CERTIFY BEFORE YOU COMMIT: TEST-TIME STRUCTURAL-REGRET CERTIFICATES FOR LEARNING-AUGMENTED GRAPH OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Machine learning can speed up hard graph optimization by fixing predicted variables or restricting search, but a single wrong commitment can exclude every high-quality solution. Prediction accuracy and confidence do not quantify this objective loss. We introduce augmented-instance regret (AIR), which measures the loss caused by enforcing a partial assignment, and AIRGraph, a model-agnostic framework that certifies such assignments at test time. AIRGraph computes a safe upper bound on AIR, combines it with a problem-specific repair credit, and accepts a proposal only when the credit exceeds the approximation-scaled bound. We prove that an independent fallback preserves the original approximation guarantee for arbitrary predictions. We instantiate AIRGraph for weighted Vertex Cover, weighted Maximum Independent Set, and signed MaxCut/Max-2Lin. Our local linear-programming Vertex Cover certificate is always at least as tight as the classical Nemhauser–Trotter certificate and transfers exactly to Maximum Independent Set; our MaxCut flow certificate is always at least as tight as a closed-form regional bound. Exact audits on small, exactly solved instances find no implementation violations. On 500 small Barabási–Albert (BA-small) graphs, the tighter MaxCut certificate changes 541 gate decisions and improves 149 outputs relative to the closed-form certificate, with no degradation. On 480 large Barabási–Albert (BA-large) graphs, evaluating the cheap MaxCut certificate for an existing proposal takes a median of 1.53 milliseconds (interquartile range 1.05–1.97 milliseconds). AIRGraph gives learned partial assignments explicit, instance-specific guarantees, independently of how the predictor was trained.

1 INTRODUCTION

Learning-augmented graph optimization uses machine-learning predictions to help classical algorithms solve hard graph problems. This approach is useful when many related instances must be solved: a learned model can reuse recurring structure, while the optimization algorithm remains responsible for feasibility and solution quality (Khalil et al., 2017; Bengio et al., 2021; Sun & Yang, 2023). A prediction can enter the solver in two ways. It can provide *soft guidance*, such as a warm start or branching priority, which the solver may revise. Or it can become a *hard commitment* that fixes predicted variables or restricts the feasible region. Neural Diving and Predict-and-Search use this second strategy to reduce the remaining search (Nair et al., 2020; Han et al., 2023). We call the predicted decisions that the solver must obey a *partial commitment*; before enforcement, it is a proposed commitment, or simply a *proposal*.

Hard commitments save computation by removing choices, but they can also remove good solutions. Once a restriction excludes a solution, even an exact residual solver cannot recover it. The cost depends on where the committed decisions occur, not only on how many predictions are wrong. Consider unweighted Vertex Cover on a graph formed by a k-leaf star and an isolated edge. Excluding the star center forces all k leaves into the cover and increases the optimum by $k - 1$. Excluding one endpoint of the isolated edge leaves the optimum unchanged because the other endpoint can be selected. Both proposals contain one exclusion, yet their objective effects are very different. Commitment size, prediction accuracy, and model confidence therefore do not upper-bound the objective loss of the realized proposal.

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.

To measure this loss directly, we introduce *augmented-instance regret* (AIR). For an instance I and proposal T, AIR is the change in the optimum caused by restricting the feasible set to solutions that obey T: an increase for minimization and a decrease for maximization. It is nonnegative, measured in objective units, and depends on both the instance and proposal. We use *structural regret* to emphasize that this loss depends on graph structure and on where the committed decisions occur; AIR is the formal quantity. The central question is: *before solving the restricted problem, can we upper-bound this loss?*

Several research directions address parts of this problem. Neural optimization methods learn assignments or search regions, but prediction accuracy does not certify the objective loss of enforcing one proposal (Khalil et al., 2017; Nair et al., 2020; Han et al., 2023; Sun & Yang, 2023; Wang et al., 2025b). Learning-augmented algorithms provide consistency and robustness guarantees under specified prediction models or error measures (Lykouris & Vassilvitskii, 2018; Purohit et al., 2018; Cohen-Addad et al., 2024; Antoniadis et al., 2025; Aamand et al., 2025). Safe fixing rules certify zero-loss decisions, and Certificate-Augmented Solution Pruning (CASP) verifies reductions that preserve optimality or a stated approximation guarantee (Nemhauser & Trotter, 1975; Schürmann & Mutzel, 2023; Li & Hai, 2026). What remains missing from these approaches is the contract studied here: after training, for the realized instance and proposal, allow bounded nonzero loss and accept the proposal only when a problem-specific repair compensates for that loss.

AIRGraph provides this contract (Figure 1). Exact AIR requires comparing the unrestricted and restricted optima, so AIRGraph computes a safe upper bound $U_I(T)$ instead. It also computes a repair credit $K_I(T)$: the objective benefit guaranteed by a problem-specific repair. The gate accepts the proposal only when the repair credit is strictly larger than the approximation-scaled certified loss. The solver then completes the restricted problem and applies the repair. In parallel, an independent classical fallback produces a commitment-free solution. AIRGraph returns the feasible solution with the better objective value. The accepted branch receives an instance-specific guarantee, and the fallback preserves the original approximation factor for arbitrary predictions.

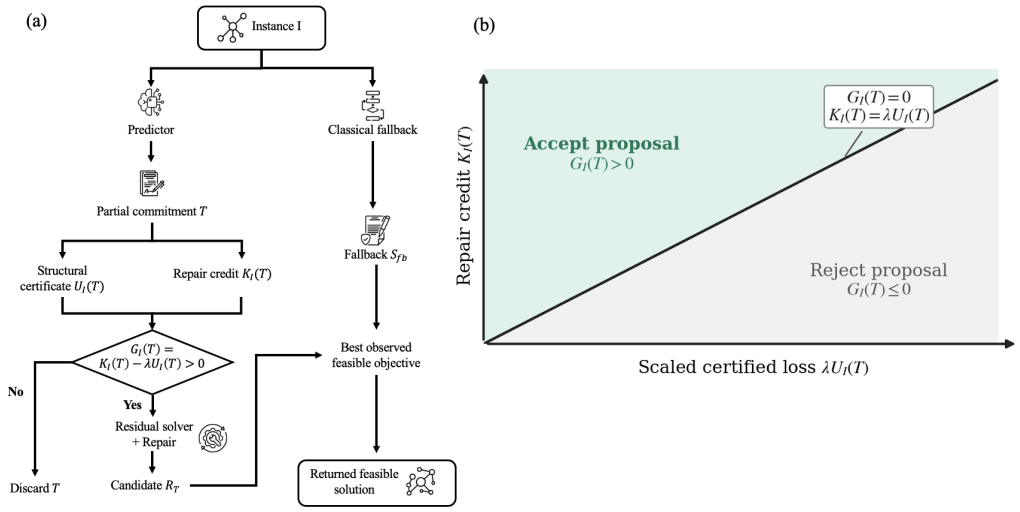

Figure 1: AIRGraph at test time. (a) A predictor proposes a partial commitment while an independent classical fallback runs in parallel. The certificate and repair credit determine whether the proposal proceeds to residual solving, and the returned solution is the better observed feasible objective. (b) The proposal is accepted above the certified-gain boundary $K_I(T) = \lambda U_I(T)$, where λ is the applicable approximation factor (ρ for minimization and α for maximization).

AIRGraph changes how a learned prediction enters an optimization algorithm. The predictor proposes a restriction, AIRGraph checks its possible objective consequence, and the solver commits only after the proposal passes the gate. Certification happens after training on the realized instance and proposal, so the framework can work with different prediction models. Because the certificate and repair credit use objective units, they connect directly to approximation guarantees. We use *branch*

for the restricted problem induced by a proposal and *arm* for any procedure that returns a feasible solution.

Contributions. We make three contributions.

(1) **A general guarantee for enforcing learned decisions.** We formalize AIR for minimization and maximization and prove a common certificate–repair rule. A safe upper bound on commitment loss, paired with sufficient repair credit, gives an explicit improvement over the classical worst-case bound. Comparing the repaired solution with an independent fallback preserves the original approximation factor for arbitrary predictions. The same guarantees compose when commitments are added and recertified in sequence.

(2) **New certificates for three graph problems.** For weighted Vertex Cover (VC), we derive a Nemhauser–Trotter (NT)-based certificate and a persistent-face local linear programming (local-LP) certificate that is always at least as tight. The additive AIR transfers exactly to weighted Maximum Independent Set (MIS). For signed MaxCut/Max-2Lin, we derive regional certificates, a tighter flow formulation, and a contraction that converts an accepted commitment into a smaller residual problem.

(3) **Proofs and decision-level evidence.** Complete proofs establish the guarantees, and finite exact checks audit the implementations. On 500 small Barabási–Albert (BA-small) graphs, the tighter MaxCut flow certificate changes 541 gate decisions and improves 149 outputs with no degradation. On 480 large Barabási–Albert (BA-large) graphs, evaluating the cheap MaxCut certificate for an existing proposal takes a median of 1.53 milliseconds. Controlled experiments further identify when graph structure adds information beyond prediction confidence and commitment size.

AIRGraph separates test-time certification from model training, so the same interface can work with different predictors. Section 2 reviews the closest directions; Sections 3–4 give the framework and problem-specific certificates; Section 5 evaluates them; and the appendices provide complete proofs and extended evidence.

2 Related Work

Learning-guided combinatorial solvers. Neural Diving fixes predicted variables and solves the remaining mixed-integer programming subproblem, while Predict-and-Search explores a neighborhood around a predicted assignment (Nair et al., 2020; Han et al., 2023). Other methods construct, refine, expand, or defer combinatorial predictions (Khalil et al., 2017; Ahn et al., 2020; Sun & Yang, 2023; Ma et al., 2025a; Wang et al., 2025b; Feng & Yang, 2025; Wang et al., 2025a). Recent benchmarks show that solver performance depends strongly on scale and evaluation protocol (Ma et al., 2025b; Nath & Kuhnle, 2025; Feng et al., 2026). This line of work improves predictions and search. AIRGraph addresses the next decision: whether a particular predicted restriction is safe enough to enforce.

Learning-augmented guarantees. Early learning-augmented algorithms study consistency with accurate advice and robustness when advice is poor (Lykouris & Vassilvitskii, 2018; Purohit et al., 2018). For graph problems, existing guarantees use noisy labels, edge predictions, predicted sets, or stochastic membership information (Cohen-Addad et al., 2024; Aamand et al., 2025; Antoniadis et al., 2025; Braverman et al., 2024). These results connect performance to a stated prediction model or an error measured against an optimum. AIRGraph asks a different test-time question: given the current graph and the commitment already proposed, how much objective value can enforcing it lose? Its answer is a deterministic, instance-specific bound; for MIS, the bound is additive.

Safe variable fixing and reduction. Nemhauser–Trotter persistency identifies Vertex Cover variables that can be fixed while keeping an optimum (Nemhauser & Trotter, 1975). Reduced-cost fixing excludes variable values using relaxation and incumbent bounds (Schürmann & Mutzel, 2023). Structure-guided hybrids and CASP also restrict learned decisions through tractable or verified reductions (Li et al., 2025; Li & Hai, 2026). CASP checks that a reduction preserves the optimum or a stated approximation guarantee. AIRGraph handles a broader case: a commitment may lose some objective value, but its repair can compensate for that loss. For Vertex Cover, our local AIR certificate is always at least as tight as the corresponding NT certificate.

Prediction quality and decision quality. Confidence calibration measures whether predicted probabilities match label frequencies (Guo et al., 2017). Decision-focused learning changes the training objective so that predictions support better downstream decisions (Wilder et al., 2019; Elmachtoub & Grigas, 2022). AIRGraph acts after training. It can therefore certify commitments produced by a calibrated model, a decision-focused model, or another black-box predictor, using the same test-time certificate–repair interface.

3 CERTIFICATE–REPAIR FRAMEWORK

A partial commitment T restricts the feasible set $\mathcal{X}(I)$ to a nonempty $\mathcal{X}(I \mid T)$. For minimization cost c_I and maximization value v_I, write $\mathrm{OPT}_{\min}(I)$ and $\mathrm{OPT}_{\max}(I)$ for the corresponding optimal objective values, and define the consequence of the restriction in the objective's native units.

Definition 3.1 (Augmented-instance regret)**.**

$$\mathrm{AIR}_I^{\min}(T) = \mathrm{OPT}_{\min}(I \mid T) - \mathrm{OPT}_{\min}(I), \quad \mathrm{AIR}_I^{\max}(T) = \mathrm{OPT}_{\max}(I) - \mathrm{OPT}_{\max}(I \mid T). \tag{1}$$

A certificate is safe *if* $\mathrm{AIR}_I(T) \le U_I(T)$ *for every commitment in its stated domain.*

AIR is nonnegative because a commitment only removes solutions. Unlike a label count, it depends on which solutions disappear. Certificate safety is deterministic and predictor-independent; tightness and runtime determine how useful the certificate is, not whether it is valid.

Proposition 3.2 (Equal error count, unequal commitment regret)**.** *For every $k \ge 2$, one unweighted graph admits (i) two feasible single-vertex VC out-commitments, each disagreeing once with the same optimum, whose regrets are $k-1$ and 0; and (ii) two full MaxCut labelings, each at Hamming distance one from the same optimum, whose regrets are k and 1.*

The construction is a star plus an isolated edge (Appendix A.8). It shows that even after error count and direction are fixed, topology can create an arbitrarily large gap in objective consequence.

