

Hodge-Guided Active Preference Elicitation for Efficient Pairwise Ranking

Eskander Bejaoui*, Mohamed Ould-Elhassen Aoueleiyine, Ridha Bouallegue

Innov'COM Laboratory, Higher School of Communication of Tunis, University of Carthage, Tunis, Tunisia

Abstract—Pairwise comparison is a standard method for eliciting user preferences in recommender systems and group decision-making. The number of required comparisons grows quadratically with the number of alternatives, creating a substantial burden for users and platforms. This paper asks whether Hodge decomposition, which splits a preference flow into gradient (consensus), curl (local inconsistency), and harmonic (global inconsistency) components, can guide which comparisons to request, rather than merely aggregate whichever comparisons arrive. Preference elicitation is formalized as a sequential acquisition problem for online HodgeRank. Seven acquisition policies are evaluated, including two novel Hodge-informed policies: HarmonicReduce, which targets edges likely to carry harmonic inconsistency, and FlowClone, a lightweight neural policy trained by behavioural cloning on a compact five-dimensional feature representation. A Bayesian Bradley-Terry active-learning baseline (BALD) is also included. Policies are evaluated on three real-world datasets under a leakage-free protocol: held-out relevance labels come from full-data batch HodgeRank, tracked via normalized discounted cumulative gain, and validated under five-fold leave-users-out cross-validation so every user is evaluated once. Across all three datasets, geometry-informed active policies cut the comparisons needed to reach a fixed recommendation-quality threshold by roughly half relative to random sampling (43–56% depending on dataset). They consistently outperform BALD, which is itself a statistically confirmed improvement over random sampling on two of the three datasets. This advantage depends on the scoring model: pairing active acquisition with a Borda count aggregator produces worse outcomes than random acquisition with HodgeRank, a phenomenon termed the Borda collapse and confirmed significant, in both directions, by paired Wilcoxon tests with Holm-Bonferroni correction. HarmonicReduce's effectiveness depends on comparison-graph density, a pattern proven exactly from the circuit rank of the comparison graph. These findings show that Hodge-theoretic structure is informative not only for aggregating collected preferences, but also for deciding which preferences to collect.

Keywords—HodgeRank; hodge decomposition; active learning; preference elicitation; pairwise comparisons; uncertainty sampling; behavioural cloning; recommender systems

I. INTRODUCTION

Pairwise comparison is a default mechanism for collecting subjective human judgments. It is used in recommender systems, crowdsourced quality assessment, and the reward-model training pipelines behind reinforcement learning from human feedback (RLHF) for large language models [1], [2], [3]. Compared with direct numerical scoring, pairwise comparisons avoid inconsistent scale interpretation across raters and are cognitively simpler to produce reliably [4]. This robustness

has a combinatorial cost: a complete comparison set over n alternatives requires $\binom{n}{2}$ judgments, growing quadratically as the item catalogue grows. Platforms therefore face a direct trade-off between recommendation quality and user burden. Too few comparisons yield an unstable ranking. Too many risk user fatigue, attrition, or abandonment before elicitation completes, especially during cold-start onboarding when no recommendation can yet be shown. The same cost-of-comparisons tension appears at a different scale in RLHF pipelines, which are explicitly constrained by the cost of human-annotated preference pairs, motivating a growing literature on which comparisons a system should request rather than passively accept [5], [6], [7], [8]. This motivates a direct question: rather than treating the arrival order of comparisons as fixed or random, can a system choose, at each step, the comparison that most reduces its uncertainty about the underlying ranking, and thereby reach an equally reliable recommendation with substantially fewer comparisons?

HodgeRank is an algebraic ranking method built on combinatorial Hodge theory. It decomposes a pairwise-comparison flow uniquely and orthogonally into a gradient component (a globally consistent ranking), a curl component (local, triangle-level inconsistency), and a harmonic component (global inconsistency that no local triangle can detect) [9], [10]. Prior work extends this decomposition from an aggregate consistency diagnostic to per-contributor behavioural signatures, deriving curl- and harmonic-energy-based features to weight or assess contributors after their comparisons are collected [11]. What remains open is whether the same orthogonal split, into consensus, local inconsistency, and global inconsistency, is informative for acquisition: deciding which comparison to request next, rather than how to weight, aggregate, or audit comparisons already collected. This is a natural question. If curl and harmonic energy identify which existing comparisons are inconsistent, the same quantities, computed incrementally as new comparisons arrive, might identify which future comparisons are most likely to resolve inconsistency, and therefore most valuable to request.

Active learning formalizes a simple intuition: a learner that chooses its own training examples can reach a target accuracy with far fewer labels than one that receives them passively [12]. For preference data specifically, Bayesian approaches select comparisons that maximize expected information gain over a Gaussian-process preference model [13], [14]. Query-complexity results show that an adaptively chosen sequence of pairwise comparisons can recover a ranking using far fewer queries than random or exhaustive comparison, when items admit a low-dimensional embedding [15]. The recent growth

*Corresponding author.

of RLHF has renewed interest in this problem under the name query-efficient preference collection, typically operating on a scalar Bradley–Terry reward model [5], [7], [6], [8]. Despite this convergence of interest, no prior work asks whether Hodge-theoretic structure specifically is informative for acquisition, rather than aggregation or quality assessment.

This paper makes four contributions. First, sequential preference elicitation for online HodgeRank is formalized under a leakage-free evaluation protocol: held-out relevance labels come from a full-data batch ranking, not from the online policy’s own trajectory. The learned policy is validated under five-fold leave-users-out cross-validation, so every user is evaluated exactly once. Second, two novel Hodge-informed acquisition policies are introduced. HarmonicReduce targets pairs likely to carry harmonic energy via a degree-weighted residual proxy. FlowClone is a lightweight neural network trained by behavioural cloning to approximate an uncertainty-sampling expert from a compact five-dimensional feature vector. A Bayesian Bradley–Terry active-learning baseline (BALD) is additionally implemented as a concrete comparison against the Bayesian preference-elicitation literature. Third, seven acquisition policies are evaluated against four scoring models on three real-world datasets of differing size and density, documenting a structural incompatibility, the Borda collapse, between gradient-residual acquisition and Borda-count scoring. Fourth, a combinatorial proposition based on the circuit rank of the comparison graph is proven. It explains exactly why HarmonicReduce’s effectiveness switches sign between dense and sparse comparison graphs.

The remainder of this paper is organized as follows. Section II reviews related work. Section III presents the mathematical framework. Section IV describes the experimental setup. Section V reports results. Section VI discusses implications and limitations. Section VII concludes.

II. RELATED WORK

Aggregating pairwise comparisons into a global ranking is classically addressed by Borda count [4], Copeland’s method, and the Bradley–Terry probabilistic model [16]. The author in [9] introduced combinatorial Hodge theory to this problem, formalizing the orthogonal gradient/curl/harmonic decomposition used throughout this paper. [10] applied it to subjective video quality assessment. Follow-up work extends the decomposition to per-contributor behavioural analysis, using curl and harmonic energy to weight or assess contributors after the fact, rather than purely for aggregate consistency assessment [11]. All of this prior work, including acquisition-agnostic evaluations of HodgeRank’s stability under sparse, randomly sampled data [17], treats the comparison set as given and fixed. None asks whether the same decomposition can inform which comparison to request next. This paper is, to our knowledge, the first to use Hodge-theoretic structure prospectively, to decide which comparison to request next, rather than retrospectively, to weight, aggregate, or audit comparisons already collected.

