THE AMORTIZATION CEILING: WHY LEARNED POLICIES DO NOT REPLACE SEARCH IN CONSTRAINT-SATISFACTION GRAPH GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

When a combinatorial objective is expensive to evaluate, a learned policy is meant to replace the evaluation: train once, decide cheaply thereafter. We study this amortization question in a controlled testbed: generate a graph whose five structural metrics (clustering, mean path length, modularity, spectral gap, degree assortativity) jointly fall within target tolerance bands, using degree-preserving edge swaps. True-metric evaluation is costly and the constraints are coupled. We benchmark eleven learned methods in thirteen configurations (imitation, value regression, policy gradients, generative flow networks, differentiable guidance, planning, a transplanted autoregressive architecture, and LLM-evolved heuristics), eleven under one unified protocol, against references at every level of the search hierarchy. Three findings follow. (1) No residual-free policy significantly beats random at any tolerance. A learner beats random only inside a narrow, protocol-sensitive niche (unimodal in tolerance tightness, concentrating on a locally computable binding constraint, and confined to small-to-mid sizes and the training family), and only for a learner that itself pays the per-decision measurement cost. (2) A ceiling: no learned method approaches true-metric greedy at any tightness or budget, and the best recovers under a quarter of the greedy–random gap on band average, at greedy's own cost. (3) The ceiling is structural: most candidate edits (96.7%) share their endpoint-degree view with another candidate, the failure localizes to assortativity, and the joint target space is thin, with a slice certified infeasible in advance by an unconditional spectral–modularity bound. A worst-case theorem shows that locally indistinguishable edits with provably different global effects exist, so bounded-radius scorers incur constant relative regret. Here, good moves are measured, not predicted; the certificate and the collision statistic are computable before any training. We distill the boundary into a testable *binding-locality criterion*: amortization can help to the extent that the binding constraint's increment is locally computable.

1 INTRODUCTION

Amortized optimization (Amos, 2023) promises an exchange: pay training compute once, act at low inference cost thereafter. The promise is realized in some domains, such as learned value functions inside tree search (Silver et al., 2018), and fails quantifiably in others: in variational inference the amortization gap can dominate the total error (Cremer et al., 2018), and in combinatorial optimization (CO) re-evaluations repeatedly find that reported wins of learned heuristics dissolve against well-tuned classical baselines (Angelini & Ricci-Tersenghi, 2023; Böther et al., 2022). What is missing is a setting where the question can be answered *mechanistically*: not whether learning wins, but why it does not, and what property decides it. We provide such a setting.

The problem. We study constraint-satisfaction graph generation: given a target vector of five structural metrics (clustering coefficient C, mean path length $\langle \ell \rangle$, modularity Q, normalized-Laplacian spectral gap λ_2, degree assortativity r) and per-metric tolerances, produce a graph satisfying all five at once, using only degree-preserving double-edge swaps (Fosdick et al., 2018). Two features make it a controlled instrument for the amortization question. First, there is no data dis-

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.

tribution to imitate, only a target vector, so distribution-learning generators (Vignac et al., 2023) answer a different question. Second, most of the metrics are global: a swap's effect is in general knowable only by re-evaluating them, at $\sim$110 ms per candidate, which is the cost truthful search pays every step and a learned policy exists to avoid. The task itself underlies controlled network experiments that move one metric while pinning the others, and fixing the degree sequence follows null-model practice (the dK-1 null model (Mahadevan et al., 2006)). Joint control is a dynamical need rather than a convenience, because process outcomes depend on several structural quantities at once (Orsini et al., 2015): on two metric–dynamics pairs, single-metric rewiring overstates the λ_2 effect on synchronization time by an order of magnitude and understates the r effect on outbreak size by up to $\sim 40\%$, while joint control pins the confounders (Appendix M).

A niche with coordinates. Against random, true-metric greedy and beam, microcanonical Metropolis–Hastings (MH) and simulated-annealing (SA) samplers, and classical dK null models, we benchmark eleven learned methods in thirteen configurations, eleven under one unified paired protocol (§3). Learning helps only for an oracle-augmented value learner, inside a region we map along four axes (§4, §7): *tightness* (unimodal, closing at both tails), *binding composition* (assortativity, closed-form in the endpoint degrees, among the binding metrics), *scale* (small-to-mid instances), and *distribution* (no transfer to an unseen verified-feasible family, while greedy transfers).

The ceiling and its structure. The central finding is a ceiling: no learned method, at any tightness or evaluation budget, approaches true-metric greedy; only deeper true-metric search (beam) exceeds it (§5). We trace the ceiling to the problem's structure (§6) and distill the boundary into the *binding-locality criterion*, a hypothesis we test three ways (§8):

Definition 1 (Binding-locality criterion)**.** *A constraint-satisfaction edit process is amortizable to the extent that the increment of its* binding *constraint is computable from the edit's local neighborhood.*

Contributions.

1. **A four-axis map of where an oracle-augmented learner beats random** (Findings 1–2, 12): no residual-free policy (one with no true-metric recomputation per candidate) does so significantly at any band; the window belongs to the value learner, which true-evaluates each candidate edit and pays greedy's per-decision cost (Finding 5). Its window is unimodal and protocol-sensitive (p=0.007 after selection correction), concentrates where the binding constraint is locally computable (31-subset intervention), and has scale and distribution boundaries.
2. **A unified negative result** (Findings 4–7): every learned configuration sits below true-metric greedy, the strongest learner buys no cost advantage, and the ceiling is robust across tested architectures, modalities, tuning, and framing.
3. **A structural explanation** (Findings 8–11). Collision statistics that complement a worst-case indistinguishability theorem with an average-case diagnostic, a state-value/action-ranking dissociation, localization to assortativity with a binding-fraction mechanism, and an unconditional Q–λ_2 exclusion bound with a partition-uniform certificate.
4. **Pre-training diagnostics and a reproducible benchmark**: the certificate and the collision screen run before any training; protocol, code, target sets, full results.

2 Related Work

Deep graph generation and property control. Deep graph generative models span autoregressive (You et al., 2018), diffusion (Vignac et al., 2023), and flow-matching (Qin et al., 2025) families, with growing attention to conditioning: metric-conditioned latent diffusion (Evdaimon et al., 2024), plug-and-play projection (Sharma et al., 2024), hard structural-class preservation (Madeira et al., 2024) (survey: Guo & Zhao, 2023). Three mismatches separate that agenda from ours: those models learn a corpus distribution, constrain by conditioning or projection rather than joint metric bands, and preserve no degree sequence. Recent work steers pretrained graph diffusion to target properties by injecting *tree search* at inference (Zhao et al., 2026) — independent evidence that property control benefits from search.

Null models and microcanonical samplers. Fixed-degree ensembles are a cornerstone of network science: the dK hierarchy, exponential random graph models (Robins et al., 2007), and double-edge-swap MCMC (Maslov & Sneppen, 2002; Fosdick et al., 2018). Our tolerance-band satisfaction is microcanonical (Squartini et al., 2015); we measure how far the dK nulls fall short of joint five-metric control. Prior attainability results cover single metrics or pairs (Cinelli et al., 2020; van der Hofstad et al., 2020), dK-2.5 targets clustering atop the joint degree distribution (Gjoka et al., 2013), and standard benchmarks control only planted communities atop prescribed degrees (Lancichinetti et al., 2008). Our certificate gives a joint-space attainability test.

Reality checks for learning-based optimization. We join a lineage of re-evaluations: greedy beating physics-inspired graph neural networks (GNNs) on maximum independent set (Angelini & Ricci-Tersenghi, 2023), GNN-guided tree search owing its gains to classical components (Böther et al., 2022), systematic TSP evaluations (Liu et al., 2023), and baseline-inflation critiques (Nath & Kuhnle, 2023). Flagship neural-CO systems approach but do not surpass strong classical search (Sun & Yang, 2023), with smooth exchange between training compute and test-time search (Jones, 2021). Boundary markers exist on both sides: action-value distillation of a search oracle succeeds in chess (Ruoss et al., 2024), where board features discriminate moves, while strong-branching imitation fails when the expert's scores stop discriminating (González-Rodríguez et al., 2025) — the small-action-gap regime (Bellemare et al., 2016). We add a thirteen-configuration benchmark (eleven under one unified protocol), a mechanism placing this problem on the failing side, and a map of the narrow region where an oracle-augmented learner beats random.

Learned graph editing and rewiring. Reinforcement learning (RL) that edits graph structure toward a global objective is an active thread (Darvariu et al., 2024); two works are closest. ResiNet (Yang et al., 2023) optimizes resilience with the *same* action space, compares against classical search (hill climbing, greedy, annealing, evolutionary), and reports that GNNs extract little signal from pure topology; but it targets one scalarized objective, not joint tolerance bands. DMGG (Choi et al., 2026) reaches a *single* target metric (assortativity) with RL over degree-preserving swaps (one of its rewiring modes) and generalizes zero-shot, but it benchmarks only against samplers, never greedy. We connect the two: we verify DMGG's architecture on its own task, show that the transplanted architecture collapses on the joint objective (§5), and explain the dissociation. With fixed degrees, Δr is an exact local function of four endpoint degrees (Lemma 1), so single-metric rewiring has dense local signal, whereas the joint objective admits no such proxy.

GNN expressiveness and amortized optimization. Message-passing limits are mature (Weisfeiler–Leman equivalence (Xu et al., 2019), depth–width bounds for global quantities (Loukas, 2020), girth- and diameter-type properties beyond message passing (Garg et al., 2020)). Our theorem adapts these symmetry principles to *ranking two edits of one graph*, a regret statement for the rewiring Markov decision process (MDP). Amortization conditions are sharpest in variational inference (Margossian & Blei, 2024). We transplant that frame to generation-as-search, with tightness as the control variable; the horizon router is per-instance algorithm selection (Kerschke et al., 2019).

3 Problem Setup and Protocol

Task. An instance is a target vector $c^\star = (C^\star, \langle \ell \rangle^\star, Q^\star, \lambda_2^\star, r^\star)$ with per-metric tolerances $\varepsilon = (0.05, 0.30, 0.05, 0.03, 0.05)$ and a band tolerance u. Starting from a rule-based, target-aware warm start (a Watts–Strogatz (WS) construction matched to the target degree regime (Watts & Strogatz, 1998), followed by rule-based refinement passes, Appendix A; shared verbatim by all swap-based methods), the generator applies degree-preserving double-edge swaps $(a, b), (c, d) \to (a, d), (c, b)$ under a simplicity mask (connectivity: Appendix A). The *worst-scaled residual* is $W(G) = \max_k |m_k(G) - c_k^\star| / \varepsilon_k$, and the instance is *satisfied at band* u iff $W(G) \le u$, so smaller u is tighter; the metric attaining the maximum is the *binding* constraint at that state. We sweep the 12-band grid $u \in \{1.0, \ldots, 0.1\}$, extended to $u \in \{1.2, 1.5, 2.0\}$ where the loose tail is at issue. The five metrics are standard top-level axes of structural characterization (da F. Costa et al., 2007) (rationale in Appendix C). Unless stated otherwise, N=200, $\langle k \rangle$=6, each episode is T=60 swaps, greedy scores K=8 candidates per step, targets are held out from training, and entries are means over 10 seeds (beam: 6). Headline deltas carry standard errors (SE), paired-bootstrap confidence intervals (CIs), and paired t-tests at the seed level, with targets averaged within each seed (Appendix A).

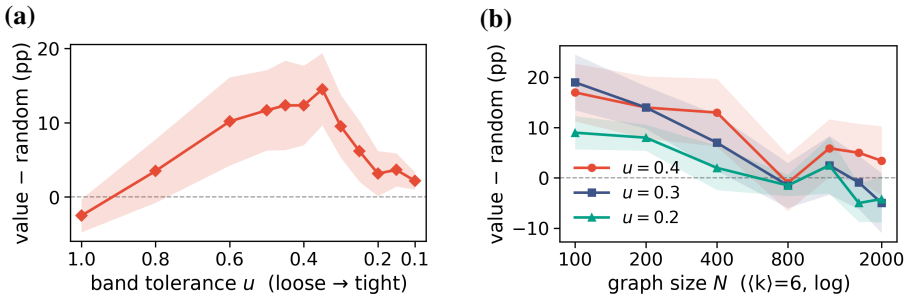


Figure 1: **Where the value learner beats random.** *(a)* $\text{value}_{\text{PE}}-\text{random}$ satisfaction gain vs. band tolerance (N=200, $\langle k\rangle$=6, 10 seeds; shading ±SE). *(b)* The same gain vs. instance size at three bands ($\langle k\rangle$=6, 10 seeds; mean ± SE; $N\geq$1200 uses accelerated metrics, Appendix A).

Reference methods. **random**: a uniform draw from the shared candidate pool (App. A). **greedy**: at each step, evaluate the *true* metrics of K candidate swaps drawn from the whole graph (sampler in Appendix A) and keep the best (weighted residual potential), an exact one-step search at ∼110 ms per candidate. **beam**: multi-step true-metric search. **MH/SA**: microcanonical samplers with true residual as energy. dK**-1**/dK**-2**: nulls preserving the degree sequence and joint degree distribution (Mahadevan et al., 2006).

Learned approaches (eleven methods, thirteen configurations). **Imitation**: behavior cloning (BC) of greedy and its online variant DAgger (Ross et al., 2011). **Value**: regression on true residuals, plain, with a spectral positional encoding (PE) recomputed on each candidate graph, or with per-step-recomputed spectral endpoint features. **Policy gradient**: REINFORCE, proximal policy optimization (PPO), self-critical sequence training (SCST), and a metric-count curriculum (coincides with SCST on this grid). **Flows**: generative flow networks (GFlowNets) (Bengio et al., 2023) with a strengthening study. **Surrogates**: differentiable-surrogate guidance, all five surrogates validated (Tsitsulin et al., 2023), and surrogate model-predictive control (MPC). **Transplants**: DMGG's factorized autoregressive policy on the five-metric objective, and swap-scoring heuristics evolved by a large language model (LLM) under Evolution of Heuristics (EoH) (Liu et al., 2024). Definitions, architectures, and budgets (matched in environment interactions) are in Appendix D, wall-clock costs appear separately (§5), and Appendix B maps each Finding below to its evidence.

4 Where an oracle-augmented learner beats random

Finding 1 (Axis 1: a unimodal tightness window). *The benefit of the oracle-augmented value learner over random is unimodal in tolerance tightness (zero at loose bands, peaking in a mid-tight window, closing again at extreme tightness) and is significant after selection correction.* (i) At loose bands the problem is one of satisficing (any in-band graph suffices): random swapping saturates, greedy−random is +0.0 at $u \geq 1.2$, and any intervention risks steering an unsatisfied start away from a band random drift would enter. At extreme tightness, feasible volume collapses and only true-metric search survives. (ii) In between, the strongest learner family (value regression, plain or with a spectral PE, statistically indistinguishable) beats random across a mid-tight window peaking at +12.0–14.5 percentage points (pp; Figure 1a, CIs in Figure 2, paired statistics in Table 5). (iii) With band-and-variant selection folded into the null, an exact max-statistic sign-flip permutation test over all 24 band×variant statistics gives p=0.007. An independent regime-grid run replicates the window (+14.0 ± 5.3 pp at u=0.4), three re-draws with 12 targets per seed are indistinguishable from zero, and a random-effects pooling over the five runs gives +6.2 ± 3.5 pp (Appendix E), so the window is real but protocol-sensitive. We report the remaining axes descriptively. This learner pays greedy's per-decision cost (Finding 5): the niche is a margin over *random*, not a cost-competitive alternative to search.

Finding 2 (Axis 2: binding composition, not count). *The niche concentrates where the binding constraints include assortativity, the locally computable metric: composition, not count, decides it.* (i) Across all 31 metric subsets (strongest learner retrained per subset), subsets containing r show

Table 1: **Unified benchmark** (satisfaction %, WS held-out targets, N=200, $\langle k \rangle$=6; 10 seeds, beam 6; representative bands; full grid and statistics in Appendix E). Bold: best learned configuration per band among the rows sharing the random floor; beam is the overall best in every band. The value rows score 6 candidates per step (greedy: 8). †separate unified-protocol run with its own paired random baseline (Table 12).

Method	u=1.0	u=0.6	u=0.4	u=0.3	u=0.2	u=0.1
random (floor)	92.3	57.5	27.7	13.5	5.3	0.0
BC	84.3	57.3	27.8	14.0	5.8	0.5
DAgger	91.0	59.3	29.8	13.0	2.8	0.0
value	**91.3**	**69.2**	39.2	**23.5**	**10.8**	1.2
value$_{PE}$	89.8	67.7	**40.0**	23.0	8.5	**2.2**
SCST	81.0	60.7	31.7	17.0	7.2	0.0
PPO	71.0	44.2	24.3	13.8	6.3	0.5
REINFORCE	76.0	45.8	25.8	13.7	3.8	0.0
GFlowNet†	82.9	48.3	25.0	14.0	4.6	0.1
MPC†	86.7	62.0	34.7	19.4	7.4	0.7
DMGG-factorized†	87.5	64.2	31.7	20.8	5.0	2.5
greedy (true-metric)	94.8	88.3	74.2	64.3	45.5	19.3
beam (true-metric)	95.0	92.5	83.3	73.3	64.2	43.3

a positive mean peak niche at every subset size and subsets without it a smaller one; the contrast survives normalizing by available headroom, and the niche tracks headroom only among r-subsets (correlation $+0.51$ vs. $+0.06$; Appendix F). (ii) Greedy leads in every subset (strictly in 351 of 372 band cells, with 18 ties and three exceptions of at most $+0.9$ pp, all in r-free subsets), so no constraint reduction opens a learning route past true-metric search. (iii) Tolerance re-weighting alone does not establish it: of four profiles each halving one metric's tolerance, none retains a significant niche at the comparison band u=0.4 (spectral-tight, nominally significant over 10 seeds, is $+6.3 \pm 3.6$ pp, p=0.100, over 20 seeds; Table 7), so the composition evidence carries the claim. The subset evidence places the niche on whether r is constrained at all (mechanism in §6); the criterion of Definition 1, if it holds, is one of *presence*. Axes 3–4 (scale, distribution) are established in Finding 12: the niche is confined to small-to-mid sizes and the training family.

Finding 3 (A predictor of search payoff, and its limited yield). *A horizon statistic predicts where search pays, and long-horizon state values are learnable, but routing on the horizon saves only ∼10% of evaluation cost.* (i) Across four task variants run under a separate protocol, the *horizon* from warm start to target predicts the greedy-over-random gap with correlation 0.989 (Appendix G), while improvement density and one-step predictability do not. Predictability extends to long-horizon *state* values (Finding 9). (ii) This predicts *whether* search will pay before it runs, but does not substitute for it: a router gating instances between low-cost random and expensive greedy on predicted failure preserves greedy-level satisfaction yet saves only ∼10% of evaluation cost over the quality-preserving bands (Appendix G).