3.1 CERTIFIED GAIN AND FALLBACK SAFETY

For minimization, suppose a repaired arm R_T obeys

$$c_I(R_T) \le \rho\,\mathrm{OPT}_{\min}(I \mid T) - K_I(T), \tag{2}$$

$$G_I^{\min}(T) = K_I(T) - \rho U_I(T), \tag{3}$$

where $\rho \ge 1$ and $K_I(T) \ge 0$. For maximization, suppose

$$v_I(R_T) \ge \alpha\,\mathrm{OPT}_{\max}(I \mid T) + K_I(T), \tag{4}$$

$$G_I^{\max}(T) = K_I(T) - \alpha U_I(T), \tag{5}$$

where $0 < \alpha \le 1$.

The repair credit is not a confidence score. It is an objective-unit saving guaranteed by the problem-specific reconstruction relative to applying the base approximation factor directly to the conditioned optimum; for example, the VC repair avoids paying twice for vertices already forced into the cover.

Theorem 3.3 (Minimization certified gain)**.** *A safe $U_I(T)$ and* (2) *imply $c_I(R_T) \le \rho\,\mathrm{OPT}_{\min}(I) - G_I^{\min}(T)$. Returning the cheaper of R_T and an independent factor-ρ fallback preserves that factor for every prediction.*

Theorem 3.4 (Maximization certified gain)**.** *A safe $U_I(T)$ and* (4) *imply $v_I(R_T) \ge \alpha\,\mathrm{OPT}_{\max}(I) + G_I^{\max}(T)$. Returning the better of R_T and an independent factor-α fallback preserves that factor for every prediction.*

Both results substitute the AIR identity and its safe upper bound into the repair inequality; Appendix A.1 gives the complete proofs. A proposal with $G_I(T) > 0$ has a strict per-instance bound beyond the classical worst-case factor before the benchmark optimum is known. For a finite proposal family $\mathcal{Q}(I)$, AIRGraph evaluates repairs for the accepted set $\mathcal{A}(I) = \{T \in \mathcal{Q}(I) : G_I(T) > 0\}$ and returns the best feasible objective in $\{S_{\mathrm{fb}}\} \cup \{R_T : T \in \mathcal{A}(I)\}$. We call the acceptance rule $G_I(T) > 0$ the *gate*. It identifies proposals with a positive certified margin and may avoid residual completions. Taking the better feasible output with the independent fallback preserves the base factor. Any additional feasible heuristic arm can enter the final comparison without weakening that guarantee, although evaluating it consumes additional computation.

3.2 CONDITIONAL COMPOSITION

Certificates computed on the original instance cannot in general be added across overlapping commitments. Let $I_0 = I$ and $I_j = I_{j-1} \mid T_j$ be nested feasible restrictions.

Theorem 3.5 (Sequential conditional composition)**.** *If each $U_{I_{j-1}}(T_j)$ bounds $\mathrm{OPT}_{\min}(I_j) - \mathrm{OPT}_{\min}(I_{j-1})$ for minimization, or $\mathrm{OPT}_{\max}(I_{j-1}) - \mathrm{OPT}_{\max}(I_j)$ for maximization, then*

$$\mathrm{AIR}_I(T_{1:k}) \le \sum_{j=1}^{k} U_{I_{j-1}}(T_j)$$

for either objective direction.

The proof telescopes conditional optimum differences (Appendix A.2). Each certificate is recomputed on the current restricted instance I_{j-1}; the theorem therefore applies to this nested sequence rather than to an unconditional sum of overlapping proposals.

4 GRAPH-PROBLEM INSTANTIATIONS

The generic theorems leave $U_I(T)$ and $K_I(T)$ problem-specific. We use the subscripts VC, MIS, and MC for Vertex Cover, Maximum Independent Set, and MaxCut, respectively. Vertex Cover charges forced neighborhoods through LP persistency, MIS inherits additive regret by complementation, and signed MaxCut charges regional parity conflict and boundary exposure. The following subsections derive each construction; Table 2 in Appendix B gives a compact summary.

4.1 WEIGHTED VERTEX COVER

Let $G = (V, E, w)$ have nonnegative vertex weights and optimum cover weight $\tau(G)$. A commitment $T = (A, B)$ fixes A in the cover and an independent set B out of it, with $A \cap B = \varnothing$. Write $\tau_T(G) = \mathrm{OPT}_{\min}(G \mid T)$. Every consistent cover must contain $N(B)$; define

$$F = A \cup N(B), \qquad H = G[V \setminus (F \cup B)]. \tag{6}$$

Lemma 4.1 (Conditioned VC identity)**.** *Every feasible T satisfies $\tau_T(G) = w(F) + \tau(H)$.*

Let $x^{\mathrm{LP}} \in \{0, \frac{1}{2}, 1\}^V$ be an optimal extreme point of the standard VC LP and $P_i = \{v : x_v^{\mathrm{LP}} = i\}$ (Nemhauser & Trotter, 1975). The NT certificate is

$$U_{\mathrm{VC}}^{\mathrm{NT}}(A, B) = [w(F \setminus P_1) - w(B \cap P_1)]_+ . \tag{7}$$

Theorem 4.2 (NT certificate)**.** *For every feasible commitment, $\mathrm{AIR}_{\mathrm{VC}}^{\min}(A, B) \le U_{\mathrm{VC}}^{\mathrm{NT}}(A, B)$.*

The NT bound uses only integral LP coordinates. To retain the same persistent decisions while accounting for fractional mass on the commitment-touched set $Q_T = F \cup B$, we keep the global cover constraints but localize the LP objective:

$$\mathcal{P}_{\mathrm{NT}}(G; x^{\mathrm{LP}}) = \{y \in [0,1]^V : y_u + y_v \ge 1 \ (\forall uv \in E),\ y_v = 0 \ (\forall v \in P_0),\ y_v = 1 \ (\forall v \in P_1)\}, \tag{8}$$

$$L_{\mathrm{local}}(T) = \min_{y \in \mathcal{P}_{\mathrm{NT}}(G; x^{\mathrm{LP}})} \sum_{v \in Q_T} w_v y_v, \qquad U_{\mathrm{VC}}^{\mathrm{local}}(T) = [w(F) - L_{\mathrm{local}}(T)]_+ . \tag{9}$$

Theorem 4.3 (Persistent-face local-LP strengthening)**.** *For every feasible T and the chosen x^{LP},*

$$\mathrm{AIR}_{\mathrm{VC}}^{\min}(T) \le U_{\mathrm{VC}}^{\mathrm{local}}(T) \le U_{\mathrm{VC}}^{\mathrm{NT}}(T).$$

The theorem holds for any chosen optimal half-integral extreme point x^{LP}. Our implementation uses the optimum returned by HiGHS; safety and dominance are independent of this choice, although certificate tightness can vary when the VC LP has multiple optimal extreme points.

To see the mechanism, NT persistency supplies an optimum cover $S^\star$ whose incidence vector is feasible for $\mathcal{P}_{\mathrm{NT}}$, so $L_{\mathrm{local}}(T) \le w(S^\star \cap Q_T)$. Replacing $B \cap S^\star$ by F then bounds AIR by

$w(F) - L_{\text{local}}(T)$. Conversely, the fixed coordinates give $L_{\text{local}}(T) \geq w(Q_T \cap P_1)$, which yields pointwise dominance over (7); Appendix A.3 gives all steps. If a factor-2 algorithm returns a cover C_H of H, then $R_T = F \cup C_H$ satisfies $c(R_T) \leq 2\tau_T(G) - w(F)$.

Corollary 4.4 (VC repair guarantee)**.** *For either certificate* $U_{\text{VC}}(T) \in \{U_{\text{VC}}^{\text{NT}}(T), U_{\text{VC}}^{\text{local}}(T)\}$,

$$c(R_T) \leq 2\tau(G) - G_{\text{VC}}(T), \qquad G_{\text{VC}}(T) = w(F) - 2U_{\text{VC}}(T).$$

Together with an independent factor-2 fallback, this is a factor-2 approximation for every prediction.

4.2 Maximum Independent Set

Let $W_V = \sum_{v \in V} w_v$. Since an independent set is the complement of a cover, $\text{OPT}_{\text{MIS}}(G) = W_V - \tau(G)$. The paired MIS commitment fixes B in the independent set and A out of it.

Theorem 4.5 (Exact VC–MIS AIR equivalence)**.** *For paired commitments* T^{VC} *and* T^{MIS}*,*

$$\text{AIR}_{\text{MIS}}^{\max}(T^{\text{MIS}}) = \text{AIR}_{\text{VC}}^{\min}(T^{\text{VC}}).$$

Hence both VC certificates transfer as safe additive *MIS certificates.*

The result supplies an additive AIR certificate, but a general weighted-greedy MIS fallback has no constant approximation factor on unrestricted graphs. Therefore the generic multiplicative gain theorem is not invoked for the MIS experiments. For proposal ranking we use the explicitly observable forced-weight score

$$\kappa_{\text{MIS}}(T^{\text{MIS}}) = w(B), \tag{10}$$

the total weight already fixed into the independent set. The top-3 arm ranks proposals by κ_{MIS}, completes those three with weighted greedy, and returns the best observed feasible set together with the fallback. The NT certificate is evaluated after this selection as an additive diagnostic; Appendix A.4 proves the AIR identity.

4.3 Signed MaxCut/Max-2Lin

For nonnegative edge weights and parity labels $b_e \in \{0, 1\}$, let $v_G(x) = \sum_{uv \in E} w_{uv} \mathbf{1}[x_u \oplus x_v = b_{uv}]$; ordinary MaxCut has $b_e = 1$. A proposal fixes only relative labels $\widehat{x}_W$ within $W \subseteq V$. Let P_W and N_W be the internal edges respectively satisfied and violated by these labels. For boundary weight $d_v = w(E(\{v\}, V \setminus W))$, write $D = \sum_{v \in W} d_v$, $d(S) = \sum_{v \in S} d_v$, and let $\delta_W(S)$ be the internal cut of $S \subseteq W$. The joint certificate is

$$U_W^{\text{joint}} = \max_{S \subseteq W} \left\{ w(N_W \cap \delta_W(S)) - w(P_W \cap \delta_W(S)) + \min\{d(S), D - d(S)\} \right\}. \tag{11}$$

Theorem 4.6 (Joint MaxCut certificate)**.** *For every signed instance and regional relative-label commitment,* $\text{AIR}_{\text{MC}}^{\max}(W, \widehat{x}) \leq U_W^{\text{joint}}$.

The proof converts an optimum to one of the two globally equivalent regional orientations: the first two terms exactly price internal changes, and the last charges the cheaper boundary exposure (Appendix A.5). Exact subset maximization is expensive, so we derive safe relaxations.

Theorem 4.7 (Computable MaxCut bounds)**.** *The binary formulation in Appendix A.6 computes* U_W^{joint}*, and the implemented polynomial-time bounds satisfy*

$$U_W^{\text{joint}} \leq U_W^{\text{flow}} \leq U_W^{\text{cheap}} = U_W^{\text{LP}}, \qquad U_W^{\text{cheap}} = w(N_W) + D/2,$$

where U_W^{LP} *is the continuous relaxation of the exact binary formulation. Any pointwise minimum of the exact, flow, and cheap bounds remains safe.*

The bare continuous relaxation is degenerate: setting every membership variable to $1/2$ attains the cheap bound, so it cannot recover good-edge cancellation. The flow construction does recover cancellation by routing bad-edge and boundary demand through predicted-good capacity, and exact optimization is reserved for small committed regions or audits. Thus the nontrivial implemented hierarchy is joint $\leq$ flow $\leq$ cheap. Finally, relative-label commitment fixes internal value $C_W = w(P_W)$, and parity-aware contraction creates a residual signed Max-2Lin instance.

Theorem 4.8 (MaxCut contraction and gain)**.** *If an α-approximation solves the contracted residual and U_W is any safe hierarchy member, then*

$$v(R_W) \geq \alpha\,\mathrm{OPT}_{\max}(G) + (1-\alpha)C_W - \alpha U_W, \tag{12}$$

$$G_{\mathrm{MC}}(W) = (1-\alpha)C_W - \alpha U_W, \qquad K_{\mathrm{MC}}(W) = (1-\alpha)C_W. \tag{13}$$

The better of R_W and an independent α-approximate fallback preserves factor α for arbitrary predictions.

The factor $(1-\alpha)$ is essential: C_W is fixed value in the conditioned instance, but only the part not already counted by the residual approximation becomes repair credit. Appendix A.7 proves the value-preserving contraction.

5 Experiments

Proposition 3.2 shows that error count need not track decision consequence. The experiments ask when inexpensive proxies nevertheless suffice, when structural certificates change a decision, and what certification contributes relative to ordinary search. MaxCut with fixed Graph Neural Network for Combinatorial Optimization (GNN4CO) predictions is primary; COExpander, MIS, VC, and transfer studies probe scope.