Active learning rests on a simple observation: a learner able to select its own training instances can reach a target performance with substantially fewer labels than passive sampling [12]. For preference data, [13] model preferences with

a Gaussian process and select comparisons by expected information gain. [14] generalize this to Bayesian active learning (BALD). These approaches typically require a probabilistic model with a tractable posterior over item utilities. HodgeRank, a deterministic weighted least-squares estimator, does not directly provide one. Our Entropy policy (Section III-C) is a deterministic analogue of uncertainty sampling that uses the current model’s residual as an informativeness proxy. We additionally implement a genuine Bayesian baseline, an online Bayesian Bradley–Terry model with a Gaussian-probit moment update adapted from [18] for pairwise outcomes, combined with BALD-style acquisition, to give an empirical rather than purely narrative comparison. In the active-ranking theory literature, [15] show that if items admit a low-dimensional Euclidean embedding, an adaptive algorithm can recover a ranking using far fewer comparisons than passive sampling. HodgeRank’s gradient component is precisely such a low-dimensional embedding, giving a theoretical anchor for the empirical query-complexity reduction reported in Section V.

Recent literature studies active or query-efficient preference collection motivated by RLHF annotation cost. This includes acquisition strategies for direct preference optimization [5], [3], provable query-complexity RLHF algorithms under a Bradley–Terry reward model [7], sample-efficient active preference optimization [6], and experimental-design criteria balancing exploration against informativeness [8]. These works share our motivation: preference labels are expensive and acquisition order matters. But they operate on a scalar Bradley–Terry-style reward model rather than the geometric, decomposition-based representation used here. Our scorer ablation (Section V-C) directly probes this distinction. Finally, behavioural cloning trains a policy by supervised imitation of an expert’s choices rather than by reward maximization. It remains an efficient alternative to reinforcement learning when the reward signal is sparse. We adopt it for FlowClone because a policy directly trained on rewards, optimized against per-step nDCG improvement, suffers from high-variance, sparse-reward updates.

III. METHODOLOGY

A. Problem Formulation

The acquisition question this paper asks, which comparison should be requested next, is meaningful only relative to a single respondent’s own trajectory of answers. We therefore frame everything at the level of one user u , not the full population, and defer aggregation across users to the evaluation protocol (Section III-E). Fix an item set $X = \{x_1, \dots, x_n\}$ for that user. Comparisons arrive one at a time as ordered pairs (w, ℓ) , read as “ w was chosen over ℓ .” We reserve 20% of each user’s available comparisons, denoted $P_{\text{held}}^{(u)}$, purely as an evaluation target never exposed to any policy. This leaves a train pool $P_{\text{train}}^{(u)}$ from which an acquisition policy π sequentially draws. At each step t , π picks a still-unqueried pair (w_t, ℓ_t) from $P_{\text{train}}^{(u)}$, receives the outcome, and folds it into a running per-user score estimate. This estimate is updated after every single comparison, not recomputed in batch.

The candidate action space at step t is therefore $P_{\text{train}}^{(u)} \setminus Q_t$, the still-unqueried comparisons that already appear with their observed winner in that user’s historical record, not the full $\binom{n}{2}$ set of possible item pairs. This is an offline-replay

setting: a policy cannot request a pair the user never actually compared, and it cannot observe an outcome other than the one the user originally gave. Offline replay approximates genuine online elicitation only to the extent the original data-collection process is representative of how the same user would answer if queried live, in the same order a policy proposes. None of the used datasets carries contextual or presentation-order effects (e.g., item position, screen layout, or fatigue over the course of a session) because the source data records only comparison outcomes, not the context under which each comparison was made; a genuinely live deployment could exhibit such effects, and offline replay on either dataset is silent on them. This scopes offline-replay validity to the acquisition-order question this paper asks.

B. HodgeRank and Hodge Decomposition

We adopt the standard HodgeRank estimator to obtain a reference (non-sequential) score for a user once all of that user's comparisons are in hand. Write $c_{ij}^{(u)}$ for how many times user u chose item i over item j . Each compared pair is summarized by a normalized margin $\hat{Y}_{ij}^{(u)} = (c_{ij}^{(u)} - c_{ji}^{(u)}) / (c_{ij}^{(u)} + c_{ji}^{(u)}) \in [-1, 1]$ and a confidence weight $\omega_{ij}^{(u)} = c_{ij}^{(u)} + c_{ji}^{(u)}$, reflecting how often that pair was tested. The batch score vector $s^{(u)} \in \mathbb{R}^n$ is the vector that best reproduces these margins as score differences, in a weighted least-squares sense:

$$\min_{s^{(u)}} \sum_{\{i,j\}} \omega_{ij}^{(u)} (s_i^{(u)} - s_j^{(u)} - \hat{Y}_{ij}^{(u)})^2 + \lambda \|s^{(u)}\|^2, \quad (1)$$

regularized by $\lambda = 10^{-5}$ and solved with sparse LSQR [19]. This estimator is useful beyond curve-fitting because its residual is not arbitrary noise. The observed margin vector y splits uniquely into three pairwise-orthogonal pieces, $y = y^g + y^c + y^h$. The first, $y_{ij}^g = s_i^* - s_j^*$, is exactly what the fitted score vector predicts: the part of the data consistent with a single global ranking. The remainder splits further into a piece explainable by local disagreement confined to individual triangles of items (y^c), and a piece that cannot be pinned to any single triangle at all (y^h). This harmonic piece is simultaneously curl-free and divergence-free, and reflects inconsistency distributed around longer cycles in the comparison graph. It is the piece this paper cares about most for acquisition: it vanishes identically whenever the comparison graph has no cycles, and only becomes possible once enough comparisons close a loop.

For elicitation, we cannot wait for a full comparison set before producing a score. We instead maintain $s^{(u)}$ online, updating it after each single comparison rather than solving Eq. (1) from scratch. With warm-up constant γ_0 , step index t , and a $1/(t + \gamma_0)$ learning-rate schedule, each newly observed pair (w, ℓ) triggers:

$$\eta_t = \frac{1}{t + \gamma_0}, \quad r_t = s_w^{(u,t)} - s_\ell^{(u,t)} - 1, \quad s_{w,\ell}^{(u,t+1)} = s_{w,\ell}^{(u,t)} \mp \eta_t r_t, \quad (2)$$

The residual r_t plays a dual role. It drives the score update above, and it also serves as a live signal for how surprising

the just-observed comparison was, a signal several acquisition policies below repurpose as a proxy for informativeness. γ_0 is set per dataset to one-third of that dataset's median per-user comparison count.

Eq. (2) is a stochastic-approximation scheme, in the sense of Robbins and Monro [20], for the same objective Eq. (1) minimizes in the unregularized limit: each step moves $s_w^{(u,t)}$ and $s_\ell^{(u,t)}$ along the negative gradient of a single comparison's squared-residual term, scaled by a learning rate $\eta_t = 1/(t + \gamma_0)$ satisfying the standard Robbins–Monro conditions $\sum_t \eta_t = \infty$ and $\sum_t \eta_t^2 < \infty$ once t is large enough. It is therefore an approximation that converges toward the batch minimizer as more comparisons accumulate, not an independent heuristic unrelated to Eq. (1), though at any finite t , and especially early in a trajectory, its running estimate need not agree with the batch solution over the same comparisons. Section V-G verifies this relationship empirically.

C. Acquisition Policies

Let $c_i^{(t)}$ be the number of times item i has appeared in a query so far and $\mathcal{N}_i^{(t)}$ its already-compared neighbours. We evaluate seven policies:

Random selects uniformly from the remaining pool, and serves as the baseline.

Popularity-LB concentrates queries on already-popular items, $\arg \max_{(i,j)} (c_i^{(t)} + c_j^{(t)})$, a simple heuristic baseline distinct from the geometry-informed policies.

Entropy is a deterministic analogue of uncertainty sampling, selecting the pair with the largest current residual, $\arg \max_{(i,j)} |s_i^{(t)} - s_j^{(t)} - 1|$.