5 The amortization ceiling

Finding 4 (The amortization ceiling). *No learned method, at any tightness, approaches true-metric greedy. The best recovers 23.5% of the greedy–random gap on band average.* (i) True-metric greedy dominates every learned approach at every band, with paired significance for the headline comparisons (Figure 2, Table 5). Imitation, policy-gradient, flow, and planning methods sit at or near the random floor; no residual-free swap policy exceeds the trajectory-random floor significantly at any band (largest on the shared floor $+6.7 \pm 3.3$ pp, SCST at u=0.45, p=0.08). (ii) Differentiable guidance, a one-shot generator, is -26.3 pp against its paired one-shot random baseline at u=1.0 and at or below the trajectory-random floor everywhere (Appendix H). (iii) Classical nulls are lower still (dK-1 satisfies at most 10.8% of the joint band, dK-2 at most 15.8%; Appendix H), giving the hierarchy dK nulls < random < samplers < greedy < beam.

Finding 5 (The strongest learner buys no cost advantage). *The cheap-and-good quadrant is empty: the strongest learner pays greedy's per-decision cost for lower quality, residual-free policies are low-cost but sit at the random floor, and no budget accounting, matched horizon, or learned screen reverses the ordering.* (i) One true-residual evaluation costs 111.5 ms and dominates every

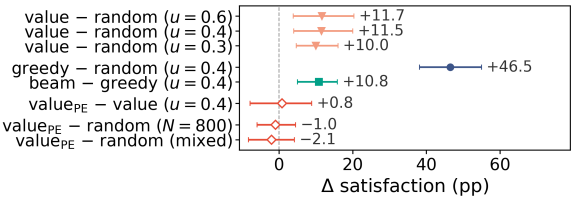


Figure 2: **Headline paired deltas at a glance** (seed-level paired means, 95% bootstrap CIs, pp; 10 seeds, beam 6, mixed family 4; filled markers: CI excludes zero; open: CI includes zero; marker color and shape distinguish the comparison groups). Groups, top to bottom: the learner's niche over random; the search references; the spectral-PE effect and the niche outside its region (N=800 and mixed family at u=0.4; protocol in Appendix A). Exact values, SEs, and p-values in Table 5.

budget. The value policy feeds each candidate's true post-swap residual to a learned lookahead (Appendix D), so it pays greedy's cost per candidate for lower quality (scoring 6 candidates per step to greedy's 8, Appendix A); residual-free policies run at most ∼18% of the cost, at the random floor (Table 10). (ii) Budget accounting agrees: true-evaluating 6 candidates per step, the value policy reaches 35.0% at ≈248 calls where greedy at K=8 reaches 81.7% at 172; greedy at every $K \leq 8$ dominates it on both axes, and larger K buys satisfaction for more calls. (iii) A longer horizon does not reverse it either: residual-free policies decide at least ∼5.6× faster by component costs, and a 5.7× longer horizon (T=340) adds at most +1.3 pp averaged over the tight window, on par with random's +0.8. (iv) The standard screen-then-verify hybrid fails too: ranking 64–256 candidates with a residual-free scorer (BC, SCST, or the LLM-evolved function) and true-evaluating its top 8 lands 19–45 pp *below* plain greedy at the same 8 calls per step at u=0.4 (all p<0.05), where an oracle screen of the same pool adds +13 to +26 pp over u=0.4–0.2: the headroom exists and no learned scorer captures it (Appendix H).

Finding 6 (Deeper search beats greedy, distilling it does not). *Deeper true-metric search exceeds greedy, while no amortized method reaches even the one-step reference, and distilling that deeper search into a policy lands back at the random floor.* (i) Beam, deeper true-metric lookahead with no learning, exceeds greedy by +10.8 pp at u=0.4 (Figure 2) and more than doubles it at the extreme band (unpaired; Table 4): the reference is a function of search depth and, in part, scoring objective (beam scores by the worst-scaled residual, Appendix A). (ii) Distilling beam into a policy lands back at the random floor: the obstacle is not distribution shift but the absence of a learnable ranking signal (§6).

Finding 7 (The ceiling is robust across tested architectures, modalities, tuning, and framing). *Architecture transplant, LLM-evolved heuristics, strengthened GFlowNet training, and a sampler framing all leave the conclusion unchanged* (four probes; Appendix H). (i) *Architecture*: the released DMGG codebase reproduces 100% held-out satisfaction on its native single-assortativity task here, yet the transplanted architecture (uniform candidate pool) on the five-metric objective is significantly below its paired random baseline at loose bands, indistinguishable at every tighter band, and far below greedy (31.7 vs. 83.4 at u=0.4 in that paired run). (ii) *Modality*: swap-scoring heuristics evolved by a 32B-parameter code LLM raise the rate at which their top pick lands in greedy's true top three by +0.005, converging to a low-cost assortativity term. (iii) *Tuning*: the standard sparse-reward prescriptions for GFlowNets leave the best configuration 6.4 pp below random, with reward mass collapsing at tight bands. (iv) *Framing*: against samplers, MH beats SA and random but collapses at tight bands, greedy exceeds MH by up to +47 pp, and a sampling-frame generator fails to beat even MH.

6 Why the Ceiling Exists: Signal, Binding, and Geometry

6.1 Local Information Cannot Rank Candidate Edits

Finding 8 (No low-cost local ranking signal). *The one-step improvement is unpredictable (R^2<0.01) from the local features available to a low-cost policy, and locally indistinguishable candidates dominate the benchmark distribution.* Candidates are *proposed* with a spectral bias (Appendix A); deployed policies also see a spectral PE of the current graph (Appendix D), and the

diagnostic asks whether local candidate features can rank them. (i) Regressing the one-step change of the worst-scaled residual on those features yields $R^2 = 0.009$ along optimization trajectories (Appendix J.3). (ii) Collision statistics (candidates with identical local view) point to the source, and show that the limitation is distributional: among 16,000 candidate swaps drawn from WS optimization states, 96.7% share their endpoint-degree view with another candidate in the pooled sample (70.7% of single states contain such ties with different true improvements; 33.7% of candidates on a separate 14,400-candidate greedy-trajectory set), and clustering candidates by their local features (endpoint degrees, Δr, Fiedler endpoint components; 40 k-means clusters) leaves 93.1% of the improvement variance *within* clusters. Both persist across the warm-start heterogeneity sweep (collision 0.72–0.98 across rungs, within-cluster variance 0.88–0.95 on the WS and power-law rungs, Appendix I). (iii) A 40-cluster selector on those features (fit in sample) still forfeits 0.20 of the attainable per-step improvement.

Scale and architecture do not recover the signal. A supervised signal-recovery study separates "does the signal exist?" from "can RL find it?". Graph-transformer regressors (Rampasek et al., 2022) trained to predict true per-swap improvements across model-size (15k–24.7M) and dataset-size grids saturate at $R^2 \approx 0.40$ with data, with weak ranking quality at the optimum. Tuned learning rates and parameter-proportional budgets do not lift it above a 0.24–0.29 test plateau, below the 0.40 saturation of the data-scaling grid, and Laplacian encodings add nothing (Appendix I). The saturation level is likely attenuated by label noise: the labels inherit the relabeling sensitivity of the Clauset–Newman–Moore (CNM) modularity term (Clauset et al., 2004) (per-swap noise variance on the order of the signal variance), and re-collecting a matched cell under the less noisy Louvain scorer (Blondel et al., 2008) (trajectories change too) raises held-out R^2 from 0.10 to 0.26. Greedy uses the same noisy oracle, so this qualifies the diagnostic, not the comparison. Per-step-recomputed spectral features carry a real but tiny ranking signal that fails in closed-loop control: the resulting policy ties random and loses to greedy by 66 pp on this probe.

Finding 9 (State values are predictable, action ranking is not). *Long-horizon state values are increasingly learnable with horizon, yet ranking adjacent edits remains unpredictable.* The dissociation of §4 sharpens: long-horizon cumulative improvement, a *state* value, grows predictable with horizon (R^2 0.06 at h=1 to 0.83 at h=50 with global spectral features; Appendix I), because aggregate progress is smooth in global structure once per-step noise averages out. But adjacent candidates differ by one double-edge swap and their predicted values nearly coincide, so choosing *among* them is the unpredictable ranking problem of Finding 8.

6.2 Failure localizes to assortativity, and the mechanism

Finding 10 (Localization to assortativity, via binding). *The learning failure concentrates on assortativity, and the mechanism is binding: when r binds, an edit's effect is its local Δr until a non-local metric becomes active, and the ranking then turns on the metrics that must be preserved.* Across the 31 constraint subsets, the imitation learner captures 11.8% of the greedy–random headroom when r is present versus 50.4% over the r-free subsets (9-seed capture ablation, Table 8), while the strongest value learner's niche concentrates in r-subsets in the 10-seed sweep (Appendix F). Consistently, assortativity is the standard convergence diagnostic for swap Markov-chain Monte Carlo (MCMC) because it mixes slowly (Dutta et al., 2025), and degree sequences tightly constrain attainable mixing (van der Hofstad et al., 2020). That Δr is the one increment closed-form in the endpoint degrees (Lemma 1; ΔC is computable from the radius-2 ball at $O(d^2)$ cost, while $\langle \ell \rangle$, Q, and λ_2 are global) looks paradoxical, and *binding* resolves it: when r binds the worst-scaled objective, an edit's effect is its local Δr only until a non-local metric becomes active; the ranking then turns on the metrics that must stay in band, not on r itself. The evidence for binding: (i) In the 31-subset intervention, the niche concentrates in subsets containing r (r-free niches do not track headroom). (ii) In the binding-fraction decomposition, the share of steps on which r binds rises with N, and local features are 5–20× more predictive on exactly those steps. (iii) The increment itself is closed-form:

Lemma 1 (Exact assortativity increment; standard, cf. Van Mieghem et al., 2010)**.** *For a non-regular degree sequence, the swap $(a,b),(c,d) \to (a,d),(c,b)$ changes Newman's assortativity by exactly $\Delta r = (d_a - d_c)(d_d - d_b)/(m\,\sigma_e^2)$, where $\sigma_e^2 > 0$ is the edge-perspective degree variance.*

6.3 THE TARGET SPACE IS STRUCTURALLY THIN

Finding 11 (A thin, coupled target space). *The five metrics are not independent axes: the sampled WS target ensemble has effective dimension ≈ 2, 99% of independently sampled targets are unreached by any tested solver under the benchmark degree sequence, and a thin slice is certified infeasible in advance by an unconditional Q–λ_2 bound.* We prove (Appendix J) the exclusion bound: for every partition with $s = \sum_c p_c^2 < 1$,

$$\lambda_2(G) \;\le\; \frac{1 - Q - s}{1 - s} \qquad \text{and} \qquad \lambda_2(G) \;\le\; 2\,(1 - Q) \tag{1}$$

(for $Q \ge 0$ the first bound is the sharper one and implies $\lambda_2 \le 1 - Q$), verified exhaustively on all connected atlas graphs (771,016 partitions, zero violations). The bound holds for *every* partition, hence verbatim for any heuristic community-detection output. The resulting band-infeasibility certificate $([\lambda_2^\star - (1-Q^\star)]_+/(u(\varepsilon_\lambda+\varepsilon_Q)) > 1$ certifies the target unreachable whenever $Q^\star - u\varepsilon_Q > 0$, which every target and band here satisfy) is valid exactly as benchmarks report Q (Appendix J). Empirically: (i) the sampled WS target ensemble has effective dimension ≈ 2, and 99% of 300 targets sampled independently and uniformly over each metric's marginal range are unreached by a union of solvers under the benchmark degree sequence (N=200, $\langle k\rangle$=6). (ii) The certificate fires in advance on 10% of that box at u=1.0 (15% at u=0.4) and never on a reached target, but explains only 10–12% of unreached targets across the full 747-target pool; most of the wall is uncertified: unreached by every tested solver, not proven infeasible (Appendix K). (iii) Within the tested budgets and solvers the wall does not move: beam recovers one of 26 greedy-unreached targets, a 10^5-swap simulated-annealing re-search of 100 unreached targets ($\sim$666$\times$ the ladder's SA budget) recovers 3% (95% CI 1–8%, Appendix K), richer actions expand feasibility only from 17% to 19%, and widening the oracle to a solver union adds one target.

6.4 A WORST-CASE THEOREM FOR THE AVERAGE-CASE FAILURE

The average-case collisions of §6.1 have a worst-case counterpart.

Theorem 1 (Local-action indistinguishability, informal, with the formal statement and proof in Appendix J)**.** *For every radius ρ and degree $d \ge 4$ there are arbitrarily large d-regular graphs G and two valid degree-preserving swaps s_1, s_2 such that (i) the marked ρ-neighborhoods of s_1 and s_2 are isomorphic; (ii) s_1 halves the conductance of G while s_2 leaves it unchanged. Consequently every radius-ρ scorer (including every $L \le \rho$-layer message-passing GNN without global or positional features) assigns them equal scores. When choosing between the pair, the induced policy with symmetric tie-breaking succeeds with probability $\frac{1}{2}$ on a separating conductance band versus 1 for the true-metric oracle, a relative regret of $\frac{1}{2}$.*

The construction places two expander copies behind four bridges: rewiring two bridges halves the bottleneck, rewiring two interior edges leaves conductance unchanged, and high girth makes every bounded-radius view an identical regular tree. The theorem is a worst-case existence result, since *exact* marked-ball isomorphism collisions are rare on benchmark states (0.1% at ρ=1, 0% at ρ=2, Appendix I), and the average-case evidence is the coarse-view collision statistic of §6.1 (96.7% pooled), a diagnostic of local information, not of the deployed scorers' inputs. The rigorous separation is for conductance (λ_2: correct order proven, strict constant open); scorers with global inputs are probed empirically (Appendix J).

7 GENERALIZATION AND EXTERNAL VALIDITY

Finding 12 (The ceiling travels, the niche is bounded on every axis). *Across a 20-fold size range, three degree regimes, and a feasibility-verified unseen target family, greedy dominates random and the strongest learner everywhere, while the niche becomes indistinguishable from zero at large N and fails to transfer across families.* (i) *Regimes*: across $N \in \{100, \ldots, 2000\}$, $\langle k\rangle \in \{4, 6, 8\}$, and the mixed target family, greedy dominance is uniform (Table 13; Appendix K). (ii) *Scale*: the niche's gains at $N \le 200$ (double-digit at $\langle k\rangle$=6) become indistinguishable from zero by N=800 at the benchmark degree (Figures 1b and 2). At $\langle k\rangle$=8 the niche stays positive at u=0.4 (+3 to +9 pp) while at u=0.3 it is sign-unstable for $N \ge 1200$ (Figure 6); no large-N cell survives family-wise correction. The greedy–random gap itself compresses at large N, but band-selectively: its peak migrates toward tighter bands while the gap at the tightest band does not shrink,

so search dominance concentrates exactly where the niche vanishes (Appendix K). (iii) *Distribution*: the niche fails to transfer to the mixed family even though every target in it is verified feasible (Figure 2). Leave-one-family-out is most direct: trained on two families and evaluated on held-out feasibility-filtered stochastic-block-model (SBM) targets, greedy generalizes ($+19$ to $+42$ pp over random, all $p<10^{-3}$) while the learner sits at or below random everywhere and significantly below at $u=0.6$. A filter-free own-degree-sequence protocol reproduces both conclusions (Appendix K). This is consistent with the binding-locality criterion: the transferable part of a policy is the part with local signal, absent here.

Real-world targets. On target vectors from 17 classic real networks, a greedy probe from a WS warm start at each network's own N and $\langle k\rangle=6$ reaches none of the 15 within its size range (Appendix L). The binding constraint is $\langle \ell \rangle$ for 9, C for 4, and λ_2 for 2; the certificate fires on none (uncertified, §6.3). A larger-budget run at native mean degree reaches two (word-adjacency 100%, football 90% at $u=1.0$), on which greedy degrades gracefully with tightness while learned policies fall to zero. Warm-starting instead from each network's *own* degree sequence rescues three networks (58–98%) and partially a fourth, while seven remain at 0% (Figure 5b, Appendix K).

Warm-start heterogeneity control. Sweeping warm-start degree heterogeneity to real-network levels (σ_e^2 $2 \rightarrow 355$; targets feasible by construction; protocol fixed), the fraction of greedy-rollout steps on which r binds falls $62\% \rightarrow 8\%$, consistent with the inverse-variance scaling of Lemma 1, yet greedy dominance ($+12$ to $+30$ pp at every rung) and the null niche are invariant (Figure 5a, Appendix K). The ceiling is therefore not an artifact of the near-regular WS ensemble: it persists whether r binds on 62% or 8% of steps.

8 Discussion

The problem is hard where amortization cannot reach: the objective can be measured on an edit at a fixed price but not predicted without measuring, and the edits to be ranked are locally alike. A policy trained to convergence, scaled, or re-framed does not close the gap: what it lacks, on our evidence, is measurement, a property of the constraint set, not of the learner.

What would it take for learning to win here? The natural remedies fail: larger models (Appendix I), alternative architectures, regimes, and paradigms (Finding 7), and low-cost global features (§6.1). The one opening is the state-value/action-ranking dissociation: predictable long-horizon values can prioritize *instances* and *budgets* (Finding 3) but not select *moves*.

Which constraints are amortizable? Single-metric assortativity is learnable and transfers zero-shot; the joint five-metric objective is neither. The positive pole of Definition 1 is established for r only (Lemma 1): ΔC is also computable from a radius-2 ball, yet C-binding subsets show no consistent niche (Appendix F), so locality appears necessary but evidently not sufficient. The criterion is therefore a hypothesis: of three direct tests, only the 31-subset intervention supports it, while the assortativity-tight profile and the warm-start sweep are null (Appendices E and K). For another constrained-generation domain, the criterion gives a recipe. (1) Screen targets with the certificate first: unverified multi-metric benchmarks can confound method quality with target attainability (99% of independent draws are unreached here). (2) Measure local-feature R^2 and collision statistics on one-step improvements: if the binding increment is not locally computable, amortization is unlikely to pay, and locality alone is, on our evidence, not sufficient. (3) Otherwise spend the budget on true-metric search, allocated across instances by predicted state values.

9 Limitations

(1) Conclusions are for the degree-preserving swap space and worst-scaled five-metric objective; richer action spaces were probed, not exhausted; the warm start is controlled (§7); the cost constant is implementation-dependent. (2) The signal diagnostics are average-case statements for the tested features and distributions (worst-case scope: §6.4). (3) The soft random-geometric-graph (soft-RGG) leave-one-family-out fold is uninformative: no method reaches a soft-RGG target (Appendix K). (4) Large-N ($\geq$ 1200) cells use accelerated metrics and reduced budgets, so absolute values are not comparable; all conclusions concern one problem domain.

Reproducibility statement

Protocol details, method definitions and budgets, and complete per-experiment tables are in Appendices A–N. Code, target sets with train/evaluation splits, per-cell result files, and the verification scripts for the theoretical identities and constructions (atlas checks of equation 1 under exact and heuristic partitions, machine-precision checks of Lemma 1, and the gadget of Theorem 1) are provided as supplementary material.

