges the estimated geometry only at the selected node and the optimum, and the resulting simple-regret bound depends on structural access only through the identification time. Experiments on protein interaction networks, a molecular similarity graph, a Wikipedia page graph, and synthetic graphs answer three questions at matched node and edge budgets: the global surrogate has lower final regret than random search, local search, and local graph BO on all ten objectives, full-information controls separate the representation from its estimation, and the knowledge-gradient rule has lower final regret than uniform pairs at every budget.

Our main contributions are summarized as follows:

- **Global BO under partial topology.** We formulate Bayesian optimization on graph nodes when topology is incomplete and revealed through budgeted edge queries of adjacency entries or rows, supporting globally defined uncertainty propagation and acquisition over all nodes.
- **Surrogate geometry via a graph-valid low-rank class and its exact spectrum.** We construct a globally defined GP kernel by fitting a graph-valid low-rank surrogate to the revealed entries and exploiting its exactly computed Laplacian spectral geometry, which makes the GP a linear model in node features.
- **Decision-aware structure acquisition and identification guarantees.** We price each edge query by its expected effect on the next decision, yielding a knowledge-gradient rule for pair queries and a pin-then-uniform rule for rows, and we prove that $d + 1$ rows identify the surrogate while d do not suffice in general, with a simple-regret bound in which structural access enters only through the identification time.
- **Empirical validation on real and synthetic graphs.** We demonstrate lower final regret than random search, local search, and local graph BO on protein interaction networks, a molecular similarity graph, a Wikipedia page graph, and synthetic graphs under the same node and edge query budgets.

2 Related Work

Bayesian Optimization on Graphs. BO on graphs places a GP prior on the nodes through a graph kernel or an embedding (Wan et al., 2023; Cui & Yang, 2018; Cui et al., 2019; Ru et al., 2021), or exploits an explicit combinatorial construction such as a product graph (Oh et al., 2019; Ramachandram et al., 2018; White et al., 2021; Xie et al., 2025; Liang et al., 2024; Yuan et al., 2024; Xie et al., 2026), and assumes the topology known. The closest analysis is the spectral bandit (Valko et al., 2014), a linear bandit in the Laplacian eigenbasis of a known graph with regret governed by an effective dimension. Our GP is such a linear model, but its basis is learned from budgeted edge

queries that the learner chooses. Under partial topology the closest prior method explores the revealed neighborhood locally, with row access at every visited node (Wan et al., 2023).

Low-Rank Graph Models and Structure Recovery. Matrix completion recovers a low-rank matrix from uniform samples under incoherence (Candès & Recht, 2012; Candès & Tao, 2010; Keshavan et al., 2010) or from deterministic patterns (Lee et al., 2023), and adaptive row or column sampling removes the incoherence requirement and the logarithmic factor (Krishnamurthy & Singh, 2013; Mahoney & Drineas, 2009). Our surrogate class is the step-graphon or mixed-membership model (Lovász, 2012; Hoff et al., 2002; Airoldi et al., 2008; Karrer & Newman, 2011), whose spectral estimation is well understood on a fully observed graph (Chatterjee, 2015; Xu, 2018; Mao et al., 2021). Recovery of A is not our goal. The class is what makes every spectral kernel of the surrogate a floor plus a rank-d term, and its landmark structure is what makes a revealed entry a linear measurement of one node, enabling decision-aware allocation of the edge budget.

Spectral Kernels, Uncertain Inputs, and Value of Information. Laplacian and diffusion kernels give GP priors on fully observed graphs (Kondor & Lafferty, 2002; Smola & Kondor, 2003; Borovitskiy et al., 2021), and regression on Laplacian eigenmaps is their finite-rank form (Belkin & Niyogi, 2003; 2004). GPs with uncertain inputs (Girard et al., 2002; Mchutchon & Rasmussen, 2011) and BO under input uncertainty (Oliveira et al., 2019) take the input distribution as given, whereas here the input is a feature whose uncertainty the learner can reduce by a query. Selecting structural queries is a value-of-information problem, connecting our method to the knowledge gradient (Frazier et al., 2008; Frazier & Powell, 2010) and multi-fidelity BO (Kandasamy et al., 2016). Here, budgeted structural observations refine the kernel geometry used for objective prediction.

3 Problem Statement

We consider the problem of optimizing an expensive-to-evaluate black-box function defined over nodes of a large, partially observed graph. Formally, let $G = (\mathcal{V}, \mathcal{E})$ denote an undirected and weighted graph, where $\mathcal{V} = \{1, \dots, n\}$ is the node set, and edges $\mathcal{E} = \{(i, j) \in \mathcal{V} \times \mathcal{V} \mid A_{ij} > 0\}$ are determined by an unknown symmetric adjacency matrix $A \in [0, 1]^{n \times n}$. $A_{ij} > 0$ indicates that nodes i and j are connected and $A_{ij} = 0$ otherwise. A black-box function $y : \mathcal{V} \to \mathbb{R}$ is defined on the graph nodes.

Evaluating $y(v)$ is expensive, so we can query $y(v)$ only at a small subset of nodes. Our goal is to identify the globally optimal node

$$v^* = \arg\max_{v \in \mathcal{V}} y(v), \tag{1}$$

using as few function evaluations as possible.

BO proceeds iteratively. Let $\mathcal{V}^{(t)} \subseteq \mathcal{V}$ be the set of nodes evaluated up to iteration t. From the observed pairs $\{(v, y(v))\}_{v \in \mathcal{V}^{(t)}}$ it forms a posterior over y and evaluates the node $v^{(t+1)}$ that maximizes an acquisition function, which trades off exploration and exploitation under that posterior. We consider exact observations of y and measure performance by the simple regret $r_T = y(v^*) - \max_{v \in \mathcal{V}^{(T)}} y(v)$ after T iterations.

In parallel with node function evaluations, the learner acquires the graph structure through *edge queries*, which are limited by a per-iteration budget. Depending on applications, we consider two query types: (i) A *pair query* reveals a single entry A_{uv} of a chosen node pair, and the learner may issue Q_E such queries per iteration. This is the natural unit when affinities are computed or assayed one pair at a time, such as similarities between molecules (Rogers & Hahn, 2010; Bajusz et al., 2015). (ii) A *row query* reveals all entries $A_{u,\cdot}$ of a chosen node u, and the learner may issue Q_R such queries per iteration. This is how relational data arrive when one measurement returns all partners of one entity, e.g., an affinity-purification experiment with bait protein u (Huttlin et al., 2021), or a call to a social-network interface that returns the contact list of one account (Leskovec & Mcauley, 2012). Let $\Omega^{(t)} \subseteq \mathcal{V} \times \mathcal{V}$ be the set of node pairs revealed up to iteration t, and let $A_\Omega^{(t)} = \mathcal{P}_{\Omega^{(t)}}(A)$ be the partially observed adjacency, where $\mathcal{P}_\Omega$ keeps the entries in Ω and sets the others to zero. Self-pairs are never queried, and we write A_{off} for A with its diagonal removed, the object the queries can reveal. Revealed entries are exact, and a revealed zero $A_{ij} = 0$ is itself an observation, since it rules out an edge.

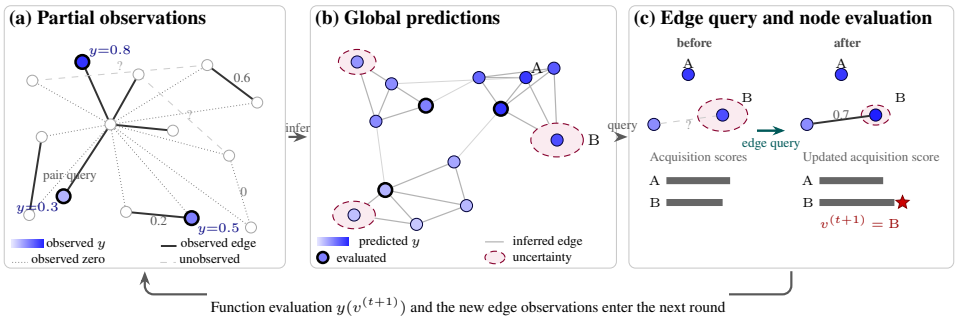

Figure 1: Global BO on a partially observed graph. (a) Two observations with two budgets: a few expensive values of the objective y, and entries of the adjacency matrix revealed by edge queries. (b) A graph-valid low-rank surrogate fitted to the revealed entries places every node in one global geometry and gives every node a posterior and an acquisition score. A node that no query has touched has an uncertain feature and hence an uncertain score. (c) Structure is queried where it can change the next decision. An edge query against a landmark sharpens the feature of candidate B, its score moves above that of A, and B is evaluated next. The new value and the new entries enter the next round.

4 Methodology

We propose an iterative framework that alternates between (i) inference of a global surrogate geometry from the revealed entries and (ii) GP-based BO over the nodes on the induced geometry. At iteration t the method holds the evaluated set $\mathcal{V}^{(t)}$ and the revealed set $\Omega^{(t)}$, and one round is the pipeline

$$\Omega^{(t)} \to \tilde{A}^{(t)} \to \tilde{L}_{\zeta}^{(t)} \to \tilde{K}^{(t)} \to \text{GP posterior} \to v^{(t+1)} \to \Omega^{(t+1)}.$$

The revealed entries are fitted by a surrogate adjacency $\tilde{A}^{(t)}$ inside a graph-valid low-rank class, whose regularized normalized Laplacian $\tilde{L}_{\zeta}^{(t)}$ has a spectrum that is computed exactly in closed form. A spectral filter of this Laplacian is the kernel $\tilde{K}^{(t)}$ of a GP over all nodes, the next node $v^{(t+1)}$ maximizes an acquisition function of the posterior, and the edge queries of the round are chosen by their value for that decision, which closes the loop from the kernel back to $\Omega^{(t+1)}$. We query one node value per iteration, and the framework extends to batches.

The key insight is the following. As the surrogate is a valid rank-d graph, every spectral kernel of its Laplacian is a white floor plus a rank-d term, so the GP is a Bayesian linear model in d-dimensional node features with a coefficient shared by all nodes. Function evaluations learn the shared coefficient, edge queries learn the features, and the value of an edge query is the improvement of the next decision that its answer brings. Section 4.1 learns the geometry from the revealed entries. Section 4.2 runs BO on the geometry. Section 4.3 states what each revealed entry determines about the features and turns the value of a query into the edge-query rule of each query type, and Section 4.4 states what estimated features cost one decision, when the revealed entries identify the geometry, and sums the cost over rounds. Algorithm 1 (Appendix B) summarizes one round.

4.1 Surrogate Graph Learning from Partial Observations

We fit the revealed entries within a *graph-valid low-rank class*. Each node has a membership $\pi_i \in \Delta_d := \{b \geq 0 : \mathbf{1}^\top b \leq 1\}$, and memberships interact through a symmetric core:

$$\mathcal{C}_d = \left\{\Pi C \Pi^\top : \pi_i \in \Delta_d,\ C = C^\top \in [0,1]^{d\times d}\right\}. \tag{2}$$

Every member is symmetric, has entries in $[0,1]$, and has rank at most d. Thus every member is a valid weighted graph by construction. The parameter d controls the surrogate's resolution, without requiring the unknown graph itself to have rank d. Given revealed pairs Ω, we solve

$$\min_{\pi_i\in\Delta_d,\ C=C^\top\in[0,1]^{d\times d}} \left\|\mathcal{P}_\Omega(\Pi C\Pi^\top - A)\right\|_F^2. \tag{3}$$

We use projected alternating least squares with spectral initialization and warm starts (Appendix E). Only the revealed entries enter the fit, and self-pairs never do.

Exact spectral geometry. For the fitted $\tilde{A}$, let $\tilde{D} = \mathrm{diag}(\tilde{A}\mathbf{1})$ and $\tilde{L}_\zeta = I - (\tilde{D} + \zeta I)^{-1/2}\tilde{A}(\tilde{D} + \zeta I)^{-1/2}$, where $\zeta > 0$ regularizes low degrees (Chaudhuri et al., 2012; Qin & Rohe, 2013). The normalized adjacency has rank at most d, so $\tilde{L}_\zeta$ has at most d eigenvalues different from 1. A thin QR factorization of the normalized memberships reduces the spectral solve to a $d \times d$ eigendecomposition, at total cost $\mathcal{O}(nd^2 + d^3)$ (Lemma B.2). This computes the surrogate's spectrum exactly without forming an $n \times n$ matrix. Write its nontrivial eigenpairs as $(\tilde{\lambda}_i, \tilde{u}_i)$.

4.2 Global Bayesian Optimization on the Surrogate

The eigenvalue 1 on the orthogonal complement provides a natural bulk level for a spectral filter (Smola & Kondor, 2003). With a positive decreasing response r, we retain the positive contrasts $r(\tilde{\lambda}_i) - r(1)$ and give the structured and bulk components separate scales:

$$\tilde{K} = \sigma_b^2 I + \sigma_s^2 \tilde{U}_+ D_+ \tilde{U}_+^\top, \qquad D_+ = \mathrm{diag}\big(r(\tilde{\lambda}_i) - r(1)\big)_{\tilde{\lambda}_i < 1}. \tag{4}$$

Here $\tilde{U}_+$ contains the $d_+ \le d$ retained modes. This *structured-plus-bulk kernel* is positive definite for $\sigma_b^2 > 0$. Retaining positive contrasts permits independent Gaussian coefficients. A one-scale filter ties their variances to the bulk level, whereas separate scales let the data tune the balance between cross-node generalization and node-specific variation. We use the resolvent $r(\lambda) = (\beta + \lambda)^{-1}$ and select $(\sigma_b^2, \sigma_s^2, \beta, \zeta)$ by marginal likelihood. The exact one-scale alternative is given in Appendix B.

Let ϕ_v be the transposed v-th row of $\tilde{U}_+$ and $D = \sigma_s^2 D_+$. Conditional on the fitted geometry, the kernel is equivalent to $y(v) = \phi_v^\top \theta + \varepsilon_v$, with $\theta \sim \mathcal{N}(0, D)$ and independent $\varepsilon_v \sim \mathcal{N}(0, \sigma_b^2)$. The residual is drawn once per node, and observations of y are exact. If w_t, Ξ_t are the posterior mean and covariance of θ, then for an unevaluated node

$$\mu_t(v) = \phi_v^\top w_t, \qquad s_t^2(v) = \sigma_b^2 + \phi_v^\top \Xi_t \phi_v. \tag{5}$$

The coefficient update uses only a $d_+ \times d_+$ system (Lemma B.1). This representation separates the two learning tasks: evaluations update the shared coefficient, and structural queries refine node features. We select the next node using the plug-in GP-UCB score on the current estimated features,

$$U_t(v) = \mu_t(v) + \kappa s_t(v), \qquad v^{(t+1)} = \arg\max_{v \notin \mathcal{V}^{(t)}} U_t(v). \tag{6}$$

Here $\kappa > 0$ controls exploration.

4.3 Decision-Aware Structure Acquisition

From entries to feature uncertainty. Conditional on the core and a reference node's membership, an entry in the surrogate model is a linear measurement, $A_{vu} = g_u^\top \pi_v$ with $g_u = C\pi_u$. We call these reference nodes *landmarks*. They are the revealed rows under row access and the initial block's nodes under pair access. Their memberships are fitted from the available entries. In the local acquisition model, we hold the fitted global quantities fixed and use a Gaussian working covariance Σ_v to track the unmeasured membership directions. A node is *pinned* when its measurement directions span $\mathbb{R}^d$, making $\Sigma_v = 0$ conditional on that geometry. The feature map is differentiable in π_v, giving $\Sigma_v^\phi = J_v \Sigma_v J_v^\top$, where $J_v = \partial\phi_v / \partial\pi_v$. The map and conditioning formulas are in Appendix B.

With feature mean $\hat{\phi}_v$ and covariance Σ_v^ϕ independent of the coefficient posterior in this working model, total variance separates three contributions:

$$\tilde{s}_t^2(v) = \underbrace{\sigma_b^2}_{\text{bulk}} + \underbrace{\hat{\phi}_v^\top \Xi_t \hat{\phi}_v}_{\text{coefficients}} + \underbrace{w_t^\top \Sigma_v^\phi w_t + \mathrm{tr}(\Sigma_v^\phi \Xi_t)}_{\text{features}}. \tag{7}$$

Function observations inform the second term, while structural measurements reduce the third under the fixed local model. The latter weights feature uncertainty by its relevance to the objective. This uncertainty is conditional on the fitted global quantities. We use it to value structural queries, retain the plug-in node score (6), and account for the actual structural score error separately in Section 4.4. Because evaluated-node features affect all predictions through the shared coefficient posterior, both query rules prioritize them within a pinning allowance ρ_{pin}. Until pinned, their coefficient updates use the moment adjustment in Eq. (17).

Pair queries. A useful pair should reveal a direction that matters to a competitive node's score. Holding the geometry and coefficient state fixed, let $c_v = J_v^\top(w_t + \kappa \Xi_t \hat{\phi}_v / s_t(v))$ be its membership sensitivity. A query along g changes the linearized score mean by a zero-mean Gaussian increment with standard deviation

$$\delta(v, g) = \frac{|c_v^\top \Sigma_v g|}{\sqrt{g^\top \Sigma_v g}}, \qquad g^\top \Sigma_v g > 0. \tag{8}$$

Let $\mathrm{gap}_t(v) = |U_t(v) - \max_{w \neq v} U_t(w)|$, with the maximum over other unevaluated nodes. The one-query knowledge gradient (Frazier et al., 2008) is $\delta f(-\mathrm{gap}_t(v)/\delta)$, where $f(z) = z\Phi(z) + \varphi(z)$ and Φ, φ are the standard normal distribution and density. It is the expected reduction in selection loss for the local linearized scores (Proposition C.1). It increases with the informative score variation δ and decreases with the competitive gap.

To account for measurements whose value emerges jointly, we use a KG($*$) lookahead (Frazier & Powell, 2010):

$$\mathrm{KG}_t^*(v) = \max_{1 \leq m \leq m_v} \frac{\delta_m(v)}{m} f\left(-\frac{\mathrm{gap}_t(v)}{\delta_m(v)}\right), \quad \delta_m(v)^2 = c_v^\top(\Sigma_v - \Sigma_v^{(m)})c_v. \tag{9}$$

Here $\Sigma_v^{(m)}$ follows m greedy best-direction queries with c_v fixed, and m_v is the number of available independent directions, at most $d - \operatorname{rank} G_v$, where G_v stacks the measured directions. Covariance updates do not depend on the answers, so plans can be scored before querying. Among unevaluated nodes with an informative unobserved landmark pair, we choose the largest score, execute its first query, and replan after the answer. Zero-variance plans have value zero. Algorithm 4 gives the allocation details.

Row queries. A revealed row measures every other node along a common direction, so all unrevealed nodes share the same membership covariance, although its effect on their features and scores is node dependent (Proposition B.1). Before the revealed rows span the membership space a row adds one direction to every unrevealed node at once, and afterwards it adds nothing to the working covariance, which leaves a row rule little to allocate beyond the node it pins. We therefore reveal the rows of evaluated nodes first, whose feature errors enter every prediction through the shared coefficient, and sample the remaining rows uniformly without replacement (Algorithm 3). Section 5.4 compares this rule with revealing the row of the current leader, which targets the pinned node, and with local rows.

4.4 Theoretical Guarantees

We analyze how estimated geometry affects optimization relative to a fixed reference surrogate: a full-information minimizer of (3), with fixed parameters in (4). Let (μ_t°, s_t°) and $U_t^\circ = \mu_t^\circ + \kappa s_t^\circ$ be its posterior and score on the actual evaluated set. Define $\Delta_t(v) = |U_t(v) - U_t^\circ(v)|$, which includes errors in both the candidate features and the shared coefficient estimate.

Assumption 1 (Confidence for the reference model)**.** For every $t < T$ and unevaluated v, $|y(v) - \mu_t^\circ(v)| \leq \kappa s_t^\circ(v)$.

For a draw from the fixed reference GP this holds with probability at least $1 - \delta_0$ for $\kappa^2 = 2\log(2nT/\delta_0)$, whatever the evaluation and query rules, since edge queries are functions of the history and of the fixed matrix A (Srinivas et al., 2010). Appendix C(H) also gives a deterministic sufficient condition.

When the optimum is unevaluated, maximizing U_t incurs the per-round cost

$$y(v^*) - y(v^{(t+1)}) \leq 2\kappa s_t^\circ(v^{(t+1)}) + \Delta_t(v^{(t+1)}) + \Delta_t(v^*). \tag{10}$$

Thus the geometry is charged through score errors at the selected node and the optimum. These errors include the influence of evaluated features on all predictions. With the reference bulk variance and coefficient state, they reduce to feature errors via $\Delta_t(v) \leq \Lambda_t \|\hat{\phi}_v - \phi_v\|_2$, with $\Lambda_t = \|w_t\|_2 + \kappa \|\Xi_t\|_2^{1/2}$ in a common feature basis (Lemma C.2). This links the local score sensitivity used for acquisition to the decision error in the analysis.