Coverage expands the comparison graph structurally, preferring pairs with few shared already-queried neighbours, with the residual used only to break ties.

HarmonicReduce, a novel policy, targets comparisons likely to carry harmonic energy via a query-count-weighted residual proxy,

$$\pi_{\text{hr}}(t) = \arg \max_{(i,j)} |r_{ij}^{(t)}| \cdot \log \left(1 + \frac{c_i^{(t)} + c_j^{(t)} + 2}{2} \right). \quad (3)$$

The harmonic subspace of a graph has dimension exactly its circuit rank $\beta_1 = |E| - |X| + c$, where c is the number of connected components; harmonic energy is nonzero only when $\beta_1 > 0$ [9], [21].

Proposition 1. *Let $G_t = (X, E_t)$ be the comparison graph after t steps and $(i, j) \notin E_t$ a candidate edge. Adding (i, j) increases β_1 by exactly 1 if i, j already lie in the same connected component, and by 0 otherwise.*

Proof: Adding an edge changes $|E|$ by +1. If i, j are already connected, c is unchanged, so $\beta_1 = |E| - |X| + c$ increases by 1. If they lie in different components, the edge merges them, so c decreases by 1 and β_1 is unchanged. ■

An edge that merely extends the spanning forest cannot carry harmonic energy; only an edge that closes

a cycle can. Query-count $c_i^{(t)}$ is a cheap proxy for $\Pr[i, j \text{ already connected}]$. Eq. (3)'s degree weighting targets exactly the precondition Proposition 1 identifies. On a sparse graph this proxy is discriminative, since most pairs remain unconnected for much of the elicitation budget. On a dense, near-complete graph, almost every pair is already connected after a handful of queries, so the proxy degenerates to a near-constant multiplier. This explains HarmonicReduce's density-dependent performance in Section V-D.

FlowClone is a 5-64-64-1 multilayer perceptron f_θ with tanh activations, scoring candidate pairs from a compact feature vector

$$\mathbf{x}_{ij}^{(t)} = [\rho_{ij}^{(t)}, \sigma_{ij}^{(t)}, \delta_{ij}^{(t)}, \kappa_{ij}^{(t)}, \tau_t]^\top, \quad (4)$$

where,

$$\rho_{ij}^{(t)} = |r_{ij}^{(t)}|, \sigma_{ij}^{(t)} = s_i^{(t)} + s_j^{(t)}, \delta_{ij}^{(t)} = |s_i^{(t)} - s_j^{(t)}|, \kappa_{ij}^{(t)} = \log(1 + |\mathcal{N}_i^{(t)} \cap \mathcal{N}_j^{(t)}|), \tau_t = t/|\mathcal{A}_0|.$$

f_θ is trained offline by behavioural cloning (supervised imitation learning) [22]. Rollouts follow Entropy with probability 0.75 and a random action otherwise; the supervised label is always Entropy's greedy choice, optimized by cross-entropy loss. We chose behavioural cloning over directly optimizing per-step nDCG improvement via reinforcement learning because the latter's reward is sparse and high-variance early in an elicitation trajectory, when most of the item ranking is still unresolved. Behavioural cloning instead supplies a dense supervised signal at every step by imitating Entropy's already-validated choice, at the cost of being bounded above by the expert it imitates. The mixed-rollout scheme (0.75 expert, 0.25 random) exposes the network to states an always-expert rollout would never visit, mitigating covariate shift between training and deployment [22]. The five-dimensional input in Eq. (4) was deliberately kept small. It excludes the raw item scores $s_i^{(t)}, s_j^{(t)}$, using only their sum and absolute difference, so the network cannot simply memorize item identities. This forces it to learn a transferable function of comparison state, a design choice consistent with the near-zero seen/unseen generalization gap reported in Section V-E. FlowClone tests whether Entropy's informativeness can be recovered from a compact, five-dimensional feature vector, and whether it generalizes to unseen users.

BALD is a Bayesian active-learning baseline [14]. Each item i has a Gaussian belief $\mathcal{N}(\mu_i^{(t)}, \sigma_i^{(t)^2})$ under an online Bayesian Bradley-Terry model, updated after each comparison by a pairwise Gaussian-probit moment update adapted from [18], with fixed observation-noise scale $\beta = 0.5$. Expected information gain under this model is monotone in predictive variance, so BALD queries the pair of maximal combined variance,

$$\pi_{\text{bald}}(t) = \arg \max_{(i,j)} \sigma_i^{(t)^2} + \sigma_j^{(t)^2}. \quad (5)$$

This is a standard tractable approximation that avoids Monte Carlo mutual-information estimation. We emphasize that we do not use the full TrueSkill system (teams, draws,

or factor-graph scheduling); we only reuse its scalar Gaussian-probit update idea in a pairwise Bradley-Terry setting.

D. Scoring Models

Given acquired comparisons Q_t , item scores can be computed by **HodgeRank** (the online update of Eq. (2)), **Borda count** (a simple win-counting aggregator that ignores graph structure), **Win-Ratio** (the fraction of comparisons won, normalizing for count imbalance without exploiting Hodge structure), or **Bayesian Bradley-Terry** (the posterior mean $\mu_i^{(t)}$ from BALD's belief model). Separating acquisition policy from scoring model lets us isolate the contribution of each.

E. Evaluation Protocol

An early version of this protocol measured elicitation quality by Kendall correlation between the online ranking and a full-data batch ranking computed from the same user's comparisons. We identify this as circular: it measures how quickly a policy helps HodgeRank reconstruct its own eventual answer, not genuine held-out quality. We replace it as follows. Each user's comparisons are split 80/20 into P_{train} and P_{held} . Held-out relevance labels come from a full-data batch HodgeRank ranking over all of that user's comparisons, train and held-out combined, so ground truth is never a function of the online trajectory. This full-data ranking doubles as the protocol's implicit performance ceiling: nDCG@5 is computed against it directly, so a policy that recovered the full-data ranking exactly from the train pool alone would score $\text{nDCG@5} = 1$, the same reference point every policy in Section V is measured against; no separate oracle policy is required to state this ceiling; it is already the metric's normalization. The item ranked r -th receives graded relevance $n_h - r$. After every elicitation step, the online score ranks the held-out items, and quality is measured by nDCG at cutoff $k = 5$ [23]. We summarize a trajectory by the trapezoidal area under the per-step nDCG@5 curve (AUC-nDCG) and by Comparisons-to-Threshold (CTT), the percentage of the train pool consumed before nDCG@5 first reaches $\tau = 0.80$.

IV. EXPERIMENTAL SETUP

A. Datasets

We evaluate on three datasets (Table I) spanning a range of item-set size and comparison-graph density. Car1 users rank all 10 items ($\binom{10}{2} = 45$ implied comparisons); Sushi-10 users each rank a fixed 10-item subset of a shared 100-item universe; Sushi-X users rank a variable-length subset (5–10 items, mean 7.4). Under the five-fold leave-users-out protocol (Section IV-B), every user is evaluated exactly once, so the counts in Table I reflect the full population evaluated for every dataset.

TABLE I. EVALUATION DATASETS, ITEM-SET SIZES, USER COUNTS, AND PER-USER COMPARISON-GRAPH CHARACTERISTICS

Dataset	Items	Users	Characteristics
Car1 [24]	10	60	Dense, near-complete per-user graphs
Sushi-10 [25]	100	500	Sparse, uniform comparison count
Sushi-X [25]	100	500	Sparse, heterogeneous density

These three datasets span an order of magnitude in item-set size (10 to 100 items) and range from dense, near-complete per-user comparison graphs (Car1) to sparse, variable-length ones (Sushi-X), giving a deliberately heterogeneous test bed for the acquisition policies evaluated in Section V.