AI use statement

In this work, we used generative AI tools to implement methods (the experiment code). We have not used generative AI tools to help develop theoretical models or conceptual frameworks, to formulate mathematical claims, to provide critical ingredients for proving mathematical claims, to assist in the writing of proofs, to propose or refine hypotheses, to design or provide feedback on research methodology or experiments, to support qualitative and thematic data analysis, or to interpret results; generating synthetic data sets and cleaning or reformatting datasets are not applicable to this work. Additionally, we used generative AI tools to manage the experiment schedule, to refine the design of existing figures (layout, fonts, and legends; no figure content was generated), to assist with literature retrieval and the discovery of related work (all cited works were read, verified, and selected by the authors), to assist with translation, and to edit the manuscript's English for readability. We have reviewed all AI-assisted work: AI-generated code was reviewed and tested by the authors, and the verification scripts released with the code check the stated identities and constructions. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

References

Brandon Amos. Tutorial on amortized optimization. *Foundations and Trends in Machine Learning*, 16(5):592–732, 2023.

Maria Chiara Angelini and Federico Ricci-Tersenghi. Modern graph neural networks do worse than classical greedy algorithms in solving combinatorial optimization problems like maximum independent set. *Nature Machine Intelligence*, 5(1):29–31, 2023.

Marc G. Bellemare, Georg Ostrovski, Arthur Guez, Philip S. Thomas, and Rémi Munos. Increasing the action gap: New operators for reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2016.

Emmanuel Bengio, Moksh Jain, Maksym Korablyov, Doina Precup, and Yoshua Bengio. Flow network based generative models for non-iterative diverse candidate generation. In *Advances in Neural Information Processing Systems*, 2021a.

Yoshua Bengio, Andrea Lodi, and Antoine Prouvost. Machine learning for combinatorial optimization: A methodological tour d'horizon. *European Journal of Operational Research*, 290(2): 405–421, 2021b.

Yoshua Bengio, Salem Lahlou, Tristan Deleu, Edward J. Hu, Mo Tiwari, and Emmanuel Bengio. GFlowNet foundations. *Journal of Machine Learning Research*, 24(210):1–55, 2023.

Federico Berto, Chuanbo Hua, Junyoung Park, Laurin Luttmann, Yining Ma, Fanchen Bu, Jiarui Wang, Haoran Ye, Minsu Kim, et al. RL4CO: an extensive reinforcement learning for combinatorial optimization benchmark. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2025.

Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

Charles Bordenave. A new proof of Friedman's second eigenvalue theorem and its extension to random lifts. *Annales scientifiques de l'École normale supérieure*, 53(6):1393–1439, 2020.

Maximilian Böther, Otto Kißig, Martin Taraz, Sarel Cohen, Karen Seidel, and Tobias Friedrich. What's wrong with deep learning in tree search for combinatorial optimization. In *International Conference on Learning Representations (ICLR)*, 2022.

Hoyun Choi, Junghyo Jo, and Deok-Sun Lee. Reinforcement learning for microcanonical graph ensemble with assortativity constraints. *arXiv preprint arXiv:2605.23285*, 2026.

Fan R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

Matteo Cinelli, Leto Peel, Antonio Iovanella, and Jean-Charles Delvenne. Network constraints on the mixing patterns of binary node metadata. *Physical Review E*, 102:062310, 2020.

Aaron Clauset, M. E. J. Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical Review E*, 70(6):066111, 2004.

Chris Cremer, Xuechen Li, and David Duvenaud. Inference suboptimality in variational autoencoders. In *International Conference on Machine Learning (ICML)*, 2018.

Luciano da F. Costa, Francisco A. Rodrigues, Gonzalo Travieso, and Paulino R. Villas Boas. Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1):167–242, 2007.

Victor-Alexandru Darvariu, Stephen Hailes, and Mirco Musolesi. Graph reinforcement learning for combinatorial optimization: A survey and unifying perspective. *Transactions on Machine Learning Research (TMLR)*, 2024.

Upasana Dutta, Bailey K. Fosdick, and Aaron Clauset. Sampling random graphs with specified degree sequences. *Journal of Computational and Graphical Statistics*, 34(3):934–947, 2025.

Iakovos Evdaimon, Giannis Nikolentzos, Christos Xypolopoulos, Ahmed Kammoun, Michail Chatzianastasis, Hadi Abdine, and Michalis Vazirgiannis. Neural graph generator: Feature-conditioned graph generation using latent diffusion models. *arXiv preprint arXiv:2403.01535*, 2024.

Dario Fasino and Francesco Tudisco. An algebraic analysis of the graph modularity. *SIAM Journal on Matrix Analysis and Applications*, 35(3):997–1018, 2014.

Miroslav Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(2): 298–305, 1973.

Bailey K. Fosdick, Daniel B. Larremore, Joel Nishimura, and Johan Ugander. Configuring random graph models with fixed degree sequences. *SIAM Review*, 60(2):315–355, 2018.

Linton C. Freeman. Centrality in social networks: Conceptual clarification. *Social Networks*, 1(3): 215–239, 1978.

Joel Friedman. A proof of Alon's second eigenvalue conjecture and related problems. *Memoirs of the American Mathematical Society*, 195(910), 2008.

Vikas K. Garg, Stefanie Jegelka, and Tommi Jaakkola. Generalization and representational limits of graph neural networks. In *International Conference on Machine Learning (ICML)*, 2020.

Minas Gjoka, Maciej Kurant, and Athina Markopoulou. 2.5K-graphs: From sampling to generation. In *Proceedings of IEEE INFOCOM*, pp. 1968–1976, 2013.

Brais González-Rodríguez, Ignacio Gómez-Casares, Bissan Ghaddar, Julio González-Díaz, and Beatriz Pateiro-López. Learning in reformulation-linearization technique-based spatial branching: Limitations of strong branching imitation. *INFORMS Journal on Computing*, 2025.

Xiaojie Guo and Liang Zhao. A systematic survey on deep generative models for graph generation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5370–5390, 2023.

Mark S Handcock. Assessing degeneracy in statistical models of social networks. *Working Paper 39, Center for Statistics and the Social Sciences, University of Washington*, 2003.

David R Hunter. Curved exponential family models for social networks. *Social Networks*, 29(2): 216–230, 2007.

Mark Jerrum, Jung-Bae Son, Prasad Tetali, and Eric Vigoda. Elementary bounds on Poincaré and log-sobolev constants for decomposable Markov chains. *Annals of Applied Probability*, 14(4): 1741–1765, 2004.

Andy L. Jones. Scaling scaling laws with board games. *arXiv preprint arXiv:2104.03113*, 2021.

Pascal Kerschke, Holger H. Hoos, Frank Neumann, and Heike Trautmann. Automated algorithm selection: Survey and perspectives. *Evolutionary Computation*, 27(1):3–45, 2019.

Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. Benchmark graphs for testing community detection algorithms. *Physical Review E*, 78(4):046110, 2008.

Fei Liu, Xialiang Tong, Mingxuan Yuan, Xi Lin, Fu Luo, Zhenkun Wang, Zhichao Lu, and Qingfu Zhang. Evolution of heuristics: Towards efficient automatic algorithm design using large language model. In *International Conference on Machine Learning (ICML)*, 2024.

Shengcai Liu, Yu Zhang, Ke Tang, and Xin Yao. How good is neural combinatorial optimization? a systematic evaluation on the traveling salesman problem. *IEEE Computational Intelligence Magazine*, 18(3):14–28, 2023.

Baptiste Louf, Colin McDiarmid, and Fiona Skerman. Modularity and graph expansion. In *Innovations in Theoretical Computer Science (ITCS)*, 2024.

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

Kanika Madan, Jarrid Rector-Brooks, Maksym Korablyov, Emmanuel Bengio, Moksh Jain, Andrei Cristian Nica, Tom Bosc, Yoshua Bengio, and Nikolay Malkin. Learning GFlowNets from partial episodes for improved convergence and stability. In *International Conference on Machine Learning (ICML)*, 2023.

Manuel Madeira, Clement Vignac, Dorina Thanou, and Pascal Frossard. Generative modelling of structurally constrained graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Priya Mahadevan, Dmitri Krioukov, Kevin Fall, and Amin Vahdat. Systematic topology analysis and generation using degree correlations. In *ACM SIGCOMM*, 2006.

Charles C. Margossian and David M. Blei. Amortized variational inference: When and why? In *Uncertainty in Artificial Intelligence (UAI)*, 2024.

Sergei Maslov and Kim Sneppen. Specificity and stability in topology of protein networks. *Science*, 296(5569):910–913, 2002.

Brendan D. McKay, Nicholas C. Wormald, and Beata Wysocka. Short cycles in random regular graphs. *The Electronic Journal of Combinatorics*, 11:R66, 2004.

Enys Mones, Lilla Vicsek, and Tamás Vicsek. Hierarchy measure for complex networks. *PLoS ONE*, 7(3):e33799, 2012.

Ankur Nath and Alan Kuhnle. Unveiling the limits of learned local search heuristics: Are you the mightiest of the meek? *arXiv preprint arXiv:2310.19990*, 2023.

M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, 2004.

Mark E. J. Newman. Assortative mixing in networks. *Physical Review Letters*, 89:208701, 2002.

Mark E. J. Newman. Spectral methods for community detection and graph partitioning. *Physical Review E*, 88:042822, 2013.

Andrew Y. Ng, Daishi Harada, and Stuart Russell. Policy invariance under reward transformations: Theory and application to reward shaping. In *International Conference on Machine Learning (ICML)*, 1999.

Chiara Orsini, Marija M. Dankulov, Pol Colomer de Simón, et al. Quantifying randomness in real networks. *Nature Communications*, 6:8627, 2015.

Filippo Passerini and Simone Severini. The von Neumann entropy of networks. *arXiv preprint arXiv:0812.2597*, 2008.

Nataša Pržulj. Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23(2):e177–e183, 2007.

Yiming Qin, Manuel Madeira, Dorina Thanou, and Pascal Frossard. DeFoG: Discrete flow matching for graph generation. In *International Conference on Machine Learning*, 2025.

Ladislav Rampasek, Mikhail Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. Recipe for a general, powerful, scalable graph transformer. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Garry Robins, Pip Pattison, Yuval Kalish, and Dean Lusher. An introduction to exponential random graph (p*) models for social networks. *Social Networks*, 29(2):173–191, 2007.

Bernardino Romera-Paredes, Mohammadamin Barekatain, Alexander Novikov, Matej Balog, M. Pawan Kumar, Emilien Dupont, Francisco J. R. Ruiz, Jordan S. Ellenberg, Pengming Wang, Omar Fawzi, Pushmeet Kohli, and Alhussein Fawzi. Mathematical discoveries from program search with large language models. *Nature*, 625:468–475, 2024.

Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011.

Anian Ruoss, Grégoire Delétang, Sourabh Medapati, Jordi Grau-Moya, Li Kevin Wenliang, Elliot Catt, John Reid, Cannada A. Lewis, Joel Veness, and Tim Genewein. Amortized planning with large-scale transformers: A case study on chess. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Michael Schweinberger. Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association*, 106(496):1361–1370, 2011.

Kartik Sharma, Srijan Kumar, and Rakshit Trivedi. Diffuse, sample, project: Plug-and-play controllable graph generation. In *International Conference on Machine Learning (ICML)*, 2024.

David Silver, Thomas Hubert, Julian Schrittwieser, et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, 362(6419):1140–1144, 2018.

Chaoming Song, Shlomo Havlin, and Hernán A. Makse. Self-similarity of complex networks. *Nature*, 433(7024):392–395, 2005.

Tiziano Squartini, Joey de Mol, Frank den Hollander, and Diego Garlaschelli. Breaking of ensemble equivalence in networks. *Physical Review Letters*, 115:268701, 2015.

Zhiqing Sun and Yiming Yang. DIFUSCO: Graph-based diffusion solvers for combinatorial optimization. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Anton Tsitsulin, John Palowitch, Bryan Perozzi, and Emmanuel Müller. Graph clustering with graph neural networks. *Journal of Machine Learning Research*, 24(127):1–21, 2023.

Remco van der Hofstad, Pim van der Hoorn, Nelly Litvak, and Clara Stegehuis. Limit theorems for assortativity and clustering in null models for scale-free networks. *Advances in Applied Probability*, 52(4):1035–1084, 2020.

Piet Van Mieghem, Huijuan Wang, Xin Ge, Siyu Tang, and Fernando A. Kuipers. Influence of assortativity and degree-preserving rewiring on the spectra of networks. *The European Physical Journal B*, 76:643–652, 2010.

Clement Vignac, Igor Krawczuk, Antoine Siraudin, Bohan Wang, Volkan Cevher, and Pascal Frossard. DiGress: Discrete denoising diffusion for graph generation. In *International Conference on Learning Representations (ICLR)*, 2023.

Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

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

Ramon Xulvi-Brunet and Igor M. Sokolov. Reshuffling scale-free networks: From random to assortative. *Physical Review E*, 70(6):066102, 2004.

Shanchao Yang, Kaili Ma, Baoxiang Wang, Tianshu Yu, and Hongyuan Zha. Learning to boost resilience of complex networks via neural edge rewiring. *Transactions on Machine Learning Research (TMLR)*, 2023.

Haoran Ye, Jiarui Wang, Zhiguang Cao, Federico Berto, Chuanbo Hua, Haeyeon Kim, Jinkyoo Park, and Guojie Song. ReEvo: Large language models as hyper-heuristics with reflective evolution. In *Advances in Neural Information Processing Systems*, 2024.

Jiaxuan You, Rex Ying, Xiang Ren, William L. Hamilton, and Jure Leskovec. GraphRNN: Generating realistic graphs with deep auto-regressive models. In *International Conference on Machine Learning (ICML)*, 2018.

Jiachi Zhao, Zehong Wang, Yamei Liao, Chuxu Zhang, and Yanfang Ye. Controllable graph generation with diffusion models via inference-time tree search guidance. In *The Web Conference (WWW)*, 2026.

A PROTOCOL DETAILS AND STATISTICAL METHODOLOGY

Instance definition. A target vector $c^\star = (C^\star, \langle\ell\rangle^\star, Q^\star, \lambda_2^\star, r^\star)$ is measured from a generator draw (WS family unless stated). The mixed family is a stratified round-robin over soft random-geometric-graph (soft-RGG), stochastic-block-model (SBM), and Holme–Kim generators, filtered for verified feasibility. Its target set has participation ratio 2.23 versus 1.49 for WS (higher effective dimension). Metrics: C = mean local clustering; $\langle\ell\rangle$ = mean shortest path length; Q = modularity of the partition returned by greedy Clauset–Newman–Moore (CNM) maximization (Clauset et al., 2004); λ_2 = second-smallest eigenvalue of the normalized Laplacian; r = Newman degree assortativity. Tolerances are $\varepsilon = (0.05, 0.30, 0.05, 0.03, 0.05)$. The band grid is $u \in \{1.0, 0.8, 0.6, 0.5, 0.45, 0.4, 0.35, 0.3, 0.25, 0.2, 0.15, 0.1\}$, extended by $\{1.2, 1.5, 2.0\}$ for the loose-tail study. Satisfaction at band u: $W(G) = \max_k |m_k(G) - c_k^\star|/\varepsilon_k \le u$. Because W is band-independent, one trajectory yields first-hitting times for all bands.

Environment and costs. Warm start: the generator's full rule-based construction pipeline. This is a WS-family draw with rewiring probability inverted from the target clustering, matched to $(N, \langle k\rangle)$, followed by the Stage-2–6 target-aware refinement passes (clustering, path length, community, spectral, assortativity). The warm start is therefore itself target-informed, not a generic random graph. It is computed once per instance and shared verbatim by every swap-based method (random, greedy, beam, samplers, and all swap policies, while guidance is one-shot), so no between-method comparison is confounded by it. Its heterogeneity is separately controlled in §7 (Figure 5). Action: degree-preserving double-edge swap under a simplicity mask; connectivity is checked explicitly in beam search and in the value learner's lookahead, and elsewhere a disconnecting swap is penalized implicitly (it sets λ_2 to zero, outside the band for all but the lowest-gap targets, so there is no universal connectivity guarantee on those paths). The budget is T=60 swaps per episode. Candidates: each step draws 32 valid swaps from the whole graph with no locality restriction. Half are uniformly random edge pairs and half are bridge-biased (one edge on each side of the Fiedler-vector sign cut). Every method chooses among them, except the DMGG transplant, which samples uniformly (App. D). Greedy true-evaluates K=8 of the 32, chosen uniformly, per step throughout (unified benchmark, regime grid, and large-N study alike), and the value learner true-evaluates 6 of them (the budget study of App. H charges its calls at 8 per step, an upper bound). The budget Pareto study sweeps $K \in \{2, \ldots, 32\}$ (App. H). Scoring objectives differ: greedy ranks its candidates by a weighted-L_1 residual potential (spectral gap weighted 3×), whereas beam and the value learner rank by the worst-scaled residual that also defines satisfaction. Greedy therefore optimizes a surrogate of the evaluation criterion, which makes its dominance conservative; the beam–greedy comparison varies both search depth and scoring objective. One true-metric evaluation costs 111.5 ms at N=200 (dominated by $\langle\ell\rangle$, Q, λ_2). For $N \ge 1200$, exact evaluation becomes impractical (28 s per call at N=2000), so those cells use accelerated metrics (sampled-source path lengths and shift-invert spectral solves). These were validated against exact evaluation (λ_2 exact match; $\langle\ell\rangle$ error < 0.07). Those cells also use reduced budgets (episodes 150, 12 evaluation targets, methods random/greedy/value). Signs and dominance relations are robust to this substitution. Absolute satisfaction values at $N \ge 1200$ should not be compared directly against smaller regimes.

Statistics. Main benchmark: 10 seeds (beam: 6). Held-out targets and training budget per seed are 15 targets/200 episodes for seeds 0–3 and 20/300 for seeds 4–9 (180 held-out targets in total). Within each seed the targets and budget are shared by all methods, so the seed-level pairing below is unaffected. The niche is present in both groups: $\text{value}_{\text{PE}}-\text{random}$ at u=0.4 is +13.3 over seeds 0–3 and $+11.7 \pm 4.2$ over seeds 4–9. The value learner, SCST, PPO, and REINFORCE are trained once at band u=0.3 on training targets (seed 1000+s); BC and DAgger imitate the band-agnostic greedy expert, whose training rollouts stop once residuals enter the u=1.0 box (a handicap for imitation, not an aid). Every method's trajectories are scored on all 12 bands; beam uses width 4 with 6 expansions per step. Regime study (all N, $\langle k\rangle$ cells): 10 seeds (mixed family: 4). Unified-protocol re-runs of the divergent paradigms (GFlowNet, MPC, DMGG-factorized, guidance): 10 seeds. Subset ablations: 9 seeds for the imitation-capture ablation (Table 8) and 10 seeds for the metric-count sweep (App. F). Tolerance profiles: 10 seeds, 20 for the spectral-tight re-test (Table 7). Budget Pareto: 10 seeds. Leave-one-family-out: 10 seeds (feasibility-filtered evaluation pools). The regime-grid, large-N, and budget studies use the PE variant of the value learner. The unified benchmark reports both variants (statistically indistinguishable, Table 5). All headline deltas are computed by first averaging

over the held-out targets within each seed (the same targets are shared by all methods within a seed) and then pairing at the seed level ($n=$ number of seeds). Each delta is reported with the standard error (SE), a 20,000-resample percentile bootstrap 95% confidence interval, and a paired t-test. We assert a sign only when the interval excludes zero and $p < 0.05$. Otherwise we report the delta as statistically indistinguishable from zero. Absolute satisfaction levels vary across target draws (greedy's per-run mean at u=0.4 ranges from 72.5% (6-seed beam subset) to 83.4% across the runs in this paper). This is why every inferential comparison is paired within a run and never made across runs; levels displayed from different seed sets are descriptive and marked unpaired.