5.1 Protocol

We consume fixed exports from released GNN4CO, COExpander, and StruDiCO models (Ma et al., 2025b;a; Wang et al., 2025b); training is outside the method. Every output is feasible and protected by an independent fallback. MaxCut uses 12 proposals from two growth rules and six sizes. A deterministic conditional-expectation solution gives each contracted residual its half-weight guarantee; a time-limited HiGHS incumbent replaces it when better. Strong controls share one-flip polishing. We call local search initialized by the predictor and given AIRGraph's paired post-inference budget *prediction-seeded matched-time search* (seeded search); gate ablations separately report *accounted stage runtime* for their recorded pipeline stages. Common-pool studies fix proposals, budgets, fallback, and polish, then compare gated, unfiltered, confidence-ranked, and size-based policies. We call the original, unperturbed predictor outputs the *clean condition*. The topology diagnostic flips exactly 0–20% of labels while preserving confidence magnitudes, so confidence cannot detect the injected errors. No policy reads benchmark references. The benchmark uses Barabási–Albert (BA) and Watts–Strogatz (WS) graph families; RB is the released benchmark's graph-family label, and RB800–1200 denotes its MIS instances with 800–1200 vertices. We evaluate 480 BA-large, 500 WS-small, and 100 RB800–1200 MIS graphs; unit-weight BA-large graphs average 1007.8 vertices and 4015.2 edges (ranges 801–1200 and 3188–4784). Intervals use 10,000 graph-cluster bootstrap replications. Appendix B.1 gives the size grid, full protocol, and timing scope.

5.2 Are the certificates faithful, tight, and cheap?

On 12 exactly solvable instances per problem (10–16 nodes), no implemented certificate violates exact AIR above tolerance 10^{-8}. Among 228 VC/MIS rows, 36 have zero AIR; local LP is safe on the 192 positive-AIR rows and tighter than NT on 158. In addition, 113 MaxCut and 30 VC proposals with positive AIR pass the certified-gain test. On these enumerable instances, normalizing $U - \mathrm{AIR}$ by total edge or vertex weight gives median finite-audit gaps 0.114 for the portfolio-selected MaxCut bound and 0.103 for local-LP VC/MIS; the identical VC/MIS distributions corroborate the additive transfer. Certificate tightness also changes decisions. On 500 BA-small graphs, flow is tighter than cheap on all 2,008 proposals, upgrades 541 gate decisions, and improves 149 final outputs with 351 ties and no losses. Flow costs 4.977 s per graph in this study (Appendix B.6); the BA-large cheap-bound microbenchmark instead takes a median 1.53 ms per materialized commitment (IQR 1.05–1.97 ms) across 5,760 commitments, with zero stability failures. Appendices B.2 and B.3 give the audit and timing details.

5.3 Certified margins before completion

At the 0.2 s operating point, every BA-large graph has a positive accepted gain. Since $\mathrm{OPT}_{\mathrm{MC}} \leq \sum_e w_e$, Theorem 3.4 gives the *normalized structural factor* $1/2 + G_I(T)/\sum_e w_e$, ranging from

Table 1: MaxCut controls. Effects are AIRGraph plus one-flip polish minus the comparator; positive is better. Relative effects average paired graph-level percentage differences. The table reports wins/losses/ties (W/L/T) and a 95% confidence interval (CI) for each mean effect. All arms include the same fallback. Prediction-seeded and prediction-free searches receive paired post-inference budgets.

Predictor/data	Comparator	W/L/T	Mean effect [95% CI]	Relative
GNN4CO/BA-large	prediction+polish	383/80/17	$8.56\ [7.77, 9.33]$	0.30%
	prediction-seeded search	1/479/0	$-27.75\ [-28.66, -26.86]$	-0.94%
	prediction-free search	478/0/2	$61.45\ [59.67, 63.27]$	2.14%
COExpander/WS-small	prediction+polish	189/235/76	$-0.38\ [-0.66, -0.10]$	-0.04%
	prediction-seeded search	26/457/17	$-4.22\ [-4.52, -3.92]$	-0.48%
	prediction-free search	499/1/0	$17.83\ [17.25, 18.40]$	2.10%

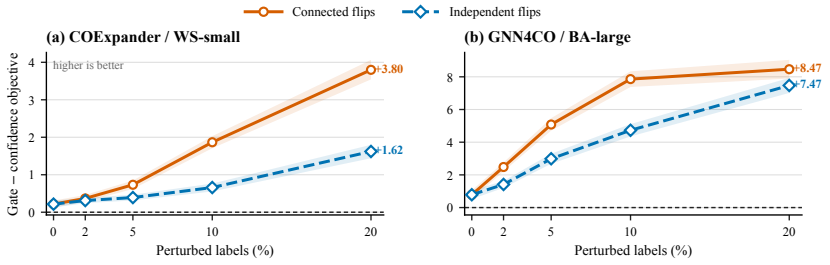

Figure 2: Controlled prediction-topology stress. Exact-count flips preserve confidence magnitudes. Curves show gate minus count-matched confidence with graph-cluster-bootstrap 95% intervals for the two predictors and graph families.

0.655 to 0.689 with median 0.673 (95% CI $[0.673, 0.674]$). This pre-completion quantity attributes a certified margin to commitment and repair. The post-completion fallback ratio is numerically stronger: fallback value divided by total edge weight averages 0.713 and exceeds the structural factor on all 480 graphs, while the final gated output averages 0.732. Appendix B.5 gives the full comparison.

On BA-large, AIRGraph beats direct GNN4CO plus identical polishing by 8.56 (0.30%) and prediction-free search by 61.45 (2.14%), while seeded search is better by 27.75 (0.94%; Table 1); 10% of the remaining budget already reverses the mean. On WS-small, direct polishing and seeded search are better by 0.38 (0.04%) and 4.22 (0.48%), while AIRGraph beats prediction-free search by 17.83 (2.10%). The final output inherits the certified inequality because all-vertex polishing and best-of fallback selection return value at least $v(R_T)$, hence at least $\alpha\,\mathrm{OPT} + G_I(T)$. These later steps may change the committed labels; applications that require those labels to remain fixed must constrain every downstream arm.

5.4 Selection under commitment-size controls

In the clean condition, the gate completes $9.031/12$ proposals and retains at least 99.85% of unfiltered improvement. Commitment size, however, accounts for this selection effect. Against the count-matched largest-size rule, the gate is 1.366 s faster but 0.096 lower in objective at the 0.05 s limit; at 0.2 and 0.5 s, the size rule is about 0.37 s faster and 0.079 higher. Deployable fixed-K controls agree: at 0.2 s, selecting the eight largest commitments saves 61.2% and differs from the gate by -0.011, while selecting nine saves 48.0% and exceeds it by 0.079. A size rule therefore reproduces the observed computation–quality frontier; certification contributes a formal bound on each accepted proposal. Appendix B.7 reports all operating points, correlations, timing scope, and intervals.

5.5 Does error topology matter beyond confidence?

This diagnostic uses conditional-expectation completion rather than the timed HiGHS protocol in Section 5.4, so its clean endpoint is a different operating point. The perturbation preserves confidence magnitudes, so confidence cannot detect the injected errors. In the clean condition, gate-minus-confidence is $0.22/0.79$ for COExpander/GNN4CO, and both effects become slightly negative against count-matched largest commitments ($-0.066/-0.210$). At 20%, GNN4CO retains connected/independent advantages of $5.15/2.82$ beyond size, whereas COExpander is mixed at $0.24/-0.23$; gate-minus-confidence reaches $3.80/1.62$ and $8.47/7.47$ (Figure 2). Commitment size is therefore an adequate proxy in the clean condition, while the certificate adds information in some predictor–graph regimes when error location is hidden from confidence. Appendix B.8 reports the full grid and graph-dependent reversal in boundary exposure.

Appendix B reports the full mechanism tables, seed-level results, transfer studies, and the additive MIS evaluation.

6 Discussion and Conclusion

What the certificate provides. AIRGraph answers Section 1's question with a pre-completion performance contract: the certified-gain theorem bounds each accepted branch before residual solving. On BA-large, an already materialized commitment has a median certificate time of 1.53 ms. In the flow study, tightening all 2,008 proposal bounds upgrades 541 gate decisions and improves 149 of 500 returned values with no losses. Selector evaluation remains separate and should control commitment size, residual difficulty, and computation.

When the commitment must remain binding. The contract applies to prediction-based variable fixing, restricted residual solving, and binding upstream or externally approved decisions. Persistency and reduced-cost fixing certify zero-loss or relaxation-excluded decisions; AIRGraph permits positive loss when repair credit covers it. In our HiGHS branch, regional labels remain fixed during residual completion. Subsequent polish and fallback inherit the value guarantee by returning an objective no worse than $v(R_T)$, although permanent label consistency requires constraining every downstream arm.

Composition with search. AIRGraph's guarantee remains valid when an additional feasible search arm enters the final best-of comparison. In the frozen paired logs, the resulting best-of outputs improve over seeded search by 0.014 and 0.088, using the sum of both runtimes. Under paired post-inference budgets, seeded search obtains raw objectives higher by 0.94% and 0.48%. The two components serve different roles: certification prices a commitment before completion, and search spends computation to improve the observed objective.

Scope and limitations. AIRGraph requires a safe, problem-specific certificate and compatible repair. Its multiplicative theory does not extend to unrestricted weighted MIS; on the RB800–1200 holdout, the matched-time classical arm is stronger by 2.85. The empirical evidence covers two MaxCut predictors, one MIS predictor, and BA, WS, and RB graph families. Exact tightness audits use 10–16-node graphs, the confidence-blind perturbations are synthetic, and reported runtimes cover stated pipeline stages rather than end-to-end deployment. Clean-condition selection can also be reproduced by simple size rules, and seeded search attains higher raw objective on the tested MaxCut predictors. These boundaries leave the paper's central claim unchanged: AIRGraph provides a pre-completion guarantee for an enforced learned commitment and can be combined with additional search.

Conclusion. The problem posed in Section 1 is how to exploit learned restrictions while quantifying the objective value they place at risk. AIRGraph answers with a model-agnostic pre-completion performance contract built from AIR certificates, repair credit, and an independent fallback. Across weighted Vertex Cover, MIS, and signed MaxCut, the theory makes this contract explicit, while the experiments identify when tighter certificates change decisions and when size or search explains observed utility. The resulting interface supports auditable learned commitments.

AI USE STATEMENT

We used LLM-based assistants (ChatGPT) during the preparation of this manuscript, in the following roles. (i) *Language editing:* improving the phrasing, grammar, and readability of author-written text. (ii) *Literature search:* suggesting candidate related work; every cited reference was subsequently retrieved and checked against its original source by the authors, and all characterizations of prior work are our own. (iii) *Review-style feedback:* pointing out unclear arguments, inconsistencies between the text and our result logs, and missing experimental controls; the resulting analyses, experiments, and revisions were designed and carried out by the authors. (iv) *Engineering support:* formatting experiment logs and assisting with anonymization of the supplementary archive, which the authors then verified manually against the double-blind policy. All definitions, theorems, and proofs were developed and verified by the authors, and no LLM output was used as evidence for a mathematical claim. The authors reviewed and edited all AI-assisted text and take full responsibility for the content, originality, and accuracy of this manuscript.

REPRODUCIBILITY STATEMENT

Appendices A–B give proofs, full protocols, and bounded post-hoc diagnostics. The anonymized supplement provides implementations, semantic entry points, provenance, and machine-readable summaries.

REFERENCES

Anders Aamand, Justin Y. Chen, Siddharth Gollapudi, Sandeep Silwal, and Hao Wu. Improved approximations for hard graph problems using predictions. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 73–101. PMLR, 2025. URL https://proceedings.mlr.press/v267/aamand25c.html.

Sungsoo Ahn, Younggyo Seo, and Jinwoo Shin. Learning what to defer for maximum independent sets. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 134–144. PMLR, 2020. URL https://proceedings.mlr.press/v119/ahn20a.html.

Antonios Antoniadis, Marek Eliáš, Adam Polak, and Moritz Venzin. Approximation algorithms for combinatorial optimization with predictions. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/7a8d388b7a17df480856dff1cc079b08-Abstract-Conference.html.

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

Vladimir Braverman, Prathamesh Dharangutte, Vihan Shah, and Chen Wang. Learning-augmented maximum independent set. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2024)*, volume 317 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 24:1–24:18. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024. doi: 10.4230/LIPIcs.APPROX/RANDOM.2024.24. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.APPROX/RANDOM.2024.24.

Vincent Cohen-Addad, Tommaso d'Orsi, Anupam Gupta, Euiwoong Lee, and Debmalya Panigrahi. Learning-augmented approximation algorithms for maximum cut and related problems. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-0816. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/2db08b94565c0d582cc53de6cee5fd47-Abstract-Conference.html.

Adam N. Elmachtoub and Paul Grigas. Smart "predict, then optimize". *Management Science*, 68(1):9–26, 2022. doi: 10.1287/mnsc.2020.3922. URL https://pubsonline.informs.org/doi/10.1287/mnsc.2020.3922.

Shengyu Feng and Yiming Yang. Regularized langevin dynamics for combinatorial optimization. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 16977–16992. PMLR, 2025. URL https://proceedings.mlr.press/v267/feng25r.html.

Shengyu Feng, Weiwei Sun, Shanda Li, Ameet Talwalkar, and Yiming Yang. FrontierCO: Real-world and large-scale evaluation of machine learning solvers for combinatorial optimization. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/c131c8875c7b1133ffdad2b53cb10e91-Abstract-Conference.html.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1321–1330. PMLR, 2017. URL https://proceedings.mlr.press/v70/guo17a.html.

Qingyu Han, Linxin Yang, Qian Chen, Xiang Zhou, Dong Zhang, Akang Wang, Ruoyu Sun, and Xiaodong Luo. A GNN-guided predict-and-search framework for mixed-integer linear programming. In *International Conference on Learning Representations*, 2023. URL https://mlanthology.org/iclr/2023/han2023iclr-gnnguided/.