B. Implementation, Leave-Users-Out Protocol, and Statistics

Batch reference rankings use $\lambda = 10^{-5}$, solved via sparse LSQR. FlowClone's network is trained for 20 epochs at learning rate 10^{-3} on rollouts collected up to 60% of each user's train-pool budget. The Bayesian Bradley-Terry model uses prior variance $\sigma_0^2 = 1$ and $\beta = 0.5$, standard Gaussian-probit defaults in this pairwise setting, untuned for this study.

Each dataset's users are partitioned into five disjoint folds. For each fold, FlowClone's behavioural-cloning trajectories are collected exclusively from the other four-fifths of users, and every policy is evaluated only on that fold's disjoint held-out fifth, so every user is evaluated exactly once across folds. This gives full-population statistical power for every policy while preserving a genuine leave-users-out guarantee for FlowClone.

Each dataset $\times$ policy $\times$ scorer combination is evaluated over ten random seeds within every fold. Every user's per-seed AUC-nDCG values are averaged into a single paired observation for that user. We assess significance of each policy's advantage over Random (HodgeRank scorer) with a paired, two-sided Wilcoxon signed-rank test over these per-user values, matched by user across the full evaluated population (pooled across all five folds, since folds are disjoint in users). Ten seeds per fold comfortably clears the conventional $n = 6$ minimum for the test to have power. Pooling to the per-user level, rather than testing across seeds alone, gives the test a much larger effective sample size: one paired observation per user, not one per seed. The six policies $\times$ three datasets form a single family of 18 hypotheses, all testing the same claim (policy beats Random). We apply a Holm-Bonferroni step-down correction across all 18 raw p -values jointly, rather than per dataset, and report the Holm-adjusted p -value (p_{holm}) as our significance criterion in Table II. We also report the raw AUC-nDCG advantage Δ and a bootstrapped 95% confidence interval for Δ (10,000 resamples of the paired per-user differences), given inline in Section V wherever a result is discussed individually. We treat consistency of sign and magnitude across all three datasets as corroborating evidence alongside these statistics.

V. RESULTS

A. Active Acquisition Roughly Halves the Comparison Budget

Table II reports per-user nDCG@5 at three budget fractions, AUC-nDCG, and CTT for the HodgeRank scorer, on the full population of each dataset. Entropy, Coverage, and FlowClone substantially and consistently outperform Random on all three datasets. Popularity-LB significantly underperforms Random on Car1 and significantly outperforms it on Sushi-10. Its point estimate on Sushi-X ($\Delta = +0.008$) is not statistically significant ($p_{\text{holm}} = 1$). On no dataset does it approach the geometry-informed policies.

Entropy, Coverage, and FlowClone reach nDCG@5 ≥ 0.80 using, on average, approximately **51% fewer comparisons** than Random (53% on Car1, 56% on Sushi-10, 43% on Sushi-X, via Entropy's CTT). All three improvements are consistent

TABLE II. MAIN HODGERANK RESULTS: PER-USER NDCG@5 AT 20%, 30%, AND 50% BUDGET, PLUS AUC-NDCG AND CTT; BOLD MARKS POLICIES SIGNIFICANTLY BETTER THAN RANDOM AFTER HOLM CORRECTION

Dataset	Policy	20%B	30%B	50%B	AUC	CTT (%)
Car1	Random	0.737	0.790	0.851	0.824	26.5
	Popularity-LB	0.627	0.701	0.799	0.769	35.3
	Entropy	0.879	0.906	0.928	0.896	12.4
	Coverage	0.871	0.895	0.911	0.888	12.2
	FlowClone	0.866	0.900	0.925	0.891	13.2
	HarmonicReduce	0.676	0.767	0.822	0.800	25.5
	BALD	0.771	0.821	0.886	0.850	24.2
Sushi-10	Random	0.770	0.812	0.875	0.841	27.2
	Popularity-LB	0.743	0.870	0.941	0.886	21.1
	Entropy	0.884	0.950	0.956	0.923	12.0
	Coverage	0.884	0.947	0.969	0.926	12.1
	FlowClone	0.901	0.950	0.954	0.925	11.5
	HarmonicReduce	0.791	0.847	0.859	0.849	9.1
	BALD	0.820	0.788	0.896	0.878	19.1
Sushi-X	Random	0.832	0.866	0.904	0.886	24.6
	Popularity-LB	0.776	0.836	0.908	0.894	23.9
	Entropy	0.885	0.947	0.971	0.942	14.1
	Coverage	0.885	0.948	0.971	0.942	14.2
	FlowClone	0.896	0.953	0.971	0.944	13.7
	HarmonicReduce	0.904	0.935	0.952	0.937	12.3
	BALD	0.861	0.881	0.922	0.909	19.4

in sign and magnitude across every seed and fold on every dataset. Each remains significant after Holm-Bonferroni correction across the full 18-test family described in Section IV-B. Fig. 1 shows the full nDCG@5-versus-budget curves, Fig. 2 shows the same comparisons-to-threshold result as bars per policy and dataset, and Fig. 3 summarizes AUC-nDCG across all datasets and policies.

Reporting comparison savings only as a percentage of each user's train pool (Table II, column CTT) obscures the absolute number of judgments this represents. Car1 users have a median 39.6 train-pool comparisons; Entropy's 12.4% CTT is therefore approximately 5 comparisons, against Random's 26.5% (≈ 10 –11 comparisons), an absolute saving of roughly 5–6 judgments per user. Sushi-10 users have a fixed 36 train-pool comparisons (80% of $\binom{10}{2} = 45$); Entropy's 12.0% CTT is approximately 4 comparisons against Random's 27.2% (≈ 10), an absolute saving of roughly 6 judgments per user. Sushi-X users have a mean 18.9 train-pool comparisons (80% of a mean 23.7 implied pairs from a mean 7.4-item subset); Entropy's 14.1% CTT is approximately 3 comparisons against Random's 24.6% (≈ 5), an absolute saving of roughly 2 judgments per user. The relative savings are therefore consistent (43–56%) while the absolute number of judgments saved per user is small at these item-set sizes (2–6 comparisons); the practical benefit compounds with the number of users elicited concurrently and would need re-measurement at the thousands-of-candidates scale, which this paper does not test (Section VI).

B. BALD: Active HodgeRank Policies Outperform a Genuine Bayesian Baseline

BALD tests whether HodgeRank-based acquisition offers a real advantage over a genuine Bayesian alternative, not only over Random and the Popularity-LB baseline. Table II shows BALD is a non-trivial baseline. It significantly outperforms