B CLAIM–EVIDENCE MAP

Table 2 collects the paper's twelve Findings with their key evidence and the location of the full backing.

Table 2: **Claim–evidence map.** One row per Finding. Key numbers are representative; full tests, uncertainties, and grids are at the referenced locations.

	Claim (one line)	Key evidence	Where
F1	Only the oracle-augmented value learner beats random, in a mid-tight tolerance window (protocol-sensitive); no residual-free policy does so significantly	peak +12.0–14.5 pp; corrected p=0.007; pooled over five runs $+6.2 \pm 3.5$	§4; Fig. 1a, 2; Tab. 5
F2	The niche concentrates in subsets whose binding constraints include assortativity (composition, not count)	r-subsets: positive mean peak niche at every subset size (+7.5–10 pp); r-free: +0.9–5 pp, not tracking headroom (Table 9)	§4; App. F
F3	Horizon predicts search payoff; a router on the warm-start residuals saves little	corr 0.989; saving ∼10%	§4; App. G
F4	No learned method approaches true-metric greedy	best recovers 23.5% of the gap	§5; Tab. 1; Fig. 2
F5	The strongest learner buys no cost advantage; the cheap-and-good quadrant is empty	value 972 vs. greedy 1048 ms/decision; 35.0 vs. 81.7% at 6 vs. 8 evaluations per step; T=340 horizon $\leq$ +1.3 pp; screen-then-verify −19 to −45 pp at u=0.4	§5; Tab. 10; App. H
F6	Deeper search exceeds greedy; distilling it fails	beam−greedy +10.8 pp at u=0.4	§5; Fig. 2
F7	The ceiling is robust across architecture, modality, tuning, framing	DMGG arch. 100% native vs. 31.7% joint; EoH +0.005	§5; App. H
F8	Low-cost local features cannot rank candidate edits	R^2=0.009; 96.7% pooled collisions, 70.7% of states confusable	§6.1; App. I
F9	State values are predictable; action ranking is not	R^2 0.06 (h=1) → 0.83 (h=50), global features	§6.1; App. I
F10	Failure localizes to assortativity, via binding	r-subsets capture 11.8% vs. 50.4% r-free	§6.2; Tab. 8
F11	The target space is thin; a slice is certified infeasible	$\lambda_2 \leq 2(1-Q)$; 99% of box-sampled targets unreached (N=200, $\langle k \rangle$=6); 10^5-swap SA recovers 3%	§6.3; App. J, K
F12	The ceiling generalizes; the niche is bounded on every axis	niche −1.0 pp (N=800), −2.1 (mixed); greedy +19–42 pp on LOFO	§7; Tab. 13; Fig. 2

C METRIC SELECTION

The five descriptors span distinct top-level axes of the standard measurement taxonomy (da F. Costa et al., 2007): local transitivity (C; Watts & Strogatz, 1998), global distance ($\langle \ell \rangle$), mesoscale community (Q; Newman & Girvan, 2004), spectral connectivity (λ_2; Fiedler, 1973), and degree–degree correlation (r; Newman, 2002). Each is the standard single-scalar representative of its axis, not a complete characterization, and we make no orthogonality claim. The degree-preserving swap fixes the degree sequence (dK-1 (Mahadevan et al., 2006)). The next orders of that hierarchy are exactly the targeted axes (dK-2 determines r, dK-3 $\ni C$), so the degree distribution is controlled by design rather than omitted. Six candidate scalars from other axes, measured on 120 mixed-family graphs, were excluded. Von Neumann entropy (Passerini & Severini, 2008) and global reaching centrality (Mones et al., 2012) are weakly correlated with the five ($\max |\rho| < 0.4$) but their structural readings (regularity, hierarchy) are less direct for the five design axes. Degree entropy, spectral radius, and betweenness centralization (Freeman, 1978) are redundant with r or C ($|\rho| \geq 0.55$). Fractal dimension (Song et al., 2005) is unstable at N=200. Motif/graphlet descriptors are vector-valued (Pržulj, 2007), with the triangle already represented by C.

Empirical coupling of the five. On the WS target ensemble (300 targets, N=200, $\langle k \rangle$=6), C, $\langle \ell \rangle$, Q, and λ_2 are strongly coupled ($|\rho| \in [0.73, 0.98]$) while r is the only relatively independent axis ($|\rho| \in [0.38, 0.44]$). The effective dimension is ≈ 2 (participation ratio 1.49). The coupling has a reducible sampling component (cross-generator $|\rho|$ drops from 0.67 under WS to 0.30 under soft-RGG) and two irreducible ones: the Q–λ_2 exclusion bound (§6.3) and the fixed degree sequence, which constrains the achievable (C, Q, λ_2) region. This is why independently sampled five-metric targets are 99% unreached under the benchmark degree sequence (§6.3), and why r, the one independent axis, is also the one whose binding collapses learning (§6.2).

D LEARNED-APPROACH ROSTER

All swap policies act on the same MDP (state = current graph + target residuals, action = choice among K sampled valid swaps). The roster spans the standard neural-CO families (cf. Bengio et al., 2021b; Berto et al., 2025). Unless noted, backbones are 2–3-layer graph convolutional (GCN) or graph isomorphism (GIN) networks over low-cost node features (degree, local clustering, triangle count, neighbor degree) plus a Laplacian positional encoding of the current graph. Here "residual-free" means no true-metric recomputation per candidate at inference.

1. **BC / DAgger**: behavior cloning of greedy's per-step choice (cross-entropy over K candidates), offline and with online dataset aggregation (Ross et al., 2011).
2. **value / value$_{\text{PE}}$ / value$_{\text{dyn}}$**: regression of the search-attainable worst-scaled residual (the best residual a depth-3 beam reaches from a state). The network takes the graph and the state's true residual vector as input. For value and value$_{\text{PE}}$, inference evaluates each candidate's true post-swap residual and picks the argmin of the *predicted* reachable residual; the policy is therefore greedy plus a learned non-myopic correction, at greedy's per-decision cost. That this learned-lookahead policy, despite its oracle inputs, scores below the greedy reference (Table 1) is itself a negative result about learnable lookahead. value$_{\text{PE}}$ appends a Fiedler positional encoding, recomputed on each candidate graph (its effect is indistinguishable from zero at every band, $p > 0.6$). value$_{\text{dyn}}$ instead recomputes the Fiedler vector of the current graph once per step and scores candidates by endpoint components, residual-free at inference (App. I).
3. **REINFORCE / PPO / SCST**: policy gradients with potential-based shaping on scaled residuals (Ng et al., 1999). SCST uses a greedy-decode baseline with a residual-free scorer. Our PPO configuration performs one policy update per episode with the old log-probabilities taken from the same forward pass, so the clipped ratio is identically one and the update coincides with REINFORCE plus a learned value baseline and an entropy bonus; we keep the label for the configuration, not as a claim about clipped PPO.
4. **GFlowNet (+ strengthening study)**: trajectory-balance flow over swap sequences (Bengio et al., 2021a; 2023); the backward policy is normalized over the swaps applied so far in the trajectory rather than over all predecessors of the current state, so the objective is a trajectory-balance variant. The strengthening study adds reward tempering schedules, ε-greedy explo-

ration, prioritized resampling of recently satisfied or high-reward targets (exact trajectory replay is impossible with regenerated candidates), subtrajectory balance (Madan et al., 2023), and their combination, at budgets matched to the value learner.

5. **Guidance**: differentiable-surrogate gradients over a soft adjacency. All five surrogates are validated (path length via soft-min-plus, Spearman 0.999; modularity via DMoN (Tsitsulin et al., 2023), Spearman 0.972).
6. **MPC**: surrogate world-model planning (depth-3 beam in surrogate space, first move executed truthfully, replan).
7. **Curriculum**: metric-count schedule over SCST. It coincides with SCST on the fixed five-metric grid (noted, not double-counted).
8. **DMGG-factorized**: the factorized autoregressive edge-pair policy with action masking of Choi et al. (2026) (GIN encoder, feature-wise linear modulation (FiLM) target conditioning, two edge-selection heads; the original adds a third head choosing the rewiring mode, which our single-mode swap does not need; the transplant samples its candidate swaps uniformly rather than from the mixed uniform/spectral pool of App. A, so its paired comparison varies proposal as well as architecture), trained with potential-shaped rewards on the five-metric objective. The five-metric transplant is our faithful re-implementation of that architecture. Cross-validation: the original authors' released codebase, run unmodified on its native single-assortativity task in our setting (WS, N=200, $\langle k\rangle$=6, tolerance 0.02), reaches 200/200 held-out satisfaction. The architecture is sound in our environment, and the re-implementation preserves its ingredients.
9. **LLM-evolved scorer (EoH)**: an EoH-style evolution loop (Liu et al., 2024) with a 32B-parameter code LLM evolves interpretable low-cost scoring functions $\mathrm{score}(\text{features}) \to \mathbb{R}$ over six generations, population evaluated by agreement@3 (top pick within greedy's true top three; cf. FunSearch and ReEvo (Romera-Paredes et al., 2024; Ye et al., 2024)).

Counting: items 1–9 above comprise eleven methods in thirteen configurations (the value family counts three; the curriculum coincides with SCST and is not counted). Eleven are evaluated under the unified 12-band, 10-seed protocol, and the curriculum is not run separately. The $\text{value}_{\text{dyn}}$ variant and the EoH scorer are single-band probes (App. I, App. H). Beam distillation and value-guided beam below are additional imitation-family probes and are not counted. Beam distillation (offline and online imitation of beam labels) was additionally evaluated and lands at the random floor, completing the imitation family. Value-guided beam (wrapping the learned value *inside* a depth-2 beam to score expansions, rather than imitating beam labels) fares no better. It scores $-33/-31/-53$ pp below greedy at u=0.5/0.4/0.3 (3 seeds) and at the tightest band collapses to the value one-step level. The learned value therefore rescues search neither by ranking its expansions nor by imitating its labels.

Training-time oracle accounting. Table 3 counts the true-metric evaluations (one call = one five-metric recomputation) each learned method consumed in training (E = training episodes, T = swaps per episode, k_g=8 greedy candidates; E=300, T=60 unless noted; GFlowNet E=600, T=40; episode terms are upper bounds). A single greedy evaluation rollout costs $T(k_g{+}1) = 540$ calls, so training budgets span 0 to ∼300 greedy-rollout equivalents. Consumption does not predict final satisfaction: DAgger, the heaviest consumer, sits at the random floor, and guidance and MPC, the lightest, also fail.

E THE NICHE: UNIFIED STATISTICS, REPLICATION, AND TOLERANCE PROBES

E.1 UNIFIED BENCHMARK, ALL 12 BANDS

Table 4 gives the complete grid underlying Table 1. Table 5 reports the paired analyses cited in the main text, and Table 10 the decision costs and band-averaged gap recovery.

E.2 PAIRED STATISTICS FOR HEADLINE DELTAS

Multiple-comparison control for the niche. The niche peak is selected from 12 bands × two indistinguishable value variants, so we fold that selection into the null with an exact max-statistic sign-flip permutation test on the seed-level niche (learner−random, paired over 10 seeds). All 2^{10}

Table 3: Training-time oracle calls per trained model (unified protocol; E=300, except that seeds 0–3 used E=200, so the E-dependent rows scale by $2/3$; derivations follow from the training-loop structure described in the text).

Method	Training oracle-call formula		≈ calls
BC (offline)	$N_{\text{tr}}\, T\,(k_g{+}1)$	(12 targets)	6.5×10^3
DAgger	$E\, T\,(k_g{+}1)$		1.6×10^5
value / value$_{\text{PE}}$	$N_{\text{st}}\big[1 + k_e(1{+}(D{-}1)b)\big] + N_{\text{st}}\,\bar{c}_{\text{act}}$ (240 states; D=3, b=4, k_e=6; $\bar{c}_{\text{act}}{\approx}4$)		1.4×10^4
value$_{\text{dyn}}$	$n_{\text{states}}(k_{\text{cand}}{+}1)$	(1000 × 17)	1.7×10^4
SCST	$2E(T{+}1)$	(sample + greedy-decode rollouts)	3.7×10^4
REINFORCE / PPO	$E(T{+}1)$		1.8×10^4
GFlowNet (per config)	$E(T{+}1)$		2.5×10^4
Guidance	0	(surrogate-only optimization)	0
MPC	n_{calib}	(surrogate calibration graphs)	60
Curriculum (per rung)	$2E(T{+}1)$	(SCST loop)	3.7×10^4
DMGG-factorized	$E(T{+}1)$		1.8×10^4
EoH scorer	$n_{\text{states}}(k_{\text{cand}}{+}1)$ (200 × 17; evolution reuses the pool)		3.4×10^3

Table 4: Satisfaction (%); WS held-out targets, N=200, $\langle k\rangle$=6; 10 seeds (beam 6).

Method / u	1.0	0.8	0.6	0.5	0.45	0.4	0.35	0.3	0.25	0.2	0.15	0.1
random	92.3	79.0	57.5	44.0	34.7	27.7	20.0	13.5	9.3	5.3	2.5	0.0
greedy	94.8	92.3	88.3	82.0	77.7	74.2	68.2	64.3	55.0	45.5	35.5	19.3
beam	95.0	94.2	92.5	90.8	83.3	83.3	80.8	73.3	69.2	64.2	52.5	43.3
value	91.3	85.7	69.2	56.0	44.8	39.2	31.5	23.5	16.3	10.8	6.3	1.2
value$_{\text{PE}}$	89.8	82.5	67.7	55.7	47.0	40.0	34.5	23.0	15.5	8.5	6.2	2.2
BC	84.3	72.2	57.3	44.2	36.7	27.8	20.7	14.0	8.0	5.8	2.3	0.5
DAgger	91.0	78.0	59.3	48.0	39.7	29.8	17.8	13.0	7.2	2.8	1.3	0.0
SCST	81.0	75.0	60.7	48.0	41.3	31.7	26.3	17.0	9.8	7.2	2.2	0.0
PPO	71.0	59.3	44.2	35.7	29.0	24.3	19.2	13.8	8.8	6.3	2.2	0.5
REINFORCE	76.0	60.5	45.8	34.8	30.0	25.8	17.8	13.7	6.3	3.8	1.2	0.0

sign assignments are enumerated, and the maximum mean niche over the 24 band×variant statistics is recorded for each. The observed maximum (+14.5 pp, PE at u=0.35) gives p=0.007. Per variant, p=0.002 for PE and p=0.052 for plain (peak +12.0 at u=0.5). Per band on the PE variant, three bands survive Holm ($u \in \{0.5, 0.4, 0.35\}$) and six survive Benjamini–Hochberg FDR at 0.05 ($u \in [0.5, 0.25]$), with paired Cohen's d up to 1.6 at u=0.4. No plain-variant band survives on its own. On the large-N grid (18 cells) no cell survives Holm (nominal p=0.042 at N=2000, $\langle k\rangle$=8 becomes 0.71). The test is conditional on pre-selecting the value family, whose competitors show no statistically significant improvement over random at any band (Table 4). It quantifies band-and-variant selection within that family, not a search over all approaches.

Replication of the niche across runs. The N=200, $\langle k\rangle$=6, WS cell is re-run, with independent training, in several studies of this paper. Table 6 collects the value−random niche at u=0.4 from every such run. The two runs with the largest evaluation pools and full training budgets (the unified benchmark and the regime-grid cell) are both significant. The three re-draws with 12 targets per seed (two with the unified target sampler, one with targets drawn feasible-by-construction from the warm start's own degree sequence) are indistinguishable from zero with standard errors of 5–8 pp. They therefore neither confirm nor exclude an effect of +10 pp. The heterogeneity is band-wide: at u=0.6 the standard tolerance-profile row gives -10.9 ± 4.1 pp against the unified benchmark's $+11.7 \pm 4.5$. A DerSimonian–Laird random-effects pooling of the five run means (the unified benchmark enters as the PE variant) gives $+6.2 \pm 3.5$ pp with significant between-run heterogeneity (Q=10.1, df=4, p=0.04). The seed-level pool (n=50) gives $+4.8 \pm 2.6$. We therefore describe the window as supported in the two larger-pool runs but protocol-sensitive, with its magnitude uncertain. The 31-subset ablation (retrained at 150 episodes) is not included because its budget differs.

Table 5: Paired deltas (pp) with SE, bootstrap 95% CI, paired t-test p, and exact sign-flip p (smallest attainable 0.002 for 10 seeds, 0.031 for 6); unified benchmark. The beam−greedy delta is paired over the 6 common beam seeds, where greedy averages 72.5% at u=0.4 (vs. 74.2% over 10 seeds in Table 4), giving $83.3 - 72.5 = +10.8$. Niche p-values (value−random rows) are per-band paired tests in the mid-tight window. The two peak cells (+12.0 plain at u=0.5, +14.5 PE at u=0.35) are the statistics of the permutation test below. Their band-selection-corrected significance is established there.

Band	Comparison	Δ	SE	95% CI	p (t)	p (exact)
u=0.6	value−random	+11.7	4.5	[+3.8, +20.3]	0.028	0.031
u=0.4	value−random	+11.5	4.3	[+4.0, +20.0]	0.027	0.063
u=0.3	value−random	+10.0	3.1	[+4.7, +16.0]	0.010	0.016
u=0.4	beam−greedy	+10.8	3.0	[+5.0, +15.8]	0.015	0.063
u=0.4	greedy−random	+46.5	4.5	[+38.2, +55.0]	< 0.001	0.002
u=0.4	value_{PE}−value	+0.8	4.6	[−7.9, +8.8]	0.860	0.887
N=800 ($\langle k\rangle$=6), u=0.4	value_{PE}−random	−1.0	2.9	[−6.0, +4.5]	0.735	0.863
mixed family, u=0.4	value_{PE}−random	−2.1	4.0	[−8.3, +4.2]	0.638	1.000

Table 6: Value−random niche (pp) at u=0.4 in every run of the N=200, $\langle k\rangle$=6, WS cell (10 seeds each; mean±SE; paired t-test).