Theorem 1 (Simple regret with estimated geometry)**.** *Under Assumption 1, let* $v_{\max} = \max_v \phi_v^\top D \phi_v$ *for the fixed reference kernel. Then*

$$r_T \le 2\kappa\sigma_b + 2\kappa\sqrt{\frac{c_1\gamma_T}{T}} + \frac{1}{T}\sum_{t<T}\big(\Delta_t(v^{(t+1)}) + \Delta_t(v^*)\big),$$

where $c_1 = 2\max(\sigma_b^2, v_{\max})$ *and* $\gamma_T = d_+ \log\big(1+(T+N_0)v_{\max}/(d_+\sigma_b^2)\big)$*, with* $\gamma_T = 0$ *if* $d_+ = 0$*.*

The terms separate node-specific variation, learning in d_+ structured directions, and the cost of estimating the geometry. The bound applies to both query types. Its proof and a version for subsets of rounds are in Appendix C(H).

When structure is identified. If A_{off} has a completion in $\mathcal{C}_d$, C is nonsingular, every d memberships are independent, and $n \ge 2d+1$, then $d+1$ distinct rows identify the surrogate, whereas d rows need not suffice (Proposition A.1). With the reference kernel parameters, a zero-residual fit, and exact coefficient conditioning thereafter, $\Delta_t = 0$ after identification. Applying the bound to those rounds gives $r_T \le 2\kappa\sigma_b + 2\kappa\sqrt{c_1\gamma_T/(T-t_{\text{id}})}$, where t_{id} counts decisions before identification (Corollary A.1). Pair queries instead control node-specific measurement sets. Their identification conditions and feature-error specialization, together with the local estimation bound outside the class, are in Appendix A.

5 Experiments

We organize the experiment results around three questions. **(Q1)** When objective evaluations and structural queries are both budgeted, does a global surrogate learned from the revealed entries find high-value nodes that structure-free search and local graph BO do not? **(Q2)** How much of the remaining gap to a fully observed graph is due to the rank-d representation, and how much to estimating it from partial observations? **(Q3)** When does decision-aware acquisition of structure improve on uninformed queries? Protocols, complete sweeps, and controls are in Appendix F.

5.1 Experimental Setup

Tasks. *In-class*: dense affinities on $n = 1{,}000$ nodes drawn from the class at rank 16, with two objectives, a draw from the class-kernel prior and a latent objective defined without reference to the kernel. *Real*: ESOL, 1,117 molecules with Tanimoto affinities and measured solubility (Delaney, 2004), BioPlex 3.0 protein interaction networks assayed in HEK293T ($n = 13{,}167$) and HCT116 ($n = 9{,}609$) cells with DepMap gene-essentiality objectives (Huttlin et al., 2021; Tsherniak et al., 2017), the negated CRISPR gene effect of each protein's gene averaged over all screened cell lines (mean essentiality) or taken in one line, the colon adenocarcinoma line C2BBe1, the myeloma line OPM-2, and HCT116 itself for the HCT116 network, and Chameleon ($n = 1{,}223$), a Wikipedia page graph with log traffic as the objective (Rozemberczki et al., 2021). Sparse DCSBM and latent-position graphs are in Appendix F.2. **Protocol.** Every method starts from the same N_0 evaluated nodes within each task and seed. Our row and pair methods receive the rows or the pair block among those nodes, then alternate one evaluation with an allowance of Q_R rows or Q_E pairs per round for T rounds, at the budgets of Table 3. **Methods.** *Ours* uses the structured-plus-bulk kernel with $d = 16$ in the row studies, chosen on separate design seeds and

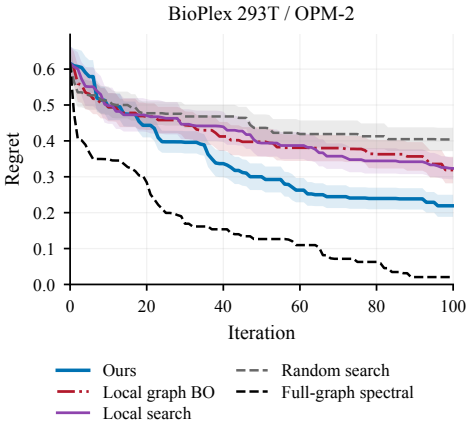

Figure 2: Row queries on BioPlex 293T with the OPM-2 essentiality objective: mean normalized simple regret with pointwise ± 1 SEM bands over 20 paired seeds, $Q_R = 10$ rows per round. All non-oracle arms come from the campaign of Table 1, and local graph BO (Wan et al., 2023) receives the same row allowance. The dashed full-graph spectral reference ($d = 16$) sees the entire graph. The dense and ESOL trajectories of the same campaign are in Figure 5.

Table 1: Row queries: final normalized simple regret, mean ± SEM over 20 paired seeds, at the budgets of Table 3. Bold marks the lowest mean among the four non-oracle methods. All methods receive the same structural allowance per round, with the neighborhood construction of local graph BO (Wan et al., 2023) described in Appendix F.1. The full-graph spectral reference ($d = 16$) observes the graph before BO.

Instance / objective	Ours	Random	Local search	Local graph BO	Full graph spectral
Dense / class prior	**.008 ± .008**	.134 ± .020	.241 ± .058	.330 ± .060	.043 ± .026
Dense / latent	**.158 ± .013**	.201 ± .027	.196 ± .024	.187 ± .023	.170 ± .026
ESOL / solubility	**.061 ± .014**	.147 ± .018	.068 ± .013	.095 ± .015	.005 ± .005
Chameleon / log traffic	**.079 ± .005**	.097 ± .012	.108 ± .024	.088 ± .023	.077 ± .001
293T / mean essentiality	**.304 ± .042**	.468 ± .032	.376 ± .035	.333 ± .045	.000 ± .000
293T / C2BBe1	**.226 ± .038**	.513 ± .034	.446 ± .032	.437 ± .034	.000 ± .000
293T / OPM-2	**.219 ± .031**	.404 ± .033	.323 ± .031	.319 ± .037	.021 ± .014
HCT116 / mean essentiality	**.323 ± .023**	.408 ± .025	.339 ± .010	.341 ± .030	.330 ± .003
HCT116 / C2BBe1	**.349 ± .010**	.456 ± .031	.363 ± .012	.400 ± .019	.330 ± .001
HCT116 / own line (HCT116)	**.359 ± .021**	.476 ± .021	.392 ± .019	.389 ± .033	.298 ± .003

held fixed across tasks, and $d = 32$ in the pair study, with pin-then-uniform rows or KG* pairs and hyperparameters by marginal likelihood on the revealed data. We compare with random node search, which buys no structure, local search with restarts over the revealed neighborhoods, and local graph BO (Wan et al., 2023), which explores the revealed neighborhood with a local kernel and receives the same structural allowance as Ours under either query type. On the dense and ESOL affinities its neighborhoods come from a precomputed kNN view of the full graph (Appendix G.3). Full-graph spectral BO on the true Laplacian sees the entire graph and is the oracle reference. **Metric.** Normalized simple regret $R_t = \big(y^* - \max_{v \in \mathcal{V}^{(t)}} y(v)\big) / \big(y^* - \mathrm{median}_v\, y(v)\big)$ after t evaluations, mean ± SEM over 20 seeds that share the initial nodes.

5.2 Optimization under Budgeted Structural Access

Table 1 and Figure 2 answer Q1 with one complete, paired-seed row-query campaign, in which roughly one row in ten is revealed on the dense, ESOL, and BioPlex instances (Table 3). Ours has the lowest mean final regret among the four non-oracle methods on all ten objectives, including the dense latent objective, which is not drawn from the kernel. On the dense in-class objective the fit to the revealed rows is within seed variability of the class fitted to the entire graph and below the full-graph spectral reference at the same d (Table 5), so the revealed rows already carry the geometry. Ours closes most of the distance from random search to the full-graph spectral reference on the dense, Chameleon, and HCT116 objectives, and between a third and two thirds of it on ESOL and the 293T objectives, where Section 5.3 decomposes the remaining gap. On BioPlex the advantage appears within the first twenty evaluations and persists (Figure 2), and Figure 6 places the 1,000-row column inside the full row-budget sweep.

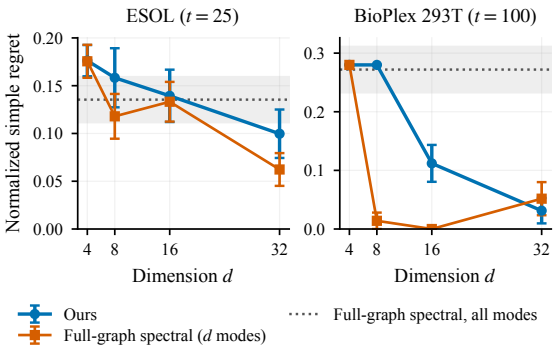

Figure 3: Representation under full graph access, mean ±1 SEM over 20 seeds: the class fitted to the entire graph against the d smoothest modes of the true Laplacian, with the all-mode kernel, tuned on its own grid, as a reference. BioPlex at $t = 100$, ESOL at $t = 25$, before the regret floor is reached.

5.3 Surrogate Representation and Resolution

Full graph access separates the two parts of the gap to the oracle. The estimation gap of a task is the regret of Ours on the revealed rows minus that of the same class fitted to the entire graph, and the representation gap is the latter minus the full-graph spectral reference at the same d (Tables 1 and 5, Appendix F.2). On the dense instance and Chameleon the partial fit is at or below the full fit, so what remains is representation. On ESOL and on 293T with mean essentiality estimation dominates, on the other two 293T objectives the two gaps are comparable, and on HCT116 both are small. Figure 3 sweeps the resolution d under full graph access. On the sparse binary BioPlex graph the eigenbasis reaches its floor at $d = 8$ while the class needs $d = 32$, so the class pays for not having the eigenbasis with a higher resolution, on ESOL the class improves across the sweep, and on DCSBM both arms improve up to $d = 16$ (Figure 7). At $d = 32$ every truncated arm is below the all-mode kernel on all three instances, and under the row-study protocol the all-mode kernel is also above the class fitted to the entire graph on the dense and ESOL instances (Table 5), so at these resolutions truncation costs no regret relative to the untruncated kernel.

5.4 Value of Decision-Aware Structure Acquisition

Under pair queries each revealed entry informs one node, and Section 4.3 prices it by its expected effect on the next decision. Figure 4 tests this on ESOL with $d = N_0 = 32$ and $T = 200$. All methods start from the same evaluated nodes and the pair block among them, and acquire $Q_E \in \{2, 4, 8, 16, 32\}$ new pairs before every evaluation. Ours (KG) has the lowest final regret at every budget. Its gap to local graph BO, adapted to the pair oracle, opens immediately and is significant at every $Q_E \geq 4$ after adjusting for the five budgets. Its gap to uniform pairs, which share the surrogate and the kernel, opens after about fifty evaluations, has the same sign at every budget, and is largest at the two largest budgets, where the regret of Ours (KG) is less than half that of uniform pairs (Table 8). The difference between the two Ours arms is the acquisition-and-conditioning procedure of Section 4.3, which spends each query on a candidate whose feature is undetermined and whose score is close to the top.

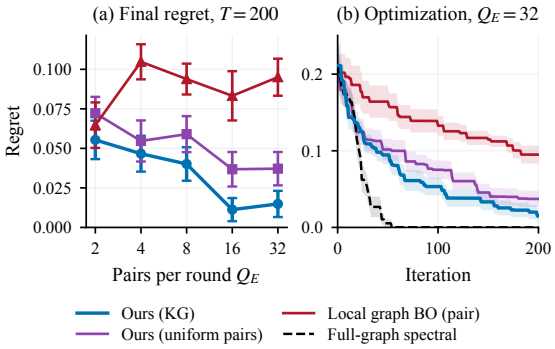

Figure 4: Pair queries on ESOL, $d = N_0 = 32$, $T = 200$, 20 shared seeds. (a) Final normalized regret against the pairs acquired per round, ± 1 SEM. (b) Regret trajectories at $Q_E = 32$ with ± 1 SEM bands. The dashed full-graph spectral reference sees the entire graph. All five budgets are in Figure 9.

Under row queries, neither UCB-leader rows nor local BFS rows improves on pin-then-uniform on ESOL consistently, with differences comparable to the SEM (Table 9). Together with the shared measurement pattern of Proposition B.1, these results support the simple coverage-based default. The pair study evaluates the benefit of targeting structure at competitive candidates.

6 Conclusion

We studied Bayesian optimization on graphs whose affinities are revealed only through budgeted edge queries, and replaced local exploration of the revealed subgraph by a global geometry inferred from it. A graph-valid rank-d surrogate with an exactly computable spectrum makes the GP a linear model in node features, in which evaluations learn the shared coefficient and edge queries learn the features, so that each edge query is priced by its expected effect on the next decision. The analysis charges the estimated geometry at two nodes, $d + 1$ rows identify an in-class surrogate, and structural access enters the regret bound only through the identification time. Across protein interaction networks, a molecular similarity graph, a Wikipedia page graph, and synthetic graphs, the method attains the lowest final regret among the non-oracle methods on all ten row-query objectives, the decision-aware pair rule improves on uniform pairs at every budget, and full-information controls separate the representation gap from the estimation gap.

AI USE STATEMENT

In this work, we used generative AI tools for writing and coding assistance: polishing prose, improving clarity of exposition, and assisting with experiment code. We have reviewed all AI-assisted work and take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work studies black-box optimization on the nodes of a partially observed graph. It uses only public data, the BioPlex 3.0 protein interaction networks with DepMap gene-effect scores, the ESOL solubility set with computed molecular fingerprints, the Chameleon Wikipedia page graph, and synthetic graphs, none of which contains personal information, and it releases no new dataset. Sample-efficient identification of high-value nodes supports beneficial uses such as target prioritization in biology and chemistry and infrastructure hardening, and could in principle support targeted influence on social networks. The setting of expensive node evaluations and partially observed topology limits but does not remove this concern. The authors declare no competing interests.

REPRODUCIBILITY STATEMENT

The assumptions and proofs underlying our theoretical results are provided in Section 4.4 and Appendices A to C. Section 4 and Appendices B, D, and E specify the algorithm, structural acquisition rules, solver, initialization, and refit schedule. Section 5 and Appendices F and G describe the datasets, graph construction, objective functions, baselines, hyperparameters, and evaluation protocols. Table 3 summarizes the initialization, evaluation horizons, and structural query budgets. Appendix F.1 details the repeated-run protocol and statistical reporting. Anonymized source code and experiment configurations are provided in the supplementary material and in the anonymized repository linked in the abstract.

REFERENCES

Yasin Abbasi-yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger (eds.), *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011.

Edoardo M. Airoldi, David M. Blei, Stephen E. Fienberg, and Eric P. Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9(65):1981–2014, 2008.

Dávid Bajusz, Anita Rácz, and Károly Héberger. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of cheminformatics*, 7(1):20, 2015.

Paul A. Bekker and Jos M.F. ten Berge. Generic global indentification in factor analysis. *Linear Algebra and its Applications*, 264:255–263, 1997. ISSN 0024-3795.

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

Mikhail Belkin and Partha Niyogi. Semi-supervised learning on riemannian manifolds. *Machine learning*, 56(1):209–239, 2004.

Viacheslav Borovitskiy, Iskander Azangulov, Alexander Terenin, Peter Mostowsky, Marc Deisenroth, and Nicolas Durrande. Matérn gaussian processes on graphs. In *International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pp. 2593–2601. PMLR, 2021.

Emmanuel Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Communications of the ACM*, 55(6):111–119, June 2012.

Emmanuel J. Candès and Terence Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, 2010.

Sourav Chatterjee. Matrix estimation by universal singular value thresholding. *The Annals of Statistics*, 43(1):177–214, 2015.

Kamalika Chaudhuri, Fan Chung, and Alexander Tsiatas. Spectral clustering of graphs with general degrees in the extended planted partition model. In Shie Mannor, Nathan Srebro, and Robert C. Williamson (eds.), *Proceedings of the 25th Annual Conference on Learning Theory*, volume 23 of *Proceedings of Machine Learning Research*, pp. 35.1–35.23, 2012.

Jiaxu Cui and Bo Yang. Graph bayesian optimization: Algorithms, evaluations and applications. *arXiv preprint arXiv:1805.01157*, 2018.

Jiaxu Cui, Bo Yang, and Xia Hu. Deep bayesian optimization on attributed graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 1377–1384, 2019.

John S. Delaney. ESOL: estimating aqueous solubility directly from molecular structure. *Journal of Chemical Information and Computer Sciences*, 44(3):1000–1005, 2004.

Peter I Frazier. A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*, 2018.

Peter I. Frazier and Warren B. Powell. Paradoxes in learning and the marginal value of information. *Decision Analysis*, 7(4):378–403, 2010.

Peter I. Frazier, Warren B. Powell, and Savas Dayanik. A knowledge-gradient policy for sequential information collection. *SIAM Journal on Control and Optimization*, 47(5):2410–2439, 2008.

Agathe Girard, Carl Rasmussen, Joaquin Quiñonero Candela, and Roderick Murray-Smith. Gaussian process priors with uncertain inputs application to multiple-step ahead time series forecasting. In S. Becker, S. Thrun, and K. Obermayer (eds.), *Advances in Neural Information Processing Systems*, volume 15, 2002.

Peter D Hoff, Adrian E Raftery, and Mark S Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.

Edward L. Huttlin, Raphael J. Bruckner, Jose Navarrete-Perea, Joe R. Cannon, Kurt Baltier, Fana Gebreab, Melanie P. Gygi, Alexandra Thornock, Gabriela Zarraga, Stanley Tam, John Szpyt, Brandon M. Gassaway, Alexandra Panov, Hannah Parzen, Sipei Fu, Arvene Golbazi, Eila Maenpaa, Keegan Stricker, Sanjukta Guha Thakurta, Tian Zhang, Ramin Rad, Joshua Pan, David P. Nusinow, Joao A. Paulo, Devin K. Schweppe, Laura Pontano Vaites, J. Wade Harper, and Steven P. Gygi. Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. *Cell*, 184(11):3022–3040.e28, 2021. ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2021.04.011. URL `https://www.sciencedirect.com/science/article/pii/S0092867421004463`.

Kirthevasan Kandasamy, Gautam Dasarathy, Junier B Oliva, Jeff Schneider, and Barnabás Póczos. Gaussian process bandit optimisation with multi-fidelity evaluations. volume 29, 2016.

Brian Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Phys. Rev. E*, 83:016107, Jan 2011. doi: 10.1103/PhysRevE.83.016107.

Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from a few entries. *IEEE Transactions on Information Theory*, 56(6):2980–2998, 2010.

Risi Imre Kondor and John D. Lafferty. Diffusion kernels on graphs and other discrete input spaces. In *Proceedings of the Nineteenth International Conference on Machine Learning (ICML)*, 2002.

Ksenia Korovina, Sailun Xu, Kirthevasan Kandasamy, Willie Neiswanger, Barnabas Poczos, Jeff Schneider, and Eric Xing. Chembo: Bayesian optimization of small organic molecules with synthesizable recommendations. In *International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 3393–3403. PMLR, 2020.

Akshay Krishnamurthy and Aarti Singh. Low-rank matrix and tensor completion via adaptive sampling. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

Can M. Le, Elizaveta Levina, and Roman Vershynin. Concentration and regularization of random graphs. *Random Structures & Algorithms*, 51(3):538–561, 2017.

Walter Ledermann. On the rank of the reduced correlational matrix in multiple-factor analysis. *Psychometrika*, 2(2):85–93, 1937.

Hanbyul Lee, Rahul Mazumder, Qifan Song, and Jean Honorio. Matrix completion from general deterministic sampling patterns. *arXiv preprint arXiv:2306.02283*, 2023.

Jure Leskovec and Julian Mcauley. Learning to discover social circles in ego networks. In *Advances in Neural Information Processing Systems (NIPS)*, volume 25, 2012.

Huidong Liang, Xingchen Wan, and Xiaowen Dong. Bayesian optimization of functions over node subsets in graphs. In *Advances in Neural Information Processing Systems*, volume 37, pp. 38199–38234, 2024.

László Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, 2012.

Michael W Mahoney and Petros Drineas. CUR matrix decompositions for improved data analysis. *Proceedings of the National Academy of Sciences*, 106(3):697–702, 2009.

Xueyu Mao, Purnamrita Sarkar, and Deepayan Chakrabarti. Estimating mixed memberships with sharp eigenvector deviations. *Journal of the American Statistical Association*, 116(536):1928–1940, 2021.

Andrew Mchutchon and Carl Rasmussen. Gaussian process training with input noise. In *Advances in Neural Information Processing Systems*, volume 24, 2011.