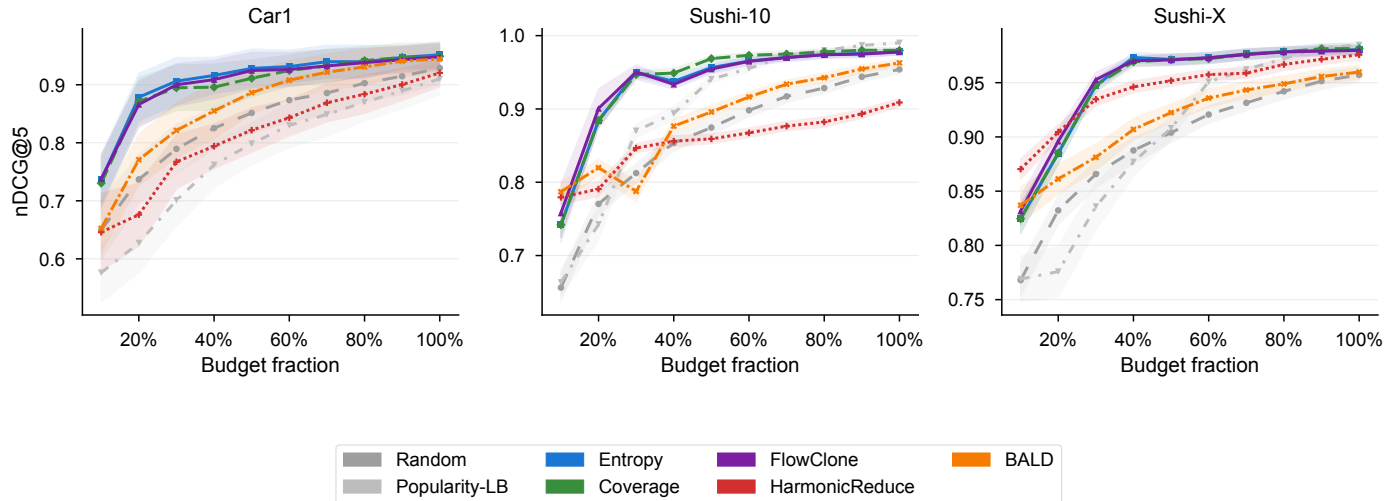


Fig. 1. nDCG@5 versus budget fraction for all acquisition policies under HodgeRank scoring (mean $\pm$ 1 standard deviation over seeds and folds).

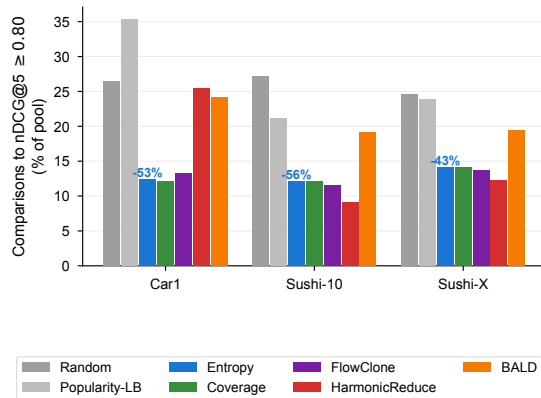


Fig. 2. Comparisons-to-threshold (CTT) to reach $\text{nDCG@5} \geq 0.80$, reported as a percentage of each user's training pool.

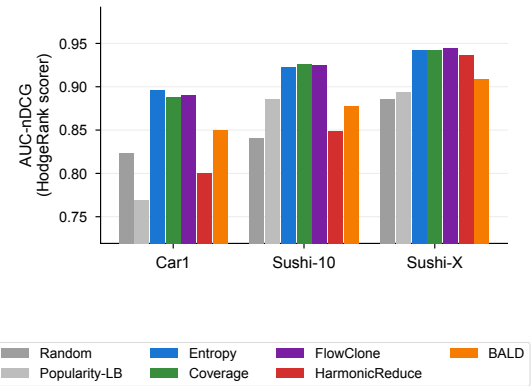


Fig. 3. AUC-nDCG summary across acquisition policies and datasets under HodgeRank scoring.

Random on Sushi-10 ($\Delta = +0.037$, $p_{\text{holm}} = 6.2 \times 10^{-17}$, $n = 500$) and Sushi-X ($\Delta = +0.023$, $p_{\text{holm}} = 9.9 \times 10^{-5}$, $n = 500$); both survive Holm-Bonferroni correction across the full 18-test family. On Car1, BALD's point estimate is smaller and does not reach significance ($\Delta = +0.026$, raw $p = 0.57$, $n = 60$), consistent with this dataset's much smaller evaluated population. Regardless of whether BALD's own advantage over Random is significant, Entropy, Coverage, and FlowClone consistently outperform BALD's point estimate on all three datasets, by roughly 0.03–0.06 AUC-nDCG and 5–12 points of CTT. We interpret this as reflecting what each criterion is sensitive to. BALD's posterior-variance criterion treats all belief uncertainty symmetrically, with no gradient/curl/harmonic distinction. Entropy and Coverage directly probe HodgeRank's structure instead.

C. HodgeRank is Necessary for Active Gains: The Borda Collapse

Table III reruns Entropy under all four scoring models. Entropy paired with Borda count is *worse* than Random paired

with HodgeRank on every dataset, and worse than Random paired with Borda on every dataset as well.

The cause is structural. Entropy selects the comparison with the largest gradient residual, i.e. the one about which the current model is most surprised. Such a comparison is, almost by construction, one for which the two items have been nearly evenly split in prior wins and losses. Its Borda contribution, a simple win increment, is close to uninformative regardless of outcome. Entropy actively steers the query budget toward exactly the comparisons Borda is least equipped to exploit, producing outcomes worse than Random. We term this the Borda collapse. Active gains are not a generic property of any acquisition-plus-scoring pipeline; they depend on compatibility between the two. The failure mode is an outright reversal, not merely a smaller gain. Win-Ratio and Bayesian Bradley-Terry remain competitive with HodgeRank throughout. On the sparse Sushi datasets they are essentially tied with it or marginally ahead: HodgeRank trails Win-Ratio by 0.004 and Bayesian-BT by 0.010 on Sushi-10, and ties both to three decimal places on Sushi-X. We attribute this to the near-acyclic structure of sparse per-user graphs, where the harmonic component is

TABLE III. SCORER ABLATION UNDER ENTROPY ACQUISITION: AUC-NDCG FOR EACH SCORING MODEL, WITH Δ COMPUTED AGAINST RANDOM USING THE SAME SCORER. BOLD MARKS SIGNIFICANCE AFTER HOLM CORRECTION IN THE RELEVANT FAMILY ($\alpha = 0.05$)

Dataset	Scorer	Entropy	Random	Δ	p_{holm}
Car1	HodgeRank	0.896	0.824	+0.072	^a
	Win-Ratio	0.860	0.827	+0.033	2.1×10^{-6}
	Bayesian-BT	0.889	0.843	+0.046	3.6×10^{-7}
	Borda	0.611	0.757	-0.146	1.2×10^{-10}
Sushi-10	HodgeRank	0.923	0.841	+0.082	^a
	Win-Ratio	0.927	0.814	+0.113	1.2×10^{-82}
	Bayesian-BT	0.933	0.861	+0.072	2.3×10^{-80}
	Borda	0.628	0.764	-0.136	8.0×10^{-79}
Sushi-X	HodgeRank	0.942	0.886	+0.056	^a
	Win-Ratio	0.942	0.861	+0.081	1.8×10^{-81}
	Bayesian-BT	0.942	0.897	+0.045	1.1×10^{-72}
	Borda	0.780	0.837	-0.057	5.0×10^{-31}

^a see the 18-test family reported in Table II, in which every Entropy-vs-Random comparison is significant after Holm correction.

close to zero and Hodge decomposition adds little beyond win-counting or a Bayesian point estimate. On the denser Car1 graph, by contrast, HodgeRank holds a clear advantage over both alternatives (+0.036 over Win-Ratio, +0.007 over Bayesian-BT). This is consistent with HodgeRank's harmonic component only becoming substantial once the comparison graph carries enough cycles to require it. Fig. 4 visualizes this pattern across all three datasets.

Extending the same Wilcoxon-plus-Holm-Bonferroni procedure used for Table II to this ablation confirms that all nine Win-Ratio, Bayesian-BT, and Borda comparisons are significant, in both directions: the positive Win-Ratio and Bayesian-BT deltas and, notably, the negative Borda deltas that define the collapse described above. The Borda collapse is therefore tested against the same statistical standard as the paper's other headline comparisons; it is the largest-magnitude, most significant effect in this table on every dataset.