Elias B. Khalil, Hanjun Dai, Yuyu Zhang, Bistra Dilkina, and Le Song. Learning combinatorial optimization algorithms over graphs. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/hash/d9896106ca98d3d05b8cbdf4fd8b13a1-Abstract.html.

Haifeng Li and Mo Hai. CASP: Learning-augmented offline approximation with verifiable certificates and bounded-loss PAC guarantees, 2026. URL https://arxiv.org/abs/2607.14545.

Jialiang Li, Weitong Chen, and Mingyu Guo. Neural tractability via structure: Learning-augmented algorithms for graph combinatorial optimization, 2025. URL https://arxiv.org/abs/2511.19573.

Thodoris Lykouris and Sergei Vassilvitskii. Competitive caching with machine learned advice. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 3296–3305. PMLR, 2018. URL https://proceedings.mlr.press/v80/lykouris18a.html.

Jiale Ma, Wenzheng Pan, Yang Li, and Junchi Yan. COExpander: Adaptive solution expansion for combinatorial optimization. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 42130–42164. PMLR, 2025a. URL https://proceedings.mlr.press/v267/ma25r.html.

Jiale Ma, Wenzheng Pan, Yang Li, and Junchi Yan. ML4CO-Bench-101: Benchmark machine learning for classic combinatorial problems on graphs. In *Advances in Neural Information Processing Systems*, volume 38, 2025b. doi: 10.52202/085713-3921. URL https://papers.nips.cc/paper_files/paper/2025/hash/aa9340420e8d78021c35ae984e1bda85-Abstract-Datasets_and_Benchmarks_Track.html. Datasets and Benchmarks Track.

Vinod Nair, Sergey Bartunov, Felix Gimeno, Ingrid von Glehn, Pawel Lichocki, Ivan Lobov, Brendan O'Donoghue, Nicolas Sonnerat, Christian Tjandraatmadja, Pengming Wang, Ravichandra Addanki, Tharindi Hapuarachchi, Thomas Keck, James Keeling, Pushmeet Kohli, Ira Ktena, Yujia Li, Oriol Vinyals, and Yori Zwols. Solving mixed integer programs using neural networks, 2020. URL https://arxiv.org/abs/2012.13349.

Ankur Nath and Alan Kuhnle. MaxCutBench: Revisiting and benchmarking graph neural networks for maximum cut. *Transactions on Machine Learning Research*, 2025. URL https://openreview.net/forum?id=322PpCGAX8.

George L. Nemhauser and Leslie E. Trotter. Vertex packings: Structural properties and algorithms. *Mathematical Programming*, 8(1):232–248, 1975. doi: 10.1007/BF01580444.

Manish Purohit, Zoya Svitkina, and Ravi Kumar. Improving online algorithms via ML predictions. In *Advances in Neural Information Processing Systems*, volume 31, pp. 9684–9693, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/73a427badebe0e32caa2e1fc7530b7f3-Abstract.html.

Lukas Schürmann and Petra Mutzel. A reduced cost-based model strengthening method. In *Proceedings of the 2023 SIAM Conference on Applied and Computational Discrete Algorithms*, pp. 75–86. SIAM, 2023. doi: 10.1137/1.9781611977714.7. URL https://epubs.siam.org/doi/10.1137/1.9781611977714.7.

Zhiqing Sun and Yiming Yang. DIFUSCO: Graph-based diffusion solvers for combinatorial optimization. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-0164. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/0ba520d93c3df592c83a611961314c98-Abstract-Conference.html.

Shiyue Wang, Ziao Guo, Changhong Lu, and Junchi Yan. Fractional langevin dynamics for combinatorial optimization via polynomial-time escape. In *Advances in Neural Information Processing Systems*, volume 38, 2025a. doi: 10.52202/085713-3542. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/989a74f28a550b71c77067ba04e9d081-Abstract-Conference.html.

Yu Wang, Yang Li, Junchi Yan, and Yi Chang. StruDiCO: Structured denoising diffusion with gradient-free inference-stage boosting for memory and time efficient combinatorial optimization. In *Advances in Neural Information Processing Systems*, volume 38, 2025b. doi: 10.52202/085713-2394. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/6728fcf94660c59c938319a6833a6073-Abstract-Conference.html.

Bryan Wilder, Bistra Dilkina, and Milind Tambe. Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 1658–1665, 2019. doi: 10.1609/aaai.v33i01.33011658. URL https://ojs.aaai.org/index.php/AAAI/article/view/3982.

A Proofs and Extended Technical Detail

A.1 Generic certificate-to-approximation proofs

Proof of Theorem 3.3. By Definition 3.1 and certificate safety,

$$\mathrm{OPT}_{\min}(I \mid T) = \mathrm{OPT}_{\min}(I) + \mathrm{AIR}_I^{\min}(T) \le \mathrm{OPT}_{\min}(I) + U_I(T).$$

Substitution into (2) gives

$$\begin{aligned} c_I(R_T) &\le \rho\,\mathrm{OPT}_{\min}(I) + \rho U_I(T) - K_I(T) \\ &= \rho\,\mathrm{OPT}_{\min}(I) - G_I^{\min}(T). \end{aligned}$$

Let S_{out} be the cheaper of R_T and a fallback S_{fb} satisfying $c_I(S_{\mathrm{fb}}) \le \rho\,\mathrm{OPT}_{\min}(I)$. Then

$$c_I(S_{\mathrm{out}}) \le \min\{\rho\,\mathrm{OPT}_{\min}(I) - G_I^{\min}(T), \rho\,\mathrm{OPT}_{\min}(I)\} \le \rho\,\mathrm{OPT}_{\min}(I).$$

This conclusion holds for every proposal because it does not impose a distributional condition on T. □

Proof of Theorem 3.4. Certificate safety implies

$$\mathrm{OPT}_{\max}(I \mid T) = \mathrm{OPT}_{\max}(I) - \mathrm{AIR}_I^{\max}(T) \ge \mathrm{OPT}_{\max}(I) - U_I(T).$$

Applying (4),

$$\begin{aligned} v_I(R_T) &\ge \alpha\,\mathrm{OPT}_{\max}(I) - \alpha U_I(T) + K_I(T) \\ &= \alpha\,\mathrm{OPT}_{\max}(I) + G_I^{\max}(T). \end{aligned}$$

Let S_{out} be the better of R_T and an independent fallback satisfying $v_I(S_{\mathrm{fb}}) \ge \alpha\,\mathrm{OPT}_{\max}(I)$. Then

$$v_I(S_{\mathrm{out}}) \ge \max\{\alpha\,\mathrm{OPT}_{\max}(I) + G_I^{\max}(T), \alpha\,\mathrm{OPT}_{\max}(I)\} \ge \alpha\,\mathrm{OPT}_{\max}(I).$$

□

A.2 Conditional composition proof

Proof of Theorem 3.5. For minimization, the nested restrictions give

$$\begin{aligned} \mathrm{AIR}_I^{\min}(T_{1:k}) &= \mathrm{OPT}_{\min}(I_k) - \mathrm{OPT}_{\min}(I_0) \\ &= \sum_{j=1}^{k} [\mathrm{OPT}_{\min}(I_j) - \mathrm{OPT}_{\min}(I_{j-1})] \\ &\le \sum_{j=1}^{k} U_{I_{j-1}}(T_j). \end{aligned}$$

For maximization, reverse the differences:

$$\begin{aligned} \mathrm{AIR}_I^{\max}(T_{1:k}) &= \mathrm{OPT}_{\max}(I_0) - \mathrm{OPT}_{\max}(I_k) \\ &= \sum_{j=1}^{k} [\mathrm{OPT}_{\max}(I_{j-1}) - \mathrm{OPT}_{\max}(I_j)] \\ &\le \sum_{j=1}^{k} U_{I_{j-1}}(T_j). \end{aligned}$$

No independence assumption is used. The argument requires each jth certificate to concern I_{j-1}, not the original instance, which is precisely why unconditional summation of overlapping certificates is not implied. □

A.3 VERTEX COVER DETAILS

Proof of Lemma 4.1. Every cover that excludes B contains $N(B)$, and every cover that includes A contains A; hence every cover consistent with T contains $F = A \cup N(B)$. Because B is independent and $F \cap B = \varnothing$, every edge with an endpoint in B is covered by $N(B) \subseteq F$, every edge incident to F is already covered, and the only edges still requiring a choice are those induced by $V \setminus (F \cup B)$. These are exactly the edges of H. Therefore a conditioned cover has weight at least $w(F) + \tau(H)$, while the union of F and a minimum cover of H attains this value. □

Proof of Theorem 4.2. Nemhauser–Trotter persistency implies that there is a minimum cover $S^\star$ with $P_1 \subseteq S^\star$ and $P_0 \cap S^\star = \varnothing$. Construct

$$S' = (S^\star \setminus B) \cup F.$$

Every edge incident to a removed vertex of B is covered by $N(B) \subseteq F$, so S' is a cover respecting (A, B). Since $F \cap B = \varnothing$,

$$\begin{aligned} \mathrm{AIR}_{\mathrm{VC}}^{\min}(A, B) &= \tau_T(G) - \tau(G) \\ &\le w(S') - w(S^\star) \\ &= w(F \setminus S^\star) - w(B \cap S^\star) \\ &\le w(F \setminus P_1) - w(B \cap P_1). \end{aligned}$$

The last inequality uses $P_1 \subseteq S^\star$. Since $[z]_+ \ge z$ for every real z, the definition in (7) is also a valid upper bound. □

Proof of Theorem 4.3. By Nemhauser–Trotter persistency, choose a minimum cover $S^\star$ satisfying $P_1 \subseteq S^\star$ and $P_0 \cap S^\star = \varnothing$. Its incidence vector belongs to $\mathcal{P}_{\mathrm{NT}}(G; x^{\mathrm{LP}})$, and therefore

$$L_{\mathrm{local}}(T) \le w\big(S^\star \cap Q_T\big). \tag{14}$$

As in the proof of Theorem 4.2, the set $S' = (S^\star \setminus B) \cup F$ is a cover consistent with (A, B). Because $F \cap B = \varnothing$ and $Q_T = F \cup B$,

$$\begin{aligned} \mathrm{AIR}_{\mathrm{VC}}^{\min}(A, B) &\le w(S') - w(S^\star) \\ &= w(F \setminus S^\star) - w(B \cap S^\star) \\ &= w(F) - w\big(S^\star \cap Q_T\big) \\ &\le w(F) - L_{\mathrm{local}}(T) \\ &\le U_{\mathrm{VC}}^{\mathrm{local}}(T). \end{aligned} \tag{15}$$

This proves safety. To prove dominance, every $y \in \mathcal{P}_{\mathrm{NT}}(G; x^{\mathrm{LP}})$ satisfies $y_v = 1$ for $v \in P_1$. Nonnegative weights give

$$\sum_{v \in Q_T} w_v y_v \ge w(Q_T \cap P_1),$$

and hence $L_{\mathrm{local}}(T) \ge w(Q_T \cap P_1)$. Using again that $F \cap B = \varnothing$,

$$\begin{aligned} w(F) - L_{\mathrm{local}}(T) &\le w(F) - w\big((F \cup B) \cap P_1\big) \\ &= w(F \setminus P_1) - w(B \cap P_1). \end{aligned}$$

The map $z \mapsto [z]_+$ is monotone, so (9) and (7) imply $U_{\mathrm{VC}}^{\mathrm{local}}(T) \le U_{\mathrm{VC}}^{\mathrm{NT}}(T)$. □

Proof of Corollary 4.4. If C_H is a factor-2 cover of H, then

$$w(F \cup C_H) \le w(F) + 2\tau(H) = 2\tau_T(G) - w(F).$$

This is (2) with $\rho = 2$ and $K_{\mathrm{VC}}(T) = w(F)$. Theorem 4.2 and Theorem 4.3 show that either permitted choice of $U_{\mathrm{VC}}(T)$ is safe, so Theorem 3.3 yields

$$c(R_T) \le 2\tau(G) - \big(w(F) - 2U_{\mathrm{VC}}(T)\big) = 2\tau(G) - G_{\mathrm{VC}}(T).$$

The fallback statement follows from the final part of Theorem 3.3. □

A.4 MIS COMPLEMENT PROOF

Proof of Theorem 4.5. Complementation is a bijection between covers and independent sets, including under the paired commitments. Thus

$$\mathrm{OPT}_{\mathrm{MIS}}(G) = W_V - \tau(G), \qquad \mathrm{OPT}_{\mathrm{MIS}}(G \mid T^{\mathrm{MIS}}) = W_V - \tau_T(G).$$

Subtracting yields

$$\begin{aligned}\mathrm{AIR}^{\max}_{\mathrm{MIS}}(T^{\mathrm{MIS}}) &= \mathrm{OPT}_{\mathrm{MIS}}(G) - \mathrm{OPT}_{\mathrm{MIS}}(G \mid T^{\mathrm{MIS}}) \\ &= \tau_T(G) - \tau(G) = \mathrm{AIR}^{\min}_{\mathrm{VC}}(T^{\mathrm{VC}}).\end{aligned}$$

If either $U^{\mathrm{NT}}_{\mathrm{VC}}(T^{\mathrm{VC}})$ or $U^{\mathrm{local}}_{\mathrm{VC}}(T^{\mathrm{VC}})$ upper-bounds the final expression, it upper-bounds the MIS AIR by equality. Because the constant W_V is subtracted before taking a ratio, this additive identity does not imply a multiplicative MIS approximation. □

A.5 MAXCUT JOINT-CERTIFICATE PROOF

Proof of Theorem 4.6. Let x^{opt} be an optimal assignment. For an orientation $\sigma \in \{0,1\}$ of the regional prediction, define $y^\sigma_v = \widehat{x}_v \oplus \sigma$ for $v \in W$, retain x^{opt} outside W, and let $S_\sigma = \{v \in W : y^\sigma_v \neq x^{\mathrm{opt}}_v\}$. The two orientations satisfy $S_1 = W \setminus S_0$ and induce the same internal cut $\delta_W(S_0) = \delta_W(S_1)$.