Changyong Oh, Jakub Tomczak, Efstratios Gavves, and Max Welling. Combinatorial bayesian optimization using the graph cartesian product. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Rafael Oliveira, Lionel Ott, and Fabio Ramos. Bayesian optimisation under uncertain inputs. In *The 22nd international conference on artificial intelligence and statistics*, pp. 1177–1184. PMLR, 2019.

Tai Qin and Karl Rohe. Regularized spectral clustering under the degree-corrected stochastic blockmodel. In *Advances in Neural Information Processing Systems*, volume 26, 2013.

Dhanesh Ramachandram, Michal Lisicki, Timothy J. Shields, Mohamed R. Amer, and Graham W. Taylor. Bayesian optimization on graph-structured search spaces: Optimizing deep multimodal fusion architectures. *Neurocomputing*, 298:80–89, 2018. ISSN 0925-2312.

David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of chemical information and modeling*, 50(5):742–754, 2010.

Benedek Rozemberczki, Carl Allen, Rik Sarkar, and xx Thilo Gross. Multi-scale attributed node embedding. *Journal of Complex Networks*, 9(1):1–22, 2021.

Binxin Ru, Xingchen Wan, Xiaowen Dong, and Michael Osborne. Interpretable neural architecture search via bayesian optimisation with weisfeiler–lehman kernels. In *International Conference on Learning Representations*, 2021.

Alexander Shapiro. Identifiability of factor analysis: Some results and open problems. *Linear Algebra and its Applications*, 70:1–7, 1985.

Alexander J Smola and Risi Kondor. Kernels and regularization on graphs. In *Learning theory and kernel machines: 16th annual conference on learning theory and 7th kernel workshop*, pp. 144–158. Springer, 2003.

Jasper Snoek, Hugo Larochelle, and Ryan Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems*, volume 25, 2012.

Niranjan Srinivas, Andreas Krause, Sham Kakade, and Matthias Seeger. Gaussian process optimization in the bandit setting: no regret and experimental design. In *Proceedings of the 27th International Conference on Machine Learning*, pp. 1015–1022, 2010.

Aviad Tsherniak, Francisca Vazquez, Phil G. Montgomery, Barbara A. Weir, Gregory Kryukov, Glenn S. Cowley, Stanley Gill, William F. Harrington, Sasha Pantel, John M. Krill-Burger, Robin M. Meyers, Levi Ali, Amy Goodale, Yenarae Lee, Guozhi Jiang, Jessica Hsiao, William F.J. Gerath, Sara Howell, Erin Merkel, Mahmoud Ghandi, Levi A. Garraway, David E. Root, Todd R. Golub, Jesse S. Boehm, and William C. Hahn. Defining a cancer dependency map. *Cell*, 170(3):564–576.e16, 2017. ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2017.06.010.

Michal Valko, Rémi Munos, Branislav Kveton, and Tomáš Kocák. Spectral bandits for smooth graph functions. In *International conference on machine learning*, pp. 46–54. PMLR, 2014.

Xingchen Wan, Pierre Osselin, Henry Kenlay, Binxin Ru, Michael A Osborne, and Xiaowen Dong. Bayesian optimisation of functions on graphs. volume 36, pp. 43012–43040, 2023.

Colin White, Willie Neiswanger, and Yash Savani. Bananas: Bayesian optimization with neural architectures for neural architecture search. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 10293–10301, 2021.

Yilin Xie, Shiqiang Zhang, Jixiang Qing, Ruth Misener, and Calvin Tsay. Global optimization of graph acquisition functions for neural architecture search. *arXiv preprint arXiv:2505.23640*, 2025.

Yilin Xie, Shiqiang Zhang, Jixiang Qing, Ruth Misener, and Calvin Tsay. BoGrape: Bayesian optimization over graphs with shortest-path encoded. In *International Conference on Learning Representations*, volume 2026, pp. 46561–46590, 2026.

Jiaming Xu. Rates of convergence of spectral methods for graphon estimation. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, pp. 5433–5442, 2018.

Stephen J Young and Edward R Scheinerman. Random dot product graph models for social networks. In *Algorithms and Models for the Web-Graph*, pp. 138–149, 2007.

Zirui Yuan, Minglai Shao, and Zhiqian Chen. Graph bayesian optimization for multiplex influence maximization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 22475–22483, 2024.

A Identification and Estimation of the Surrogate

We collect the identification and estimation results used in Section 4.4, together with the row and pair specializations of its regret bound.

Corollary A.1 (Row queries in the class)**.** *Suppose A_{off} has a completion in $\mathcal{C}_d$. Let t_{id} count decisions before the observed rows identify that completion, and suppose the solver returns a zero-residual fit thereafter. Under Theorem 1, using the reference kernel parameters and exact coefficient conditioning after identification, for $T > t_{\mathrm{id}}$,*

$$r_T \le 2\kappa\sigma_b + 2\kappa\sqrt{c_1\gamma_T/(T - t_{\mathrm{id}})}.$$

If C is nonsingular, memberships are in general position, $n \ge 2d+1$, and N_0 initial rows are followed by $Q_R \ge 1$ distinct rows before each decision, then $t_{\mathrm{id}} \le (\lceil (d+1-N_0)_+/Q_R \rceil - 1)_+$. For interior representatives this threshold is also necessary in general.

The proof is in Appendix C(H). The bound concerns identification from observations and recovery by the solver as separate conditions.

Assumption A.1 (Split smoothness)**.** Let $K_s = \sigma_s^2 \tilde{U}_+ D_+ \tilde{U}_+^\top$ be the structured part of the kernel of Eq. (4) built from the reference surrogate. There are known $B_s > 0$ and $\eta \ge 0$ such that $y = y_s + y_b$ with $\|y_s\|_{\mathcal{H}(K_s)} \le B_s$ and $\|y_b\|_\infty \le \eta$.

This condition separates variation on the structured modes from a bounded node-specific residual. Appendix C(H) gives the resulting deterministic confidence condition.

Remark A.1 (Node-specific variation under a finite evaluation budget). In Assumption A.1, take $y_s = 0$ and $y_b = \eta \mathbf{1}_{v_0}$, where v_0 is uniform on $\mathcal{V}$ independently of A. Until v_0 is evaluated, the observations match those of the zero objective. With $N_0 + T < n$ evaluations, the probability of finding it is at most $(N_0 + T)/n$, regardless of structural queries. Hence expected simple regret is at least $(1 - (N_0 + T)/n)\eta$, and some fixed spike location has at least this expected regret over the algorithm's randomness. The bound holds for every structural budget and every adaptive query rule, since edge queries reveal A and not y.

Under pair queries the errors are node specific, and Theorem 1 specializes as follows.

Corollary A.2 (Pair queries)**.** *Under the conditions of Theorem 1, if at every decision the algorithm uses the reference kernel parameters and the evaluated nodes are pinned with exact features, then every pair-query rule satisfies*

$$r_T \;\le\; 2\kappa\,\sigma_b \;+\; 2\kappa\sqrt{\frac{c_1\,\gamma_T}{T}} \;+\; \frac{1}{T}\sum_{t<T}\Lambda_t\big(e_t(v^{(t+1)}) + e_t(v^*)\big),$$

with $e_t(v) = \|\hat{\phi}_v - \phi_v\|_2$ in a common feature basis at decision time. If moreover $A_{\mathrm{off}} \in \mathcal{C}_d$, the landmark skeleton is identified with exact global quantities (R, V, h), and the spanning and pinning-allowance conditions of Lemma D.1 hold, Algorithm 4 makes both errors vanish in every round with $Q_E \ge d(m_t + 1)$, where m_t is the number of unpinned unevaluated nodes at the start of the round.

The proof is in Appendix C(L).

Identification. Two conventions. A set of membership vectors is in *general position* if every d of them are linearly independent. A representative (Π, C) of a class member is *interior* if every π_i lies in the interior of Δ_d and C is nonsingular with entries in $(0, 1)$. Throughout, $M = \Pi C \Pi^\top$ denotes the surrogate including its diagonal, which is the object the kernel uses, and the revealed entries are entries of $M_{\mathrm{off}} = A_{\mathrm{off}}$.

Proposition A.1 (Identification)**.** *Let $A_{\mathrm{off}} = M_{\mathrm{off}}$ with $M = \Pi C \Pi^\top$, $\pi_i \in \Delta_d$, $C = C^\top$ nonsingular, and* $\operatorname{rank}\Pi = d$. *(i)* Count. *The set of such M contains a nonempty open subset of the manifold of symmetric matrices of rank d, whose dimension is $nd - \binom{d}{2}$, so no set of fewer than $nd - \binom{d}{2}$ entries determines M for every member of the class. (ii)* Rows. *Let $\mathcal{R}$ be the revealed rows and $r = |\mathcal{R}|$. If* $\operatorname{rank}\Pi_{\mathcal{R}\setminus\{u\}} = d$ *for every $u \in \mathcal{R}$ and* $\operatorname{rank}\Pi_{\mathcal{V}\setminus\mathcal{R}} = d$, *then M is the only symmetric matrix of rank at most d that agrees with A on the revealed entries, and every zero-residual point of Problem* (3) *equals M. Both conditions hold when $r \ge d + 1$, $n - r \ge d$, and the memberships are in general position, and for pure memberships $\pi_i = \theta_i e_{k(i)}$ with $\theta_i > 0$ they held exactly when every community has two revealed rows and one unrevealed node. Conversely, if $r \le d$,* $\operatorname{rank}\Pi_{\mathcal{R}} = r$, *and the representative is interior, the revealed rows do not determine M. (iii)* Pairs. *Let $\mathcal{R}$ be a set of r nodes with all pairs among them revealed. If $r < r_0(d) := \lceil (2d + 1 + \sqrt{8d+1})/2 \rceil$, the revealed block does not determine $M_{\mathcal{R}\mathcal{R}}$ for every member of the class. If $\mathcal{R}$ contains disjoint $\mathcal{R}_1, \mathcal{R}_2$ with* $\operatorname{rank}\Pi_{\mathcal{R}_1\setminus\{u\}} = d$ *for every $u \in \mathcal{R}_1$ and* $\operatorname{rank}\Pi_{\mathcal{R}_2} = d$, *which holds for $r \ge 2d + 1$ in general position, then the block determines $M_{\mathcal{R}\mathcal{R}}$. A node $v \notin \mathcal{R}$ whose entries against $R_v \subseteq \mathcal{R}$ are revealed with* $\operatorname{rank}\Pi_{R_v} = d$ *has $M_{v,\mathcal{R}}$ determined, and M is determined once every node is so measured.*

The proof is in Appendix C(G). The row design with $r = d + 1$ reveals $(d + 1)(n - 1) - \binom{d+1}{2}$ distinct entries and the pair design $\binom{r}{2} + (n - r)d$, both of order nd, so both are order-optimal by (i). The count in (iii) is Ledermann's bound of factor analysis (Ledermann, 1937): $\binom{r}{2} \ge rd - \binom{d}{2}$ is the condition under which a symmetric $r \times r$ matrix of rank d with unknown diagonal has no more free parameters than observed entries. For positive semidefinite low-rank parts, generic identifiability of the low-rank part from the off-diagonal entries is classical when Ledermann's bound is strictly satisfied (Shapiro, 1985; Bekker & ten Berge, 1997). Our core C is indefinite in general, and we use only the constructive condition $r \ge 2d + 1$. Completion from uniformly sampled entries needs of the order of $\mu_0 nd \log n$ entries under an incoherence assumption with parameter μ_0 (Candès & Recht, 2012), whereas the landmark designs need no logarithmic factor and no incoherence, because every node is measured against the same landmarks and the problem decouples into n small linear systems. The proof of (ii) is constructive and exhibits the estimator as a skeleton decomposition $M = M_{\cdot,\mathcal{R}} M_{\mathcal{R}\mathcal{R}}^{+} M_{\mathcal{R},\cdot}$ (Mahoney & Drineas, 2009), once the r diagonal entries of the landmark block are filled in from the column space of $M_{\mathcal{R},\mathcal{V}\setminus\mathcal{R}}$.

Estimation error outside the class. Let $(\Pi^\star, C^\star)$ be the reference fit, a minimizer of Eq. (11), with features $\phi_v^\star$, and let $(\hat{\Pi}, \hat{C})$ be a fit to the revealed rows $\mathcal{R}$. Write $B^\star = \Pi^\star_{\mathcal{R}} C^\star$ and $\hat{B} = \hat{\Pi}_{\mathcal{R}} \hat{C}$ for the two $r \times d$ matrices of landmark directions, and $\rho_v = A_{\mathcal{R},v} - B^\star \pi_v^\star$ for the residual of the reference fit on the revealed entries of an unrevealed node v, which is gauge invariant and vanishes in the class.

Proposition A.2 (Per-node estimation error under row queries)**.** *Let $\hat{B}$ have rank d and let $\hat{\pi}_v$, $v \notin \mathcal{R}$, minimize Problem* (3) *over $\pi_v \in \Delta_d$ with $(\hat{\Pi}_{\mathcal{R}}, \hat{C})$ fixed, as every minimizer of Problem* (3) *does, since the objective is separable across unrevealed nodes. Then for every $v \notin \mathcal{R}$,*

$$\|\hat{\pi}_v - \pi_v^\star\|_2 \;\le\; \frac{\|P_{\hat{B}}\,\rho_v\|_2 + \|\hat{B} - B^\star\|_2}{\sigma_{\min}(\hat{B})} \;\le\; \frac{\|\hat{B}^\top \rho_v\|_2}{\sigma_{\min}(\hat{B})^2} + \frac{\|\hat{B} - B^\star\|_2}{\sigma_{\min}(\hat{B})},$$

where $P_{\hat{B}}$ is the orthogonal projector onto the column space of $\hat{B}$. With the reference feature map $\psi(\pi) = F\pi/\sqrt{\pi^\top h + \zeta}$ of Eq. (14) *held fixed, ψ is Lipschitz on Δ_d with $L_\phi \le \|F\|_2 \zeta^{-1/2}(1 + \|h\|_2/(2\zeta))$. Thus $\|\psi(\hat{\pi}_v) - \psi(\pi_v^\star)\|_2$ obeys the same bound multiplied by L_ϕ. The statement holds for every gauge-equivalent representative $(\Pi^\star S, S^{-1} C^\star S^{-\top})$ of the reference fit whose memberships lie in Δ_d.*

Proof. Fix $v \notin \mathcal{R}$ and write $a = A_{\mathcal{R},v}$. The terms of Problem (3) that involve π_v are proportional to $f(\pi) = \frac{1}{2}\|a - \hat{B}\pi\|_2^2$, so $\hat{\pi}_v \in \arg\min_{\pi \in \Delta_d} f(\pi)$. Since $\hat{B}$ has full column rank, $f(\pi) = \frac{1}{2}\|\hat{B}(\pi - \tilde{\pi})\|_2^2 + \frac{1}{2}\|(I - P_{\hat{B}})a\|_2^2$ with $\tilde{\pi} = \hat{B}^+ a$, so $\hat{\pi}_v$ is the projection of $\tilde{\pi}$ onto the closed convex set Δ_d in the norm $\|x\|_{\hat{B}} = \|\hat{B}x\|_2$. Metric projections onto convex sets are nonexpansive and fix the points of the set, so for every $\pi \in \Delta_d$, $\|\hat{B}(\hat{\pi}_v - \pi)\|_2 \le \|\hat{B}(\tilde{\pi} - \pi)\|_2 = \|P_{\hat{B}}(a - \hat{B}\pi)\|_2$. Take $\pi = \pi_v^\star$ and write $a - \hat{B}\pi_v^\star = \rho_v + (B^\star - \hat{B})\pi_v^\star$. Since $\|\pi_v^\star\|_2 \le \|\pi_v^\star\|_1 \le 1$ and $\|P_{\hat{B}}\|_2 \le 1$, we get $\|\hat{B}(\hat{\pi}_v - \pi_v^\star)\|_2 \le \|P_{\hat{B}}\rho_v\|_2 + \|\hat{B} - B^\star\|_2$, and dividing by $\sigma_{\min}(\hat{B})$ gives the first bound. The second follows from $P_{\hat{B}} = \hat{B}(\hat{B}^\top \hat{B})^{-1}\hat{B}^\top$ and $\|\hat{B}(\hat{B}^\top \hat{B})^{-1}\|_2 = 1/\sigma_{\min}(\hat{B})$. For the feature map, Eq. (15) gives the Jacobian $Fs\,(I - \frac{1}{2}s^2 \pi h^\top)$ with $s = (\pi^\top h + \zeta)^{-1/2} \le \zeta^{-1/2}$, because $h \ge 0$ entrywise and $\pi \ge 0$, and $\|\pi h^\top\|_2 \le \|h\|_2$ on Δ_d, so its operator norm is at most L_ϕ on the convex set Δ_d and the mean value inequality applies. Replacing $(\Pi^\star, C^\star)$ by $(\Pi^\star S, S^{-1} C^\star S^{-\top})$ maps $\pi_v^\star \mapsto S^\top \pi_v^\star$ and $B^\star \mapsto B^\star S^{-\top}$, leaves ρ_v and $P^\star$ unchanged, and the argument applies verbatim. □

The first term is specific to the node: its numerator $\hat{B}^\top \rho_v = \sum_{u \in \mathcal{R}} \hat{g}_u \rho_{uv}$ sums the residual of the reference fit at the revealed entries of v against the landmark directions, and its denominator $\sigma_{\min}(\hat{B})^2 = \lambda_{\min}\big(\sum_{u \in \mathcal{R}} \hat{g}_u \hat{g}_u^\top\big)$ is nondecreasing when directions are added with the existing directions held fixed. In the class $\rho_v = 0$, and after identification $\hat{B} = B^\star S^{-\top}$ for a gauge S by Proposition A.1, so both terms vanish. The second term, the error of the landmark directions, is shared by all nodes, as is the error of the global quantities (R, V, h) that the proposition holds fixed. Under pair queries the same argument applies with $\hat{B}$ replaced by the $|R_v| \times d$ matrix of directions against which v was probed. With exactly d probes that matrix is square, $P_{\hat{B}} = I$, and the bound reads $\|\rho_v\|_2/\sigma_{\min}(\hat{B})$, so the residual passes through without projection, whereas every additional row projects it onto a d-dimensional subspace of a growing space (Appendix F.4). Section 5.3 measures the total gap, shared errors included.

The oracle approximant and its bias. Nothing about A is assumed by the method. The price of finite resolution is measured by the deterministic bias functional

$$\varepsilon_d(A) := \min_{X \in \mathcal{C}_d} \|(A - X)_{\mathrm{off}}\|_F, \qquad P^\star \in \arg\min_{X \in \mathcal{C}_d} \|(A - X)_{\mathrm{off}}\|_F, \tag{11}$$

where $(\cdot)_{\mathrm{off}}$ zeroes the diagonal, since self-pairs are never queried. The minimum is attained (Appendix C(B)), and a fixed minimizer $P^\star$ is the reference surrogate of Section 4.4. Proposition A.2 shows how its residuals enter the features. The bias is non-increasing in d, vanishes when A_{off} lies in the class, and decays as $d^{-\alpha/\ell}$ when $A_{ij} = f(x_i, x_j)$ for α-Hölder f and latent positions in $[0,1]^\ell$ (Lemma B.4, Appendix B). For a sparse binary network the relevant comparison is between $L_\zeta(\tilde{A})$ and $L_\zeta(A)$ in operator norm, which degree regularization with ζ of the order of the mean degree keeps controlled (Chaudhuri et al., 2012; Qin & Rohe, 2013; Le et al., 2017).

Algorithm 1 Global BO on a Partially Observed Graph (one campaign)

Input: $\mathcal{V}$, T, N_0, resolution d, per-round budget of Q_R rows or Q_E pairs, pin share ρ_{pin}
Initialize: evaluate y on N_0 random nodes and reveal their rows or their pair block (the landmarks)

for $t = 0, \dots, T-1$ **do**
  Pin: reveal the rows, or d landmark probes, of unpinned evaluated nodes within the share ρ_{pin}
  Rows: spend the remaining rows uniformly at random over unrevealed nodes (Algorithm 3)
  Fit: solve (3) and compute the spectrum, features ϕ_v, and kernel (4) exactly (Lemma B.2)
  Posterior: $\Sigma_v, \Sigma_v^{\phi}$ by (16), (w_t, Ξ_t) by Lemma B.1 with (17) at unpinned evaluated nodes, U_t by (6)
  Pairs: spend the remaining pairs one at a time by KG_t^* (9), refreshing the queried node (Algorithm 4)
  Evaluate: $v^{(t+1)} = \arg\max_{v \notin \mathcal{V}^{(t)}} U_t(v)$, observe $y(v^{(t+1)})$, and add it to $\mathcal{V}^{(t+1)}$
end for
return $\hat{v} = \arg\max_{v \in \mathcal{V}^{(T)}} y(v)$

B Method Details

This appendix gives Algorithm 1, the round summarized in Section 4, together with the coefficient update, local feature model, exact spectral solve, and query implementation used there.