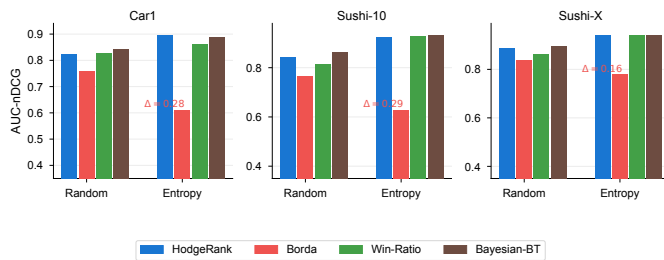


Fig. 4. Acquisition-scorer interaction for Random and Entropy; active gains persist for HodgeRank, Win-Ratio, and Bayesian-BT, while Borda exhibits the collapse.

D. HarmonicReduce: A Density-Dependent Hodge-Theoretic Finding

HarmonicReduce underperforms Random on Car1: a negative point estimate ($\Delta = -0.024$), significant under an uncorrected test (raw $p = 0.02$) but not significant once the full 18-test family is Holm-corrected ($p_{\text{holm}} = 0.08$). It shows dataset-density-dependent behavior on the sparse Sushi datasets, and this behavior differs between the two evaluation metrics. On

Sushi-10, it reaches the $\text{nDCG}@5 \geq 0.80$ threshold faster than every other policy tested (CTT 9.1% versus Random's 27.2%, a $3.0\times$ reduction). But its AUC-nDCG advantage over Random is small and not statistically significant ($\Delta = +0.008$, raw $p = 0.67$, $p_{\text{holm}} = 1$, $n = 500$), and it trails Entropy's AUC-nDCG (0.849 versus 0.923) by a substantial margin. We read this as fast early convergence to a moderately good ranking that does not translate into a reliably higher AUC-nDCG over the full trajectory on this dataset. On Sushi-X, by contrast, HarmonicReduce's AUC-nDCG advantage over Random is both large and significant ($\Delta = +0.051$, $p_{\text{holm}} = 6.5 \times 10^{-28}$, $n = 500$). It closely matches Entropy (0.937 versus 0.942, a gap of only 0.005) while also reaching threshold faster (CTT 12.3% versus 14.1%), making it the clearly favorable policy on that dataset by both metrics. This is a direct, theoretically predicted consequence of Proposition 1: the degree-weighted proxy in Eq. (3) is discriminative only when item degree varies substantially, which occurs precisely when the graph is sparse. On Car1's near-complete graphs, item degrees are nearly uniform, and the proxy degenerates to a near-constant multiplier on the residual. HarmonicReduce therefore collapses toward a noisier Entropy without Entropy's benefit. Even this underperformance, while consistent in direction across every seed and fold we tested, is only raw-significant, not Holm-significant. This underscores how close to indifferent HarmonicReduce is to Random on this dense dataset once multiple comparisons are accounted for. On the sparse Sushi graphs, degree variance is substantial, and harmonic inconsistency is concentrated on specific high-degree edges. This makes the proxy genuinely discriminative early in elicitation and explains the CTT advantage on both Sushi datasets. Whether this early advantage survives to a significant full-trajectory AUC-nDCG gain differs between Sushi-10 and Sushi-X, an open question we do not resolve here. We report this transparently, including the two non-significant results on Car1 and Sushi-10. The explanation itself, not merely a single aggregate number, is the contribution: a Hodge-decomposition-derived heuristic's practical value, and its limits, can be predicted in advance from approximate comparison-graph completeness. Its statistical reliability should be assessed metric by metric, under appropriate multiple-comparisons correction, rather than assumed to transfer from one metric or dataset to another.

Fig. 5 shows this contrast directly.

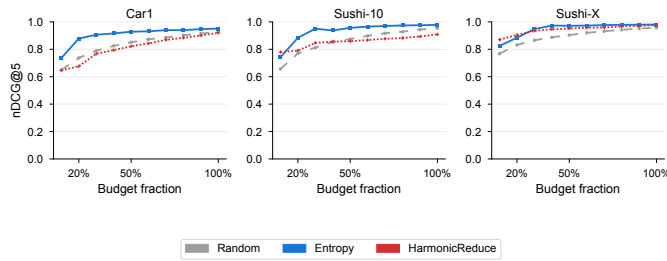


Fig. 5. HarmonicReduce compared with Random and Entropy across datasets, showing its density-dependent behavior in both early and full-budget regimes.

E. FlowClone: An Offline-Trainable Approximation of Entropy

Despite its compact five-dimensional feature representation, FlowClone’s AUC-nDCG is statistically indistinguishable from Entropy’s on all three datasets under our leakage-free, five-fold leave-users-out protocol (-0.005 on Car1, $+0.002$ on Sushi-10, $+0.002$ on Sushi-X; Table II). FlowClone is the only policy that consumes training data. We therefore additionally compared, pooled across all five folds, its AUC-nDCG on users it was trained on versus the disjoint users it was evaluated on in the same fold. The seen-versus-unseen gap, shown in Fig. 7, shrinks to near zero and changes sign across datasets (Car1: $\Delta = +0.0035$; Sushi-10: $\Delta = -0.0005$; Sushi-X: $\Delta = -0.0017$, seen minus unseen). Every gap is smaller than the corresponding standard deviation of per-user AUC-nDCG. This is read as population-level evidence that FlowClone does not overfit to its training users: what it learns transfers, within noise, to entirely unseen users in every fold and dataset tested. Fig. 6 shows the corresponding training-loss curves converge smoothly on all three datasets, with no sign of instability that would cast doubt on the trained network used for this comparison.

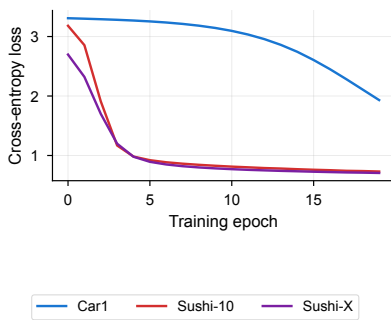


Fig. 6. FlowClone cross-entropy training loss by epoch (fold 1) for Car1, Sushi-10, and Sushi-X.

F. Robustness to the Convergence-Threshold Definition

CTT and the headline efficiency figure both depend on the choice of convergence threshold $\tau = 0.80$. We therefore swept τ from 0.65 to 0.90 and recomputed CTT for Random, Entropy, and Coverage on all datasets, pooled across all five folds (Fig. 8). The Random-to-Entropy efficiency ratio remains within $1.3 \times -2.8 \times$ across the entire tested range and

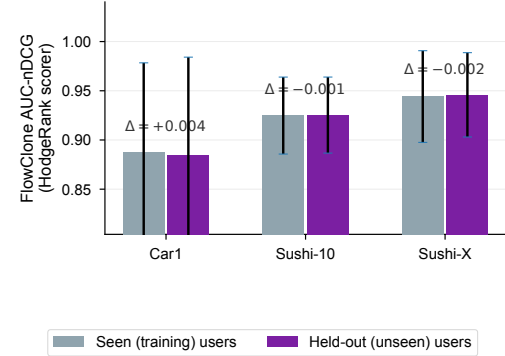


Fig. 7. FlowClone AUC-nDCG on seen (training) versus unseen (held-out) users, pooled across five folds, with per-dataset generalization gaps.

all datasets. This confirms the central efficiency finding is not an artefact of the particular threshold chosen for the headline result.

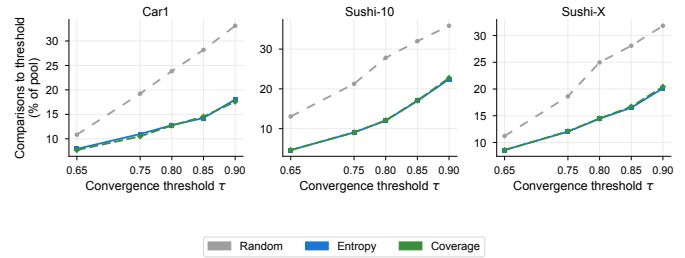


Fig. 8. CTT sensitivity to the convergence threshold τ for Random, Entropy, and Coverage across all datasets.