Run	Protocol difference	niche	SE	p
unified benchmark, value_{PE}	15–20 targets/seed, 200–300 episodes	+12.3	2.4	0.001
unified benchmark, value (plain)	same run, plain variant	+11.5	4.3	0.027
regime grid, N=200 cell	separate run, 20 targets/seed	+14.0	5.3	0.027
tolerance-profile standard row	12 targets/seed	+0.8	5.3	0.88
budget study (value_{PE}, K=6)	12 targets/seed	+0.0	5.4	0.997
warm-start sweep, WS rung	feasible-by-construction targets	−3.3	8.3	0.70
random-effects pooling (5 runs)		+6.2	3.5	

E.3 Loose-band extension

At $u \in \{1.2, 1.5, 2.0\}$ (4 seeds), random saturates at 93.8%, greedy−random is +0.0, and value−random is $0.0/-2.1/-4.2$ pp at u=2.0/1.5/1.2. Because satisfaction is first-hitting, this is not a matter of leaving a band already entered: at very loose bands a policy's swaps can only steer a still-unsatisfied warm start away from a band that random drift would have entered, which closes the loose tail of the unimodal window.

E.4 Tolerance-profile probe

Five profiles are run: the standard one and four that each halve one metric's tolerance to tighten that metric relative to the others (10 seeds, 12 evaluation targets per seed). The standard row is the within-table control (Table 6). The peak column is a maximum over 12 bands and therefore selection-biased upward. At the comparison band u=0.4, only the spectral-tight cell is nominally significant over 10 seeds ($+14.2 \pm 5.1$, p=0.02). Over 20 seeds it is $+6.3 \pm 3.6$ pp (p=0.100 paired t; 0.073 Wilcoxon; -1.7 ± 3.9 on the 10 added seeds alone). Across the full 5×12 profile-by-band grid, three cells are nominally positive (uncorrected p<0.05: spectral-tight at u=0.35 and 0.30, clustering-tight at u=0.25), and none survives correction over the 60 cells. No profile therefore establishes a niche by tolerance re-weighting alone, and the binding-locality criterion (Definition 1) remains a hypothesis supported only by the 31-subset intervention (Appendix F).

F Metric-count ablation: all 31 subsets

To separate the effect of the *number* of constrained metrics from their *composition*, we re-ran the benchmark on every non-empty subset of the five metrics ($\sum_k \binom{5}{k} = 31$ subsets, $k = 1, \ldots, 5$; 10 seeds; N=200, $\langle k\rangle$=6, 12 bands). Metrics outside the subset are masked from the worst-scaled

Table 7: Tolerance profiles and the niche (value−random in pp, mean±SE over 10 seeds). ‡Over 20 seeds this cell is $+6.3 \pm 3.6$ pp (p=0.100; see text).

Tolerance profile	Tightened metric	Peak band	Peak value−random	at u=0.4	p (u=0.4)
standard	mixed	u=0.25	$+5.0 \pm 4.3$	$+0.8 \pm 5.3$	0.88
spectral-tight	λ_2	u=0.6	$+16.7 \pm 8.3$	$+14.2 \pm 5.1$‡	0.02‡
clustering-tight	C	u=0.25	$+9.2 \pm 3.6$	$+5.8 \pm 3.1$	0.09
modularity-tight	Q	u=0.25	$+5.8 \pm 3.5$	-0.8 ± 6.0	0.89
assortativity-tight	r	u=0.25	$+6.7 \pm 5.2$	$+0.8 \pm 7.3$	0.91

Table 8: **Constraint-subset ablation** (31 subsets, 9 seeds): fraction of the greedy–random headroom captured by the imitation (behavior-cloning) policy at u=0.4, averaged over the 16 subsets containing each metric (E=80 episodes, T=20, K=4, 12 evaluation targets per seed). Bold marks the outlying metric; the groups overlap, and the 15 r-free subsets average 50.4%. The strongest value learner is also composition-sensitive over all 31 subsets, its niche concentrating in r-subsets (Table 9).

metric in subset	$\langle\ell\rangle$	Q	C	λ_2	r
BC capture rate (%)	36.2	27.7	39.3	22.8	**11.8**

objective (their scale factor is set to 10^9, removing them from the max), so random, greedy, and the value learner (retrained per subset, 150 episodes) all operate subset-aware without code changes. Because the per-subset learner is retrained at this reduced budget, the full five-metric row is not a replication of the unified benchmark's headline niche. The controlled quantity is the contrast *across* subsets within this table.

Table 9 reports, per subset, the peak niche (value−random, maximum over the 12-band sweep), the niche and greedy−random gap at u=0.3, and the three satisfaction rates at u=0.3. Five regularities emerge. (i) *Composition, not count*: the niche concentrates in subsets containing assortativity. Group means of the peak niche are $+7.5$ to $+10$ pp at every k with r included, vs $+0.9$ to $+5$ pp without. Normalizing by headroom (niche/gap at u=0.3, cells with gap $>$ 10 pp) gives $+0.09 \pm 0.03$ with r vs -0.04 ± 0.07 without, so the effect is not explained by the larger gaps of r-subsets. (ii) *Greedy dominance is composition-invariant*: value is strictly below greedy in 351 of 372 subset×band cells, ties it in 18 (three of these are $+0.01$ pp differences), and exceeds it in 3, by at most $+0.9$ pp, all in r-free subsets at loose-to-mid bands. The mean deficit is -25.6 pp (from -9.5 at k=1 to -36.3 at k=5), including the single-metric r subset (67% vs 100% at u=0.3). The conclusion is unchanged: reducing the metric count does not open a learning route past true-metric search. (iii) *r is the dominant binding metric*: relief is measured leave-one-out, as the gain in greedy satisfaction when one metric is removed from a subset, averaged over $u \in \{0.3, 0.25, 0.2\}$. Removing r yields the largest mean relief ($+12.6$ pp, vs $+10.7$ for C, $+9.5$ for $\langle\ell\rangle$, $+6.0$ for λ_2, $+5.7$ for Q; $+18$ pp at k=5). (iv) *The niche tracks headroom only among r-subsets*: across the 31 subsets at u=0.3, the correlation between niche and greedy−random gap is $+0.51$ with r vs. $+0.06$ without (Figure 3b; the two groups' headroom ranges barely overlap, so the contrast is between regimes rather than slopes at equal headroom). Without r, even subsets with 30–40 pp of headroom yield only small, non-significant niches ($+6.7 \pm 4.9$ and $+1.7 \pm 5.7$ pp), although five of the 15 r-free point estimates are positive. (v) *Global co-constraints erode the niche from the tight end*: the single-metric r subset keeps a positive niche down to the tightest band ($+5.9$ pp at u=0.3, $+8$ pp at u=0.1), whereas at k=5 the peak shifts to $u \approx 0.45$ and the tight end ($u \in [0.35, 0.25]$) turns negative (Figure 3c). The tolerance-profile probe (Table 7) shows no corresponding significant negative niche, so this tight-end negativity stands on its own.

The boundary of the niche thus tracks whether r is constrained at all: no headroom-tracking niche when r is absent (the surviving binding constraints are global, or radius-2 for C), and a niche when r binds on a fraction of steps, consistent with the binding-fraction mechanism of §6. Whether the niche would invert when r-binding dominates is not established: the assortativity-tight profile shows no significant niche in either direction ($+0.8 \pm 7.3$ pp at u=0.4; Table 7).

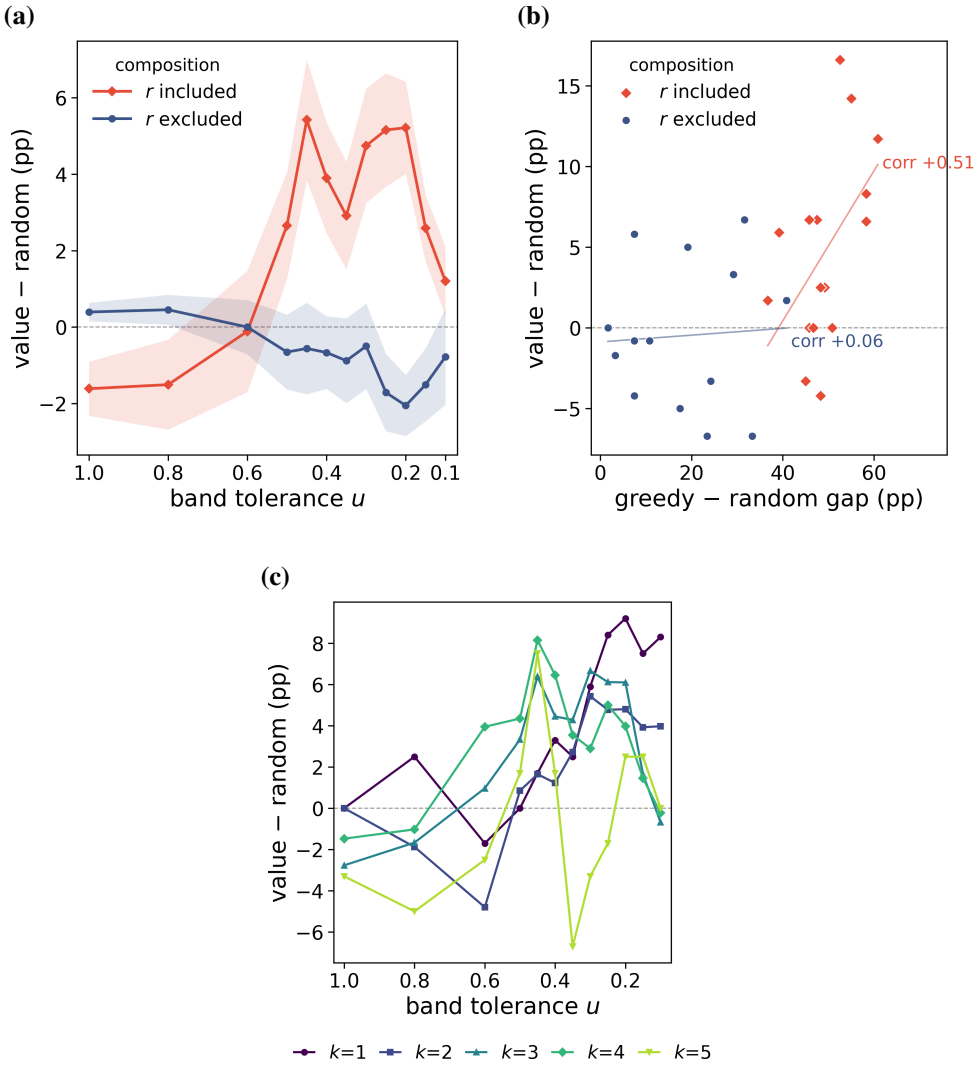


Figure 3: **Niche vs. constraint composition** (31 subsets, 10 seeds, N=200). *(a)* value−random vs. band; group means ±SE over subsets with and without r. *(b)* Niche vs. greedy−random headroom at u=0.3, one point per subset, with per-group least-squares lines. *(c)* r-containing subsets grouped by metric count k (group means; SE omitted for clarity). Interpretation in the text.

G HORIZON PREDICTOR AND ROUTER

A logistic gate on warm-start features (the five warm-start residuals, i.e. the per-metric horizon, and the target vector) predicts "will random fail?" and routes to greedy on predicted failure (12 bands × 4 seeds). At u=1.0 the router matches greedy within −1.7 pp at 86% evaluation saving. At $u \in [0.6, 0.8]$ it over-trusts random (−14 to −17 pp quality). At $u \in \{0.45, 0.4\}$ it saves $\leq 4\%$; from u=0.35 down it routes everything to greedy (0% saving); quality is preserved throughout. Averaged over the bands where quality is preserved (within 5 pp of greedy), the saving is ∼10%. A cost–quality plot (Pareto view) places the router between random and greedy but below the greedy/oracle frontier. Learned gating reduces cost only where the problem is easy enough for random.

Horizon predictor (separate protocol). Across four task variants run under a separate protocol (5 seeds × 300 targets, default band; different band grid and budgets from the unified protocol), the horizon from warm start to target predicts the greedy−random gap with correlation 0.989. Im-

Table 9: All 31 metric subsets (10 seeds, N=200), grouped by metric count k and ordered canonically ($C, \langle\ell\rangle, Q, \lambda_2$) with r-free subsets before r-containing ones. Columns: satisfaction rates of random / value / greedy at u=0.3; niche (value−random) and headroom (greedy−random) at u=0.3; peak niche over the 12-band sweep (band in parentheses). Differences are computed before rounding. **Bold** value entries exceed their paired random baseline (a positive niche at u=0.3): 11 of 16 r-containing subsets versus 5 of 15 r-free ones. Greedy leads in every row.

subset	k	r	$\text{rnd}_{0.3}$	$\text{val}_{0.3}$	$\text{grd}_{0.3}$	$\text{niche}_{0.3}$	$\text{gap}_{0.3}$	peak niche (u)
C	1	–	91.7	90.0	95.0	−1.7	+3.3	+1.7 (0.1)
$\langle\ell\rangle$	1	–	88.3	88.3	90.0	+0.0	+1.7	+1.7 (0.2)
Q	1	–	92.5	91.7	100.0	−0.8	+7.5	+0.0 (0.15)
λ_2	1	–	88.3	87.5	99.2	−0.8	+10.8	+0.0 (0.35)
r	1	✓	60.8	**66.7**	100.0	+5.9	+39.2	+9.2 (0.2)
C+$\langle\ell\rangle$	2	–	81.7	**87.5**	89.2	+5.8	+7.5	+5.8 (0.3)
C+Q	2	–	66.7	63.3	90.8	−3.3	+24.2	+1.7 (0.45)
C+λ_2	2	–	67.5	60.8	90.8	−6.7	+23.4	+3.3 (0.6)
$\langle\ell\rangle$+Q	2	–	80.0	75.8	87.5	−4.2	+7.5	+4.2 (0.5)
$\langle\ell\rangle$+λ_2	2	–	82.5	81.7	90.0	−0.8	+7.5	+8.3 (0.1)
Q+λ_2	2	–	76.6	**81.7**	95.8	+5.0	+19.2	+5.0 (0.3)
C+r	2	✓	35.8	35.8	81.7	+0.0	+45.8	+5.0 (0.45)
$\langle\ell\rangle$+r	2	✓	46.7	**48.3**	83.3	+1.7	+36.7	+8.3 (0.15)
Q+r	2	✓	39.2	**47.5**	97.5	+8.3	+58.3	+8.3 (0.25)
λ_2+r	2	✓	34.2	**45.8**	95.0	+11.7	+60.8	+11.7 (0.3)
C+$\langle\ell\rangle$+Q	3	–	58.3	**61.7**	87.5	+3.3	+29.2	+3.3 (0.3)
C+$\langle\ell\rangle$+λ_2	3	–	55.0	**61.7**	86.7	+6.7	+31.6	+12.5 (0.45)
C+Q+λ_2	3	–	45.8	**47.5**	86.7	+1.7	+40.8	+4.2 (0.8)
$\langle\ell\rangle$+Q+λ_2	3	–	68.3	63.3	85.8	−5.0	+17.5	+0.0 (1.0)
C+$\langle\ell\rangle$+r	3	✓	34.2	**40.8**	81.7	+6.7	+47.5	+7.5 (0.2)
C+Q+r	3	✓	30.8	26.7	79.2	−4.2	+48.3	−3.3 (0.1)
C+λ_2+r	3	✓	24.2	24.2	75.0	+0.0	+50.8	+2.5 (0.45)
$\langle\ell\rangle$+Q+r	3	✓	32.5	**39.2**	78.3	+6.7	+45.8	+6.7 (0.3)
$\langle\ell\rangle$+λ_2+r	3	✓	28.4	**45.0**	80.8	+16.7	+52.5	+16.7 (0.45)
Q+λ_2+r	3	✓	33.3	**47.5**	88.3	+14.2	+55.0	+19.2 (0.25)
C+$\langle\ell\rangle$+Q+λ_2	4	–	49.2	42.5	82.5	−6.7	+33.3	+2.5 (1.0)
C+$\langle\ell\rangle$+Q+r	4	✓	23.3	**25.8**	72.5	+2.5	+49.2	+15.0 (0.4)
C+$\langle\ell\rangle$+λ_2+r	4	✓	23.3	**25.8**	71.7	+2.5	+48.3	+5.0 (0.2)
C+Q+λ_2+r	4	✓	17.5	**24.2**	75.8	+6.6	+58.3	+9.2 (0.45)
$\langle\ell\rangle$+Q+λ_2+r	4	✓	28.3	28.3	75.0	+0.0	+46.7	+10.9 (0.45)
C+$\langle\ell\rangle$+Q+λ_2+r	5	✓	22.5	19.2	67.5	−3.3	+45.0	+7.5 (0.45)

provement density and one-step predictability do not. The greedy−random gap itself rises from +0.4 pp at loose to +50.4 pp at tight bands under that protocol. The same protocol evaluated encoder (GCN/GIN/GAT, random-walk PE), policy-gradient, preference-conditioned, and lookahead-feature variants, all at or below random (77–89% vs. random 90.1%, greedy 94.5%, $p < 10^{-3}$). Its imitation results are not comparable to the unified benchmark's BC row.

H THE CEILING: COST, BUDGETS, AND PARADIGM VARIANTS

H.1 COST ACCOUNTING AND GAP RECOVERY

H.2 BUDGET PARETO

The budget axis is the number of true-metric (oracle) evaluations at u=0.4 (10 seeds, 12 held-out targets per seed). Greedy with K=2/4/8/16/32 candidates spends 75/126/172/318/566 calls for 47.5±4.8 / 62.5±5.0 / 81.7±2.7 / 83.3±2.2 / 85.8±2.5% satisfaction. The value policy, true-evaluating 6 candidates per step, is charged 331 calls at 8 per step (≈248 at its actual 6) for 35.0±4.1%. That is not distinguishable from random (35.0% at zero calls; paired +0.0±5.4, p=0.997) and −46.6±4.5 pp below greedy at K=8 ($p<10^{-4}$), while spending more calls than

Table 10: Left: wall-clock cost per decision (N=200, greedy profiled at K=8 and the value learner at K=6, as in the benchmark; both scale linearly in K). A true-residual evaluation costs 111.5 ms per candidate; the random, greedy, and value rows are timed end to end, whereas the residual-free and beam rows are composed from the profiled primitives (feature plus scoring cost per candidate times K; 24 true evaluations plus overhead); the residual-free figure is an upper bound, since the deployed rollout encodes the graph once per step and scores all candidates together. Right: fraction of the greedy−random gap recovered, averaged over bands with gap $\geq$ 10 pp.