Coefficient posterior. The structured-plus-bulk kernel belongs to

$$\mathcal{K}(\tilde{U}) = \{\sigma_b^2 I + \tilde{U}\,\mathrm{diag}(a)\tilde{U}^\top : \sigma_b^2 > 0,\ a \geq 0\}. \tag{12}$$

Lemma B.1 (Reduction to a linear model in node features)**.** *Let $\tilde{K} = \sigma_b^2 I + \tilde{U} D \tilde{U}^\top \in \mathcal{K}(\tilde{U})$ with $D \succ 0$ diagonal, $\tilde{U}$ collecting the d_+ modes with $a_i > 0$, let ϕ_v be the transposed v-th row of $\tilde{U}$, $S = \mathcal{V}^{(t)}$, and $G_S = \tilde{U}_S^\top \tilde{U}_S$. For every $v \notin S$,*

$$\begin{aligned} \mu_t(v) &= \phi_v^\top w_t, & w_t &= \big(\sigma_b^2 D^{-1} + G_S\big)^{-1} \tilde{U}_S^\top \mathbf{y}_S, \\ s_t^2(v) &= \sigma_b^2 + \phi_v^\top \Xi_t \phi_v, & \Xi_t &= \sigma_b^2 \big(\sigma_b^2 D^{-1} + G_S\big)^{-1}, \end{aligned} \tag{13}$$

the posterior of the Bayesian linear model $y(v) = \langle \phi_v, \theta \rangle + \varepsilon_v$, $\theta \sim \mathcal{N}(0, D)$, $\varepsilon_v \sim \mathcal{N}(0, \sigma_b^2)$. Exact inference over all n nodes costs $\mathcal{O}(|S|d^2 + nd^2 + d^3)$ per round (Appendix C(D)).

Feature map and local covariance. Use the factors of Lemma B.2, and let $V_+, \tilde{\Gamma}_+$ retain the positive eigenvalues of the normalized adjacency. With fitted global quantities fixed,

$$\phi_v = \frac{F \pi_v}{\sqrt{\pi_v^\top h + \zeta}}, \qquad F = \tilde{\Gamma}_+^{-1} V_+^\top R C, \qquad h = C \Pi^\top \mathbf{1}. \tag{14}$$

Indeed $\tilde{U}_+ = M C R^\top V_+ \tilde{\Gamma}_+^{-1}$. If R is invertible, $F = V_+^\top R^{-\top}$, recovering the full-rank form. The first expression also applies when the memberships are rank deficient. Its local Jacobian is

$$J_v = F s_v \big(I - \tfrac{1}{2} s_v^2 \pi_v h^\top\big), \qquad s_v = (\pi_v^\top h + \zeta)^{-1/2}. \tag{15}$$

For $G_v = [g_u^\top]_{u \in R_v}$, with R_v the measured landmarks of v, Gaussian conditioning gives

$$\Sigma_v = \Sigma_0 - \Sigma_0 G_v^\top (G_v \Sigma_0 G_v^\top)^+ G_v \Sigma_0, \qquad \Sigma_v^{\phi} = J_v \Sigma_v J_v^\top. \tag{16}$$

Here Σ_0 is the empirical covariance of the landmark memberships plus a positive ridge. These covariances describe unmeasured directions conditional on the fit. In the exact, consistent measurement model, the conditional mean can be projected onto $\{\pi \in \Delta_d : G_v \pi = A_{R_v, v}\}$ when this set is nonempty. For general observations the constrained residual fit (3) supplies the membership estimate.

Exact spectral solve.

Lemma B.2 (Exact spectral solve)**.** *Let $\tilde{A} = \Pi C \Pi^\top \in \mathcal{C}_d$ with $d \le n$ and $\zeta > 0$. Let $M = (\tilde{D} + \zeta I)^{-1/2}\Pi$ with thin QR factorization $M = QR$, and let $RCR^\top = V\tilde{\Gamma}V^\top$ be a $d \times d$ eigendecomposition. Then $\tilde{U} = QV$ collects orthonormal eigenvectors of $\tilde{L}_\zeta$ with eigenvalues $\tilde{\lambda}_i = 1 - \tilde{\gamma}_i$, and all remaining eigenvalues of $\tilde{L}_\zeta$ equal 1. The total cost is $\mathcal{O}(nd^2 + d^3)$.*

Proof. $\tilde{L}_\zeta = I - (\tilde{D} + \zeta I)^{-1/2}\Pi C \Pi^\top (\tilde{D} + \zeta I)^{-1/2} = I - MCM^\top = I - QRCR^\top Q^\top = I - QV\tilde{\Gamma}V^\top Q^\top$. The columns of $\tilde{U} = QV$ are orthonormal, so $\tilde{L}_\zeta \tilde{U} = \tilde{U}(I - \tilde{\Gamma})$, which gives the d eigenpairs. For x orthogonal to $\mathrm{col}(Q)$ we have $M^\top x = R^\top Q^\top x = 0$, so $\tilde{L}_\zeta x = x$ and every remaining eigenvalue equals 1. The thin QR costs $\mathcal{O}(nd^2)$, the $d \times d$ eigendecomposition $\mathcal{O}(d^3)$, and the product QV another $\mathcal{O}(nd^2)$. □

Predictive variance under feature uncertainty. Equation (7) follows from a moment calculation conditional on the current data. Under the working independence of ϕ_v and θ,

$$\mathrm{Var}(\phi_v^\top \theta) = \mathbb{E}[\phi_v^\top \Xi_t \phi_v] + \mathrm{Var}(\phi_v^\top w_t).$$

The first term is $\hat{\phi}_v^\top \Xi_t \hat{\phi}_v + \mathrm{tr}(\Sigma_v^\phi \Xi_t)$ and the second is $w_t^\top \Sigma_v^\phi w_t$. Adding the variance σ_b^2 of an independent residual gives the three terms in the main text. Conditional independence and the local propagation $\Sigma_v^\phi = J_v \Sigma_v J_v^\top$ are working approximations that hold the fitted global quantities fixed. Node selection uses the plug-in score of Eq. (6), and this variance prices structural queries. For an evaluated node whose feature is still uncertain, we instead use the moment-based errors-in-variables adjustment

$$\sigma_b^2(v) = \sigma_b^2 + w_t^\top \Sigma_v^\phi w_t + \mathrm{tr}\big(\Sigma_v^\phi \Xi_t\big) \tag{17}$$

in place of σ_b^2 when estimating the shared coefficients. This is an approximate treatment of feature error, whose effect on the coefficient estimate is described in Appendix C(I).

Heteroscedastic form of Lemma B.1. With node-wise variances $N_S = \mathrm{diag}(\sigma_b^2(v))_{v \in S}$ from Eq. (17), the GP state is $\Xi_t = (D^{-1} + \tilde{U}_S^\top N_S^{-1} \tilde{U}_S)^{-1}$ and $w_t = \Xi_t \tilde{U}_S^\top N_S^{-1} \mathbf{y}_S$, which reduces to Eq. (13) when $N_S = \sigma_b^2 I$.

Rank-one update of the design covariance. One more measurement of v along a direction g updates Σ_v by

$$\Sigma_v' = \Sigma_v - \frac{\Sigma_v g g^\top \Sigma_v}{g^\top \Sigma_v g}, \tag{18}$$

which depends on g only through its component outside the row space of G_v and is void along an already measured direction.

Spectral containment.

Lemma B.3 (Spectral containment)**.** *If $\tilde{A} \ge 0$ entrywise and $\zeta > 0$, then* $\mathrm{spec}(\tilde{L}_\zeta) \subset [\zeta/(\widetilde{\deg}_{\max} + \zeta),\ 2 - \zeta/(\widetilde{\deg}_{\max} + \zeta)]$. *In particular $\tilde{L}_\zeta \succ 0$. For the resolvent $r(\lambda) = (\beta + \lambda)^{-1}$ with $\beta > 0$, $r(\tilde{L}_\zeta) \succeq (\beta + 2)^{-1} I$. The structured-plus-bulk kernel of Eq.* (4) *satisfies $\tilde{K} \succeq \sigma_b^2 I$.*

The lower end is the quadratic-form argument of Appendix C(A) applied to $\tilde{A}$, whose diagonal is included in $\widetilde{\deg}$. For the upper end, with $\breve{x} = (\tilde{D} + \zeta I)^{-1/2} x$, $x^\top (2I - \tilde{L}_\zeta) x = \|x\|_2^2 + \breve{x}^\top \tilde{A} \breve{x} = \frac{1}{2}\sum_{i,j} \tilde{A}_{ij}(\breve{x}_i + \breve{x}_j)^2 + \zeta \|\breve{x}\|_2^2 \ge \zeta(\widetilde{\deg}_{\max} + \zeta)^{-1} \|x\|_2^2$, since $\|x\|_2^2 = \breve{x}^\top (\tilde{D} + \zeta I)\breve{x}$. The resolvent bound follows from $\tilde{L}_\zeta \preceq 2I$, and $\tilde{K} \succeq \sigma_b^2 I$ because the structured term is positive semidefinite.

One-scale kernels and the linear model. For a response positive on the spectrum, the exact one-scale kernel has the identity

$$K_{\mathrm{one}} = r(\tilde{L}_\zeta) = r(1) I + \tilde{U} D_r \tilde{U}^\top, \qquad D_r = \mathrm{diag}(r(\tilde{\lambda}_i) - r(1)). \tag{19}$$

It uses the same surrogate and spectral solve, and supports standard GP conditioning. When $D_r \succeq 0$, removing zero corrections gives Lemma B.1 with bulk variance $r(1)$. If D_r has negative entries, the kernel remains positive definite but these entries cannot be variances of independent coefficients.

Algorithm 2 Surrogate Reconstruction and Kernel Construction

Input: observed entries Ω with landmarks $\mathcal{R}$, resolution d, regularizers (ζ, β) and scales (σ_b^2, σ_s^2), previous fit (Π, C) if available, refit flag

Output: kernel $\tilde{k}(\cdot,\cdot)$ in factored form and feature covariances $\{\Sigma_v^\phi\}$

1: **Structured initialization** (first call, or refit flag set): $(\Pi_{\mathcal{R}}, C) \leftarrow$ rank-d factorization of the observed landmark entries by truncated SVD or successive projection, and each remaining $\pi_v \leftarrow$ projected least squares on its observed entries ▷ $\mathcal{O}(|\Omega|d)$

2: **repeat**

3: one alternating least-squares sweep on Problem (3): each membership by projected least squares over Δ_d given the others and C, then C by box-constrained least squares given Π (Appendix E) ▷ $\mathcal{O}(|\Omega|d + nd^2)$ per sweep under rows, $\mathcal{O}(|\Omega|d^2 + nd^2)$ under pairs

4: **until** the sweep budget is spent or the residual decrease falls below a tolerance

5: If the refit flag is set, run the sweeps from the previous fit as well and keep the solution with the smaller residual

6: $\widetilde{\deg} \leftarrow \Pi(C(\Pi^\top \mathbf{1}))$ ▷ $\mathcal{O}(nd^2 + |\Omega|d)$

7: $M \leftarrow (\tilde{D} + \zeta I)^{-1/2}\Pi$, thin QR $M = QR$, eigendecomposition $RCR^\top = V\tilde{\Gamma}V^\top$, $\tilde{U} \leftarrow QV$, $\tilde{\Lambda} \leftarrow I - \tilde{\Gamma}$; keep the d_+ directions with $\tilde{\lambda}_i < 1$ ▷ Lemma B.2, $\mathcal{O}(nd^2 + d^3)$

8: $\Sigma_v \leftarrow$ Eq. (16) with $g_u = C\pi_u$, and $\Sigma_v^\phi \leftarrow J_v \Sigma_v J_v^\top$ ▷ membership covariance shared under rows (Proposition B.1)

9: **return** $\tilde{k}$ via Eq. (4) as $(\sigma_b^2, \tilde{U}_+, \sigma_s^2 D_+)$, and $\{\Sigma_v^\phi\}$ ▷ $\mathcal{O}(nd)$ storage

Its positive decomposition instead uses the correlated residual $r(1)(I - \tilde{U}\tilde{U}^\top)$ and structured term $\tilde{U}\,\mathrm{diag}(r(\tilde{\lambda}_i))\tilde{U}^\top$.

Clipping negative corrections gives $K_{\mathrm{clip}} = r(1)I + \tilde{U}_+ D_+ \tilde{U}_+^\top$, the tied-scale case of Eq. (4). Our chosen kernel additionally separates the scales. This allows the strength of the structured component to vary independently of the bulk, whose relative average correction in the one-scale kernel is $\mathcal{O}(d/n)$ when spectral contrasts are bounded (Appendix C(E)). Clipping and scale separation are modeling choices, while Eq. (19) is an identity. Appendix H.1 compares the chosen kernel with K_{one}.

Bias under smooth latent interaction. The following lemma gives an interpretable sufficient condition for a small bias $\varepsilon_d(A)$, namely smooth latent interaction, with a fully deterministic proof in Appendix C(C). It is the noise-free weighted analogue of the eigenvalue-tail estimates behind spectral graphon estimation (Xu, 2018; Chatterjee, 2015).

Lemma B.4 (Bias under smooth latent interaction)**.** *Suppose $A_{ij} = f(x_i, x_j)$ for all $i \neq j$, with latent positions $x_i \in [0,1]^\ell$ and $f : [0,1]^\ell \times [0,1]^\ell \to [0,1]$ symmetric and α-Hölder with constant L in each coordinate, $\alpha \in (0, 2]$ (for $\alpha \in (1, 2]$: differentiable with $(\alpha-1)$-Hölder partial derivatives). Then for every $d \geq 2^\ell$ of the form $d = (k+1)^\ell$,*

$$\varepsilon_d(A) \;\leq\; 2^{\alpha+1}\ell\, L\, n\, d^{-\alpha/\ell}.$$

The witness is explicit and lies in the class. The row π_i collects the tensor hat-function (degree-one B-spline) basis evaluated at x_i, which is nonnegative and sums to one, and $C_{ab} = f(t_a, t_b) \in [0,1]$ are function values at the knots. Hard-membership blocks, the step-graphon subclass, already achieve the rate for $\alpha \leq 1$, and the soft-membership construction extends it to $\alpha \leq 2$. For general d, monotonicity of ε_d in d gives the same rate up to the constant factor 2^α by rounding down to the nearest ℓ-th power.

The hard-membership subclass of $\mathcal{C}_d$ consists of d-block step matrices, which connects the class to stochastic block models and step graphons (Airoldi et al., 2008; Lovász, 2012) as descriptions of the approximant rather than as assumptions on A. An indefinite core C represents disassortative (heterophilous) structure.

Row rule.

Proposition B.1 (Shared membership uncertainty under row queries)**.** *With common reference covariance Σ_0 and fitted directions held fixed, let $G_{\mathcal{R}} = [g_u^\top]_{u \in \mathcal{R}}$ for the revealed rows. Every*

Table 2: Per-round complexity with $m = |\mathcal{V}^{(t)}|$ queried nodes and $r = |\mathcal{R}^{(t)}|$ landmarks. By Lemma B.1 the GP state is $(G_S, \tilde{U}_S^\top \mathbf{y}_S)$, updated in $\mathcal{O}(d^2)$ per new evaluation while features are fixed. After one $d \times d$ solve, posterior means over all n nodes cost $\mathcal{O}(nd)$ and exact variances $\mathcal{O}(nd^2)$, independent of m.

storage (factors, spectrum, GP state)	$\mathcal{O}(nd + \|\Omega\| + d^2)$
surrogate update (per ALS sweep, rows / pairs)	$\mathcal{O}(\|\Omega\|d + nd^2)$ / $\mathcal{O}(\|\Omega\|d^2 + nd^2)$
spectral refresh (exact)	$\mathcal{O}(nd^2 + d^3)$
GP state update (Lemma B.1)	$\mathcal{O}(md^2 + d^3)$
acquisition: means (given w_t)	$\mathcal{O}(nd)$
acquisition: exact variances	$\mathcal{O}(nd^2)$
local covariance propagation under rows (Eq. (16))	$\mathcal{O}(rd^2 + nd^2)$
row acquisition (Algorithm 3)	$\mathcal{O}(nrd)$
pair acquisition, per query (Algorithm 4)	$\mathcal{O}(rd^3 + n)$

Algorithm 3 Row Acquisition (one round)

Input: revealed rows $\mathcal{R}$, evaluated nodes $\mathcal{V}^{(t)}$, number of rows q_t available in this round, pin share ρ_{pin}, fit (Π, C), resolution d
Output: updated $\mathcal{R}$ and Ω

$q \leftarrow 0$
while $q < q_t$, the pin share is not exhausted, and some evaluated node has an unrevealed row **do**
  reveal the row of the earliest such node and set $q \leftarrow q + 1$ ▷ accurate features at evaluated nodes
end while
while $q < q_t$ **do**
  draw w uniformly at random from $\mathcal{V} \setminus \mathcal{R}$ ▷ uniform rows, Proposition B.1
  reveal the row of w, set $\mathcal{R} \leftarrow \mathcal{R} \cup \{w\}$ and $q \leftarrow q + 1$, and update the fit by warm-started sweeps of Algorithm 2
end while
return $\mathcal{R}$ and the corresponding Ω

unrevealed node has the same covariance $\Sigma_{\mathcal{R}}$ given by Eq. (16)*. If* $\operatorname{rank} G_{\mathcal{R}} = d$*, this covariance is zero.*

The proof is in Appendix C(F). Feature and score uncertainties also depend on J_v and c_v. Algorithm 3 first spends its pinning allowance on evaluated nodes, then draws uniformly from the remaining rows. The uniform component provides broad coverage, while the pinning component improves the features used to estimate the shared coefficient.

C Derivations for the Methodology

(A) Dirichlet-energy identity. Let $A \geq 0$ be symmetric with $A_{ii} = 0$, $W = (D_A + \zeta I)^{-1/2}$, $\breve{y} = Wy$. Then $\|y\|^2_{\mathcal{H}(K(A))} = y^\top K(A)^{-1} y = y^\top(\beta I + L_\zeta(A))y = \beta\|y\|_2^2 + y^\top L_\zeta(A) y$. Since $y^\top L_\zeta(A) y = \|y\|_2^2 - \breve{y}^\top A \breve{y}$ and $\|y\|_2^2 = \breve{y}^\top(D_A + \zeta I)\breve{y} = \breve{y}^\top D_A \breve{y} + \zeta\|\breve{y}\|_2^2$, we get $y^\top L_\zeta(A) y = \breve{y}^\top(D_A - A)\breve{y} + \zeta\|\breve{y}\|_2^2$. With $\deg_i = \sum_j A_{ij}$, $\breve{y}^\top D_A \breve{y} = \sum_{ij} A_{ij}\breve{y}_i^2$, so by symmetry $\breve{y}^\top(D_A - A)\breve{y} = \frac{1}{2}\sum_{ij} A_{ij}(\breve{y}_i - \breve{y}_j)^2 = \frac{1}{2}\sum_{i \neq j} A_{ij}(\breve{y}_i - \breve{y}_j)^2$ (diagonal terms vanish, so the identity also holds when A has a nonnegative diagonal that D_A includes, as for the surrogate $\tilde{A}$). The same computation applied to $x^\top L_\zeta(A) x$ shows $L_\zeta(A) \succeq \zeta/(\deg_{\max}(A) + \zeta)\, I \succ 0$. Applied to the surrogate, the identity shows that a signal x with $\breve{x}^\top \tilde{A}\breve{x} = 0$, uncorrelated along the edges of the surrogate, has Rayleigh quotient $x^\top \tilde{L}_\zeta x / \|x\|^2 = 1$, which is the reading of the eigenvalue 1 used in Section 4.2.

(B) Attainment of the bias minimum (11)**.** $(\Delta_d)^n \times \{C = C^\top \in [0,1]^{d \times d}\}$ is compact, $(\Pi, C) \mapsto \Pi C \Pi^\top$ is continuous, so $\mathcal{C}_d$ is compact. $X \mapsto \|(A - X)_{\text{off}}\|_F$ is continuous, so the minimum is attained by the Weierstrass theorem.