G. The Online Update Tracks the Batch Estimator

Section III-B frames the online update (Eq. (2)) as a stochastic-approximation scheme for the batch objective (Eq. (1)). This is verified empirically as follows: for a random (non-adaptive) query order, holding out no items, the online running score after each step is compared by Kendall τ against the batch HodgeRank solution computed over the same user’s full comparison set. Random order isolates the online/batch relationship itself from any acquisition-policy effect. Averaged over 40 randomly sampled users per dataset and 5 seeds each, τ rises monotonically with the fraction of comparisons observed and reaches 0.90 (Car1), 0.95 (Sushi-10), and 0.95 (Sushi-X) once all comparisons are incorporated, up from 0.38–0.43 at 10% of comparisons on all three datasets. The online estimate does not reach exact agreement with the batch solution even at 100% of comparisons, which is expected: the online update has no explicit ℓ_2 regularization term and follows a single online pass rather than jointly solving the weighted least-squares system, so the two estimators are not guaranteed to coincide exactly at any finite step, only to converge together as the stochastic-approximation conditions on η_t are satisfied. This confirms the online update is a genuine, increasingly faithful approximation of the batch estimator.

H. Computational Overhead of Acquisition Policies

Reporting only relative comparison savings leaves open whether an acquisition policy is itself cheap enough to deploy at the per-decision latency a live system requires. Table IV reports the wall-clock cost of the selection step alone (excluding the shared, policy-independent cost of maintaining Online-HodgeRank and OnlineBayesianBT, which every policy pays identically), averaged over 200 timed calls per candidate-pool size after 20 untimed warm-up calls, at pool sizes spanning this paper's datasets (10 items $\Rightarrow$ up to 45 candidate pairs at Carl's density) up to a synthetic 2,000-candidate pool representative of a substantially larger catalogue.

TABLE IV. MICROSECONDS PER ACQUISITION DECISION (SELECTION STEP ONLY) BY CANDIDATE-POOL SIZE

Policy	@10	@45	@100	@500	@2000
Entropy	2.8	12.3	25.4	127.4	493.1
Coverage	9.6	32.4	63.3	292.7	1720.5
HarmonicReduce	9.7	35.7	67.5	289.1	1168.1
BALD	2.3	9.5	20.0	98.2	392.4
FlowClone	35.0	100.7	201.0	908.8	4885.3
FlowClone / Entropy	12.4 $\times$	8.2 $\times$	7.9 $\times$	7.1 $\times$	9.9 $\times$

FlowClone is measured to be 7–12 $\times$ slower per decision than Entropy at every pool size tested (single-run wall-clock measurement). The exact ratio fluctuates modestly across repeated runs due to timing noise at microsecond scale, but the qualitative result is stable across runs. The measured cost does not come from the network itself, a 5-64-64-1 MLP forward pass is a handful of small matrix multiplications, but from constructing its five-dimensional feature vector (Eq. (4)) for every candidate pair, which includes a common-neighbour set intersection per candidate, the same operation that makes Coverage and HarmonicReduce several times slower than Entropy in the same table. FlowClone pays this cost per candidate in addition to, not instead of, Entropy's own residual computation, since its feature vector includes the residual as one input. This is an implementation-dependent cost of per-candidate feature construction, not an inherent property of the distilled network, and a vectorized batch implementation of the feature computation may substantially close this gap. FlowClone is not currently a faster substitute for Entropy, and its practical justification rests on distillation fidelity and cross-user generalization (Section V-E), not on deployment latency.

VI. DISCUSSION

The roughly 51% comparison-budget reduction is consistent across three structurally different datasets. This suggests a concrete deployment guideline: systems collecting pairwise preferences for HodgeRank-based recommendation should adopt gradient-residual (Entropy) or common-neighbour (Coverage) acquisition by default. Density-dependent policies such as HarmonicReduce should be reserved for settings where the comparison graph is known in advance to be far from complete. The BALD baseline confirms this advantage is not merely because HodgeRank-based policies beat weak or naive baselines: a properly implemented Bayesian active learner is also outperformed. Coverage additionally requires no trained model, making it a natural default where implementation

simplicity matters as much as query efficiency. FlowClone is attractive as a distilled approximation of Entropy that preserves near-identical ranking quality under leave-users-out evaluation, though Section V-H shows this distillation is not currently a deployment-latency win: at every pool size tested, FlowClone is 7–12 $\times$ slower per decision than the Entropy policy it approximates, so its practical justification rests on model compactness and cross-user transfer rather than on speed. These properties map onto distinct deployment scenarios. Cold-start onboarding, where a new user's item preferences must be elicited under a strict comparison budget before a first recommendation can be produced, benefits directly from the roughly halved comparisons-to-threshold shown in Table II. Real-time or high-throughput settings, where many users are elicited concurrently, benefit from Coverage's simple hand-crafted rule; FlowClone remains attractive if a compact, transferable model of Entropy's behaviour is independently valuable, for instance for on-device deployment where Entropy's own residual computation is unavailable.

The Borda collapse documents the case of an active acquisition policy that is actively harmful, not merely less beneficial, when paired with an incompatible scoring model. This has a direct methodological implication: reporting an acquisition policy's gain over random selection is insufficient without specifying, and justifying, the compatibility of the downstream aggregation model. The HarmonicReduce density-dependence finding similarly shows that a heuristic's practical value can be predicted, not merely observed post hoc, from graph-theoretic properties that Hodge decomposition makes explicit. The concurrent RLHF-motivated active preference literature operates on scalar Bradley–Terry-style reward models. Our direct implementation of a representative Bayesian Bradley–Terry active learner (BALD), which HodgeRank-based acquisition outperforms, together with our scorer ablation, suggests that acquisition-scorer compatibility deserves explicit examination rather than being assumed.

The BALD comparison points to a plausible explanation for why geometry-informed acquisition holds an advantage over a well-specified Bayesian alternative. BALD's posterior-variance criterion treats belief uncertainty as a single scalar per item. It therefore cannot distinguish a comparison that would resolve a locally inconsistent triangle (curl) from one that would resolve a globally unresolvable disagreement (harmonic) from one that would simply sharpen an already largely settled ranking (gradient). All three reduce posterior variance and are treated interchangeably. Entropy and Coverage, by contrast, operate directly on the residual of the gradient fit. This residual is itself informative about where the current consensus ranking is least supported by observed comparisons. This is consistent with the pattern in Table II: BALD's advantage over Random is real but consistently smaller than Entropy's or Coverage's, across every dataset where BALD itself is unambiguously informative.

We highlight several limitations. This paper targets the item-set scale typical of recommender systems and group decision-making (10–100 items), not the thousands-of-candidates scale of some large language model reward-modeling pipelines. Within this scale, our mechanism and theoretically grounded predictions (density dependence, scorer compatibility) are not directly validated beyond the 100-item Sushi setting. Evaluation follows a leakage-free offline-

replay protocol on previously collected comparison datasets (Section III), consistent with prior active-ranking work [15]. The HarmonicReduce density-dependence claim rests on a two-point (dense-versus-sparse) contrast, not a continuous density sweep. This is an opportunity for dedicated follow-up. FlowClone’s feature representation is intentionally compact, to test whether a handful of simple statistics suffice. A larger feature set might close its small residual gap to Entropy on Sushi-X, but would weaken the distillation claim, so we did not pursue it. Several hyperparameters, most notably γ_0 (the online update’s warm-up constant), λ (HodgeRank’s regularization strength), and the Bayesian Bradley-Terry model’s prior and noise parameters, are fixed per dataset rather than swept systematically; a targeted stability check would confirm the reported comparison-budget reductions are not artifacts of these particular settings, and is left to future work rather than attempted here for every hyperparameter family.