Method	ms/decision	uses true residuals	Method	band-avg. recovery
random	144	one per step (trajectory record)	value	23.5%
greedy	1048	yes	value$_{\text{PE}}$	21.6%
value	972	yes	BC	−4.0%
residual-free policies	∼188	no	beam	134.6%

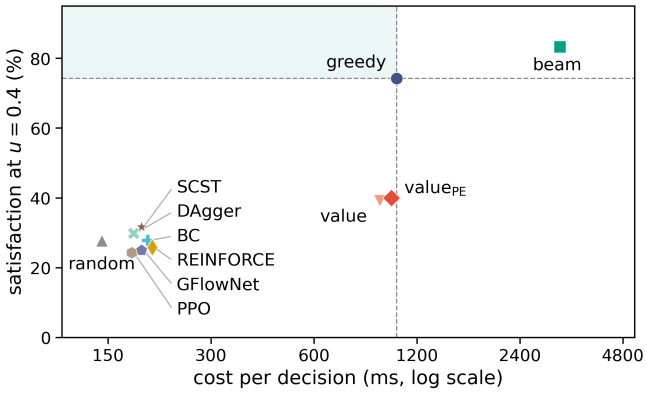


Figure 4: Cost per decision versus satisfaction at u=0.4 (unified benchmark; Tables 1 and 10). Beam's cost is computed from its 24 true evaluations per step (width 4, 6 expansions) at 111.5 ms each, with non-oracle overhead scaled accordingly (3144 ms); the value pair and the residual-free cluster are dodged horizontally by a few percent for marker visibility. The shaded region is the cheap-and-good quadrant of Finding 5: cheaper than greedy and at least as good. It is empty.

greedy at K=8. Greedy at every $K \leq 8$ dominates the learner on both axes under either accounting; K=16 and K=32 buy more satisfaction for more calls. Figure 4 plots the two axes jointly.

H.3 Horizon-matched evaluation of residual-free policies

Residual-free policies decide at least ∼5.6× faster than greedy by component costs (Table 10; not an end-to-end wall-clock match), so their evaluation horizon is extended from T=60 to T=340 swaps (5.7×) without retraining. BC, DAgger, SCST, and a random control are rolled out for 340 swaps, and the T=60 score is read off the prefix of the same trajectory (10 seeds, 12 bands, held-out WS targets, paired within seed). Averaged over $u \in [0.5, 0.2]$, the paired gain $\text{sat}(T{=}340) - \text{sat}(T{=}60)$ is +0.0 (BC), +1.3 (DAgger), and +0.4 pp (SCST) against +0.8 pp for random. No method×band cell reaches p<0.05 (largest +2.5 ± 1.3 pp, DAgger at u=0.6, p=0.10).

H.4 Screen-then-verify hybrid

The standard amortization hybrid ranks a large candidate pool with a residual-free scorer and spends the oracle only on the top few. Under the unified protocol (10 seeds, 15 held-out WS targets, T=60), a pool of $K_{\text{screen}} \in \{64, 256\}$ swaps is drawn per step with the benchmark's sampler. A scorer (BC or SCST retrained with the unified recipe, or the LLM-evolved function of the LLM-evolved scoring study below) ranks it at zero oracle cost. Exactly the top 8 are true-evaluated and the best applied, so every screened arm spends greedy's 8 calls per step. Controls: greedy on 8 random candidates of the same pool (*pool-8*) and an oracle screen that evaluates all 64 (*pool-64*, 64 calls per step). Table 11 shows that every learned screen is far *below* plain greedy at equal cost (−19 to −45 pp at

u=0.4, all p<0.05) and degrades with the larger pool, while the oracle screen would add +13 to +26 pp over u=0.4–0.2. The pool thus contains better swaps than the 8 random ones, but a learned ranking selects a worse subset than chance. This is consistent with the absence of a local ranking signal (Appendix I).

Table 11: Screen-then-verify (satisfaction %, 10 seeds; paired delta vs. greedy with K=8 in pp, mean±SE, all p<0.05 for the learned screens). Each screened arm true-evaluates 8 candidates per step; pool-64 evaluates 64.

	calls/step	u=0.6	u=0.4	u=0.3	u=0.2	Δ vs. greedy at u=0.4
random	0	64.0	27.3	12.0	5.3	-47.3 ± 2.9
greedy, K=8	8	87.3	74.7	61.3	45.3	—
pool-8 (random 8 of the pool)	8	89.3	76.7	66.0	46.7	$+2.0 \pm 4.1$
pool-64 (oracle screen)	64	91.3	88.0	78.7	71.3	$+13.3 \pm 3.7$
screen BC, pool 64 / 256	8	72.0 / 70.7	55.3 / 52.7	42.7 / 38.7	27.3 / 22.7	-19.4 ± 7.1 / -22.0 ± 8.4
screen SCST, pool 64 / 256	8	70.0 / 63.3	46.7 / 42.0	34.0 / 29.3	17.3 / 12.0	-28.0 ± 5.9 / -32.7 ± 5.5
screen EoH, pool 64 / 256	8	62.0 / 50.7	38.0 / 29.3	29.3 / 20.7	19.3 / 8.7	-36.7 ± 5.6 / -45.3 ± 4.6

H.5 Divergent paradigms and architecture transplant under the unified protocol

Table 12 reports the three divergent runs of Table 1 on the full 12-band grid (10 seeds, held-out WS targets), each with its own paired random and greedy baselines. GFlowNet (best strengthened configuration) and DMGG-factorized train once at u=0.3 and are scored at all bands, while MPC replans per band.

Table 12: Unified-protocol divergent runs (satisfaction %, 10 seeds, all 12 bands): each method column is flanked by the paired random and greedy baselines of its own run. DMGG-factorized is the transplanted architecture of Choi et al. (2026).

	GFlowNet run			MPC run			DMGG-factorized run		
Band	random	GFlowNet	greedy	random	MPC	greedy	random	DMGG	greedy
u=1.0	94.2	82.9	95.8	89.3	86.7	93.3	94.2	87.5	95.0
u=0.8	75.8	69.6	95.8	76.0	79.3	91.3	85.8	77.5	93.3
u=0.6	54.2	48.3	92.5	54.7	62.0	86.7	61.7	64.2	90.8
u=0.5	43.3	37.4	84.2	44.7	52.0	82.0	46.7	48.3	89.2
u=0.45	35.8	31.0	82.5	40.0	40.0	79.3	37.5	37.5	86.7
u=0.4	27.5	25.0	79.2	34.7	34.7	75.3	26.7	31.7	83.4
u=0.35	20.0	18.9	71.7	28.7	25.3	68.7	22.5	25.0	75.0
u=0.3	12.5	14.0	64.2	18.0	19.4	62.7	13.3	20.8	67.5
u=0.25	9.2	9.4	55.8	12.7	10.7	55.3	5.0	10.0	59.2
u=0.2	2.5	4.6	48.3	3.3	7.4	44.7	4.2	5.0	45.8
u=0.15	1.7	1.4	39.2	0.7	4.0	32.0	2.5	3.3	30.8
u=0.1	0.0	0.1	22.5	0.0	0.7	16.0	0.0	2.5	20.0

Paired statistics (per-seed, 10 seeds): greedy exceeds each of the three methods with $p \leq 0.01$ at every band (most $p < 0.001$). GFlowNet is significantly below its paired random at u=1.0 (-11.3 ± 3.3, p=0.008) and statistically indistinguishable from it at every other band (-6.2 to $+2.1$ pp). MPC is indistinguishable from its paired random throughout (-3.4 to $+7.3$ pp, all $p \geq 0.09$). DMGG-factorized is significantly below its paired random at loose bands (-6.7 pp, p=0.036 at u=1.0; -8.3 pp, p=0.023 at u=0.8) and indistinguishable from it at every tighter band (up to $+7.5$ pp at u=0.3, p=0.11).

Differentiable guidance, a one-shot generator, is scored by the final graph's worst-scaled residual (10 seeds, 12 bands, held-out WS targets; disconnected outputs count as unsatisfied) against a paired one-shot random baseline. It reaches 60.8% versus 87.1% at u=1.0 (-26.3 ± 3.0 pp, p<0.001) and at most +7.3 pp above it at mid bands (significant only at u=0.5, $+5.7 \pm 2.3$, p=0.037). It reaches ≤1% at tight bands, and sits at or below the benchmark's trajectory random floor at every band (Table 4).

H.6 GFlowNet strengthening study

Six configurations (base, tempering, ε-exploration, prioritized target resampling, subtrajectory balance, combination; three bands, 2 seeds, budgets matched to the value learner) give mean satisfaction minus random of -6.4 (explore), -6.7 (resampling), -8.9 (subtrajectory balance and combination), -11.6 (tempering), and -12.8 pp (base). All remain far below greedy, with reward mass collapsing toward zero at tight bands.

H.7 LLM-evolved scoring

Six generations (60 candidate functions) raise the rate at which the scorer's top pick lies in the true top three (agreement@3, measured on the evolution fitness pool) from 0.255 (best seed function) to 0.260. The best evolved function is dominated by the closed-form assortativity increment (Lemma 1).

H.8 Classical null models

dK-1 (configuration model) and dK-2 (joint-degree) samples (5 seeds $\times$ 12 bands $\times$ 4 target families) satisfy at most 10.8% and 15.8% at the loosest band (dK-1: 0% for WS and soft-RGG targets), collapsing to 0–8% at u=0.4.

Exponential random graph models (ERGMs). A degree-only ERGM constrains degrees only in expectation and is not the hard fixed-degree ensemble above. The standard extension that targets clustering (edges, triangles, two-stars, assortativity numerator) is degenerate on these clustered targets. Fitted by stochastic approximation to WS targets, its expected triangle count stays 80–89% below target while the natural parameter diverges (Handcock, 2003; Schweinberger, 2011). No satisfaction rate is therefore reported for that family. The curved family designed for this degeneracy (Hunter, 2007) uses edges plus geometrically weighted edgewise-shared-partner (GWESP) and degree (GWDEGREE) terms (decay 0.5; `ergm` 4.12). It converges on $20/20$ WS-family targets (simulated edge counts within $\pm 3\%$), but 30 draws per fitted model satisfy the five-metric bands at 1.3% (u=1.0), 0.3% (u=0.6), and 0.0% (u=0.4) versus greedy's $94.8/88.3/74.2\%$. A maximum-entropy ensemble around its sufficient statistics is not a band-targeted generator.

H.9 Samplers and the sampling-frame generator

Microcanonical samplers with the true residual as energy (12 seeds) order MH $>$ SA $>$ random except at the most extreme band, where all collapse to 1–2%. At u=0.2, MH reaches 11.1%, SA 5.6%, random 4.9%, and greedy 58.3%. The greedy-over-MH gap peaks at $+47.2$ pp while greedy spends 7–10$\times$ more evaluations. A residual-free learned generator trained in the sampling frame (potential shaping, terminal on satisfaction, 6 seeds) is below MH at every band (-2 to -13 pp).

I Signal, scaling, and feature studies

Collision statistics. On 16,000 candidate swaps from WS optimization states, the collision rate is 0.967 (share of candidates whose endpoint-degree local view coincides with another candidate's). The confusable-state rate is 0.707 (states containing locally identical candidates with different true improvements). The within-cluster variance fraction is 0.931 (improvement variance unexplained after clustering the standardized local-feature vectors into 40 k-means clusters), and the local-Bayes top-1 regret is 0.2006 (loss of the cluster-mean selector, fit in sample, against the true best). Grouping by exact endpoint-degree signature instead gives a within-class variance fraction of 0.13 on the greedy-trajectory set of the marked-ball audit below, because exact classes are small; the two statistics are not comparable. Along the warm-start heterogeneity sweep of Appendix K (10 seeds $\times$ 3,200 candidates per rung at u=0.4, same statistics), the endpoint-degree collision rate is 0.96 (WS, $\sigma_e^2 \approx 2$), 0.72 (Erdős–Rényi, ER), and $0.98/0.91/0.80$ (power-law γ=3.0/2.5/2.1, $\sigma_e^2 \approx 128$–355; SE ≤ 0.04). The within-cluster variance fraction is 0.88–0.95 on the WS and power-law rungs (0.66 ± 0.14 on ER). Even under heavy-tailed degrees, low-degree endpoints dominate candidate edge pairs, so the endpoint-degree view remains non-unique and the local-Bayes regret (0.11–0.37) does not vanish.

Exact marked-ball audit. Under the theorem's own equivalence (exact isomorphism of role-marked radius-ρ balls, verified within Weisfeiler–Leman (WL) hash groups), 14,400 candidates from greedy-trajectory states collide at rate 0.001 (ρ=1) and 0.000 (ρ=2) (confusable-state rates 0.011 and 0.000), while the endpoint-degree signature on the same states collides at 0.337. Exact bounded-radius indistinguishability is therefore rare on the benchmark. The theorem is the worst-case counterpart of the feature-level statistics above, not the typical mechanism.

Supervised signal-recovery scaling. Two grids regress true per-swap improvements on GraphGPS representations (Rampasek et al., 2022). A pilot grid (64 cells; 15k–3.8M parameters $\times$ 500–9,000 pairs $\times$ attention on/off $\times$ Laplacian PE on/off) peaks at the 80k size (mean R^2 0.070) and decreases with size (3.8M: 0.006). A ranking-aware grid (40 cells; up to 6.5M parameters, 19,000 pairs) rises with data and saturates at $R^2 \approx 0.40$. At the best cell Spearman is 0.18, Kendall 0.13, pairwise accuracy 0.57 (random 0.5), top-1 regret 0.19, and Laplacian encodings match no encoding. Sizes of 11.2M and 24.7M parameters under the fixed protocol leave R^2 at zero, and rolling the scorer out as a closed-loop policy gives satisfaction scattered around the random baseline (42–83% across sizes vs. random 50%, greedy 100% on that probe).

Controlled scaling (budget and learning rate). Larger models in the ranking grid (2.2M, 6.5M) collapse to constant outputs under the fixed 150-epoch budget and shared learning rate. On the same 19,000-pair dataset we therefore sweep the learning rate ($\{10^{-3}, 3\times10^{-4}, 10^{-4}\}$, selected on validation R^2) under a fixed budget and under a parameter-proportional budget (150/748/4,214/12,245 epochs for 80k/0.4M/2.2M/6.5M). With a tuned learning rate no size collapses (2.2M: R^2=0.242; 6.5M: 0.260 at the fixed budget). With parameter-proportional budgets the best test R^2 is 0.239/0.262/0.285/0.255, a 0.24–0.29 plateau with no size trend (Δ at 6.5M: -0.006 despite an 82$\times$ budget). No size at any learning rate or budget exceeds the $R^2 \approx 0.40$ saturation of the data-scaling grid.

Label noise in the supervised targets. The regression labels inherit heuristic noise through the modularity term. Q is scored by CNM, whose output depends on node order, so re-scoring the same graph under node relabelings moves the label (a relabeling sensitivity, not repeated-measurement noise under the benchmark's fixed labels). Over 50 graphs the per-graph Q standard deviation is 0.0045 (max 0.0093; $\approx$ 0.09 residual units at ε_Q=0.05). Propagated to per-swap improvement labels, the noise variance is of the same order as the label variance, so the supervised R^2 values are likely attenuated by it. Re-collecting one matched cell (300 states, 2,000 pairs) with Louvain in place of CNM (Q sd 0.0027; trajectories change with the metric, so this is not a pure relabeling) raises held-out R^2 from 0.099 to 0.256. Greedy consumes the same noisy oracle at every step and dominates every learner (Table 1), so the noise qualifies the diagnostic scale, not the comparative conclusions.

Per-step-recomputed spectral features. Scoring candidates by the current graph's Fiedler-vector endpoint components (recomputed once per step, $\sim$11 ms, versus greedy's $K \times 111$ ms) adds a small supervised ranking signal over local features (R^2 0.004 $\rightarrow$ 0.025; Spearman $-0.051 \rightarrow +0.043$; pairwise accuracy 0.484 $\rightarrow$ 0.515). The closed-loop policy, however, satisfies 27% versus random 27%, a local-features policy 33%, and greedy 93% at u=0.4 on that probe. The signal is measurable but policy-inert.

Horizon sweep (state values versus action ranking). Predicting the cumulative improvement of an h-step greedy run from the starting state (n=400, held-out $\approx$ 80), R^2 at h=1/5/10/30/50 is 0.05/0.14/0.29/0.57/0.60 from target-distance features alone and 0.06/0.23/0.50/0.80/0.83 with global spectral features. Long-horizon state values are predictable, while the action-ranking quantities above are not (§6).

J THEORY

Notation: G simple, undirected, connected; $\mathcal{L} = I - D^{-1/2}AD^{-1/2}$; λ_2 its second-smallest eigenvalue; for $S \subseteq V$, $\mathrm{vol}(S) = \sum_{v\in S} d_v$, $\mathrm{cut}(S)$ the boundary edge count; conductance

$h(G) = \min_{0<\mathrm{vol}(S)\le m} \mathrm{cut}(S)/\mathrm{vol}(S)$; partition $\mathcal{P} = \{C_c\}$, $p_c = \mathrm{vol}(C_c)/2m$, modularity $Q(\mathcal{P}) = \sum_c [m_c/m - p_c^2]$. All numerical checks ship as scripts with seeds and expected outputs.

J.1 Unconditional exclusion bound and partition-uniform certificate

Lemma 2 (Unconditional exclusion bound). *For every partition $\mathcal{P}$ with $s := \sum_c p_c^2 < 1$,*

$$\lambda_2(G) \;\le\; \frac{1 - Q(\mathcal{P}) - s}{1 - s}.$$

Moreover $\lambda_2(G) \le 2\big(1 - Q(\mathcal{P})\big)$ for every partition; when $Q(\mathcal{P}) \ge 0$ the first bound gives the sharper $\lambda_2(G) \le 1 - Q(\mathcal{P})$ (since $\frac{1-Q-s}{1-s} \le 1 - Q \iff Qs \ge 0$), and the $2(1-Q)$ form, which also covers $Q < 0$, is the conservative one used by the certificate below.

Proof. For each part set $y_c = \mathbf{1}_{C_c} - p_c\mathbf{1}$ and $x_c = D^{1/2}y_c$; then $x_c^\top D^{1/2}\mathbf{1} = \mathrm{vol}(C_c) - 2mp_c = 0$, so x_c is orthogonal to the trivial eigenvector and Rayleigh–Ritz gives $\lambda_2 \le (x_c^\top \mathcal{L} x_c)/(x_c^\top x_c)$. Compute $x_c^\top x_c = 2m\,p_c(1-p_c)$ and $x_c^\top \mathcal{L} x_c = \mathrm{cut}(C_c)$. Averaging these bounds with weights $p_c(1-p_c)$, using $\sum_c \mathrm{cut}(C_c) = 2E_{\mathrm{inter}}$, the modularity identity $\sum_c m_c/m = Q+s$, and $\sum_c p_c(1-p_c) = 1 - s$ yields the first inequality. For the second bound: when $Q \ge 0$, $2(1-Q) - \frac{1-Q-s}{1-s} = \frac{(1-Q-s)+2Qs}{1-s} \ge 0$ (the numerator is $1 - Q - s(1-2Q) \ge \min(Q, 1-Q) \ge 0$ for $Q \in [0,1]$, $s < 1$), so it follows from the first; when $Q < 0$ it holds directly from $\lambda_2 \le 2 < 2(1-Q)$. □

The bound is a task-specific averaged Rayleigh-quotient corollary. Connections among modularity, normalized cut, and expansion are classical (Fasino & Tudisco, 2014; Newman, 2013; Louf et al., 2024). Exhaustive verification on the 995 connected atlas graphs (771,016 partitions): zero violations (maximal numerical excess 2.2×10^{-16}).