For an internal edge $e \in P_W$, y^σ satisfies e. Its status relative to x^{opt} changes exactly when $e \in \delta_W(S_\sigma)$, in which case the contribution to $v_G(x^{\mathrm{opt}}) - v_G(y^\sigma)$ is $-w_e$. Similarly, an edge $e \in N_W$ contributes $+w_e$ exactly when it crosses the disagreement set. Hence the internal objective difference is exactly

$$w(N_W \cap \delta_W(S_\sigma)) - w(P_W \cap \delta_W(S_\sigma)).$$

On the boundary of W, replacing x^{opt} by y^0 can lose at most $d(S_0)$, while replacing it by the complementary orientation y^1 can lose at most $d(W \setminus S_0) = D - d(S_0)$. Choose the orientation with smaller exposed boundary weight. The resulting feasible committed assignment y therefore satisfies

$$v_G(x^{\mathrm{opt}}) - v_G(y) \le w(N_W \cap \delta_W(S)) - w(P_W \cap \delta_W(S)) + \min\{d(S), D - d(S)\}$$

for one $S \subseteq W$. Because $\mathrm{OPT}_{\max}(G \mid W, \widehat{x}) \ge v_G(y)$, maximizing the right-hand side over all S proves (11). □

A.6 MAXCUT CERTIFICATE HIERARCHY

Exact binary formulation and degenerate LP relaxation. For $S = \{v : z_v = 1\}$, use a binary membership variable z_v for every $v \in W$, a binary exclusive-or (XOR) variable t_e for every $e = uv \in E(W)$, and a continuous boundary variable $h \ge 0$. Impose

$$\begin{aligned} t_{uv} &\ge z_u - z_v, & t_{uv} &\ge z_v - z_u, \\ t_{uv} &\le z_u + z_v, & t_{uv} &\le 2 - z_u - z_v, \\ h &\le \sum_{v \in W} d_v z_v, & h &\le D - \sum_{v \in W} d_v z_v. \end{aligned} \tag{16}$$

Together with $z_v, t_e \in \{0,1\}$, maximizing

$$\sum_{e \in N_W} w_e t_e - \sum_{e \in P_W} w_e t_e + h \tag{17}$$

computes U^{joint}_W. Replacing the binary domains by $z_v, t_e \in [0,1]$ gives the bare relaxation U^{LP}_W.

Flow upper bound. Fix any root $r \in W$ and replace every $g \in P_W$ by its two directed arcs, obtaining A_P. The implementation chooses $r \in \arg\max_{v \in W} d_v$, breaking ties by the stored regional order; the validity proof does not depend on this choice. There are two types of commodities:

$$\mathcal{C}_{\mathrm{bad}} = N_W, \qquad \mathcal{C}_{\mathrm{bnd}} = \{v \in W \setminus \{r\} : d_v > 0\}.$$

For each bad edge $e \in N_W$, orient its endpoints arbitrarily as (σ_e, τ_e), choose $0 \le q_e \le w_e$, and set demand $\theta_e = q_e$. For every boundary commodity $v \in \mathcal{C}_{\mathrm{bnd}}$, set $(\sigma_v, \tau_v) = (v, r)$ and demand $\theta_v = \eta d_v$, where $0 \le \eta \le 1$. Let $f_a^c \ge 0$ be the flow of commodity c on arc $a \in A_P$. Flow conservation requires, for every c and $u \in W$,

$$\sum_{a \in \delta^+_{A_P}(u)} f_a^c - \sum_{a \in \delta^-_{A_P}(u)} f_a^c = \begin{cases} \theta_c, & u = \sigma_c, \\ -\theta_c, & u = \tau_c, \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

With overload variables $o_g \ge 0$, the undirected good-edge capacities are

$$\sum_{c \in \mathcal{C}_{\mathrm{bad}} \cup \mathcal{C}_{\mathrm{bnd}}} \left(f^c_{u \to v} + f^c_{v \to u} \right) \le w_g + o_g, \qquad g = \{u, v\} \in P_W. \tag{19}$$

The flow certificate is the value of the linear program

$$U_W^{\mathrm{flow}} = \min_{q,\eta,o,f} \left\{ \sum_{e \in N_W} (w_e - q_e) + \sum_{g \in P_W} o_g + (1-\eta)\frac{D}{2} \,:\, \substack{(18),(19),\\ 0 \le q_e \le w_e \ (e \in N_W),\\ 0 \le \eta \le 1,\ o_g \ge 0,\ f_a^c \ge 0} \right\}. \tag{20}$$

The program is valid whether or not (W, P_W) is connected. The implementation solves this single global LP over all of W; it does not decompose good-edge components. Any componentwise variant would have to charge every bad edge crossing two components in full. In the global program, unroutable demands are automatically suppressed by feasibility: for example, if a vertex with $d_v > 0$ lies outside the root component, then $\eta = 0$. The choice $q = 0$, $\eta = 0$, $o = 0$, and zero flow is always feasible.

Proof of Theorem 4.7. For binary z, the XOR-hull constraints force $t_{uv} = \mathbf{1}[z_u \ne z_v]$. For the corresponding set $S = \{v : z_v = 1\}$, maximizing h gives $h = \min\{d(S), D - d(S)\}$. Thus the binary program is exactly (11). Relaxing integrality enlarges the feasible region of a maximization problem, so $U_W^{\mathrm{joint}} \le U_W^{\mathrm{LP}}$. In every relaxed solution,

$$\sum_{e \in N_W} w_e t_e \le w(N_W), \qquad -\sum_{e \in P_W} w_e t_e \le 0, \qquad h \le \frac{D}{2}.$$

These inequalities prove $U_W^{\mathrm{LP}} \le U_W^{\mathrm{cheap}}$. Equality always holds: set $z_v = 1/2$ for every $v \in W$, $t_e = 1$ for $e \in N_W$, $t_e = 0$ for $e \in P_W$, and $h = D/2$. This point satisfies (16) and attains $w(N_W) + D/2$. Hence $U_W^{\mathrm{LP}} = U_W^{\mathrm{cheap}}$ for every instance.

For the flow inequality, fix any $S \subseteq W$ and write $m(S) = \min\{d(S), D - d(S)\}$. Every routed bad-edge commodity whose endpoints are separated by S sends q_e units across $P_W \cap \delta_W(S)$. The side of the cut not containing r originates boundary demand equal to either $\eta d(S)$ or $\eta(D - d(S))$, and hence at least $\eta m(S)$ units must cross the same good-edge cut. Summing (19) over that cut yields

$$\sum_{e \in N_W \cap \delta_W(S)} q_e + \eta m(S) \le w(P_W \cap \delta_W(S)) + \sum_{g \in P_W \cap \delta_W(S)} o_g.$$

Rearranging, enlarging the two nonnegative sums to all bad and good edges, and using $(1-\eta)m(S) \le (1-\eta)D/2$ gives

$$\begin{aligned} &w(N_W \cap \delta_W(S)) - w(P_W \cap \delta_W(S)) + m(S) \\ &\quad \le \sum_{e \in N_W} (w_e - q_e) + \sum_{g \in P_W} o_g + (1-\eta)D/2. \end{aligned}$$

This holds for every S and every feasible flow solution. Maximizing over S and then minimizing over feasible flow solutions gives $U_W^{\mathrm{joint}} \le U_W^{\mathrm{flow}}$. Finally, the zero-demand solution above has objective $w(N_W) + D/2$, proving $U_W^{\mathrm{flow}} \le U_W^{\mathrm{cheap}} = U_W^{\mathrm{LP}}$. Theorem 4.6 makes the joint bound safe, and each displayed relaxation upper-bounds it; the pointwise minimum of safe upper bounds remains safe. □

A.7 MaxCut parity contraction and repair credit

Proof of Theorem 4.8. Under the regional commitment, every internal edge in P_W is satisfied and every internal edge in N_W is unsatisfied, independent of the chosen global orientation. Hence the internal contribution is exactly $C_W = w(P_W)$. Parity-aware contraction records the predicted parity between every $v \in W$ and one representative and updates the signs of incident boundary edges accordingly. This gives a value-preserving bijection between assignments consistent with the regional commitment and assignments of the residual signed Max-2Lin instance G_W^{res}. Therefore

$$\mathrm{OPT}_{\max}(G \mid W) = C_W + \mathrm{OPT}_{\max}(G_W^{\mathrm{res}}).$$

If the residual solver is an α-approximation, reconstruction returns a committed assignment with value

$$\begin{aligned} v(R_W) &\geq C_W + \alpha\,\mathrm{OPT}_{\max}(G_W^{\mathrm{res}}) \\ &= \alpha\,\mathrm{OPT}_{\max}(G \mid W) + (1-\alpha)C_W. \end{aligned}$$

This is the generic repair inequality (4) with $K_{\mathrm{MC}}(W) = (1-\alpha)C_W$. For any safe selected hierarchy member U_W,

$$\begin{aligned} v(R_W) &\geq \alpha\big(\mathrm{OPT}_{\max}(G) - \mathrm{AIR}_{\mathrm{MC}}^{\max}(W, \widehat{x})\big) + (1-\alpha)C_W \\ &\geq \alpha\,\mathrm{OPT}_{\max}(G) + (1-\alpha)C_W - \alpha U_W, \end{aligned}$$

which is (12). Equation (13) and the fallback conclusion now follow from Theorem 3.4. □

A.8 Structural-separation proofs

Proof of Proposition 3.2. Take the disjoint union of a star $K_{1,k}$ with center c and a single edge ab. For VC, fix the optimum cover $S^\star = \{c, a\}$, of cost 2, and compare the partial out-commitments $T_c = (\varnothing, \{c\})$ and $T_a = (\varnothing, \{a\})$. Each makes exactly one decision contrary to $S^\star$. Excluding c forces all k leaves into the cover, so the conditioned optimum is $k+1$ and AIR is $k-1$. Excluding a merely replaces it by b, leaving the conditioned optimum at 2 and AIR at zero.

For MaxCut, choose an optimal labeling that satisfies all $k+1$ edges. Flipping only the star center relative to this labeling makes all k star edges unsatisfied, so the full commitment has regret k. Flipping one endpoint of the isolated edge loses exactly that edge and has regret 1. Both predicted labelings have Hamming distance one from the same optimum. Global-complement symmetry changes neither distance modulo orientation nor cut value. □

B EXTENDED EXPERIMENTAL EVIDENCE

Table 2: AIRGraph instantiations. The framework shares the AIR–gain–fallback contract; commitments, certificates, and repairs remain problem-specific.

Problem	Commitment	Safe certificate	Repair or downstream use
Weighted VC	A in, independent B out	NT, or its dominating local-LP strengthening	force F, solve H, factor-2 fallback
Weighted MIS	complementary VC commitment	exact additive transfer of the VC bound	complete a feasible set; compare with fallback
Signed MaxCut	relative labels on region W	exact, flow, or cheap bound (bare LP equals cheap)	parity contraction; α-approximate residual and fallback

This appendix gives the full protocol and numerical results behind Section 5. It separates finite implementation checks, strong-baseline comparisons, mechanism studies, and transfer analyses. We abbreviate Machine Learning for Combinatorial Optimization Benchmark 101 as ML4CO-Bench-101 (Ma et al., 2025b).

B.1 STUDIES, OPERATING POINTS, AND STATISTICAL UNITS

Table 3: Experimental studies and statistical units. "Graph" is the independent unit unless otherwise stated; results are not pooled across problems.