Several extensions follow naturally. A continuous comparison-graph-density sweep rather than the two-point dense/sparse contrast used here would more precisely characterize where the HarmonicReduce transition occurs, and test whether it follows a sharp threshold or a gradual trend, guided directly by Proposition 1. Extending to batch, rather than single-comparison, acquisition would require selecting a diverse set of informative pairs per round instead of the single greedy pair each policy currently selects; this matters for item catalogues large enough that per-comparison model updates are impractical. A systematic study of acquisition-scorer compatibility across the broader family of pairwise aggregation models used in recommender systems and group decision-making would extend the four-scorer ablation in Section V-C to rank-centrality and matrix-factorization-based aggregators. A further direction, suggested by the BALD comparison, is a hybrid acquisition rule that combines HodgeRank’s gradient residual with an explicit posterior-variance term. Such a rule could capture both the structural informativeness Entropy and Coverage exploit and the calibrated uncertainty estimate BALD provides.

VII. CONCLUSION

This paper asked whether the Hodge-theoretic structure used elsewhere for aggregation and quality assessment is also informative for deciding which pairwise comparisons to actively request. We use a leakage-free, five-fold leave-users-out evaluation protocol and seven acquisition policies, evaluated across three real-world datasets and four scoring models. Geometry-informed active acquisition (Entropy, Coverage, and its lightweight behavioural-cloning approximation FlowClone) reduces the comparison budget required to reach a fixed recommendation-quality threshold by approximately 51% relative to random sampling. It consistently outperforms a properly implemented Bayesian Bradley-Terry active learner (BALD) on all three datasets; BALD itself is a statistically confirmed improvement over random sampling on two of the three. A novel harmonic-energy-targeted policy, HarmonicReduce, exhibits a dataset-density dependence that we both observe empirically and prove exactly (Proposition 1) from the circuit rank of the comparison graph. FlowClone is validated under five-fold cross-validation, with a seen-versus-unseen generalization gap that is near zero and changes sign across datasets, indicating no detectable overfitting. Taken together, these findings

indicate that combinatorial Hodge theory provides a coherent geometric language across the pairwise-preference pipeline, from deciding what to ask to aggregating what was said.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest.

FUNDING

This research received no external funding.

DECLARATION ON GENERATIVE AI

Generative AI was used solely for language refinement and grammatical improvement. The authors take full responsibility for the final content of the manuscript.

ACKNOWLEDGMENTS

The authors thank the maintainers of the Car Preference [24] and Sushi [25] datasets used throughout this work.

REFERENCES

- [1] P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei, “Deep reinforcement learning from human preferences,” *Advances in Neural Information Processing Systems*, vol. 30, pp. 4299–4307, 2017.
- [2] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray *et al.*, “Training language models to follow instructions with human feedback,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 27 730–27 744, 2022.
- [3] R. Rafailov, A. Sharma, E. Mitchell, S. Ermon, C. D. Manning, and C. Finn, “Direct preference optimization: Your language model is secretly a reward model,” *Advances in Neural Information Processing Systems*, vol. 36, 2023.
- [4] J. Rothe, “Borda count in collective decision making: A summary of recent results,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 9830–9836, 2019.
- [5] W. Muldrew, P. Hayes, M. Zhang, and D. Barber, *Active preference learning for large language models*, ser. ICML’24. JMLR.org, 2024.
- [6] N. Das, S. Chakraborty, A. Pacchiano, and S. R. Chowdhury, *Active Preference Optimization for Sample Efficient RLHF*, R. P. Ribeiro, B. Pfahringer, N. Japkowicz, P. Larrañaga, A. M. Jorge, C. Soares, P. H. Abreu, and J. Gama, Eds. Cham: Springer Nature Switzerland, 2026.
- [7] K. Ji, J. He, and Q. Gu, “Reinforcement learning from human feedback with active queries,” *arXiv preprint arXiv:2402.09401*, 2024.
- [8] Y. Shen, H. Sun, and J.-F. Ton, “Reviving the classics: Active reward modeling in large language model alignment,” *arXiv preprint arXiv:2502.04354*, 2025.
- [9] X. Jiang, L.-H. Lim, Y. Yao, and Y. Ye, “Statistical ranking and combinatorial Hodge theory,” *Mathematical Programming*, vol. 127, no. 1, pp. 203–244, 2011.
- [10] Q. Xu, Y. Yao, X. Jiang, and Q. Huang, “HodgeRank on random graphs for subjective video quality assessment,” *IEEE Transactions on Multimedia*, vol. 14, no. 3, pp. 844–857, 2012.
- [11] E. Bejaoui, M. O.-E. Aoueilayine, R. Bouallegue, and A. Yazidi, “Hodge-based methods for contributor quality assessment in crowd-sourced preferences,” in *Advanced Information Networking and Applications (AINA 2026)*, L. Barolli, W. K. G. Seah, and I. Woungang, Eds. Cham: Springer Nature Switzerland, 2026, pp. 260–272.
- [12] B. Settles, “Active learning literature survey,” *Computer Sciences Technical Report 1648*, University of Wisconsin–Madison, 2009.
- [13] W. Chu and Z. Ghahramani, “Preference learning with Gaussian processes,” *Proceedings of the 22nd International Conference on Machine Learning (ICML)*, pp. 137–144, 2005.

- [14] N. Houlisby, F. Huszár, Z. Ghahramani, and M. Lengyel, "Bayesian active learning for classification and preference learning," *arXiv preprint arXiv:1112.5745*, 2011.
- [15] K. G. Jamieson and R. D. Nowak, "Active ranking using pairwise comparisons," *Advances in Neural Information Processing Systems*, vol. 24, pp. 2240–2248, 2011.
- [16] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. the method of paired comparisons," *Biometrika*, vol. 39, no. 3–4, pp. 324–345, 1952.
- [17] E. Bejaoui, M. O.-E. Aoueleiyine, R. Bouallegue, and A. Yazidi, "Hodgerank for group decision-making: Evaluating recommendation stability with incomplete preference data," in *2025 2nd International Conference on Innovative and Intelligent Information Technologies (IC3IT)*. IEEE, 2026, pp. 1–6.
- [18] R. Herbrich, T. Minka, and T. Graepel, "Trueskill™: a bayesian skill rating system," in *Proceedings of the 20th International Conference on Neural Information Processing Systems*, ser. NIPS'06. Cambridge, MA, USA: MIT Press, 2006, p. 569–576.
- [19] C. C. Paige and M. A. Saunders, "LSQR: An algorithm for sparse linear equations and sparse least squares," *ACM Transactions on Mathematical Software*, vol. 8, no. 1, pp. 43–71, 1982.
- [20] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, 1951.
- [21] R. Diestel, *Graph Theory*, 5th ed. Berlin, Heidelberg: Springer, 2017.
- [22] S. Ross, G. Gordon, and D. Bagnell, "A reduction of imitation learning and structured prediction to no-regret online learning," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, vol. 15, 2011, pp. 627–635.
- [23] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques," *ACM Transactions on Information Systems*, vol. 20, no. 4, pp. 422–446, 2002.
- [24] E. Abbasnejad, S. Sanner, E. V. Bonilla, and P. Poupart, "Learning community-based preferences via Dirichlet process mixtures of Gaussian processes," *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1213–1219, 2013.
- [25] T. Kamishima, "Nantonac collaborative filtering: Recommendation based on order responses," *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 583–588, 2003.