(C) Proof of Lemma B.4. *Univariate pieces.* Let $g : [0,1] \to [0,1]$ and let I_h denote piecewise-linear (hat-function) interpolation at knots $t_a = a/k$, $h = 1/k$. (i) If $|g(s) - g(t)| \le L|s-t|^\alpha$, $\alpha \in (0,1]$: for $x \in [t_a, t_{a+1}]$, $I_h g(x) = \lambda g(t_a) + (1-\lambda) g(t_{a+1})$ is a convex combination, so $|g(x) - I_h g(x)| \le \lambda |g(x) - g(t_a)| + (1-\lambda)|g(x) - g(t_{a+1})| \le L h^\alpha$. (ii) If g is differentiable with $|g'(s) - g'(t)| \le L|s-t|^{\alpha-1}$, $\alpha \in (1,2]$: on $[a,b]$ with $h = b-a$, the mean value theorem gives $g(b) - g(a) = h g'(\xi)$ for some $\xi \in (a,b)$, hence $g(x) - I_h g(x) = \int_a^x \big(g'(t) - g'(\xi)\big)\, dt$ and $|g(x) - I_h g(x)| \le (x-a)\, L\, h^{\alpha-1} \le L h^\alpha$.

Tensorization. View f as a function of 2ℓ scalar coordinates and let $I^\otimes = I_1 \circ \cdots \circ I_{2\ell}$ interpolate each coordinate on the same knot grid. Each coordinate operator is positive with partition of unity, hence sup-norm nonexpansive, so telescoping $f - I^\otimes f = \sum_{j=1}^{2\ell} (I_1 \cdots I_{j-1})(f - I_j f)$ yields $\|f - I^\otimes f\|_\infty \le \sum_j \|f - I_j f\|_\infty \le 2\ell\, L h^\alpha$.

Class membership. $I^\otimes f(x,y) = \sum_{a,b} f(t_a, t_b)\, \varphi_a(x) \varphi_b(y)$, where $\{\varphi_a\}$ is the tensor hat basis over $[0,1]^\ell$ with $(k+1)^\ell$ elements: $\varphi_a \ge 0$, $\sum_a \varphi_a \equiv 1$. Setting $\pi_i = (\varphi_a(x_i))_a \in \Delta_d$ and $C_{ab} = f(t_a, t_b) \in [0,1]$ (symmetric since f is) exhibits $X = \Pi C \Pi^\top \in \mathcal{C}_d$ with $d = (k+1)^\ell$ and $X_{ij} = I^\otimes f(x_i, x_j)$.

Constants. For $d \ge 2^\ell$, $k = d^{1/\ell} - 1 \ge d^{1/\ell}/2$, so $h \le 2 d^{-1/\ell}$ and the entrywise error is at most $2\ell L (2d^{-1/\ell})^\alpha \le 2^{\alpha+1} \ell L\, d^{-\alpha/\ell}$. Summing over at most n^2 off-diagonal entries gives $\varepsilon_d(A) \le 2^{\alpha+1} \ell L n\, d^{-\alpha/\ell}$. For $\alpha \le 1$, replacing hats by indicators of $d = k^\ell$ congruent cells (hard memberships, $C_{ab} = f$ at cell centers) gives $\varepsilon_d \le 2^{1-\alpha} \ell L n\, d^{-\alpha/\ell}$, proving the step-graphon claim. □

(D) Proof of Lemma B.1. Let $\tilde{K} = \sigma_b^2 I + \tilde{U} D \tilde{U}^\top$ and $v \notin S$, so that $\tilde{K}_{vS} = \phi_v^\top D \tilde{U}_S^\top$, $\tilde{K}_{vv} = \sigma_b^2 + \phi_v^\top D \phi_v$, and $\tilde{K}_{SS} = \sigma_b^2 I_m + \tilde{U}_S D \tilde{U}_S^\top$. The push-through identity $B(\sigma_b^2 I_m + AB)^{-1} = (\sigma_b^2 I_d + BA)^{-1} B$ with $B = D\tilde{U}_S^\top$ and $A = \tilde{U}_S$ gives $\tilde{K}_{vS} \tilde{K}_{SS}^{-1} = \phi_v^\top (\sigma_b^2 I_d + D G_S)^{-1} D \tilde{U}_S^\top = \phi_v^\top (\sigma_b^2 D^{-1} + G_S)^{-1} \tilde{U}_S^\top$, hence $\mu_t(v) = \phi_v^\top w_t$. By the Woodbury identity, $D - D\tilde{U}_S^\top (\sigma_b^2 I_m + \tilde{U}_S D \tilde{U}_S^\top)^{-1} \tilde{U}_S D = (D^{-1} + G_S/\sigma_b^2)^{-1} = \sigma_b^2 (\sigma_b^2 D^{-1} + G_S)^{-1} = \Xi_t$, hence $s_t^2(v) = \sigma_b^2 + \phi_v^\top \Xi_t \phi_v$. For $D \succ 0$ the same expressions are the posterior of θ in $y = \langle \phi, \theta \rangle + \varepsilon$ with prior $\mathcal{N}(0, D)$ and per-node bulk variance σ_b^2. For fixed features, G_S and $\tilde{U}_S^\top \mathbf{y}_S$ are updated in $\mathcal{O}(d^2)$ per new evaluation. One $d \times d$ solve then gives w_t and Ξ_t, and each node costs $\mathcal{O}(d)$ for the mean and $\mathcal{O}(d^2)$ for the variance. □

(E) Contrast bound.

Lemma C.1 (Contrast bound)**.** *For $\tilde{K} = r(\tilde{L}_\zeta)$ with $r > 0$ nonincreasing, $\frac{1}{n}\sum_v (\tilde{K}_{vv} - r(1))/r(1) \le \frac{d}{n}\,(r(\tilde{\lambda}_{\min}) - r(1))/r(1)$. For the resolvent the right-hand side is at most $\frac{d}{n}(1 - \tilde{\lambda}_{\min})/\tilde{\lambda}_{\min}$ for every $\beta > 0$, and for the heat filter $e^{-\tau\lambda}$ it equals $\frac{d}{n}(e^{\tau(1-\tilde{\lambda}_{\min})} - 1)$, which is unbounded in τ.*

Proof: $\tilde{K}_{vv} = r(1) + \sum_{i \le d} (r(\tilde{\lambda}_i) - r(1)) \tilde{U}_{vi}^2$ and $\sum_v \tilde{U}_{vi}^2 = 1$. For the resolvent, $(r(\tilde{\lambda}) - r(1))/r(1) = (1 - \tilde{\lambda})/(\beta + \tilde{\lambda})$. □

Example: four disconnected cliques of 250 nodes, $\zeta = 1$, true graph. Within a clique $\tilde{L}_\zeta$ has the eigenvalue $1 - 249/250 = 0.004$ on the clique indicator and $1 + 1/250 = 1.004$ on its complement. With $r(\lambda) = (\beta + \lambda)^{-1}$ at $\beta = 0.2$, $r(0.004) = 4.90$ and $r(1.004) = 0.831$, so the within-clique kernel correlation is $(4.90 - 0.831)/(4.90 + 249 \cdot 0.831) = 0.019$. As $\beta \to 0$ it tends to $(250 - 0.996)/(250 + 249 \cdot 0.996) = 0.50$, and with the heat filter $e^{-10\lambda}$ it is 0.989. The same d/n dilution holds for the true graph and for the surrogate, since it comes from having few structured modes and many bulk modes and not from the low-rank class.

(F) Shared covariance and the weighted query gain. For Proposition B.1, Eq. (16) depends on a node only through G_v, which is $G_\mathcal{R}$ for every unrevealed node. If $G_\mathcal{R}$ has full column rank, exact conditioning leaves zero covariance. A further direction g induces Eq. (18), which depends only on its component outside the row space of $G_\mathcal{R}$. These statements concern membership covariance conditional on the fit, while its propagation to scores also uses each node's sensitivity. □

The gain $g \mapsto (c^\top \Sigma g)^2/(g^\top \Sigma g)$ of Appendix D is not submodular as a set function of the probes. With $d = 2$, $\Sigma = I$, $c = e_1$, $g_1 = e_2$, $g_2 = e_1 + e_2$, the marginal gain of g_1 is 0 alone and $1/2$ after g_2, because after g_2 one has $\Sigma' = I - g_2 g_2^\top/2$, $c^\top \Sigma' g_1 = -1/2$, and $g_1^\top \Sigma' g_1 = 1/2$.

(G) Proof of Proposition A.1. Write $\mathcal{S}_d$ for the set of symmetric $n \times n$ matrices of rank exactly d, a smooth manifold of dimension $nd - \binom{d}{2}$ whose tangent space at $X = FGF^\top$, with $F \in \mathbb{R}^{n\times d}$ of rank d and G symmetric nonsingular, is $\{\dot F G F^\top + F\dot G F^\top + FG\dot F^\top\}$.

(i). At an interior representative with $\operatorname{rank}\Pi = d$ the differential of $(\Pi, C) \mapsto \Pi C \Pi^\top$ is $(\dot\Pi, \dot C) \mapsto \dot\Pi C\Pi^\top + \Pi\dot C\Pi^\top + \Pi C\dot\Pi^\top$, which is onto the tangent space of $\mathcal{S}_d$ at M. By the submersion theorem, the image of a neighborhood of the representative, which stays inside the class because the representative is interior, contains a neighborhood of M in $\mathcal{S}_d$. A set Ω of entries defines the continuous map $X \mapsto X_\Omega$ from that neighborhood, homeomorphic to an open subset of $\mathbb{R}^{nd-\binom{d}{2}}$, into $\mathbb{R}^{|\Omega|}$. If $|\Omega| < nd - \binom{d}{2}$ this map is not injective, by invariance of domain, so two members of the class share the observations.

(ii), sufficiency. Let $B = \Pi_{\mathcal{R}} C \in \mathbb{R}^{r\times d}$, which has rank d, and let X be symmetric of rank at most d with $X_{uv} = M_{uv}$ for all $u \in \mathcal{R}$ and $v \neq u$. First, $X_{\mathcal{R},\mathcal{V}\setminus\mathcal{R}} = M_{\mathcal{R},\mathcal{V}\setminus\mathcal{R}} = B\Pi_{\mathcal{V}\setminus\mathcal{R}}^\top$ has rank d because $\operatorname{rank}\Pi_{\mathcal{V}\setminus\mathcal{R}} = d$, and its column space is $\operatorname{col}(B)$. Since $X_{\mathcal{R},\cdot}$ has rank at most d and contains these columns, $\operatorname{col}(X_{\mathcal{R},\cdot}) = \operatorname{col}(B)$. Second, for $u \in \mathcal{R}$ the column $z = X_{\mathcal{R},u} - M_{\mathcal{R},u}$ lies in $\operatorname{col}(B)$ and vanishes off coordinate u. Writing $z = B\alpha$, the rows of B other than u give $\Pi_{\mathcal{R}\setminus\{u\}} C\alpha = 0$, and $\operatorname{rank}\Pi_{\mathcal{R}\setminus\{u\}} = d$ with C nonsingular forces $\alpha = 0$, so $X_{uu} = M_{uu}$. Hence $X_{\mathcal{R},\cdot} = M_{\mathcal{R},\cdot}$ and, by symmetry, $X_{\cdot,\mathcal{R}} = M_{\cdot,\mathcal{R}}$. Third, $X_{\mathcal{R}\mathcal{R}} = M_{\mathcal{R}\mathcal{R}} = \Pi_{\mathcal{R}} C \Pi_{\mathcal{R}}^\top$ has rank d, so X has rank exactly d and its row space equals the row space of $X_{\mathcal{R},\cdot}$. Thus $X = W X_{\mathcal{R},\cdot}$ for some $W \in \mathbb{R}^{n\times r}$, hence $X_{\cdot,\mathcal{R}} = W X_{\mathcal{R}\mathcal{R}}$, and since $X_{\mathcal{R}\mathcal{R}} X_{\mathcal{R}\mathcal{R}}^+$ is the orthogonal projector onto $\operatorname{col}(X_{\mathcal{R}\mathcal{R}}) = \operatorname{col}(X_{\mathcal{R},\cdot})$, we get $X_{\cdot,\mathcal{R}} X_{\mathcal{R}\mathcal{R}}^+ X_{\mathcal{R},\cdot} = W X_{\mathcal{R},\cdot} = X$. The same identity holds for M, and the three factors agree, so $X = M$. A zero-residual point (Π', C') of Problem (3) yields such an $X = \Pi' C' \Pi'^\top$. If the memberships are in general position, every d of them are independent, which gives both rank conditions when $r \geq d + 1$ and $n - r \geq d$. For pure memberships with $\theta_i > 0$, $\operatorname{rank}\Pi_{\mathcal{I}}$ equals the number of communities represented in $\mathcal{I}$, which gives the stated characterization.

(ii), necessity. Let $r \leq d$ with $\operatorname{rank}\Pi_{\mathcal{R}} = r$ and an interior representative, and let $B_0 = \Pi_{\mathcal{R}} C$, of full row rank r. The map $E \mapsto (\Pi_{\mathcal{R}} + E)C(\Pi_{\mathcal{R}} + E)^\top$ from $\mathbb{R}^{r\times d}$ to the symmetric $r \times r$ matrices has differential $E \mapsto EB_0^\top + B_0 E^\top$ at $E = 0$, which is onto: for a symmetric target Z, the choice $E = \frac{1}{2} Z (B_0^+)^\top$ gives $EB_0^\top = \frac{1}{2} Z (B_0 B_0^+)^\top = \frac{1}{2} Z$ because $B_0 B_0^+ = I_r$. By the implicit function theorem, for every sufficiently small diagonal matrix Δ there is a small E with $(\Pi_{\mathcal{R}} + E)C(\Pi_{\mathcal{R}} + E)^\top = M_{\mathcal{R}\mathcal{R}} + \Delta$. Set $\Pi'_{\mathcal{R}} = \Pi_{\mathcal{R}} + E$ and, for $v \notin \mathcal{R}$, let π'_v be the solution of $\Pi'_{\mathcal{R}} C \pi = M_{\mathcal{R},v}$ closest to π_v, which exists because $\Pi'_{\mathcal{R}} C$ has full row rank and lies within $\mathcal{O}(\|E\|)$ of π_v. For small $\Delta \neq 0$ all memberships stay in Δ_d, and $M' = \Pi' C \Pi'^\top$ agrees with M on every revealed entry while $M'_{uu} \neq M_{uu}$ for some $u \in \mathcal{R}$.

(iii), necessity. By the argument of (i) applied to $\Pi_{\mathcal{R}}$, the block $M_{\mathcal{R}\mathcal{R}} = \Pi_{\mathcal{R}} C \Pi_{\mathcal{R}}^\top$ ranges over a set containing an open subset of the symmetric $r \times r$ matrices of rank $\min(r, d)$, a manifold of dimension $rd - \binom{d}{2}$ if $r \geq d$ and $\binom{r+1}{2}$ if $r < d$. Both exceed $\binom{r}{2}$ when $r < r_0(d)$, since $\binom{r}{2} \geq rd - \binom{d}{2}$ is equivalent to $r^2 - (2d+1)r + d^2 - d \geq 0$, whose larger root is $(2d+1+\sqrt{8d+1})/2$, and invariance of domain applies as in (i).

(iii), sufficiency. The off-diagonal entries of $M_{\mathcal{R}\mathcal{R}}$ form the row pattern of (ii) inside the block, with revealed rows $\mathcal{R}_1$ and unrevealed nodes $\mathcal{R} \setminus \mathcal{R}_1 \supseteq \mathcal{R}_2$, so (ii) gives $X_{\mathcal{R}\mathcal{R}} = M_{\mathcal{R}\mathcal{R}}$ for every symmetric X of rank at most d with the same revealed entries. Let $v \notin \mathcal{R}$ be measured against R_v. Since X has rank at most d and $X_{\mathcal{R},\cdot}$ contains the rank-d block $X_{\mathcal{R}\mathcal{R}}$, X has rank d and its row space is that of $X_{\mathcal{R},\cdot}$, so $X_{v,\mathcal{R}}$ lies in the row space of $X_{\mathcal{R}\mathcal{R}} = M_{\mathcal{R}\mathcal{R}}$, a d-dimensional subspace of $\mathbb{R}^r$. The coordinate projection of this subspace onto R_v is injective if and only if $(M_{\mathcal{R}\mathcal{R}})_{\cdot,R_v} = \Pi_{\mathcal{R}} C \Pi_{R_v}^\top$ has rank d, that is, $\operatorname{rank}\Pi_{R_v} = d$, and then the revealed coordinates $X_{v,R_v} = M_{v,R_v}$ determine $X_{v,\mathcal{R}} = M_{v,\mathcal{R}}$. Once this holds for every $v \notin \mathcal{R}$, $X_{\cdot,\mathcal{R}} = M_{\cdot,\mathcal{R}}$ by symmetry, and the skeleton identity of (ii) gives $X = M$. □

(H) Proofs of Theorem 1 and Corollary A.1, and a deterministic instance of Assumption 1. *Theorem 1.* We prove the stronger subset form. For any nonempty $I \subseteq \{0, \dots, T-1\}$ with $m = |I|$, replace T in the denominator of the theorem by m and its error average by $m^{-1} \sum_{t \in I} (\Delta_t(v^{(t+1)}) + \Delta_t(v^*))$, retaining γ_T. If $v^* \in \mathcal{V}^{(t)}$ for some $t < T$ then $r_T = 0$. Otherwise Lemma C.2 applies in every round $t \in I$, and the simple regret is at most the average over I of its right-hand side. It remains to bound $\frac{1}{m} \sum_{t \in I} s_t^\circ(v^{(t+1)})$. By Lemma B.1, $s_t^\circ(v)^2 - \sigma_b^2 = \phi_v^\top \Xi_t \phi_v$, where Ξ_t is the posterior covariance of a d_+-dimensional Bayesian linear model with prior covariance D and per-node bulk variance σ_b^2 on the evaluated set. Since $\Xi_t \preceq D$, one has $\phi_v^\top \Xi_t \phi_v \le v_{\max}$, hence $\phi^\top \Xi_t \phi \le \max(\sigma_b^2, v_{\max}) \min(1, \phi^\top \Xi_t \phi / \sigma_b^2)$, and the elliptical potential lemma (Abbasi-yadkori et al., 2011) applied to the actual evaluation sequence gives $\sum_{t<T} \min\bigl(1, \phi_{v^{(t+1)}}^\top \Xi_t \phi_{v^{(t+1)}} / \sigma_b^2\bigr) \le 2 \log\det\bigl(I + D^{1/2} G_S D^{1/2} / \sigma_b^2\bigr) \le 2\gamma_T$, where the last step uses $\log\det(I + X) \le d_+ \log(1 + \operatorname{tr} X / d_+)$ and $\operatorname{tr}(D^{1/2} G_S D^{1/2}) = \sum_{s \in S} \phi_s^\top D \phi_s \le (T + N_0)\, v_{\max}$. The sum over I is a partial sum of nonnegative terms and obeys the same bound, so $\sum_{t \in I} \phi_{v^{(t+1)}}^\top \Xi_t \phi_{v^{(t+1)}} \le c_1 \gamma_T$. By $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ and the Cauchy–Schwarz inequality, $\frac{1}{m} \sum_{t \in I} s_t^\circ(v^{(t+1)}) \le \sigma_b + \sqrt{c_1 \gamma_T / m}$. □

Corollary A.1. For $t \ge t_{\mathrm{id}}$ the surrogate is identified and the kernel parameters and conditioning agree with the reference, so the zero-residual point of Problem (3) is the reference surrogate, and the fitted features are the reference features up to an orthogonal transformation within each eigenspace of $\tilde{L}_\zeta$, which leaves the kernel, the posterior, and the scores unchanged. Hence $\Delta_t = 0$ for $t \ge t_{\mathrm{id}}$, and the subset form just proved with $I = \{t_{\mathrm{id}}, \dots, T-1\}$ gives the bound. Under Algorithm 1 the rows of the N_0 initial nodes are revealed at the start and Q_R rows before each decision, so $N_0 + (t+1)Q_R$ rows are revealed at the decision of round t. In general position any $d+1$ distinct rows satisfy Proposition A.1(ii) when $n \ge 2d+1$, so identification occurs no later than $(\lceil (d+1-N_0)_+ / Q_R \rceil - 1)_+$ decisions. Later observations preserve uniqueness. For interior representatives, d or fewer rows leave nonunique completions, giving the matching threshold in general. □

Deterministic instance of Assumption 1. Let $y = y_s + y_b$ as in Assumption A.1, and write the posterior mean as $\mu_t^\circ(v) = a_t(v)^\top \mathbf{y}_S$ with weights $a_t(v) = \tilde{U}_S (\sigma_b^2 D^{-1} + G_S)^{-1} \phi_v$, by Lemma B.1 and the push-through identity. On $\mathcal{H}(K_s)$ the functional $f \mapsto f(v) - a_t(v)^\top f_S$ is bounded, with squared norm $K_s(v,v) - 2a_t(v)^\top K_{s,Sv} + a_t(v)^\top K_{s,SS} a_t(v) = \phi_v^\top \Xi_t \phi_v - \sigma_b^2 \|a_t(v)\|_2^2 \le \phi_v^\top \Xi_t \phi_v$, the kernel ridge regression identity with regularization σ_b^2. Hence $|y_s(v) - a_t(v)^\top \mathbf{y}_{s,S}| \le B_s (\phi_v^\top \Xi_t \phi_v)^{1/2}$, while $|y_b(v) - a_t(v)^\top \mathbf{y}_{b,S}| \le \eta(1 + \|a_t(v)\|_1)$. Therefore

$$|y(v) - \mu_t^\circ(v)| \;\le\; B_s\, s_t^\circ(v) + \eta\bigl(1 + \|a_t(v)\|_1\bigr) \qquad \text{for every } v \text{ and } t,$$

deterministically, so Assumption 1 holds with $\kappa = B_s + \eta(1+\alpha)/\sigma_b$ whenever the posterior weights satisfy $\max_{t,v} \|a_t(v)\|_1 \le \alpha$, because $s_t^\circ \ge \sigma_b$.