Study	Role	Data and statistical unit	Test-time interface
Exact implementation check	correctness	12 solvable graphs per problem; sizes 10, 12, 14, 16	exact AIR against NT/local-LP VC–MIS and exact/flow/cheap MaxCut bounds; bare-LP identity check
MaxCut BA-large holdout	main MaxCut	480 ML4CO-Bench-101 BA-large graphs, indices 20–499	fixed GNN4CO predictions; cheap-certificate 12-proposal portfolio
GNN4CO common polishing	strong baseline	same 480 BA-large graphs; one observation per graph after five solver seeds	identical one-flip polish; prediction-free and prediction-seeded search with paired post-inference budgets
COExpander replication	predictor generality	500 WS-small graphs; one observation per graph after five solver seeds	eight-sample COExpander marginals; common polish; both matched-time search controls
GNN4CO topology stress	structural mechanism	same 480 BA-large graphs; three perturbation seeds clustered within graph	exact-count connected/independent label flips with confidence magnitudes preserved; gated, unfiltered, and confidence-count-matched policies
COExpander topology stress	predictor replication	same 500 WS-small graphs; three perturbation seeds clustered within graph	identical perturbation and policy controls applied to fixed COExpander predictions
MIS five-seed holdout	main MIS	100 held-out RB800–1200 graphs under five inference seeds; clustered by graph	fixed StruDiCO predictions; forced-weight top-3; NT evaluated after selection
MIS matched-time baseline	scope check	same 100 graphs and five prediction seeds; clustered by graph	prediction-free classical search with paired post-inference budgets
MaxCut certificate comparison	mechanism	500 GNN4CO BA-small sparse graphs	cheap, bare-LP, and flow bounds
MaxCut controlled gate ablation	confirmatory mechanism	480 BA-large graphs	common prediction, 12-proposal pool, residual budgets, fallback, and one-flip polish; gated, unfiltered, and prediction-only count-matched policies
External-commitment timing	incremental certificate cost	same 480 BA-large graphs; 5,760 commitments, 30 timed repetitions after five warm-ups	cheap certificate evaluated after the commitment is materialized; prediction, proposal construction, and completion excluded
Boundary diagnostics	post-hoc boundary analysis	frozen gate, exact-audit, and paired-search logs; graph is the resampling cluster	count-matched and fixed-K size controls, normalized additive slack, within-size correlation, and best-of composition; no new solver calls
MaxCut WS-small transfer	transfer	500 independent WS-small graphs	flow-certificate portfolio versus confidence control
VC RB-large study	exploratory VC	500 ML4CO-Bench-101 RB-large graphs	certified structural repair
VC RB-giant scale transfer	transfer	20 RB-giant graphs	same structural repair
VC gate analysis	mechanism	same 20 RB-giant graphs	certified versus unfiltered completion
VC size-control study	transfer follow-up	30 fresh RB-giant graphs, indices 20–49	fixed-size versus fractional-size commitments

All large-scale experiments consume fixed prediction artifacts, which isolates test-time certification and repair from predictor training. MaxCut and VC use one observation per graph; intervals for the main results resample graphs $10{,}000$ times. We report percentile intervals from a fixed bootstrap seed of 2027. MIS resamples the 100 graph clusters while preserving all five prediction-seed rows within each cluster. A *committed region* is the set of vertices fixed by a proposal. Fractional-size studies specify its size as a fraction of $|V|$; the fixed-size VC study instead uses 50, 75, 100, 150, 200, 250, 300, or 400 vertices. For MaxCut, confidence-frontier growth repeatedly adds the highest-confidence vertex adjacent to the current region; certificate-guided growth uses the same connected expansion with a structural gain proxy. The 12-proposal portfolio is the union of fractions $0.25, 0.375, 0.50, 0.625, 0.75$, and 0.875 under these two growth rules. Parity contraction converts an accepted regional labeling into a residual Max-2Lin instance. Deterministic conditional expectation supplies a half-total-weight feasible residual solution, and a time-limited HiGHS incumbent replaces it only when its reconstructed value is larger. Common-polish comparisons then apply the same deterministic all-vertex one-flip routine and unrestricted fallback to the AIRGraph and direct-prediction arms; these operations need not preserve the proposal labels. The prediction-free matched-time arm starts without learned labels; the stronger prediction-seeded arm starts from the commonly polished prediction and alternates incumbent perturbations with random restarts. Both use the recorded AIRGraph post-inference budget paired to each instance and search seed and neither reads certificate or reference values. In the controlled gate ablation, all 12 polished proposal completions are solved once and reused for paired policy accounting. The unfiltered policy completes all 12; the certificate-gated policy completes proposals with positive robust gain; and the confidence control receives the same number of completion slots but ranks the pool by mean committed confidence. Two count-matched post-hoc controls inherit the gate's per-instance allocation: one ranks the largest proposals, while the size-profile control also matches the gate's accepted count at each size and then ranks by confidence. Separately, fixed-K largest-size policies always select 8, 9, or 10 proposals and are deployable without certificate values. None of these post-hoc controls makes new solver calls. The topology study applies exact-count flips to the fixed predictor output, preserves each vertex's confidence magnitude, and uses no reference solution for perturbation or policy decisions. Connected sets are generated by breadth-first growth from random roots and independent sets by sampling without replacement; the clean condition has one row per graph, while each nonzero condition has three within-graph corruption seeds. For MIS, forced-weight top-3 and size-ranked top-3 sort the same 12 proposals by decreasing $w(B)$ and decreasing commitment size, respectively; ties are broken by proposal strategy and identifier, after which the selected three proposals use the same weighted-greedy completion and fallback.

Hardware and software. Experiments ran on one x86_64 node with one Advanced Micro Devices (AMD) EPYC 9754 processor, 128 physical cores, 256 hardware threads, one socket, and frequency boost enabled. The strong-baseline extensions used AIRGraph 0.1.0, Python 3.10.21, NumPy 2.2.6, SciPy 1.15.2, and PyYAML 6.0.3. Optional neural-export dependencies and the environment lock are included in the artifact. Timing comparisons are made only between arms of the same study. Matched-time budgets account for the recorded post-inference proposal, certificate, fallback, completion, and common-polish stages; neural inference is recorded separately and excluded from both compared arms. Controlled-gate runtime is the sum of fallback, proposal construction, certificate evaluation, the inner residual-solver timer, and polishing for the completions charged to a policy. The inner timer includes residual model construction and presolve, but outer regional contraction, assignment reconstruction, and original-graph objective reevaluation are not separately timed. The external-commitment microbenchmark instead isolates one cheap-certificate call after a commitment is materialized; its numbers are not substituted into another study's policy accounting.

B.2 EXACT IMPLEMENTATION CHECKS

Table 4: Finite exact implementation check. The largest positive numerical residual is $\mathrm{AIR} - U$; none exceeds the 10^{-8} tolerance. The singleton column compares a joint NT certificate with the sum of NT certificates computed separately for every committed decision. "Grouped" counts the 48 multi-vertex proposals.

Problem	Exact graphs	Largest $\mathrm{AIR} - U$	Joint NT $\leq$ singleton sum	Local LP versus NT	Positive AIR accepted
VC	12	5.33×10^{-15}	48 safe; 38 strict	192 safe; 158 strict (40 grouped)	30
MIS	12	3.55×10^{-15}	48 safe; 38 strict	192 safe; 158 strict (40 grouped)	additive audit only
MaxCut	12	2.66×10^{-15}	—	—	113

For a multi-vertex commitment, the *singleton-sum diagnostic* adds the certificates obtained after isolating each committed decision and ignores their joint overlap. Joint NT is no worse than this diagnostic on all 48 grouped proposals and is strictly tighter on 38. In the separate local-LP extension, the local certificate is safe on all 192 rows and strictly tighter than NT on 158; among the 48 grouped proposals, it is strictly tighter than NT on 40. Stress proposals are generated only for this check: starting from an exact optimum, we flip one of up to four highest weighted-degree vertices, place that vertex first, grow a deterministic structural order to commitment fractions $0.25, 0.5, 0.75, 1.0$, and for VC/MIS project out-cover labels to an independent set while retaining the flipped root. Exact optima are never used by the deployed proposal pipeline. Figures 3 and 4 visualize these checks.

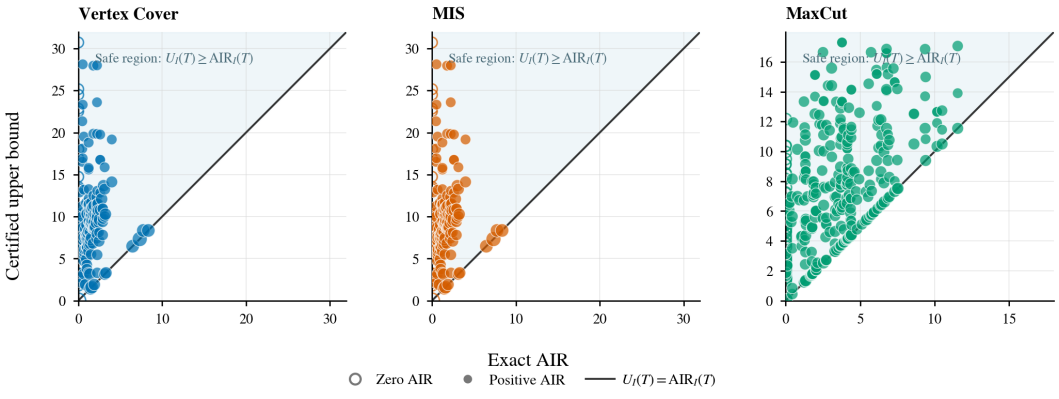

Figure 3: Finite certificate safety check. Each panel plots exact AIR against the implemented certificate upper bound on enumerated small instances; the diagonal marks equality and the shaded half-plane is the safe region. Hollow and filled markers denote zero- and positive-AIR commitments. Coincident coordinate pairs are collapsed, with marker area indicating repeated rows.

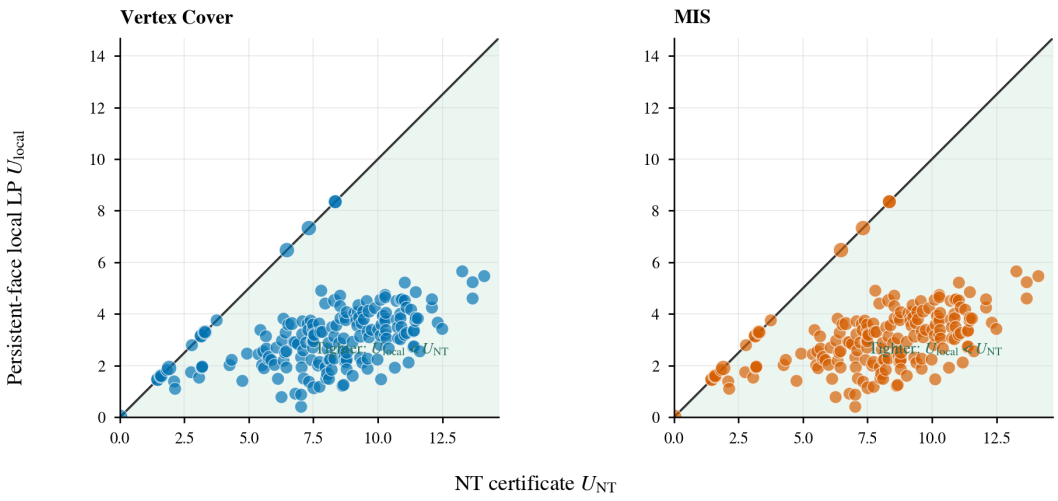

Figure 4: Persistent-face local-LP tightening check for VC and its additive MIS transfer. Points compare U_{local} with U_{NT}; the shaded region below the diagonal is strictly tighter, while the diagonal denotes equality. Coincident coordinate pairs are collapsed, with marker area indicating repeated rows.

B.3 Certificate tightness and isolated evaluation cost

Table 5 reports two complementary finite-audit quantities. The normalized additive slack $(U - \text{AIR})/W$ remains defined when exact AIR is zero, where W is total edge weight for MaxCut and total vertex weight for VC/MIS. The ratio U/AIR is reported only on positive-AIR rows because it is unstable near zero. The audit covers the portfolio-selected joint bound for MaxCut (224 exact-joint and 6 flow rows) and the persistent-face local-LP bound for VC/MIS on enumerable instances; it does not estimate BA-large cheap-bound tightness. Identical VC and MIS distributions up to floating-point precision provide an implementation cross-check of Theorem 4.5.

Table 5: Finite certificate-tightness diagnostics on 12 exactly solved graphs per problem. MaxCut uses the selected exact-joint/flow bound; VC/MIS use local LP. Brackets are interquartile ranges across proposal rows, and ratios exclude zero-AIR rows.

Problem	Rows	Zero AIR	Median $(U - \text{AIR})/W$	Median U/AIR
MaxCut	230	38	$0.114\ [0.008, 0.195]$	$1.87\ [1.00, 3.23]$
VC	228	36	$0.103\ [0.058, 0.147]$	$2.08\ [1.50, 3.21]$
MIS	228	36	$0.103\ [0.058, 0.147]$	$2.08\ [1.50, 3.21]$

We separately time the cheap MaxCut certificate on the 480 BA-large holdout graphs after each commitment has been materialized. The 5,760 graph–commitment pairs each receive five warm-up calls and 30 timed calls. The per-commitment median runtime is 1.528 ms, with IQR $[1.052, 1.966]$ ms and 95th percentile 2.526 ms; no repeated call changes the certificate output. Summing the 12 independently timed per-commitment medians gives a graph-level median 18.359 ms (IQR $[16.655, 20.538]$ ms). In the same run, constructing the 12 proposals takes median 0.740 s. The isolated timer covers certificate evaluation after materialization; prediction inference, proposal construction, fallback, and residual completion lie outside its scope.

B.4 COMMON POLISHING, MATCHED TIME, AND PREDICTOR GENERALITY

Table 6: MaxCut objective and post-inference runtime under common polishing and matched-time controls. AIRGraph+polish and prediction+polish use the same one-flip routine. Prediction-seeded search uses a live remeasured paired budget; other runtimes are retained from their source studies and should be compared only within the stated pairing. The benchmark reference is a supplied feasible solution, not a certified optimum.