Corollary 1 (Partition-uniform band-infeasibility certificate). *Let $Q_{\mathrm{alg}}(G)$ be the modularity of the partition returned by* any *fixed algorithm. With $W_{Q\lambda}(G) = \max\{|Q_{\mathrm{alg}}(G) - Q^\star|/(u\varepsilon_Q),\ |\lambda_2(G) - \lambda_2^\star|/(u\varepsilon_\lambda)\}$ (the $\{Q, \lambda_2\}$-restriction of the scaled residual at band u; satisfaction requires $W_{Q\lambda} \le 1$), unconditionally*

$$W_{Q\lambda}(G) \;\ge\; \frac{[\lambda_2^\star - 2(1 - Q^\star)]_+}{u\,(\varepsilon_\lambda + 2\varepsilon_Q)}.$$

If the right-hand side exceeds 1, the target is infeasible for the band. This is decidable before any search, and valid exactly as benchmarks report Q, because Lemma 2 holds for the returned partition itself (the sharper $1 - Q$ form requires $Q_{\mathrm{alg}}(G) \ge 0$ on the realized graph, guaranteed when $Q^\star - u\varepsilon_Q > 0$, which holds for every target and band here). It is a sufficient test (a right-hand side ≤ 1 does not imply feasibility). For bands with $Q^\star - u\varepsilon_Q > 0$ (every satisfying graph then has $Q > 0$, hence a nontrivial partition with $s < 1$), the same argument with $1 - Q^\star$ in place of $2(1 - Q^\star)$ gives the sharper certificate $[\lambda_2^\star - (1 - Q^\star)]_+/(u(\varepsilon_\lambda + \varepsilon_Q)) > 1$. Every target in this paper has $Q^\star > 0$ (minimum 0.235), so the sharper form is the one reported in the main text and in Appendix K. It fires on no target that any solver reaches.

Verification under the CNM heuristic: zero violations of $\lambda_2 \le 2(1 - Q_{\mathrm{CNM}})$ across the 994 ($n \ge 3$) atlas graphs (minimum slack $+0.500$) and 17 real networks (minimum slack $+0.272$; the smallest slacks belong to high-modularity sparse networks, consistent with their observed near-unreachability).

J.2 Assortativity increment and step-budget bound

Lemma 3 (Restatement of Lemma 1; standard, cf. Van Mieghem et al., 2010). *With $\sigma_e^2 > 0$, the swap $(a,b),(c,d) \to (a,d),(c,b)$ gives $\Delta r = (d_a - d_c)(d_d - d_b)/(m\sigma_e^2)$.*

Proof. Degrees are preserved, so only the edge degree-product sum changes: $(d_a d_d + d_c d_b) - (d_a d_b + d_c d_d) = (d_a - d_c)(d_d - d_b)$. □

Verified against NetworkX over 6,405 valid swaps (max error 5.6×10^{-15}). A caution: small degree *spread* does not imply small per-swap mobility, because σ_e^2 in the denominator can be small as well

(a cycle with two chords has spread 1 yet $|\Delta r| \to 1/6$). Mobility is governed jointly by endpoint degree differences and σ_e^2.

Corollary 2 (Step-budget lower bound). *With $\delta_r^{\max} = (d_{\max} - d_{\min})^2/(m\sigma_e^2)$, reaching $|r - r^\star| \le u\varepsilon_r$ from r_0 requires at least $\lceil [\,|r^\star - r_0| - u\varepsilon_r\,]_+/\delta_r^{\max} \rceil$ swaps.*

J.3 Average-case diagnostic and a cautionary counterexample

The one-step diagnostic regresses the change of the worst-scaled residual on candidate-local features: $R^2 = 0.009$ along trajectories (per-seed $0.006/0.010/0.012$), 0.115 at the warm start; Spearman 0.26; median argmax gain over random -0.01; median regret versus best 0.082. Three scope notes. (i) These characterize the tested feature sets. The feature-optimal correlation ratio $\eta^2 = \mathrm{Var}(\mathbb{E}[\Delta \mid F])/\mathrm{Var}(\Delta)$ is not upper-bounded by a measured R^2. The collision statistics of Appendix I give a complementary, model-free view (the regret of an in-sample 40-cluster selector on those features), which bounds only that quantized selector, not the feature-optimal one. (ii) Pearson correlation alone cannot bound selection gain: for $S \sim \mathrm{U}(-1,1)$, $\Delta = S^2$, $\mathrm{Corr}(S,\Delta) = 0$ exactly, yet argmax selection with $K{=}3$ gains $\mathbb{E}[(\max_i S_i)^2] - \frac{1}{3} = \frac{1}{15}$. Correlation-based bounds are therefore avoided throughout. (iii) All label-dependent quantities here (per-swap R^2, argmax gain, regret) inherit the CNM label noise quantified in Appendix I; R^2 is likely attenuated by it, whereas the regret can move in either direction. The label-free collision statistics are unaffected.

J.4 Local-action indistinguishability (formal version of Theorem 1)

Definition 2 (Marked ρ-ball; radius-ρ scorers). *For a valid swap $s = ((a,b),(c,d))$ on G, $F_\rho(G,s)$ is the isomorphism class of the subgraph induced by vertices within distance ρ of $\{a,b,c,d\}$, endpoints marked by role. A* radius-ρ scorer *assigns each candidate a score depending only on $F_\rho(G,s)$. The induced policy selects uniformly among score-maximal candidates (symmetric tie-breaking), and Π_ρ denotes such policies. Improvement is $I(s) = \Phi(G) - \Phi(G \circ s)$ and* $\mathrm{regret}(s) = \max_{s'} I(s') - I(s)$.

Lemma 4 (Ties). *If $F_\rho(G,s_1) = F_\rho(G,s_2)$, every radius-$\rho$ scorer scores them equally, and on the two-candidate decision $A = \{s_1,s_2\}$ every $\pi \in \Pi_\rho$ picks each with probability $\frac{1}{2}$; hence $\mathbb{E}[\mathrm{regret}] = \frac{1}{2}|I(s_1) - I(s_2)|$, and if exactly one candidate reaches the target band, π succeeds with probability $\frac{1}{2}$ versus 1 for the oracle. (A deterministic tie-break by enumeration order could evade the $\frac{1}{2}$ conclusion. The equal-score statement holds for any radius-ρ scorer.)*

Construction (two-expander bridge gadget). Fix $d \ge 4$ and radius ρ. Let H be d-regular on n vertices with girth exceeding $\max(2\rho{+}2,\, c_1)$ and edge expansion $\beta > 0$, where $c_1 = (\kappa{+}E_0{+}1)/\beta$ is the constant of Lemma 6 (so the girth requirement is a constant depending only on d, ρ, β). This discharges the girth-$> c_1$ hypothesis of that lemma, since deleting edges does not decrease girth. Fix a spanning tree T and choose four non-tree edges $e = (u,v)$, $f = (x,y)$, (p,q), (w,z), pairwise at distance $\ge \max(2\rho+3, c_1)$ with the same c_1. A greedy choice succeeds for large n, since non-tree edges number $\Omega(n)$ while each choice excludes $O((d-1)^{c_1})$ others. Build G: two copies H_1, H_2; delete e, f in both; add bridges $u_1u_2, v_1v_2, x_1x_2, y_1y_2$. G is d-regular and connected (each copy retains T), with girth $\ge 2\rho+3$. Set $s_1 = ((u_1,u_2),(v_2,v_1))$ (rewires two bridges back into the deleted internal edges; copy-to-copy cut $4 \to 2$) and $s_2 = ((p_1,q_1),(w_1,z_1))$ (two interior non-tree edges of H_1; cut unchanged). Both swaps are valid and connectivity-preserving. Write $G_1 := G \circ s_1$ and $G_2 := G \circ s_2$.

Lemma 5 (Local indistinguishability). *$F_\rho(G,s_1) = F_\rho(G,s_2)$; all valid candidates whose two edges are $\ge 2\rho+2$ apart share the same F_ρ: two disjoint marked depth-ρ d-regular edge-trees.*

Proof. Girth $\ge 2\rho+3$ makes the ρ-ball of every edge the complete depth-ρ d-regular tree; distance $\ge 2\rho+2$ makes the two balls disjoint and non-adjacent; roles match root-to-root. □

Lemma 6 (Bottleneck localization). *Let κ be the bridge count (4 for G, G_2; 2 for G_1), $E_0 \le 4$ the per-copy edge removals, $c_1 = (\kappa + E_0 + 1)/\beta$. If $n > 2c_1$, each edited copy has girth $> c_1$, and marked vertices are pairwise ≥ 4 apart within their copy, then $h(G') = \kappa/(dn)$ for $G' \in \{G, G_1, G_2\}$, attained exactly by the two copy sides; every other S with $\mathrm{vol}(S) \le dn$ has $\mathrm{cut}(S) \ge \kappa + 1$ or a strictly larger ratio.*

Proof. Write $S = A \cup B$ per copy, $a = |A|$, $b = |B|$, $\alpha_i = \min(|S_i|, n - |S_i|)$. *Case 1* ($\alpha_1 > c_1$ or $\alpha_2 > c_1$): the internal boundary of that copy gives $\mathrm{cut}(S) \geq \beta c_1 - E_0 = \kappa + 1$. *Case 2a* ($a, b \leq c_1$): both sides induce forests ($a, b <$ girth). If one side is empty, $\mathrm{cut}(S) \geq (d-2)a + 2$ with ratio $> \kappa/(dn)$; if both are nonempty, bridges inside S are vertex-disjoint (each marked vertex lies on one bridge), so at most $|S|/2$ of them, and $\mathrm{cut}(S) \geq (d-3)|S| + 4 \geq 4 \geq \kappa$ with ratio $> \kappa/(dn)$ since $|S| \leq 2c_1 < n$. *Case 2b* ($a \geq n - c_1, b \leq c_1$): let $T' = V(H_1) \setminus A$, $k = |T'|$; if $k = 0$ then $a = n$, and the volume constraint $\mathrm{vol}(S) \leq dn$ forces $b = 0$, so $S = V(H_1)$ with $\mathrm{cut} = \kappa$. For $k \geq 1$, decomposing T' (and B) into tree components and counting internal boundary plus bridge crossings gives $\mathrm{cut}(S) \geq \kappa + [(d-2)k + 2c_{T'} - 2\tau_{T'}] + [(d-2)b + 2c_B - 2\beta_B]\mathbf{1}[b \geq 1]$, with τ, β counting marked vertices and c components. Componentwise, a tree component with q vertices and j marks contributes $(d-2)q + 2 - 2j \geq 1$: j=0 gives $\geq d$; j=1 gives $(d-2)q \geq d-2 \geq 2$; for $j \geq 2$, marks are pairwise $\geq D$=4 apart, so doubling the tree yields a closed walk of length $2|E(C)|$ visiting all marks with cyclic gaps $\geq D$, whence $q \geq jD/2 + 1$ and $(d-2)q + 2 - 2j \geq j(\frac{D}{2} - 2) + 3 \geq 3$. Hence $\mathrm{cut}(S) \geq \kappa + 1$ off the two copy sides. □

Theorem 2 (Formal version of Theorem 1)**.** *For every radius ρ and degree $d \geq 4$ there exist arbitrarily large connected d-regular graphs G and valid swaps s_1, s_2 with (i) $F_\rho(G, s_1) = F_\rho(G, s_2)$; (ii) $h(G) = h(G \circ s_2) = 4/(dn)$ and $h(G \circ s_1) = 2/(dn)$, so $|I(s_1) - I(s_2)| = 2/(dn)$ for $\Phi = h$; (iii) on the two-candidate decision $\{s_1, s_2\}$, every $\pi \in \Pi_\rho$ has $\mathbb{E}[\mathrm{regret}] = 1/(dn)$ and succeeds with probability $\frac{1}{2}$ on the band "$h \in [0, 3/(dn)]$" versus 1 for the oracle. The absolute gap is $\Theta(1/n)$; the size-independent statement is the relative regret $\frac{1}{2}$.*

Proof. Combine Lemmas 5, 6, 4. Existence for every fixed (d, ρ): a uniformly random d-regular graph has, with probability bounded away from zero, expansion bounded below (Friedman's theorem (Friedman, 2008; Bordenave, 2020)) and girth exceeding $\max(2\rho + 2, c_1)$. The latter is still a fixed constant length once (d, ρ, β) are fixed, so the same Poisson-short-cycle argument applies (McKay et al., 2004). Intersecting the events and choosing the four non-tree edges greedily satisfies all hypotheses. □

Proposition 1 (λ_2 tracks the cut)**.** *If additionally H has internal normalized-Laplacian gap $\geq \gamma = \Omega(1)$ and $d \geq d_0(\gamma)$ for a constant d_0 depending only on γ, then $\lambda_2(G_\kappa) = \Theta_\gamma(\kappa/(dn))$ for $\kappa \in \{2, 4\}$.*

Proof sketch. Upper bound: the Cheeger inequality $\lambda_2 \leq 2h$ (Chung, 1997). Lower bound: decompose the lazy walk by the two copies. The projected two-state chain has gap $\Theta(\kappa/(dn))$, and each restricted chain is an expander walk with at most $E_0 = O(1)$ edited edges. Editing one edge of a d-regular graph perturbs the normalized adjacency by spectral norm $O(1/d)$ (the difference matrix has $O(1)$ entries of magnitude $1/d$), so the restricted gap is $\geq \gamma - cE_0/d \geq \gamma/2$ once $d \geq d_0(\gamma) := 2cE_0/\gamma$. The Markov-chain decomposition theorem (Jerrum et al., 2004) multiplies the two. □

Scope. (i) The rigorous separation is for conductance. Proposition 1 gives the correct order for λ_2 (observed contrast a stable factor ≈ 2.0 for $n \in \{100, \ldots, 400\}$). A strict constant-factor λ_2 inequality is open. (ii) s_2 leaves conductance unchanged. Other global metrics may move slightly. (iii) The theorem exhibits a failure mode whose prevalence on the benchmark is measured in Appendix I. Scorers with symmetry-breaking inputs (positional encodings, random identifiers, spectral features, attention, history) are outside it and are probed empirically. (iv) Bounded-radius indistinguishability on high-girth regular graphs is classical in local algorithms and GNN expressivity (Loukas, 2020; Garg et al., 2020). Here it is stated as action-ranking regret. (v) Numerical illustration (ρ=1, d=3, n=200; at d=3 the separation holds by direct check). Marked balls are isomorphic, the cut goes $4 \to 2$ versus unchanged, and the λ_2 ratio is 0.497 versus 1.000. Twenty random-weight radius-1 message-passing scorers give score differences exactly 0. Bottleneck localization is verified over all 2^{20} vertex subsets of a Petersen-based gadget for each of G, G_1, G_2 (a small instance outside the lemma's girth and separation hypotheses, checked exhaustively rather than through the lemma).

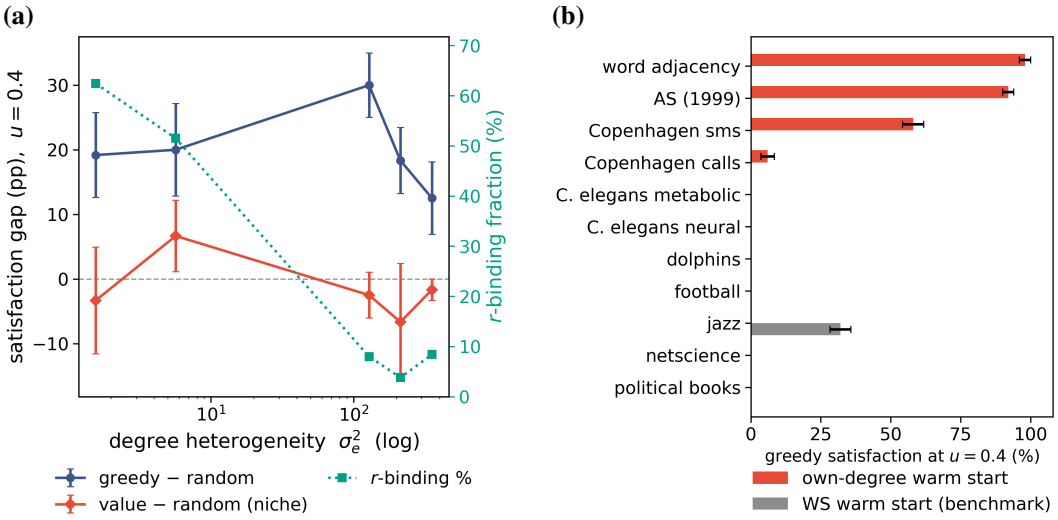


Figure 5: **Warm-start controls** (u=0.4; 10 seeds; error bars ±SE). *(a)* Greedy−random and value−random gaps (left axis) and the r-binding fraction (right axis, dotted) vs. warm-start degree heterogeneity σ_e^2. *(b)* Greedy satisfaction on the 11 real networks admitting an own-degree warm start, from their own degree sequence vs. the WS warm start; absent bars are 0%. Interpretation in the text.

K GENERALIZATION DETAILS

Warm-start heterogeneity sweep (Figure 5a). Targets are drawn from each warm start's own degree sequence, so they are feasible by construction. Across the sweep the r-binding fraction falls 62% → 8%, consistent with the inverse-variance scaling of Lemma 1; greedy−random stays +12–30 pp, and the value−random niche is indistinguishable from zero throughout (niche vs. $\log \sigma_e^2$ correlation −0.10, p=0.51). The sweep is therefore a null direct test of Definition 1: the criterion predicts a niche that scales with the r-binding fraction, and none appears. The WS rung is a separate target draw (random 31.7% vs. 27.7% in the unified benchmark at u=0.4). Its place among the re-runs of this cell is given in Table 6. In panel (b), own-degree warm starts rescue three networks (0% → 58–98%) and partially a fourth (6%), while seven remain at 0%; jazz alone is reached from the WS start under this own-degree protocol (32%), not under the $\langle k \rangle$=6 probe.

Peak migration of the search advantage. The greedy−random curve is unimodal in tolerance. As N grows its peak moves from u≈0.37 (N=100) to u≈0.26 (N=2000) (correlation with $\log N$ −0.85, p=0.015, n=7). The advantage shrinks at loose bands (+24 → +5 pp at u=0.6, N=200 → 2000, $\langle k \rangle$=6) but not at the tightest (+20 → +23 pp at u=0.1). Random keeps pace through the mid bands at large N, relocating greedy's advantage to the tight tolerances (the opposite of the vanishing learning niche, Figure 1b).