(I) Errors in variables at evaluated nodes. With $w_t = (\sigma_b^2 D^{-1} + G_S)^{-1} \tilde{U}_S^\top \mathbf{y}_S$, a perturbation $\tilde{U}_S \mapsto \tilde{U}_S + E$ of the evaluated nodes' features changes the shared coefficient at first order by $\delta w_t = (\sigma_b^2 D^{-1} + G_S)^{-1} \bigl[E^\top \mathbf{y}_S - (E^\top \tilde{U}_S + \tilde{U}_S^\top E) w_t\bigr]$, and therefore every prediction by $\delta \mu_t(v) = \phi_v^\top \delta w_t$. Making the features of evaluated nodes accurate is thus a global correction, which is why Algorithm 3 reveals the rows of evaluated nodes first.

(J) Why the unsplit smoothness constant is of order $\sqrt{n}$. A single RKHS bound $\|y\|_{\mathcal{H}(K)} \le B$ for the full-rank kernel $K = (\beta I + L_\zeta)^{-1}$ contains $\beta \|y\|_2^2$ by the identity of (A), so $B^2 \ge \beta \|y\|_2^2$. For an objective standardized to unit variance over the nodes, $\|y\|_2^2 = n$ and hence $B \ge \sqrt{\beta n}$. Any regret bound proportional to B, and any confidence width $B s_t(v)$ with $s_t(v) \ge (\beta+2)^{-1/2}$, is then at least of order $\sqrt{\beta n / (\beta + 2)}$, which is 9.5 at $n = 1000$ and $\beta = 0.2$, while the range of such an objective is about 6. Assumption A.1 removes the $\sqrt{n}$ by bounding the bulk component in the supremum norm instead of through the RKHS norm.

(K) The scalar δ, the knowledge-gradient form, and Lemma D.1.

Proposition C.1 (Value of a pair query)**.** *In the working model with the global geometry and the coefficient state held fixed and the memberships independent across nodes, let $Z_w = U_t(w) + c_w^\top(\pi_w - \hat{\pi}_w)$ be the linearized scores of the unevaluated nodes and $\mathcal{L}(\mathcal{H}) = \mathbb{E}[\max_w Z_w \mid \mathcal{H}] - \max_w \mathbb{E}[Z_w \mid$*

$\mathcal{H}$] *the expected loss of selecting by the current scores given the information* $\mathcal{H}$*. A query of* v *along* g *with answer* a *reduces this loss in expectation by*

$$\mathrm{KG}_t(v,g) = \mathcal{L}(\mathcal{H}) - \mathbb{E}_a[\mathcal{L}(\mathcal{H},a) \mid \mathcal{H}] = \delta(v,g)\, f\Big(-\frac{\mathrm{gap}_t(v)}{\delta(v,g)}\Big), \qquad f(z) = z\Phi(z) + \varphi(z),$$

where $\mathrm{gap}_t(v) = |U_t(v) - \max_{w\neq v} U_t(w)|$ *and* Φ, φ *are the standard normal distribution and density. The value increases with* δ*, decreases with the gap, and vanishes when* $\delta = 0$*.*

Condition $\pi_v \sim \mathcal{N}(\hat{\pi}_v, \Sigma)$ on $a = \langle g, \pi_v\rangle + e$ with $e \sim \mathcal{N}(0,\sigma^2)$, where the exact observations of this paper correspond to $\sigma^2 \to 0$. The predictive law of a is $\mathcal{N}(\langle g, \hat{\pi}_v\rangle,\ g^\top \Sigma g + \sigma^2)$, the posterior mean is $\hat{\pi}_v + \Sigma g\,(a - \langle g, \hat{\pi}_v\rangle)/(g^\top \Sigma g + \sigma^2)$, and in the exact-observation limit the posterior covariance is Eq. (18). Hence $c^\top \hat{\pi}'_v - c^\top \hat{\pi}_v$ is Gaussian with standard deviation $|c^\top \Sigma g|/\sqrt{g^\top \Sigma g + \sigma^2} = \delta$, and $c^\top \Sigma c - c^\top \Sigma' c = \delta^2$, which is Eq. (8). *Proposition C.1.* By the tower property, $\mathbb{E}_a[\mathbb{E}[\max_w Z_w \mid \mathcal{H}, a] \mid \mathcal{H}] = \mathbb{E}[\max_w Z_w \mid \mathcal{H}]$, so $\mathcal{L}(\mathcal{H}) - \mathbb{E}_a[\mathcal{L}(\mathcal{H},a) \mid \mathcal{H}] = \mathbb{E}_a[\max_w \mathbb{E}[Z_w \mid \mathcal{H}, a] \mid \mathcal{H}] - \max_w \mathbb{E}[Z_w \mid \mathcal{H}]$, the expected increase of the maximal conditional score mean. Before the query the conditional means are $U_t(w)$. After it, the mean of Z_v is $U_t(v) + \delta Z$ with Z standard normal under the predictive law of a, by the computation above, and the other means are unchanged because the target memberships are independent in the working model. With $A = U_t(v)$ and $B = \max_{w\neq v} U_t(w)$, direct integration gives $\mathbb{E}\max(A + \delta Z, B) - \max(A,B) = \delta\varphi(|A-B|/\delta) - |A-B|\Phi(-|A-B|/\delta) = \delta f(-|A-B|/\delta)$. Its derivative in δ is $\varphi(|A-B|/\delta) > 0$, its derivative in the gap is $-\Phi(-|A-B|/\delta) < 0$, and it tends to 0 as $\delta \to 0$. For a plan of m queries with c_v held fixed, the joint Gaussian update gives the change of the mean of Z_v the variance $c_v^\top(\Sigma_v - \Sigma_v^{(m)})c_v = \delta_m(v)^2$, and the same integral gives the value of the plan, which Eq. (9) divides by m.

Lemma D.1. A probe of v along a landmark direction g with $g^\top \Sigma_v g > 0$ lowers the rank of Σ_v by one, by Eq. (18), and Algorithm 4 probes only such directions. Under the spanning assumption, every unpinned unevaluated node has an eligible query until it is pinned, which takes at most d probes per node. The initial pin step uses at most d probes for $v^{(t)}$. Subsequently, each query reduces the sum of the remaining covariance ranks over the m_t unpinned unevaluated nodes by one, so at most dm_t further queries pin them all. Nodes already pinned remain exact under the fixed exact geometry. Thus $Q_E \geq d(m_t + 1)$ suffices to make every unevaluated node's feature exact, including the selected node and the optimum if it is still unevaluated. No confidence-based candidate exclusion is used. □

(L) Proof of Corollary A.2. With the evaluated nodes pinned, $\Sigma_v = 0$ there and the moment adjustment of Eq. (17) is void, so exact conditioning by Lemma B.1 with the reference prior D and exact evaluated features gives the reference coefficient state, so Lemma C.2 gives $\Delta_t(v) \leq \Lambda_t e_t(v)$ at every unevaluated v, and Theorem 1 gives the display. The argument uses neither $A_{\mathrm{off}} \in \mathcal{C}_d$ nor the identification of the skeleton, so for any graph the same display holds with e_t the distance to the reference features. Under an identified skeleton with exact global quantities, the spanning conditions, and the budget $Q_E \geq d(m_t + 1)$ of Lemma D.1, every unevaluated node is pinned at decision time, including v^* and $v^{(t+1)}$, so $e_t(v^{(t+1)}) = e_t(v^*) = 0$. □

(M) Decision error under an estimated geometry.

Lemma C.2 (Decision error under an estimated geometry)**.** *Under Assumption 1, if* v^* *is unevaluated and* $v^+ = v^{(t+1)}$ *maximizes* U_t*, Eq.* (10) *holds. If the algorithm and reference use the same bulk variance and coefficient state* (w_t, Ξ_t)*, then* $\Delta_t(v) \leq \Lambda_t e_t(v)$*, where* $e_t(v) = \|\hat{\phi}_v - \phi_v\|_2$ *in a common feature basis and* $\Lambda_t = \|w_t\|_2 + \kappa\|\Xi_t\|_2^{1/2}$*. Exact evaluated features, the reference prior, and exact coefficient conditioning suffice for this state agreement.*

On the event of Assumption 1, $y(v^*) \leq U_t^\circ(v^*) \leq U_t(v^*) + \Delta_t(v^*) \leq U_t(v^+) + \Delta_t(v^*) \leq U_t^\circ(v^+) + \Delta_t(v^+) + \Delta_t(v^*)$, where the third step is the maximality of v^+ under U_t, and $y(v^+) \geq \mu_t^\circ(v^+) - \kappa s_t^\circ(v^+) = U_t^\circ(v^+) - 2\kappa s_t^\circ(v^+)$. Subtracting the two displays gives Eq. (10). For the second claim, when the coefficient state is the reference state both scores are the same function $U(\phi) = \phi^\top w_t + \kappa(\sigma_b^2 + \phi^\top \Xi_t \phi)^{1/2}$, evaluated at $\hat{\phi}_v$ and at ϕ_v. Its gradient is $w_t + \kappa\,\Xi_t\phi/(\sigma_b^2 + \phi^\top \Xi_t \phi)^{1/2}$, and $\|\Xi_t\phi\|_2 \leq \|\Xi_t\|_2^{1/2}(\phi^\top \Xi_t \phi)^{1/2}$, so U is Λ_t-Lipschitz and $\Delta_t(v) \leq \Lambda_t e_t(v)$ by the mean value inequality. □

Algorithm 4 Entry Acquisition by the Knowledge Gradient KG($*$) (one round)

Input: observations Ω, landmarks $\mathcal{R}$, budget Q_E, fit (Π, C), $\{\Sigma_v\}$, GP state (w_t, Ξ_t), the node $v^{(t)}$ evaluated last
Output: $\Omega^{(t+1)}$ and the node $v^{(t+1)}$ to evaluate

1: $q \leftarrow 0$. If $v^{(t)}$ is not pinned, probe it along up to d spanning landmark directions within the pin share ρ_{pin} of Q_E, add the number of probes to q, and recompute (w_t, Ξ_t) ▷ exact features at evaluated nodes
2: compute U_t by Eq. (6) over unevaluated nodes, and the eligible set $\mathcal{N}_t$ by Eq. (21)
3: **while** $q < Q_E$ and $\mathcal{N}_t \neq \varnothing$ **do**
4: for each $v \in \mathcal{N}_t$: $u_v \leftarrow \arg\max\{\delta(v, g_u) : u \in \mathcal{R},\ (v,u) \notin \Omega,\ g_u^\top \Sigma_v g_u > 0\}$ with $c = c_v$, and $\mathrm{KG}_t^*(v)$ by Eq. (9)
5: $v^\star \leftarrow \arg\max_{v \in \mathcal{N}_t} \mathrm{KG}_t^*(v)$ with ties broken at random, and reveal $A_{v^\star u_{v^\star}}$
6: update Ω, $\hat{\pi}_{v^\star}$, $\Sigma_{v^\star}$, $\hat{\phi}_{v^\star}$, $c_{v^\star}$ and $U_t(v^\star)$; refresh $\mathcal{N}_t$ and set $q \leftarrow q + 1$ ▷ $\mathcal{O}(|\mathcal{R}|d^3)$ for the probed node, all other nodes unchanged
7: **end while**
8: $v^{(t+1)} \leftarrow \arg\max_v U_t(v)$ over unevaluated nodes, and carry $Q_E - q$ to the next round

D Pair Queries: Landmark Skeleton and Knowledge-Gradient Probes

This appendix completes the pair-query design of Section 4.3: the cold start that creates landmarks, the greedy KG($*$) procedure, and the lateration lemma that bounds its budget. For a linear functional $c^\top \pi_v$ the rank-one update of Eq. (18) reads

$$c^\top \Sigma_v' c = c^\top \Sigma_v c - \frac{(c^\top \Sigma_v g)^2}{g^\top \Sigma_v g}, \tag{20}$$

a query along an already measured direction has $g^\top \Sigma_v g = 0$ and is skipped, and d queries with spanning directions pin v conditional on the fixed landmark geometry.

Cold start: a landmark skeleton. Nothing can start without landmarks, and the averages that every feature depends on, namely Σ_0, the degree vector $h = C\Pi^\top \mathbf{1}$, and the orthonormalization of Lemma B.2, must be estimated from nodes with known memberships. The experiments create the landmarks from the block among the N_0 initial nodes, $\binom{N_0}{2}$ pairs, which identifies the landmark memberships once $N_0 \geq 2d + 1$ in general position and needs $N_0 \geq r_0(d)$ (Proposition A.1(iii)). A node that has not been probed sits at the mean landmark membership until its first query, which is why the queries of a round precede its selection. An alternative skeleton probes every node against r uniformly drawn landmarks at cost $r(n-1)$ entries, reproduces the row pattern, and identifies the surrogate once $r \geq d + 1$ in general position (Proposition A.1(ii)).

Greedy allocation. The eligible set is

$$\mathcal{N}_t = \{v \notin \mathcal{V}^{(t)} : \exists u \in \mathcal{R},\ (v,u) \notin \Omega^{(t)},\ g_u^\top \Sigma_v g_u > 0\}. \tag{21}$$

Algorithm 4 first pins the node evaluated last, forms the scores and the eligible set $\mathcal{N}_t$ of Eq. (21), and then spends the remaining budget greedily by the KG($*$) score of Eq. (9), updating eligibility and re-optimizing the plan after every answer. All unevaluated, unpinned nodes with an available informative landmark pair are eligible, and there is no confidence-radius threshold. Budget that is left when no eligible query remains is carried to the next round. Since the gains are not submodular in general (Appendix C(F)), the plan is re-optimized after every answer, and Lemma D.1 gives a budget that pins every unevaluated node, including the selected one.

Lemma D.1 (Lateration over all unevaluated nodes)**.** *Let $A_{\text{off}} \in \mathcal{C}_d$ with an identified landmark skeleton whose global quantities are exact and held fixed during the round. Suppose that for every unpinned unevaluated node, and for the node evaluated last if it is unpinned, the observed and available landmark directions together span its d membership coordinates. Run Algorithm 4 with budget Q_E and a pin share that admits d probes, and let m_t be the number of all unpinned unevaluated nodes at the start of the round. If $Q_E \geq d(m_t + 1)$, then every unevaluated node is pinned by the end of the round, so that $e_t(v^{(t+1)}) = 0$ and, if the optimum is unevaluated, $e_t(v^*) = 0$ at decision time.*

The proof is in paragraph (K) of Appendix C. Each step of Algorithm 4 costs $\mathcal{O}(|\mathcal{R}|d^3)$ for the probed node and $\mathcal{O}(n)$ for the maxima, and a round costs $\mathcal{O}(n|\mathcal{R}|d^3 + Q_E(|\mathcal{R}|d^3 + n))$ after the refresh of Algorithm 2.

E Initialization and Refit Schedule

Solver. Problem (3) is solved by projected alternating least squares. Given the core and the other memberships, the objective is a convex quadratic in each π_i, and under row queries every unrevealed node solves the same shared-design problem $\min_{\pi \in \Delta_d} \|G_{\mathcal{R}}\pi - A_{\mathcal{R},v}\|_2^2$ with $G_{\mathcal{R}} = \Pi_{\mathcal{R}} C$, for all nodes at once. Given the memberships, the core solves a box-constrained least-squares problem over symmetric $C \in [0,1]^{d \times d}$. Each subproblem is solved by accelerated projected gradient with Euclidean projection onto Δ_d or onto the symmetric box, a block update is kept only if the residual does not rise, and one sweep updates the revealed memberships, the unrevealed memberships, and the core in turn, at cost $\mathcal{O}(|\Omega|d + nd^2)$ under row queries and $\mathcal{O}(|\Omega|d^2 + nd^2)$ under pair queries, where each node carries its own design. The sub-simplex Δ_d is the simplex over $d+1$ communities with the weight of a null community, which interacts with no one, dropped: a membership has d free coordinates, d measurements pin a node, and the class contains the degree-corrected block memberships $\pi_i = \theta_i e_{k(i)}$ with $\theta_i \in [0,1]$, which the simplex $\{b \geq 0,\ \mathbf{1}^\top b = 1\}$ does not.

Structured initialization. Problem (3) is nonconvex, and the quality of a local solver depends on its starting point. Row observations supply a natural one. The revealed strip $A_{\mathcal{R},\cdot}$ is an $r \times n$ matrix that is fully observed except for its r self-pairs, and under the class it equals $\Pi_{\mathcal{R}} C \Pi^\top$ off those pairs. We compute a rank-d factorization of the strip in two ways, by a truncated SVD whose factors are rescaled and projected onto Δ_d, and by successive projection, which selects d columns of the strip as approximate pure nodes. Each candidate initializes Π. The core C is then obtained from the box-constrained least-squares problem, a fixed number of sweeps (12 in the experiments) is run from each candidate, and the candidate with the smaller residual is kept.

Warm starts and periodic refits. After every reveal the fit is updated by a few sweeps (four in the experiments) started from the previous solution, which is cheap and tracks the data when the pattern changes little. A warm-started iterate can nevertheless settle in a poor local minimum, especially during the first rounds, when each new row changes the strip substantially. Every k_{refit} revealing rounds the structured initialization is therefore recomputed on the current strip, the fit is run from it, and the better of the fresh and the warm-started solutions in terms of the residual is kept, with the warm-started solution retained on a tie. The experiments use $k_{\text{refit}} = 10$, chosen on the design seeds.

Calibration of the shared averages. The reference covariance Σ_0, the degree vector $h = C\Pi^\top \mathbf{1}$, and the orthonormalization $\sum_u M_u M_u^\top$ that every feature depends on are averages over nodes. Under row queries all n memberships are estimated from the first round on, so these averages are computed over all nodes. Their accuracy depends on the fitted memberships. The uniform component of the row policy provides coverage beyond the adaptively evaluated nodes.

F Additional Experimental Results

This appendix gives the protocols and the complete results behind the three questions of Section 5. Results from different initializations or budget schedules are never pooled.

F.1 Protocols and Structural Information Accounting

Budget conventions. The row totals of Table 3 are cumulative additional rows, excluding initialization. Revealing r distinct rows discloses $r(n-1) - r(r-1)/2$ distinct off-diagonal entries. Both nonzero and zero answers count. Entries shared by two queried rows count once in the observed-entry set, while both row queries are charged. Pair queries target previously unobserved unordered pairs. In the main row study, exactly Q_R new rows are bought before each post-initialization evaluation, TQ_R in total, and local graph BO has the same allowance and may use fewer rows. In the pair study, all three acquisition methods receive exactly Q_E new pairs before every post-initialization evaluation, including the first, following the common 496-pair initial block.

Table 3: Protocols used in Section 5. T counts evaluations after initialization, and the row totals exclude the initial rows. BioPlex targets share the indicated protocol. The full-information study is a separate campaign with the horizons below, and ESOL uses a different initialization from the main row study.

Study / instance	n	N_0	T	d	Additional structure
Main: dense (two objectives)	1,000	8	50	16	$Q_R = 2$, $TQ_R = 100$ rows
Main: ESOL	1,117	8	60	16	$Q_R = 2$, $TQ_R = 120$ rows
Main: Chameleon	1,223	8	100	16	$Q_R = 5$, $TQ_R = 500$ rows
Main: BioPlex 293T	13,167	20	100	16	$Q_R = 10$, $TQ_R = 1{,}000$ rows
Main: BioPlex HCT116	9,609	20	100	16	$Q_R = 10$, $TQ_R = 1{,}000$ rows
Main: HCT116 / own line	9,494	20	100	16	$Q_R = 10$, $TQ_R = 1{,}000$ rows
Sparse tasks (Table 4)	2,000 / 5,000	8	100	16	400 / 1,000 total rows
Scale ablation: 293T / DCSBM	13,167 / 5,000	20 / 8	100	16	250 / 1,000 total rows
Full information: BioPlex 293T	13,167	20	100	4–32	Full graph
Full information: DCSBM	$\approx 5{,}000$	8	100	4–32	Full graph
Full information: ESOL	1,117	32	100	4–32	Full graph, score at 25
ESOL per-round pair study	1,117	32	200	32	$Q_E = 2, 4, 8, 16, 32$
ESOL row-rule study	1,117	32	200	16	$Q_R = .1, .25, .5, 1, 2$

The row-rule study of Appendix F.6 uses Q_R as an average acquisition rate. The integer allocation is $q_t = \lfloor tQ_R \rfloor - \lfloor (t-1)Q_R \rfloor$, so cumulative additional rows are $\lfloor tQ_R \rfloor$. In particular, $Q_R = 0.1$ buys one complete row every ten evaluations (20 rows by $T = 200$), not a fraction of a row. This is a fixed schedule with zero-query rounds, distinct from acquiring a fixed positive integer number of rows at every iteration.