Predictor/data	Arm	Mean objective	Mean post-inference time
GNN4CO/BA-large	raw prediction	2830.39	—
	prediction+polish+fallback	2919.98	0.0103 s
	AIRGraph	2921.74	1.3933 s
	AIRGraph+polish	2928.54	1.4029 s
	prediction-seeded matched-time	2956.29	0.8722 s
	prediction-free matched-time	2867.08	1.3371 s
	benchmark reference	2939.74	—
COExpander/WS-small	raw prediction	821.28	—
	prediction+polish+fallback	863.69	0.00241 s
	AIRGraph	861.51	0.1103 s
	AIRGraph+polish	863.31	0.1126 s
	prediction-seeded matched-time	867.53	0.0850 s
	prediction-free matched-time	845.48	0.1068 s
	benchmark reference	872.12	—

The GNN4CO comparison yields 8.56 $[7.77, 9.33]$ over direct prediction plus the same polish and 61.45 $[59.67, 63.27]$ over prediction-free search, but -27.75 $[-28.66, -26.86]$ against seeded search. The corresponding graph-level paired relative effects are 0.30%, 2.14%, and -0.94%. For COExpander, the absolute effects are -0.38 $[-0.66, -0.10]$, 17.83 $[17.25, 18.40]$, and -4.22 $[-4.52, -3.92]$, with paired relative effects -0.04%, 2.10%, and -0.48%. Relative to the supplied GNN4CO benchmark reference, common polishing has a mean deficit of 19.76 and AIRGraph reduces it to 11.20, closing 43.3% of that deficit; seeded search exceeds the supplied reference and confirms that it is a feasible reference rather than a certified optimum. AIRGraph outputs themselves exceed the reference on 120 GNN4CO and 11 COExpander instances.

B.5 STRUCTURAL MARGINS, FEASIBLE-VALUE BOUNDS, AND THE PREDICTION-SEARCH FRONTIER

For each BA-large graph at the 0.2 s controlled operating point, we take the largest positive accepted gain $G_I(T)$ and use $\mathrm{OPT}_{\mathrm{MC}} \le W_I := \sum_e w_e$. Theorem 3.4 then gives

$$v(R_T) \ge \tfrac{1}{2}\,\mathrm{OPT}_{\mathrm{MC}} + G_I(T) \ge \left(\tfrac{1}{2} + \frac{G_I(T)}{W_I}\right)\mathrm{OPT}_{\mathrm{MC}}\,. \tag{21}$$

All 480 normalized structural factors exceed $1/2$; the range is $[0.655, 0.689]$, the median is 0.673, and its graph-bootstrap interval is $[0.673, 0.674]$. The quantity is computable without benchmark references and is fixed once the commitment, certificate, and repair credit are known, before running residual completion. It attributes a pre-completion margin to that commitment–repair pair. After any feasible solution S is computed, the same upper bound also gives $v(S)/\,\mathrm{OPT}_{\mathrm{MC}} \ge v(S)/W_I$. In the controlled study, the independent fallback has mean/median $v(S_{\mathrm{fb}})/W_I = 0.713/0.713$ and exceeds Equation (21) on all 480 graphs; the final gated output has mean/median $0.732/0.732$.

The strong-baseline outputs provide the corresponding observed comparison under the same total-edge-weight normalization. On GNN4CO/BA-large, the graph-level mean ratios are 0.729 for AIRGraph plus polishing, 0.727 for direct prediction plus polishing, and 0.736 for prediction-seeded matched-time search. These feasible-value ratios are stronger post-computation certificates on this benchmark, but they do not attribute any part of the value to the safety of the enforced commitment.

Figure 5a reports the seeded-search budget frontier. On GNN4CO, AIRGraph is ahead by 8.56 before extra search, but is behind by 5.98 once search receives 10% of AIRGraph's remaining budget and by 27.75 at the full paired budget. On COExpander, the commonly polished prediction is already ahead by 0.38, and the gap reaches 4.22. The full-budget paired relative effects are -0.94% and -0.48%, respectively. Panel (b) shows that the clean-condition confidence advantage disappears under size

controls, while panel (c) reports the normalized structural factors. These observations separate three quantities that a single objective table conflates: a commitment-specific guarantee, heuristic search quality, and the computational consequences of commitment size.

The same separation appears within the proposal pool. After centering certified gain and realized improvement within each graph–proposal-size cell, Pearson correlation is -0.183 with graph-cluster bootstrap interval $[-0.213, -0.152]$; Spearman correlation is -0.126 $[-0.161, -0.092]$. Commitment structure also changes residual difficulty, so these descriptive associations support no causal interpretation. Within this clean portfolio, certified gain behaves as a safety margin rather than an independent utility-ranking signal.

AIRGraph and seeded search compose by returning their better observed feasible solution. Table 7 computes this best-of arm from frozen paired logs. It weakly dominates seeded search by construction; the displayed runtime is the sum of both arms, and the result documents algebraic composability under that additional computation.

Table 7: Post-hoc best-of composition from frozen paired logs. The objective effect is best-of minus prediction-seeded search; sequential runtime sums both recorded arms.

Predictor/data	Search value	Best-of value	Effect [95% CI]	Sequential time
GNN4CO/BA-large	2956.288	2956.302	0.014 [0.003, 0.029]	1.793 s
COExpander/WS-small	867.528	867.616	0.088 [0.062, 0.117]	0.175 s

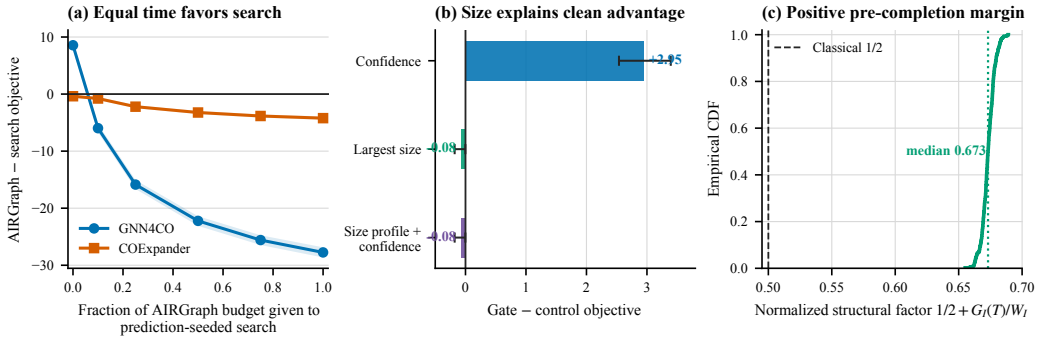

Figure 5: Certification–search–selection tradeoffs. (a) AIRGraph minus prediction-seeded search as the latter receives a fraction of AIRGraph's paired post-inference budget; bands are graph-cluster bootstrap 95% intervals. (b) Gated minus count-matched controls at the 0.2 s operating point; confidence ranking loses by 2.95, whereas the largest-commitment and size-profile controls win by 0.079. (c) Empirical distribution of the normalized structural factor in Equation (21); the dotted line marks the median. This pre-completion factor attributes certified gain to the commitment–repair pair; the text reports stronger post-completion feasible-value ratios. The size and factor results are post-hoc diagnostic analyses of frozen completions and make no new solver calls.

B.6 MaxCut certificate frontier and gate cost

Table 8: MaxCut certificate comparison on 500 GNN4CO BA-small sparse graphs and 2,008 matched proposals. Times are mean certificate time per graph; final value is the observed feasible objective selected by each pipeline.

Certificate	Strictly tighter than cheap	Gate upgrades	Accepted / graph	Certificate time	Mean final value
Cheap	—	—	0.764	0.00130 s	708.890
Bare LP (= cheap)	0/2,008	0	0.764	0.0241 s	708.890
Flow	2,008/2,008	541	1.846	4.977 s	714.566

The bare LP and cheap pipelines tie because Theorem 4.7 proves $U^{\mathrm{LP}} = U^{\mathrm{cheap}}$. Flow is strictly tighter on all 2,008 proposals and improves the final objective on 149 of 500 graphs, with 351 ties and no losses.

The certificate-frontier study isolates bound tightness, whereas the controlled ablation in Appendix B.7 isolates proposal filtering under a common candidate and completion interface. Both final outputs retain the fallback's factor; the additional guarantee is that every completed certificate-gated candidate has a positive instance-wise margin over the classical worst-case bound.

B.7 Controlled certificate-gate ablation

Table 9: Controlled MaxCut gate ablation on 480 BA-large graphs. Each policy shares the prediction, 12-proposal pool, per-proposal HiGHS limit, fallback, and deterministic one-flip polish. "Prediction" is the confidence-ranked completion-count-matched selector; it is not wall-time matched. Accounted runtime triples are gate/prediction-only/unfiltered under the stage definition in Appendix B.1. "Saved" averages the per-instance saving, so it need not equal the ratio of displayed means.

Time limit	Completed	Runtime triple	Saved	Gain retained	Gate − prediction [95% CI]
0.05 s	9.031/9.031/12	26.13/37.25/39.14 s	31.9%	99.85%	$3.18\ [2.73, 3.64]$
0.20 s	9.031/9.031/12	31.46/57.37/59.30 s	46.9%	99.88%	$2.95\ [2.52, 3.39]$
0.50 s	9.031/9.031/12	32.97/59.80/61.74 s	46.5%	99.88%	$2.80\ [2.39, 3.21]$

Against the confidence count-matched selector, the certificate gate has W/L/T counts 220/10/250, 211/8/261, and 211/8/261 as the time limit increases. Against unfiltered completion, the corresponding counts are 0/10/470, 0/8/472, and 0/8/472, with mean objective effects $-0.117\ [-0.221, -0.038]$, $-0.092\ [-0.190, -0.019]$, and $-0.092\ [-0.190, -0.019]$. The gate retains more than 99.8% of aggregate unfiltered improvement at every operating point. Its lower runtime than the confidence selector is not imposed by design: confidence ranking favors smaller commitments with larger residuals, whereas accepted commitments are larger on average.

Table 10: Post-hoc count-matched controls for the same frozen completions. Every control receives the gate's per-instance completion count. "Size profile + confidence" additionally matches the number of gate acceptances at each commitment size before ranking by confidence. Objective and accounted-runtime effects are gate minus control.

Limit	Gate − confidence (objective)	Gate − largest size (objective)	Gate − largest size (accounted time)	Gate − size profile (objective)
0.05 s	$3.18\ [2.72, 3.64]$	$-0.096\ [-0.200, -0.017]$	$-1.366\ [-1.761, -0.981]$ s	$-0.096\ [-0.196, -0.017]$
0.20 s	$2.95\ [2.52, 3.39]$	$-0.079\ [-0.183, -0.002]$	$0.370\ [0.315, 0.426]$ s	$-0.079\ [-0.179, -0.002]$
0.50 s	$2.80\ [2.40, 3.21]$	$-0.079\ [-0.179, -0.002]$	$0.363\ [0.301, 0.425]$ s	$-0.079\ [-0.179, -0.004]$

The count-matched size controls reverse the confidence comparison at all three limits, although only 7–8 graphs account for the objective losses and at least 471 tie. At 0.20 and 0.50 s, the size rule is also slightly faster; at 0.05 s, the gate is faster but loses 0.096 objective, yielding a genuine but small tradeoff. Because the count is inherited from the gate, this comparison is diagnostic rather than an independently deployable policy.

Table 11: Deployable fixed-K largest-size controls on the same frozen completions. Each policy always selects K proposals without certificate values. "Saved" is the mean per-instance accounted-runtime saving relative to unfiltered completion.

Limit	Policy	Mean objective	Mean accounted runtime	Saved
0.05 s	$K = 8$	2937.267	21.04 s	44.4%
	$K = 9$	2937.392	27.31 s	28.8%
	$K = 10$	2937.402	32.19 s	16.9%
0.20 s	$K = 8$	2938.785	22.97 s	61.2%
	$K = 9$	2938.875	30.83 s	48.0%
	$K = 10$	2938.885	39.06 s	34.1%
0.50 s	$K = 8$	2939.029	24.17 s	60.8%
	$K = 9$	2939.119	32.33 s	47.6%
	$K = 10$	2939.129	40.86 s	33.8%

For reference, gate objectives are 2937.296, 2938.796, and 2939.040 at the three limits, with mean accounted runtimes 26.13, 31.46, and 32.97 s. The fixed-K policies trace a competitive clean-condition frontier without consulting certificates, and the size controls remove the original gate–confidence advantage within this pool. Accepted proposals still carry positive certified gain. Completed proposal values are computed once and reused for all policies; runtime accounting charges each policy only for recorded stages it would execute. HiGHS time limits govern the inner optimization call; its model construction and presolve are timed, whereas outer contraction, reconstruction, and original-graph reevaluation are excluded. No policy output falls below the fallback, and no gated output exceeds its unfiltered counterpart.

B.8 Controlled prediction-topology stress

The topology stress test contains 12,000 instance rows and 144,000 proposal rows for GNN4CO, and 12,500 instance rows and 150,000 proposal rows for COExpander. The recorded hard gates verify complete condition grids, exact flip counts, preserved confidence magnitudes, matched completion counts, fallback dominance, gated–unfiltered ordering, and absence of benchmark-reference access. Table 12 reports the clean and 20% endpoints; Figure 2 gives every intermediate perturbation rate.

Table 12: Controlled prediction-topology stress endpoints. The confidence policy ranks the same 12-proposal pool without certificate values and receives exactly the gate's completion count. Effects are gated minus comparator; positive is better. "Saved" is the fraction of unfiltered proposal completions avoided.