Regime summary. Table 13 is the one-glance regime summary referenced from §7. Figure 6 expands it to all bands, including the accelerated-metric large-N cells. In the figure, the niche is unimodal in tolerance across the small-to-mid rows and decays down the size column (large-N rows are noisier); the mixed family is positive only at loose bands, where any swap passes. The N=800, $\langle k \rangle$=8 row (the degree-dependent persistence discussed in §7) stays positive through the mid-tight bands (+12.0/ + 11.0 at u=0.35/0.3), unlike N=800, $\langle k \rangle$=6. Paired per-cell tests on the large-N rows: at u=0.4 the niche is small but positive in every cell, significant only at N=2000, $\langle k \rangle$=8 (p=0.042). At u=0.3 it is sign-unstable (N=1200, $\langle k \rangle$=4 significantly negative, p=0.015). Greedy−random stays +10 to +23 pp at u=0.4 throughout.

The greedy−random gap alone was additionally verified at +33 to +50 pp at u=0.4 across all small-to-mid $(N, \langle k \rangle)$ cells (N=100–400). The one-step local-feature R^2 (warm-start states, cf. App. J.3) rises from 0.109 (N=100) to 0.397 (N=400). The binding-fraction mechanism explains

Table 13: **Generalization across regimes** (satisfaction gaps in pp at u=0.4 unless noted; 10 seeds for all WS regimes (mixed family: 4); $N \geq 1200$ uses accelerated exact-spectral/sampled-path metrics, App. A). Niche p-values are nominal per cell; no large-N cell survives family-wise correction (Appendix E). LOFO = leave-one-family-out (train on two families, evaluate on the held-out one; u=0.6–0.2 range).

Regime	greedy−random	value−random	niche verdict (95% CI)
N=100–200, $\langle k\rangle \in \{4, 6, 8\}$	+40 to +50	+3.0 to +17.5	sig. + at $\langle k\rangle$=6 ([+6.5, +27.5], N=100)
N=400	+33 to +40	+7.5 to +13.0	significant + ([+4.2, +21.8])
N=800, $\langle k\rangle$=6	+20	−1.0	0 in CI [−6.0, +4.5], p=0.74
N=1200–2000	+10 to +23	+3.3 to +8.3	small +; sig. only at N=2000, $\langle k\rangle$=8
mixed family (feasible)	+29	−2.1	0 in CI [−8.3, +4.2], p=0.64
LOFO, held-out SBM	+19 to +42	−9.2 to −0.4	learner at/below random

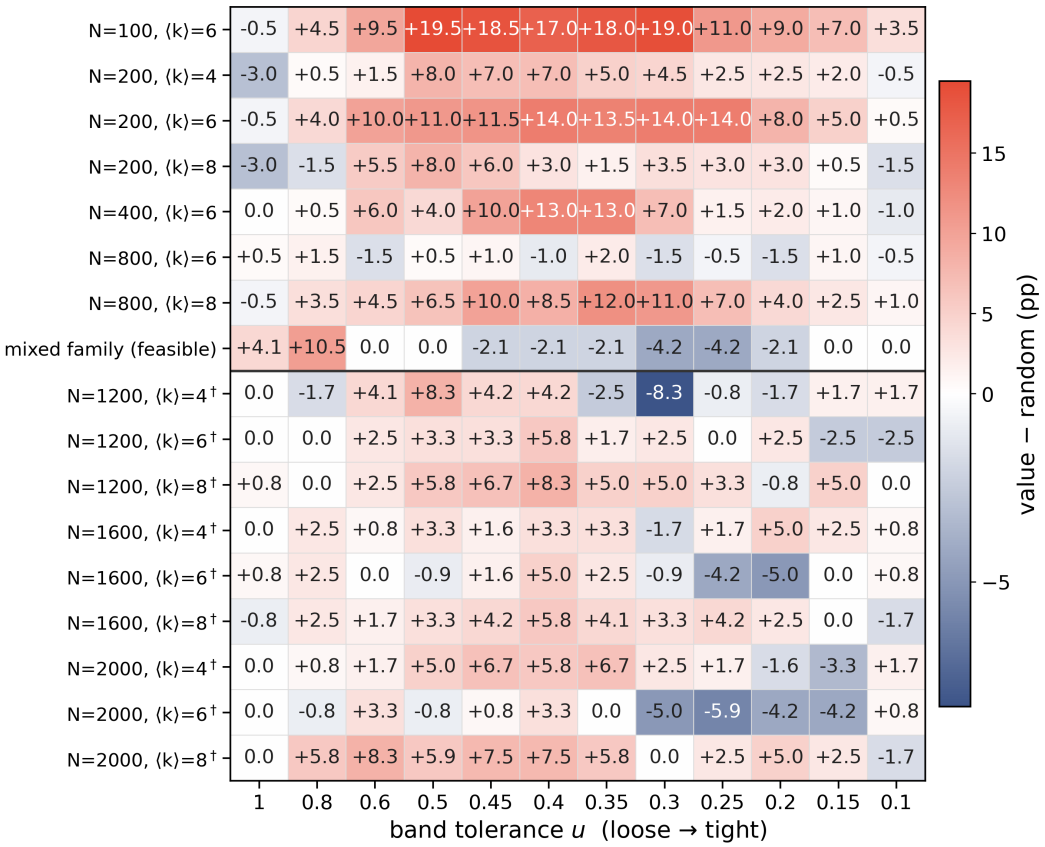


Figure 6: **Niche across regimes and bands** (value_{PE}−random, pp, all 12 bands; regime-grid protocol of App. A, not the unified-benchmark run; 10 seeds, mixed family 4; per-cell summaries in the supplementary material). Red: learner above its paired random baseline; blue: below. †Rows below the separator ($N \geq 1200$) use accelerated metrics and reduced budgets, so absolute values are not comparable across it.

this rise. The share of steps on which assortativity binds rises $0.148 \to 0.188 \to 0.307 \to 0.426$ for N=100 → 800 (measured along random-swap walks; the warm-start sweep of Figure 5a measures it along greedy rollouts, hence its larger values). On those steps local features are 5–20× more predictive (R^2 0.060–0.214 vs. 0.003–0.020 on other steps), as expected from Lemma 1.

Leave-one-family-out (LOFO). Training on two of {soft-RGG, SBM, Holme–Kim} and evaluating on the third (10 seeds; evaluation pools feasibility-filtered by a greedy probe: SBM 24/44 can-

didates pass, Holme–Kim 9/960, soft-RGG 0/960). Held-out SBM: greedy $57.1/43.7/31.7/19.6\%$ versus random $15.4/5.0/2.5/0.4$ at u=$0.6/0.4/0.3/0.2$ (paired $+41.7 \pm 3.7$ to $+19.2 \pm 2.7$, all $p<10^{-3}$). The value learner reaches $6.3/2.9/2.1/0.0$, at or below random everywhere and significantly below at u=0.6 (-9.2 ± 2.6, p=0.007). Held-out Holme–Kim is informative only at loose bands (greedy 97.8%, random 34.4%, value 10.0% at u=1.0; all 0 by u=0.4). No soft-RGG candidate passes the filter at $\langle k \rangle$=6 (Limitation 3).

Filter-free LOFO (own-degree-sequence warm starts). Warm-starting evaluation from each held-out target's own degree sequence makes every target reachable by construction, so no feasibility filter is needed (training unchanged: two families, WS warm starts; 10 seeds per fold, 12 evaluation targets, T=60). Held-out SBM: greedy−random $+18.4$ to $+29.2$ pp across u=1.0–0.2 ($+25.8 \pm 4.0$ at u=0.6), value−random -5.0 to 0.0 pp. Held-out Holme–Kim (informative at every band here; greedy $63.4/55.8/40.0/34.2/21.7\%$ at u=$1.0/0.6/0.4/0.3/0.2$): greedy−random $+20.0$ to $+36.6$ pp, value−random -13.4 to -0.8 pp. Held-out soft-RGG: every method, greedy included, satisfies 0% at every band even from the target's own degree sequence. The learner alone faces a train/evaluation warm-start mismatch under this protocol. Greedy dominance and the absence of a transfer niche do not depend on the filter. Re-running the feasibility probe at each family's native mean degree (soft-RGG 8.5, Holme–Kim 5.9) leaves feasibility unchanged ($0\% \to 0\%$; SBM $50\% \to 50\%$).

Certificate partition of the unreached set. We classify 747 targets (WS 100, generator-derived mix 300, uniform box 300, certificate-firing control 30, real networks 17) by the certificate and a solver ladder (greedy, 4-start greedy, beam, SA). WS targets are 94% reached (pipeline sanity). Uniform-box targets (N=200, $\langle k \rangle$=6) are 99% unreached, with the sharper $1 - Q$ certificate firing on $29/300$ of the box at u=1.0 ($44/300$ at u=0.4; the conservative $2(1 - Q)$ form fires on $0/300$ at u=1.0 and $3/300$ at u=0.4). The certificate-firing control is $30/30$ unreached with zero violations of $W \geq$ margin, and no reached target has a firing certificate. Of all unreached targets, the certificate explains 10% at u=1.0 ($59/596$; 12% at u=0.4). The wall is predominantly uncertified: unreached by every tested solver, not proven infeasible. This is why we report "unreached by a solver union" rather than "infeasible" wherever the certificate does not fire.

Two budget controls complete the partition. We re-search a stratified sample of 100 unreached targets (60 uniform-box, 40 generator-mix) with simulated annealing at 10^5 swaps per target (4 restarts, geometric cooling, same warm start; $\sim 666\times$ the ladder's SA budget). It recovers $3/100$ at u=1.0 (95% CI 1–8%; $0/60$ on the box; one also inside u=0.4). The median improvement of SA's best residual over the ladder's is -0.14 scaled units, and only 10% of samples improve by more than 0.5 (max 3.6). The three recoveries occur by swap $\sim 1.9 \times 10^4$ (median). Within this budget the wall does not yield slowly. The greedy probe's early-rejection constant (0.28 residual units per step) is the P98 of the per-step improvement distribution over 1,200 probe steps (P95 among improving steps). An overestimate errs toward fewer rejections.

Feasibility-oracle robustness. Widening the feasibility definition from single-start greedy to the union of 4-start greedy, beam, and SA on 30 raw mixed-family candidates adds one target ($13\% \to 17\%$; 83% unresolved by all solvers). On the same ladder single-start greedy reaches 4 of the 30, beam recovers one of the 26 it leaves unreached. A richer-action probe (adding rewiring and edge-addition/deletion moves) widens mixed-family feasibility only from 17% to 19%.

L REAL-WORLD TARGETS

Data and certificate. Seventeen classic networks (Table 14) were preprocessed to simple, undirected, largest-component graphs and their five-metric vectors used as targets. Every target passes the certificate test of Corollary 1 by construction, so real-network unreachability is uncertified: unreached, not proven infeasible. The smallest slacks $1 - Q - \lambda_2$ belong to sparse, high-modularity networks (Euroroad, netscience, Copenhagen), which are also the most tree-like.

Probe at the benchmark degree. A greedy probe (WS warm start at each network's own N with $\langle k \rangle$=6; K=8, 40 swaps, u=1.0, 2 seeds) reaches none of the 15 networks within its size range

($N \leq 2000$). The binding constraint is $\langle \ell \rangle$ for 9 (sparse, long-path networks), C for 4, and λ_2 for 2. The probe matches each network's N but neither its mean degree nor its degree sequence.

Native mean degree and larger budget. A dedicated run (greedy K=32, T=60, 4 seeds, 12 bands) on a five-network shortlist at each network's native N and mean degree reaches word-adjacency and football. Greedy attains 100%/90% at u=1.0, 70%/58% at u=0.6, and 2.5%/15% at u=0.4. Random reaches 5%/25% at u=1.0, and the value and imitation learners 0–58% at u=1.0 and $\leq$ 2.5% by u=0.6. Political books (7.5%), dolphins, and the autonomous-systems snapshot stay at or near 0%.

Own degree sequence. Warm-starting from each network's own degree sequence (Figure 5b; 11 networks) rescues three (58–98% at u=0.4) and partially a fourth (6%) while seven remain at 0%, so neither the size nor the mean-degree mismatch alone explains the wall. The certificate-partition ladder of Appendix K (run at N=200, $\langle k \rangle$=6 with a smaller per-target budget) marks only the e-mail network as reached. The verdicts differ in N, mean degree, budget, and solver ladder, not in data.

Table 14: The 17 real-network targets: size, mean degree, tree-ness (cyclomatic-based index; WS $\approx$ 0.33), certificate slack $1 - Q - \lambda_2$, the $\langle k \rangle$=6 probe's binding metric and best worst-scaled residual at u=1.0 (reached iff ≤ 1; — = outside the probe's size range), greedy satisfaction at u=1.0 in the native-degree run (five-network shortlist), and greedy satisfaction at u=0.4 from the own-degree warm start (11 networks); n/r = not run. Under the own-degree protocol's WS comparison (Figure 5b) jazz alone is reached from the WS start (32%).

Network	N	$\langle k \rangle$	tree-ness	slack	binding	best W	native-$\langle k \rangle$	own-degree
dolphins	62	5.1	0.38	0.47	$\langle \ell \rangle$	1.25	0%	0%
AS (1999)	103	4.6	0.43	0.48	C	2.00	0%	92%
political books	105	8.4	0.24	0.46	C	1.45	7.5%	0%
word adjacency	112	7.6	0.26	0.35	λ_2	1.80	100%	98%
football	115	10.7	0.19	0.31	C	2.17	90%	0%
jazz	198	27.7	0.07	0.45	$\langle \ell \rangle$	5.27	n/r	0%
C. elegans neural	297	14.5	0.14	0.44	$\langle \ell \rangle$	4.24	n/r	0%
Copenhagen calls	347	2.8	0.73	0.18	$\langle \ell \rangle$	11.5	n/r	6%
netscience	379	4.8	0.41	0.16	C	4.58	n/r	0%
C. elegans metabolic	453	8.9	0.22	0.54	$\langle \ell \rangle$	8.09	n/r	0%
Copenhagen sms	457	2.8	0.73	0.18	$\langle \ell \rangle$	10.9	n/r	58%
Euroroad	1039	2.5	0.80	0.14	$\langle \ell \rangle$	47.4	n/r	n/r
e-mail	1133	9.6	0.21	0.38	$\langle \ell \rangle$	3.87	n/r	n/r
political blogs	1222	27.4	0.07	0.49	$\langle \ell \rangle$	8.23	n/r	n/r
FAA routes	1226	3.9	0.51	0.30	λ_2	6.03	n/r	n/r
US air traffic	2274	51.2	0.04	0.72	—	—	n/r	n/r
AS (1997)	3015	3.4	0.59	0.37	—	—	n/r	n/r

M DOWNSTREAM DEMONSTRATION: WHY JOINT CONTROL MATTERS

The introduction's motivating claim, that attributing a dynamical outcome to one structural metric requires pinning the others, is itself testable. For two canonical metric–dynamics pairs we compare ensembles from our five-metric generator (*controlled*: co-metrics held in the u=0.4 band, then decorrelated by in-band Metropolis swaps; demo 1 pins C, Q, and r while path length is left free and stays within 0.40ε, demo 2 pins all four) against three single-metric rewiring baselines that represent common practice. The baselines are *n1*, biased rewiring in the style of Xulvi-Brunet–Sokolov (Xulvi-Brunet & Sokolov, 2004); *n2*, single-metric finite-temperature Metropolis–Hastings, the most charitable baseline; and *n3*, single-metric greedy. All arms use the same degree-preserving swap space, with M=40 graphs per rung per arm over a five-rung feasibility-probed ladder.

Synchronization and λ_2 (WS substrate, p=0.3). We simulate Kuramoto oscillators with degree-normalized coupling (the linearization at a phase-locked state involves the phase-weighted Laplacian L_w, $w_{ij} = A_{ij}\cos(\phi_j - \phi_i)$, which reduces near synchrony to the normalized Laplacian whose λ_2

we control). The primary outcome is time-to-synchronization. Controlled ensembles keep per-rung mean co-metric drift $\leq 0.42\varepsilon$ and yield an end-to-end time-to-sync change of -0.05 across the λ_2 ladder. Generation failed to reach the band for 18.5% (demo 1) and 7% (demo 2) of controlled graphs; these are retained in all aggregates, and restricting to in-band graphs moves the controlled contrasts to -0.09 (demo 1) and $+0.04$ (demo 2, below) without changing sign. Every naive arm yields -0.54 to -0.73, an attribution bias of -0.49 to -0.68 (bootstrap 95% CIs exclude zero for all three arms). That is, single-metric rewiring overstates the λ_2 effect by roughly an order of magnitude in slope (-19.1 vs. -1.8). The negative control is as expected: the Kuramoto onset threshold K_c, which theory ties to spectral quantities that are fixed here, is flat in the controlled ensemble ($+0.04$, CI spans zero).

Epidemic spreading and r (Holme–Kim substrate). Bond percolation (exact for Reed–Frost susceptible–infected–recovered (SIR) dynamics) at $T \in \{1.2, 1.5, 2.0\}\, T_c$ with per-graph T_c. Naive r-rewiring drifts the pinned co-metrics by up to 5.1ε (clustering collapse), versus $\leq 0.49\varepsilon$ (per-rung mean) controlled. The resulting attribution bias on outbreak size at $T{=}1.5T_c$ is -0.017 to -0.019 for the literature-standard and Metropolis arms (CIs exclude zero) and -0.008 for single-metric greedy (CI spans zero), against a controlled effect of $+0.045$ (a $\sim 40\%$ underestimate for the first two arms); the sign is stable at $T{=}1.2T_c$.

Effects here are statements *within the constrained ensemble* (degree sequence and banded co-metrics). The demonstration shows that the joint-control instrument changes the answer a practitioner would report, in one case by the sign-carrying magnitude and in the other by up to roughly 40% of the effect. Raw per-graph results are released with the code.

N Assets

The supplementary material contains the code (environment, generator, surrogates, methods, and the experiment entry points listed in its README, with smoke and reproduction scripts), the mixed-family and leave-one-family-out target pools with train/evaluation splits (WS targets are regenerated from seeds), per-seed raw results for the studies cited in this paper (the regime grid is released as per-cell summaries; unified benchmark, regime grid, tolerance profiles and re-test, 31-subset ablation, warm-start sweep, real-network runs, samplers, null models, divergent and architecture runs, GFlowNet strengthening, collision, marked-ball, horizon, scaling, label-noise, and controlled-scaling diagnostics, router, budget Pareto, horizon-matched, screen-then-verify, certificate partition, large-budget SA, feasibility union, richer actions, ERGM, EoH, loose-band, and downstream studies; the router gate module is not bundled and ships as raw results only), and theory verification scripts (exact-partition and CNM atlas checks of the exclusion bound, the Δr identity, and the gadget's marked-ball isomorphism, conductance enumeration, and random-scorer tie checks). All experiments ran on a single workstation; the dominant cost is true-metric evaluation.