Comparison axes. Local graph BO (Wan et al., 2023) receives the same cumulative row allowance as Ours and builds its neighborhoods from its own revealed rows on Chameleon and BioPlex. On the dense and ESOL affinities it uses a precomputed kNN view of the full graph, which conveys structural information beyond the allowance and is disclosed in Section 5.1. In the pair study it is adapted to the pair oracle, its local context is built from the revealed entries, and its restarts are part of the adaptation. Full-graph references receive all structural entries but still query objectives sequentially.

Statistics. The main comparisons and the full-information study use 20 evaluation seeds, the observation-pattern study and the scale ablation use 10. The design seeds, on which every configuration was chosen, are disjoint from the evaluation seeds. Seeds and initial nodes are shared within each protocol. SEM is the sample standard deviation divided by the square root of the seed count, and uncertainty bands are pointwise. For the pair study we report percentile bootstrap intervals from 10,000 resamples of paired seeds, at the nominal 95% level and at 99% for a Bonferroni adjustment across the five budgets within each comparator family. No curve is smoothed, and every displayed cell has all its seeds.

F.2 Budget Sensitivity and Additional Graphs

Row-query trajectories of the main campaign. Figure 5 gives the trajectories of the campaign of Table 1 on the dense in-class instance and on ESOL next to the BioPlex panel of Figure 2. Both dense objectives remain in Table 1, and the class-prior objective provides a controlled setting in which the surrogate family matches the graph generator: it is drawn from the class kernel of the generating matrix, so the class fitted to 108 rows already matches the full-graph references within seed variability.

Figure 6 reports every archived row budget of Ours on the six BioPlex targets, with local graph BO at the same allowances. The 1,000-row protocol of Table 1 lies inside this range. The response to the budget is not monotone at every step, since each fit changes the evaluation path.

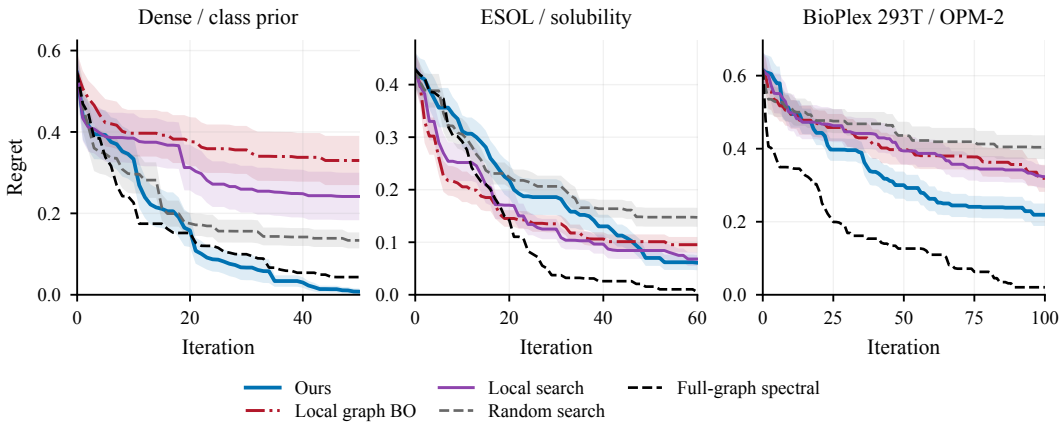

Figure 5: Row-query trajectories of the campaign of Table 1: the dense in-class instance with the class-prior objective ($n = 1{,}000$, $N_0 = 8$, $T = 50$, $Q_R = 2$), ESOL ($N_0 = 8$, $T = 60$, $Q_R = 2$), and BioPlex 293T with the OPM-2 objective ($N_0 = 20$, $T = 100$, $Q_R = 10$), all at $d = 16$. Curves show mean normalized simple regret with pointwise ± 1 SEM bands over 20 paired seeds, and the dashed spectral reference sees the entire graph. On the dense and ESOL affinities local graph BO and local search use a precomputed kNN view of the full graph. The BioPlex panel is Figure 2.

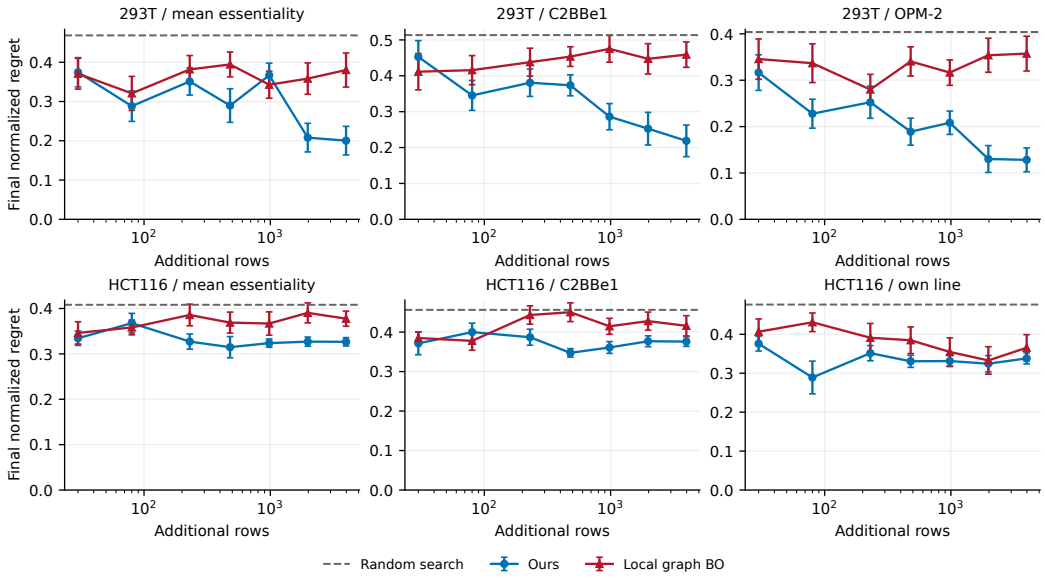

Figure 6: BioPlex row-budget sensitivity, final normalized regret after $T = 100$ evaluations following $N_0 = 20$. Additional rows are the rows revealed after initialization over all T rounds, spent at a constant rate, in a campaign separate from Table 1. Error bars are ± 1 SEM over 20 seeds. Every archived Ours row budget is shown, and local graph BO runs on the original binary graph with the same row allowance. Dashed lines show random-search means.

Table 4 extends the comparison with random search to sparse synthetic graphs. The DCSBM and latent-position examples have lower Ours mean regret than random search at the displayed budgets.

Initial structure and full-information references. Table 5 complements Table 1 at the same task-specific initialization and objective-evaluation horizon. Initial-only retains Ω_0 throughout BO: function observations still update the posterior, and periodic kernel-hyperparameter fitting remains enabled. Its comparison with Ours measures the value of acquiring structure after initialization. Ours, full graph fits the same surrogate family to all entries, whereas the all-mode reference uses the true

Table 4: Additional row tasks, final normalized regret (mean ± SEM, 20 seeds), $d = 16$, $T = 100$, and $N_0 = 8$. Rows are total revealed rows including the initial ones, spent at a constant rate over the rounds. Only random search was run on these instances, and the two full-graph columns have complete structural information.

Instance / objective	Rows	Ours	Random	Ours (full)	Full spectral
DCSBM $n \approx 2{,}000$ / latent	400	$.154 \pm .030$	$.209 \pm .027$	$.094 \pm .023$	$.088 \pm .022$
DCSBM $n \approx 5{,}000$ / latent	1000	$.165 \pm .027$	$.245 \pm .024$	$.113 \pm .023$	$.077 \pm .020$
DCSBM $n \approx 5{,}000$ / class prior	1000	$.120 \pm .028$	$.341 \pm .024$	$.020 \pm .016$	$.047 \pm .033$
Latent-position $n \approx 5{,}000$	1000	$.177 \pm .017$	$.212 \pm .018$	$.101 \pm .014$	$.099 \pm .015$

Table 5: Additional representation controls for the main row study, normalized regret (mean ± SEM, 20 seeds). The task-specific N_0 and T match Table 3. Initial-only buys no structure after initialization, and the other columns receive the entire graph. The full-graph $d = 16$ spectral reference is in Table 1. The all-mode column is measured under this protocol, at the same initialization, horizon and seeds as the other two columns.

Instance / objective	Initial-only	Ours, full graph $d = 16$	Full graph spectral, all modes
Dense / class prior	$.287 \pm .056$	$.048 \pm .027$	$.341 \pm .061$
Dense / latent	$.247 \pm .038$	$.170 \pm .026$	$.244 \pm .037$
ESOL / solubility	$.145 \pm .022$	$.005 \pm .005$	$.121 \pm .013$
Chameleon / log traffic	$.106 \pm .020$	$.095 \pm .006$	$.055 \pm .015$
293T / mean essentiality	$.401 \pm .037$	$.112 \pm .031$	$.272 \pm .039$
293T / C2BBe1	$.492 \pm .035$	$.162 \pm .037$	$.289 \pm .042$
293T / OPM-2	$.324 \pm .037$	$.126 \pm .021$	$.285 \pm .037$
HCT116 / mean essentiality	$.374 \pm .016$	$.314 \pm .006$	$.344 \pm .022$
HCT116 / C2BBe1	$.421 \pm .023$	$.358 \pm .021$	$.364 \pm .024$
HCT116 / own line (HCT116)	$.451 \pm .024$	$.309 \pm .016$	$.359 \pm .023$

graph's untruncated spectral kernel. The latter is tuned on its own normalization and hyperparameter grid and serves as the untruncated reference.

Representation and estimation gaps. Section 5.3 reads Tables 1 and 5 as a decomposition at $d = 16$. The estimation gap of a task is the regret of Ours on the revealed rows minus the regret of the same class fitted to the entire graph, and the representation gap is the latter minus the regret of the full-graph spectral reference at the same d. The estimation gaps are -0.040 and -0.012 on the dense class-prior and latent objectives, -0.016 on Chameleon, 0.056 on ESOL, 0.192, 0.064, and 0.093 on BioPlex 293T (mean essentiality, C2BBe1, OPM-2), and 0.009, -0.009, and 0.050 on HCT116. The representation gaps are 0.005 and 0.000 on the dense objectives, 0.018 on Chameleon, 0.000 on ESOL, 0.112, 0.162, and 0.105 on 293T, and -0.016, 0.028, and 0.011 on HCT116. On the dense instance and Chameleon the partial fit is at or below the full fit, so what remains is representation. On ESOL and on 293T with mean essentiality estimation dominates, on 293T with C2BBe1 and OPM-2 the two gaps are comparable, and on HCT116 both are small. The two class fits share d, filter, initialization, and horizon, so the estimation gap isolates the revealed entries, and the full-graph spectral reference, tuned on its own normalization, makes the representation gap a comparison of two representations of the same graph at the same d.

Check against the per-seed records that the dense/latent and ESOL cells of the middle column of Table 5 are not copies of the full-graph spectral column of Table 1, since the two coincide to three decimals.

F.3 Full-Information Representation Study

Question and protocol. The full-information study asks whether fitting the class to a fully observed graph yields features useful for BO, and how this compares with using the true graph's smooth spectral modes at the same nominal dimension. The class arm fits the graph-valid surrogate using the

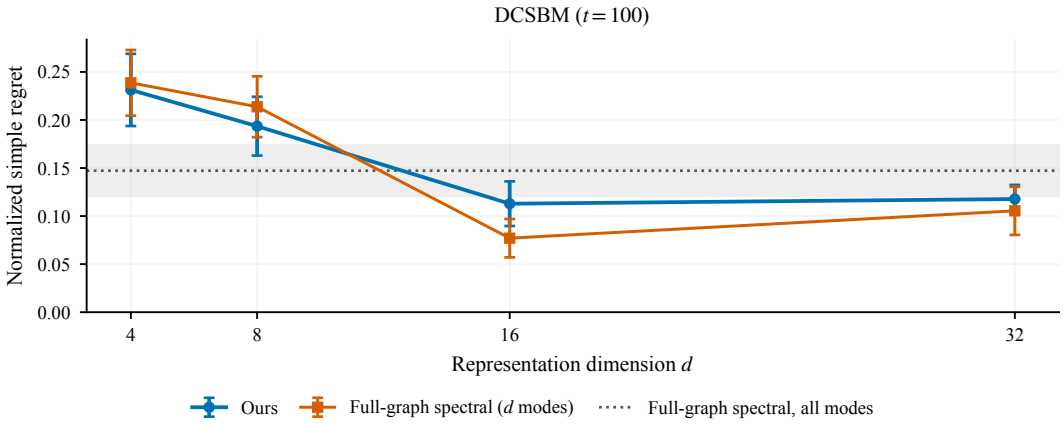

Figure 7: DCSBM full-information resolution control, complementing Figure 3. The graph generator has 16 communities and nominal $n = 5{,}000$ before giant-component restriction, and the objective is latent, with $N_0 = 8$ and $T = 100$. Error bars and the all-mode reference band are ± 1 SEM over 20 seeds. Both finite-dimensional arms improve up to $d = 16$, and larger d does not improve regret further. The all-mode kernel is tuned on its own grid.

Table 6: Full-information study: complete dimension sweep under full graph access. Entries are normalized simple regret, mean $\pm$ SEM over 20 seeds. The checkpoint t counts BO rounds after initialization. The untruncated kernel is a separate reference tuned on a wider hyperparameter grid and is listed once per instance. At $d \leq 8$ on BioPlex both arms end on the same node in every seed, which gives identical cells.

Instance / checkpoint	d	Ours, full graph	Full-graph spectral
BioPlex 293T	4	$.280 \pm .000$	$.280 \pm .000$
($N_0 = 20, t = 100$)	8	$.280 \pm .000$	$.014 \pm .014$
	16	$.112 \pm .031$	$.000 \pm .000$
	32	$.031 \pm .022$	$.052 \pm .028$
Untruncated reference		$.272 \pm .039$	
DCSBM, latent	4	$.231 \pm .038$	$.239 \pm .034$
($N_0 = 8, t = 100$)	8	$.194 \pm .031$	$.214 \pm .032$
	16	$.113 \pm .023$	$.077 \pm .020$
	32	$.118 \pm .015$	$.105 \pm .025$
Untruncated reference		$.147 \pm .026$	
ESOL	4	$.176 \pm .016$	$.176 \pm .017$
($N_0 = 32, t = 25$)	8	$.158 \pm .031$	$.118 \pm .023$
	16	$.139 \pm .027$	$.133 \pm .021$
	32	$.100 \pm .025$	$.062 \pm .017$
Untruncated reference		$.135 \pm .024$	

entire graph, and the spectral arm extracts the d smoothest true-graph Laplacian modes. We sweep $d \in \{4, 8, 16, 32\}$ without changing the initialization size within a dataset.

BioPlex 293T uses the mean-essentiality objective on 13,167 nodes, $N_0 = 20$, and $T = 100$ subsequent BO rounds. DCSBM uses a nominal 5,000 nodes before giant-component restriction, 16 generating communities, mean degree 20, and the latent objective, with $N_0 = 8$ and $T = 100$, and the synthetic graph is generated separately for each seed. ESOL uses 1,117 molecules, the real solubility objective, $N_0 = 32$, and $T = 100$. ESOL is scored at $t = 25$, before the full-information runs reach the regret floor. The configured UCB multiplier is 2, with hyperparameter updates every 10 rounds. Full-graph access does not reveal unqueried objective values to BO.

Metric and aggregation. Table 6 reports the normalized simple regret of Section 5.1 at the stated checkpoint, mean $\pm$ SEM over 20 seeds, with the complete objective used only for offline scoring.

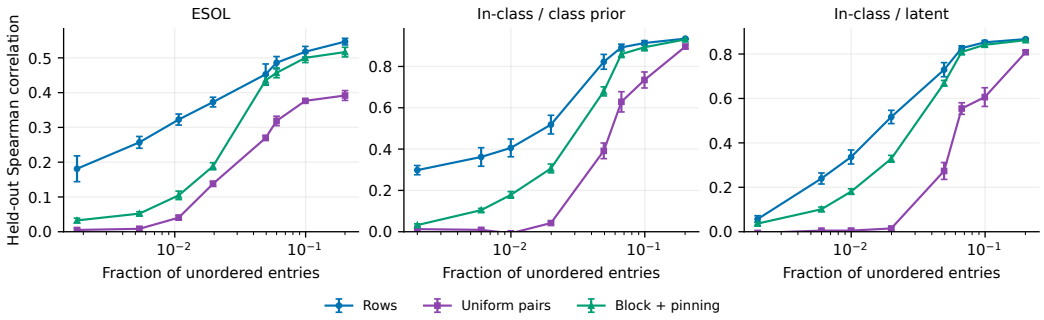

Figure 8: Observation-pattern sensitivity with the class estimator, $d = 16$. Held-out prediction Spearman correlation versus the fraction of distinct unordered entries, higher is better. Means and ± 1 SEM over 10 outer seeds. Each instance uses matched entry counts across the three displayed patterns.

Reading. The sweep shows that class representations at dimensions far below the graph size are useful without the adjacency being low rank. On BioPlex the eigenbasis reaches its floor at $d = 8$ while the class needs $d = 32$, on DCSBM both arms improve up to $d = 16$ and not beyond, and on ESOL the class improves across the sweep and stays above the eigenbasis. Full graph access removes the estimation of entries and leaves the class approximation, the solver, the kernel fit, and the policy in the comparison, so a difference between the two arms at fixed d is a difference of representations. The dimension sweep is an offline sensitivity analysis: the partial-information algorithm never inspects the full graph to select d.

F.4 Observation Patterns at Matched Entry Counts

The static observation-pattern study holds the estimator to our class surrogate and varies how a fixed number of unordered entries is observed: whole rows, uniformly sampled pairs, or a block followed by pinning. The feature map is then frozen for held-out GP prediction. Figure 8 reports rank correlation on ESOL and two dense in-class objectives at $d = 16$, with 10 outer seeds. At 12,221 observed ESOL entries, the correlations are 0.373 ± 0.014 for rows, 0.138 ± 0.007 for uniform pairs, and 0.188 ± 0.010 for block-plus-pinning. At 124,785 entries they are 0.547 ± 0.010, 0.392 ± 0.014, and 0.517 ± 0.014, respectively. Thus the pattern of observations matters even when the number of entries is matched. Rows can deliver useful geometry with fewer entries in these settings, which helps explain why the two query types need separate acquisition studies. Among the pair patterns, a landmark block followed by probes against the landmarks, the pattern that Algorithm 4 produces, is above uniform pairs at every entry count on all three instances, which is the static counterpart of the pair study of Section 5.4. The study measures held-out prediction from a frozen feature map, and Section 5.4 gives the BO trajectories.

F.5 Pair Acquisition

Protocol. The ESOL pair study uses $n = 1{,}117$, $d = N_0 = 32$, $T = 200$, and 20 shared seeds. The 496 initial pairs are common to all three acquisition methods, which then purchase exactly Q_E previously unobserved pairs before each objective evaluation. Ours (KG) combines landmark queries chosen by the KG rule with conditional membership updates and a pinning step capped at half the per-round budget. Ours (uniform pairs) samples uniformly from unrevealed unordered pairs and updates the class completion from those observations. Both use the structured-plus-bulk resolvent kernel, and this comparison evaluates their complete acquisition-and-update procedures rather than isolating the scoring rule. Local graph BO uses the same pair oracle and budget and builds its local context from the revealed entries. Its local-context threshold and restarts are part of the pair-oracle adaptation.

Budget response and optimization trajectories. Table 7 reports all five budgets. Ours (KG) has the lowest mean final regret among the three acquisition methods throughout the sweep. Increasing

Table 7: ESOL pair-query study: final normalized simple regret, mean ± SEM over 20 seeds, $d = N_0 = 32$, $T = 200$. Each method acquires Q_E pairs before every post-initialization evaluation. Bold marks the lowest mean among the three methods.

Q_E	Ours (KG)	Ours (uniform pairs)	Local graph BO
2	**.055** ± .012	.072 ± .010	.065 ± .014
4	**.047** ± .011	.055 ± .013	.105 ± .011
8	**.040** ± .011	.059 ± .011	.094 ± .010
16	**.011** ± .007	.037 ± .011	.083 ± .016
32	**.015** ± .008	.037 ± .011	.095 ± .012

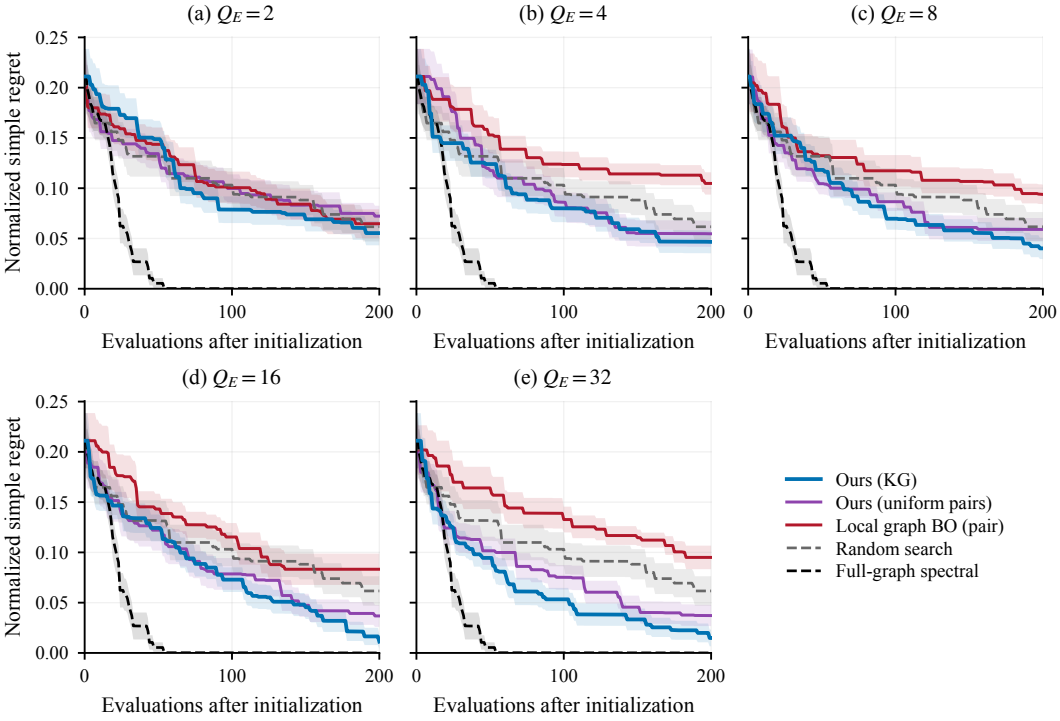

Figure 9: ESOL pair-query study: all five acquisition budgets, mean ±1 SEM over 20 seeds, $d = N_0 = 32$, $T = 200$. The three acquisition methods share the initial 496 pairs and acquire Q_E additional pairs per round. Dashed curves are random node search and full-graph spectral BO references, repeated across panels for context.

Q_E from 2 to 16 reduces its mean from 0.055 to 0.011, and the mean at 32 is 0.015. Figure 9 shows every trajectory, including the two reference methods. Random node search on the seeds of this study ends at 0.062 ± 0.015, and the full-graph spectral reference reaches zero final regret. Figure 4b shows the highest tested allocation.

Paired evidence and scope. The paired differences between Ours (KG) and local graph BO are negative with intervals excluding zero at $Q_E = 4, 8, 16, 32$ after the five-budget adjustment (Table 8). Against uniform pairs all five paired differences favor Ours (KG). The comparison evaluates the acquisition-and-conditioning procedure as a whole, with the landmark geometry estimated from the available entries.

F.6 Row Rules

The row-query study compares uniform rows with pinning, UCB-leader rows, and local-BFS rows using the same class learner on ESOL ($d = 16$, $N_0 = 32$, $T = 200$, 20 seeds). UCB-leader rows reveal the row of the unrevealed node with the largest score, and local BFS rows reveal rows in

Table 8: Paired final-regret differences, Ours (KG) minus each comparator. Brackets are 99% percentile bootstrap intervals, providing a five-budget Bonferroni adjustment within each comparator family. Negative differences favor Ours (KG). All pairs use the same 20 seeds.

Q_E	Versus uniform pairs	Versus local graph BO
2	$-.0169$ $[-.0537, .0227]$	$-.0093$ $[-.0507, .0317]$
4	$-.0081$ $[-.0503, .0361]$	$-.0581$ $[-.0918, -.0244]$
8	$-.0189$ $[-.0510, .0130]$	$-.0537$ $[-.0851, -.0174]$
16	$-.0255$ $[-.0601, .0118]$	$-.0720$ $[-.1172, -.0250]$
32	$-.0223$ $[-.0495, .0020]$	$-.0802$ $[-.1130, -.0435]$

Table 9: Row-query study: ESOL row policies, mean normalized regret $\pm$ SEM over 20 seeds, $d = 16$, $N_0 = 32$. Additional rows at checkpoint t are $\lfloor tQ_R \rfloor$.

t	Q_R	Uniform + pinning	UCB-leader	Local BFS
100	0.1	$.047 \pm .019$	$.046 \pm .017$	$.050 \pm .019$
100	0.25	$.036 \pm .015$	$.028 \pm .011$	$.043 \pm .016$
100	0.5	$.016 \pm .009$	$.017 \pm .012$	$.031 \pm .011$
100	1	$.011 \pm .008$	$.027 \pm .011$	$.008 \pm .008$
100	2	$.000 \pm .000$	$.012 \pm .008$	$.000 \pm .000$
200	0.1	$.006 \pm .005$	$.008 \pm .006$	$.011 \pm .007$
200	0.25	$.011 \pm .011$	$.005 \pm .005$	$.005 \pm .005$
200	0.5	$.005 \pm .005$	$.005 \pm .005$	$.000 \pm .000$
200	1	$.000 \pm .000$	$.000 \pm .000$	$.000 \pm .000$
200	2	$.000 \pm .000$	$.005 \pm .005$	$.000 \pm .000$

breadth-first order from the evaluated nodes along revealed edges. Table 9 gives both $t = 100$ and 200, and Figure 10 gives the trajectories. Neither alternative improves on the default consistently across budgets, and the differences are comparable to the SEM. Random node search on the seeds of this study gives 0.103 ± 0.012 at $t = 100$ and 0.055 ± 0.011 at $t = 200$, and every row rule improves on it at both checkpoints. At $t = 200$ most configurations have many zero-regret runs, which limits discrimination between policies.

Reading the row comparison. Rows provide common membership measurements across unrevealed nodes (Proposition B.1). With $N_0 = 32 > d = 16$, the local covariance is already zero whenever the fitted initial directions span the membership space. Further rows can still improve the surrogate fit on ESOL, where exact in-class identification is not assumed. The observed comparison supports pin-then-uniform as a practical default across these budgets.

G Experiment Details

G.1 Synthetic Random Graph Models

Dense in-class affinity. The dense instance of Table 1 is a member of the class. Each of $n = 1{,}000$ nodes receives one of $d = 16$ blocks uniformly at random and a degree factor $\theta_v \sim \mathrm{U}[0.5, 1]$, its membership is $\pi_v = \theta_v e_{z_v}$, the core is $C = 0.05\,\mathbf{1}\mathbf{1}^\top + 0.25\,I$ (within-block affinity 0.30, between-block 0.05), and the affinity matrix is $\Pi C \Pi^\top$ with its diagonal removed, a weighted matrix with entries in $[0, 1]$ that a row query reveals exactly. The regularizer ζ is the mean row sum.

Sparse degree-corrected stochastic block model (DCSBM). Each node v receives a community label z_v uniformly at random among $K = 16$ communities and a degree factor $\theta_v = \exp(0.5\, g_v)$ with g_v standard normal, truncated at $|g_v| \le 2.5$ and normalized to mean one. The edge intensity between $u \neq v$ is $s\,\theta_u \theta_v B_{z_u z_v}$, where B equals 10 on the diagonal and 1 elsewhere, and the scale s is set so that the expected mean degree is 20 (Karrer & Newman, 2011). The graph is sampled block pair by block pair as a Poisson multigraph collapsed to a simple graph, so that no $n \times n$ array is formed, and it is restricted to its giant component. The intensity has a 16-community factorization, but collapsing Poisson counts to binary edges changes the expectation to $1 - \exp(-s\theta_u \theta_v B_{z_u z_v})$ off

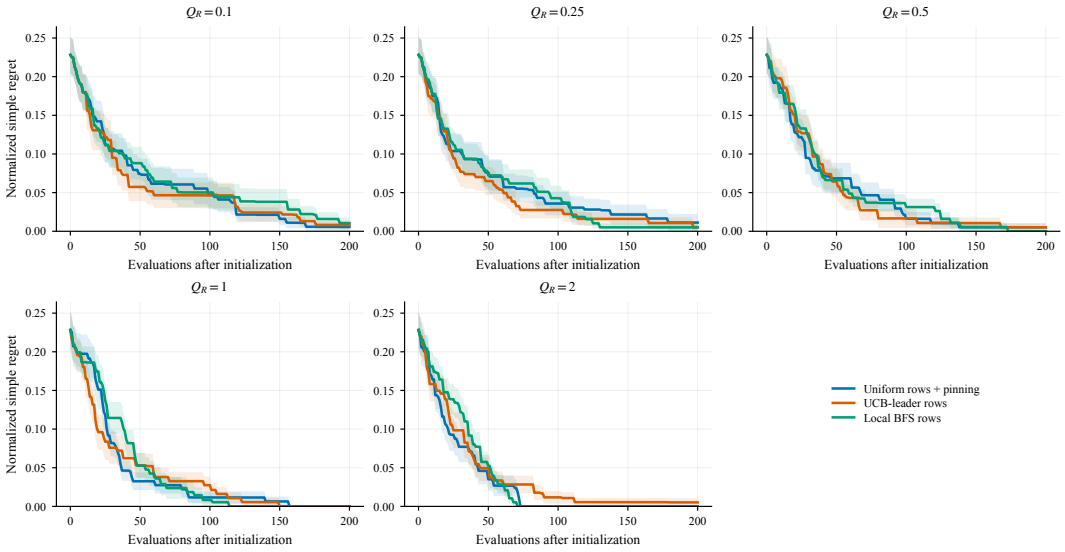

Figure 10: Row-query study: ESOL row-rule trajectories, mean ± 1 SEM over 20 seeds, $d = 16$, $N_0 = 32$, $T = 200$. The additional row count at checkpoint t is $\lfloor tQ_R \rfloor$. Later saturation motivates reporting both intermediate and final results. These are row-selection policies within Ours.

the diagonal. The sampled binary graph is therefore an out-of-class instance for the rank-16 surrogate. We use $n \in \{2000, 5000\}$.

Sparse latent-position graph (RDPG). Latent positions x_v are uniform on $[0, 1]^2$ and each pair is connected independently with probability $\rho \exp(-\|x_u - x_v\|^2/2h^2)$ with $h = 0.25$, where ρ is set so that the expected mean degree is 20 (Young & Scheinerman, 2007). The link is a fixed smooth kernel and not a shrinking geometric radius, so the expected adjacency has a smoothly decaying spectrum and the rank-d class is an approximation with a nonzero bias. We use $n = 5000$.

G.1.1 Synthetic node functions

Latent. On the DCSBM and on the dense in-class instance, $y(v) = a_{z_v} + 0.5\,\epsilon_v$ with community levels a_k and node noise ϵ_v standard normal. The objective has no degree term and is a function of the generating membership alone. On the RDPG, $y(v) = g(x_v) + 4\exp(-\|x_v - c\|^2/2w^2) + 0.5\,\epsilon_v$, where g is a standardized random trigonometric polynomial with four terms of unit integer frequency, the bump has width $w = 0.2$, and its centre c is uniform on $[0.2, 0.8]^2$. **Class-prior.** A draw from the GP prior of the structured-plus-bulk kernel built from the expected adjacency, plus independent noise whose variance is five percent of the total. All objectives are standardized to zero mean and unit variance. The parameters of the objectives were fixed on the design seeds.

G.2 Real-World Graph Models

BioPlex and DepMap. BioPlex 3.0 (Huttlin et al., 2021) reports protein interactions detected by affinity-purification mass spectrometry in HEK293T and HCT116 cells. We use each network as an undirected binary graph restricted to its largest connected component and to the genes present in the DepMap CRISPR gene-effect data (Tsherniak et al., 2017), which gives $n = 13{,}167$ proteins with mean degree 16.4 for 293T and $n = 9{,}609$ for HCT116, of which 9,494 have a gene effect measured in the HCT116 screen and form the node set of the own-line objective. The objective is the negated DepMap CRISPR gene effect of the gene encoding each protein, so that larger values mean stronger dependency, either averaged over all screened cell lines (mean essentiality) or taken in a single line: C2BBe1 (DepMap ACH-000009, colon adenocarcinoma), OPM-2 (ACH-000024, multiple myeloma), and, for the HCT116 network, HCT116 itself (ACH-000971). The two lines were fixed by their DepMap identifiers before any run, and the gene effects are those of DepMap release 24Q4 (the release name of the fourth quarter of 2024). A structural query corresponds to one bait experiment and a function evaluation to one knockout phenotype.

ESOL. The ESOL collection (Delaney, 2004) provides measured aqueous solubilities, of which we use 1,117 molecules. The affinity is the Tanimoto similarity of Morgan fingerprints of radius two, a dense matrix with entries in $[0, 1]$. The Tanimoto similarity is cheap to compute and stands in for expensive pairwise similarities, for which each entry carries a real cost.

Chameleon. The chameleon page graph of the MUSAE collection (Rozemberczki et al., 2021) has Wikipedia articles as nodes and mutual hyperlinks as edges. We symmetrize and binarize the graph, keep its largest connected component, collapse nodes with identical neighbor sets onto one representative, and keep the largest connected component again, which leaves $n = 1{,}223$ nodes. The objective is the logarithm of the average monthly traffic of the page.

Scope of the benchmark. The tasks cover dense affinities, sparse biological networks, and synthetic models. Different BioPlex objectives share a network, so their results are reported separately and never pooled into one significance test.

G.3 Baseline Algorithms and Implementation Settings

Random node search evaluates a uniformly sampled unevaluated node and buys no additional structure. **Local search** evaluates a revealed, unevaluated neighbor of its current node, moves when the value improves, discovers neighborhoods through the same charged oracle, and restarts at a uniformly drawn unevaluated node when its accessible neighborhood is exhausted. **Initial structure only** runs our procedure without buying additional rows after initialization and still updates the function posterior and kernel hyperparameters. **Ours, full graph** fits the surrogate with full structural access, whereas **full-graph spectral** uses the true graph's low-dimensional spectral features. **Untruncated** uses the full-spectrum kernel. These three references are distinct and are labeled separately throughout.

Local graph BO (Wan et al., 2023) is the `poly-default` configuration of the published method. In the row studies it runs on the original graph with neighborhood access charged against the same cumulative row allowance as Ours, and in the pair study it is adapted to the pair oracle, where a local step cannot obtain an unseen neighborhood for free and restarts are part of the adaptation. Apart from the disclosed kNN view on the dense and ESOL affinities it receives no uncharged full-graph preprocessing. On those dense affinities every revealed row is complete, and local search uses the same kNN view to define neighborhoods.

Acquisition controls retain the class learner and vary the structural queries: uniform pairs against Ours (KG) in the pair study, and uniform rows with pinning, UCB-leader rows, and local-BFS rows in the row-rule study. Membership conditioning differs between the two pair arms, so that comparison is a comparison of procedures.

Our main row configurations use $d = 16$, a structured-plus-bulk resolvent kernel, UCB multiplier $\kappa = 2$, refits and hyperparameter updates every 10 evaluations, and pinning share $\rho_{\mathrm{pin}} = 3/4$. The multiplier $\kappa = 2$ is the usual practical choice and is smaller than the union-bound schedule of Assumption 1. The ESOL pair studies use $d = 32$, $N_0 = 32$, and $\zeta = 1$. Initialization, horizon, and structural schedule vary by study as specified in Table 3. The full-information study changes d at fixed initialization within each dataset, and the scale ablation below uses its separately held-out evaluation seeds. Configuration files and numerical-source manifests are preserved in the experiment archive.

H Ablation Study

This appendix reports the kernel-scale ablation. Resolution, budgets, observation patterns, and acquisition rules are in Appendix F.

H.1 Structured and Bulk Scales

This ablation asks whether the exact one-scale resolvent kernel $r(\tilde{L}_\zeta)$ of Eq. (19) is a valid alternative to the structured-plus-bulk kernel of Eq. (4). It compares the two using the same surrogate family, under row queries with the row rule of Section 4.3, on BioPlex 293T with the mean-essentiality objective ($N_0 = 20$) and on the DCSBM with $n \approx 5{,}000$ and the latent objective ($N_0 = 8$), both at

Table 10: One-scale against structured-plus-bulk resolvent kernels: final normalized regret after 100 evaluations, mean ± SEM and median over 10 paired evaluation seeds, wins, losses and ties of the one-scale kernel, and the sign-test p-value. Budgets are total revealed rows.

Instance	Rows	One-scale	Two-scale	Median one	Median two	W/L/T	p
BioPlex 293T	250	$.300 \pm .037$	$.365 \pm .054$	.312	.364	4/1/5	.375
BioPlex 293T	1,000	$.365 \pm .038$	$.377 \pm .054$	.312	.436	5/4/1	1.000
DCSBM	250	$.151 \pm .043$	$.176 \pm .045$	.140	.127	5/3/2	.727
DCSBM	1,000	$.142 \pm .024$	$.169 \pm .043$	.144	.141	5/2/3	.453

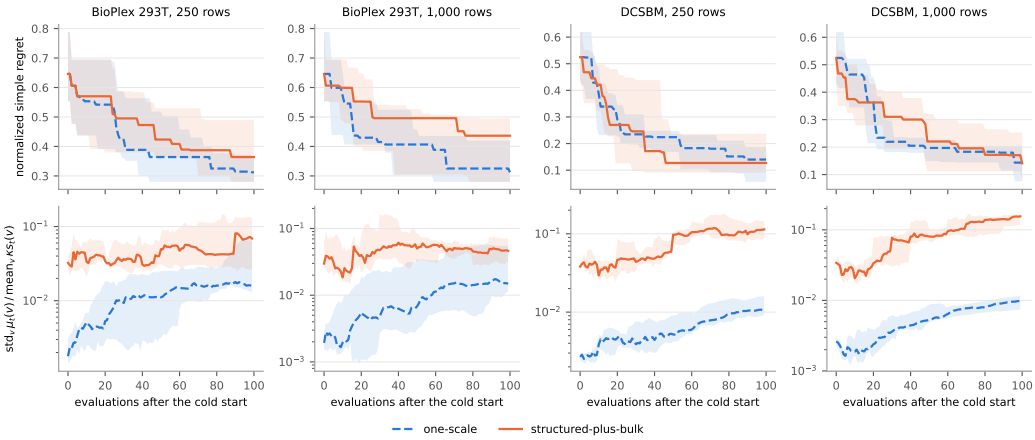

Figure 11: One-scale against structured-plus-bulk resolvent kernels using the same surrogate family, row queries, resolution 16, median with interquartile band over 10 paired evaluation seeds. Top: normalized simple regret against evaluations after the cold start. Bottom: the structured contrast of the acquisition, $\mathrm{std}_v\, \mu_t(v) / \mathrm{mean}_v\, \kappa s_t(v)$, on a logarithmic axis. For the resolvent shape, decoupling the scales raises the contrast by about one order of magnitude on these instances while the regret trajectories overlap.

resolution 16 and at 250 and 1,000 total revealed rows. The ten evaluation seeds are disjoint from the design seeds on which the default kernel was selected, and the two kernels use paired cold starts and matched structural budgets. Figure 11 shows the regret trajectories and the contrast diagnostic, and Table 10 the final normalized regret after 100 evaluations. The two kernels perform comparably: none of the four paired differences is significant at the 5% level, the medians are split, and the means differ by amounts comparable to the SEM. The one-scale kernel is therefore a valid alternative, and the main text uses the structured-plus-bulk kernel because it supports the linear representation for every core and separates the structured and bulk scales.

The ratio $\mathrm{std}_v\, \mu_t(v) / \mathrm{mean}_v\, \kappa s_t(v)$, the variation of the structured posterior mean relative to the exploration term, is about one order of magnitude larger for the structured-plus-bulk kernel than for the one-scale kernel on both instances and both budgets (Figure 11, bottom). Decoupling the two scales thus raises the structured contrast of the acquisition, as the dilution argument of Appendix B predicts, while the regret trajectories overlap. The one-scale resolvent retains all 16 modes and the structured-plus-bulk kernel 9 to 11, so the comparison changes retained directions as well as the structured-to-bulk contrast.