Predictor/data	Perturbation	Gate − confidence [95% CI]	W/L/T	Gate − unfiltered [95% CI]	Saved
COExpander/WS-small	clean	0.22 $[0.12, 0.31]$	81/19/400	-0.10 $[-0.15, -0.05]$	28.5%
	connected 20%	3.80 $[3.54, 4.07]$	454/32/14	-0.50 $[-0.59, -0.41]$	51.2%
	independent 20%	1.62 $[1.44, 1.80]$	383/67/50	-0.85 $[-0.96, -0.74]$	63.9%
GNN4CO/BA-large	clean	0.79 $[0.52, 1.05]$	117/31/332	-0.28 $[-0.42, -0.16]$	24.7%
	connected 20%	8.47 $[7.89, 9.04]$	435/18/27	-4.54 $[-4.94, -4.15]$	72.6%
	independent 20%	7.47 $[7.01, 7.94]$	454/14/12	-1.04 $[-1.22, -0.88]$	59.9%

Table 13: Size-controlled topology endpoints from the same frozen proposal completions. Effects are gate minus control. Neither diagnostic uses certificate values for ranking, but both inherit the gate's completion allocation; the size-profile control additionally conditions on the accepted size histogram.

Predictor/data	Perturbation	Gate − largest size	Gate − size profile + confidence
COExpander/WS-small	clean	-0.066 $[-0.114, -0.026]$	-0.072 $[-0.130, -0.018]$
	connected 20%	0.243 $[0.139, 0.350]$	1.448 $[1.261, 1.643]$
	independent 20%	-0.229 $[-0.330, -0.127]$	0.071 $[-0.065, 0.205]$
GNN4CO/BA-large	clean	-0.210 $[-0.346, -0.098]$	-0.210 $[-0.348, -0.096]$
	connected 20%	5.150 $[4.556, 5.728]$	5.283 $[4.671, 5.904]$
	independent 20%	2.820 $[2.514, 3.128]$	5.102 $[4.711, 5.494]$

The clean rows of Table 13 show that both size controls reverse the original gate–confidence advantage. Under perturbation, GNN4CO retains a clear effect beyond both size controls. COExpander does so against both controls for connected flips, but not for independent flips. Equal Hamming perturbations also induce different topology orderings. At 20%, connected flips expose 21.4% of total edge weight on WS-small versus 32.1% for independent flips, whereas on BA-large the ordering reverses to 39.0% versus 32.0% (Figure 6a–b). Gate acceptance responds differently as well: connected/independent acceptance is $48.8/36.1\%$ for COExpander and $27.4/40.1\%$ for GNN4CO (Figure 6c–d). Downstream consequence therefore depends jointly on the graph, prediction, and error location; error count, commitment size, and the coarse label "connected" are individually insufficient. These controlled overconfidence tests use synthetic perturbations relative to predictor outputs and do not represent natural corruptions or errors measured against an optimum.

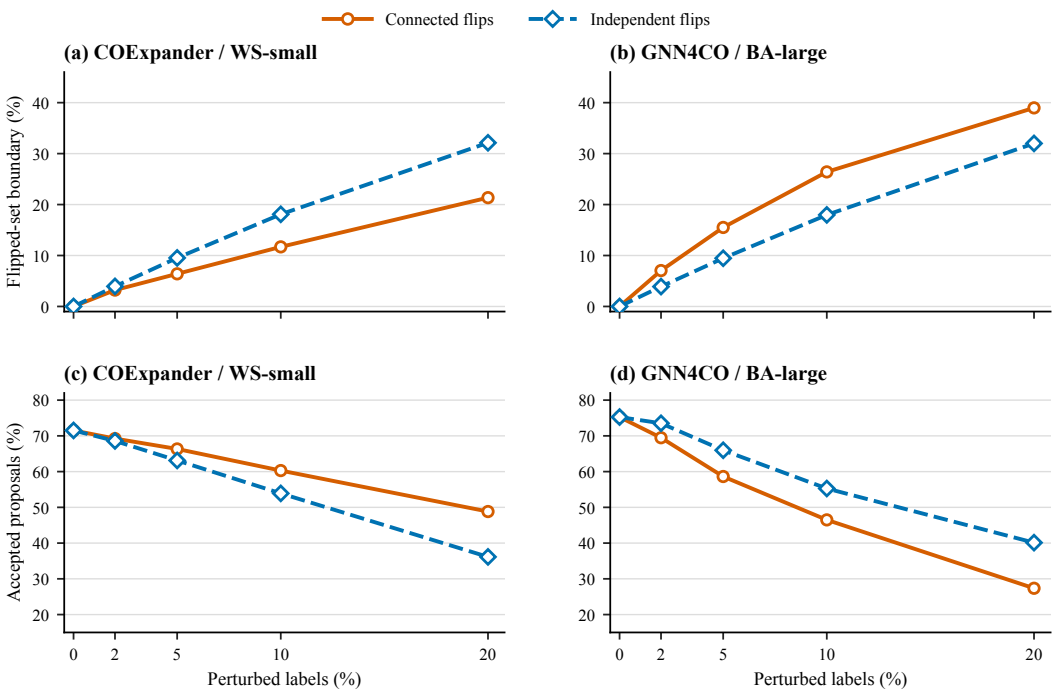

Figure 6: Structural diagnostics for the controlled perturbations. (a–b) Mean weight of edges with exactly one flipped endpoint, normalized by total edge weight. (c–d) Mean fraction of proposals accepted by the certificate gate. Connected and independent conditions use the same exact flip counts and preserve confidence magnitudes; the clean point is shared by both curves. The reversal between graph families rules out a universal connected-versus-independent ordering.

B.9 MIS COMPLETION EFFICIENCY AND SEED STABILITY

Table 14: MIS seed stability for forced-weight top-3. Retention compares its aggregate improvement with completing all 12 proposals for the same seed; runtime ratios cover the recorded completion pipeline.

Seed	Top-3 gain	Full gain	Retained	Mean top-3 − size top-3	Runtime / full
0	93	95	97.9%	0.23	49.6%
1	76	78	97.4%	0.26	44.0%
2	86	89	96.6%	0.25	48.8%
3	77	79	97.5%	0.16	47.0%
4	96	96	100.0%	0.28	47.5%
Aggregate	428	437	97.9%	0.236	—

The row-level forced-weight-versus-size comparison is 96/1/403 W/L/T over 500 graph–seed pairs. We report 39/0/61 after averaging within each of the 100 graph clusters because stochastic predictions for the same graph are not independent experimental units. Under the matched-time protocol, prediction-free search reaches mean objective 37.652 versus 34.806 for forced-weight top-3. The paired AIRGraph-minus-classical effect is -2.846 $[-3.136, -2.546]$ with 3/94/3 graph-level W/L/T. The additive AIR certificate remains valid, while the present MIS proposal and completion policy is not competitive with this matched-time classical arm.

B.10 COMPLETE EXPERIMENTAL ENDPOINTS

Table 15: Complete utility and transfer endpoints. Bootstrap intervals use graphs as clusters; effects are oriented so that positive is better and are rounded to two decimals. The WS-small certificate-frontier diagnostic reports its observed mean without a normal-approximation interval.

Problem	Comparison	Evidence role	Graphs	W/L/T	Mean effect [95% CI]
MaxCut	GNN4CO AIRGraph+polish versus prediction+polish	primary strong baseline	480	383/80/17	8.56 $[7.77, 9.33]$
MaxCut	GNN4CO AIRGraph+polish versus matched classical	primary strong baseline	480	478/0/2	61.45 $[59.67, 63.27]$
MaxCut	GNN4CO AIRGraph+polish versus prediction-seeded search	compute-matched boundary	480	1/479/0	-27.75 $[-28.66, -26.86]$
MaxCut	COExpander AIRGraph+polish versus prediction+polish	predictor replication	500	189/235/76	-0.38 $[-0.66, -0.10]$
MaxCut	COExpander AIRGraph+polish versus matched classical	predictor replication	500	499/1/0	17.83 $[17.25, 18.40]$
MaxCut	COExpander AIRGraph+polish versus prediction-seeded search	compute-matched boundary	500	26/457/17	-4.22 $[-4.52, -3.92]$
MaxCut	certified portfolio versus confidence control	structural control	480	50/0/430	0.61 $[0.42, 0.83]$
MaxCut	certified portfolio versus fallback	structural control	480	478/0/2	60.05 $[57.90, 62.14]$
MaxCut	certificate gate versus prediction-only count-matched	controlled gate, 0.20 s	480	211/8/261	2.95 $[2.52, 3.39]$
MaxCut	certificate gate versus largest-size count-matched	size control, 0.20 s	480	1/7/472	-0.08 $[-0.18, -0.00]$
MaxCut	certificate gate versus unfiltered completion	controlled gate, 0.20 s	480	0/8/472	-0.09 $[-0.19, -0.02]$
MaxCut	GNN4CO gate versus confidence, connected 20%	topology stress	480	435/18/27	8.47 $[7.89, 9.04]$
MaxCut	GNN4CO gate versus confidence, independent 20%	topology stress	480	454/14/12	7.47 $[7.01, 7.94]$
MaxCut	COExpander gate versus confidence, connected 20%	topology stress	500	454/32/14	3.80 $[3.54, 4.07]$
MaxCut	COExpander gate versus confidence, independent 20%	topology stress	500	383/67/50	1.62 $[1.44, 1.80]$
MIS	forced-weight top-3 versus fallback	internal utility	100	59/0/41	0.86 $[0.65, 1.07]$
MIS	forced-weight top-3 versus size top-3	ranking mechanism	100	39/0/61	0.24 $[0.16, 0.32]$
MIS	forced-weight top-3 versus matched classical	matched-time scope	100	3/94/3	-2.85 $[-3.14, -2.55]$
MaxCut	flow portfolio versus confidence control	WS transfer control	500	2/0/498	0.04 (observed mean)
VC	certified structural repair versus fallback	RB-large exploration	500	115/0/385	0.30 $[0.25, 0.36]$
VC	fixed structural repair versus fallback	RB-giant transfer	20	1/0/19	0.05 $[0.00, 0.15]$
VC	fixed-size regions versus fallback	RB-giant size control	30	2/0/28	0.07 $[0.00, 0.17]$

All 500 paired differences in the WS-small certificate-frontier comparison are nonnegative (2 wins, 498 ties). Because this zero-inflated sample is unsuitable for a Gaussian interval, Table 15 reports its observed mean and W/L/T count only.

B.11 VERTEX COVER SCALE-TRANSFER ANALYSIS

Table 16: VC scale-transfer sequence. Unfiltered and union rows are comparison controls. Improvements are reductions relative to the fixed fallback and are rounded to three decimals.

Study	Arm	Evidence role	Graphs	W/L/T	Mean (total) improvement
RB-large exploration	certified extended repair	exploratory	500	115/0/385	0.302 (151)
RB-large exploration	unfiltered extended repair	unfiltered control	500	172/0/328	0.462 (231)
RB-giant scale transfer	fixed structural portfolio	transferred rule	20	1/0/19	0.050 (1)
RB-giant gate analysis	unfiltered extended repair	comparison control	20	5/0/15	0.250 (5)
RB-giant size control	fixed-size regions	fixed-size rule	30	2/0/28	0.067 (2)
RB-giant size control	fractional-size regions	comparison control	30	2/0/28	0.100 (3)
RB-giant size control	union portfolio	inclusion control	30	3/0/27	0.133 (4)

The gate analysis contains 480 proposal rows: 6.25% are accepted, and 3 of 20 graphs have at least one accepted proposal. Switching from fractional-size to fixed-size regions does not increase acceptance: the fixed-size rule accepts 35 of 720 proposals (4.86%), compared with 49 of 720 (6.81%) for the fractional-size control. Changing region-size parameterization alone therefore does

not alleviate low acceptance at this scale; certificate tightness and proposal quality remain potential constraints.

C Experiment and Artifact Index

Table 17: Experiment entry points in the anonymous artifact and the results they reproduce.

Experiment entry point	Study type	Reproduced outputs
`certificate-audit`	finite implementation check	exact AIR, certificate residuals, local/NT comparisons, and positive-regret acceptance
`maxcut-certificate-frontier`	MaxCut mechanism	cheap/bare-LP/flow comparison, acceptance, runtime, and final objective
`maxcut-ws-transfer`	MaxCut transfer	independent WS-small comparison with the confidence control
`maxcut-ba-large-portfolio`	main MaxCut	BA-large portfolio, confidence, and fallback comparisons
`maxcut-gate-ablation`	MaxCut mechanism	candidate-completion savings and retained improvement
`matched-time-maxcut`	strong baseline	prediction-free search under paired post-inference budgets
`matched-time-mis`	MIS scope check	matched-time prediction-free comparison across five prediction seeds
`maxcut-expensive-completion-gate`	gate value	time-limited HiGHS completion, accounted stage-time savings, and retained improvement
`maxcut-controlled-gate-ablation`	controlled gate	common-pool gated, unfiltered, and prediction-only completion-count-matched policies with common polishing
`maxcut-gnn4co-common-polish`	strong baseline	identical one-flip polish and matched-time controls on BA-large
`maxcut-coexpander-generality`	predictor replication	COExpander common-polish and matched-time controls on WS-small
`maxcut-gnn4co-prediction-matched-time`	compute-matched boundary	prediction-seeded iterated search with live paired AIR-Graph budgets on BA-large
`maxcut-coexpander-prediction-matched-time`	compute-matched boundary	identical prediction-seeded control on WS-small
`maxcut-gnn4co-prediction-budget-frontier`	budget sensitivity	prediction-seeded refinement at six fractions of the paired post-inference budget
`maxcut-coexpander-prediction-budget-frontier`	budget sensitivity	identical budget frontier on WS-small
`maxcut-gnn4co-prediction-topology-stress`	topology stress	exact-count connected/independent perturbations and fixed-pool gate controls on BA-large
`maxcut-coexpander-prediction-topology-stress`	topology stress	identical perturbation and gate controls on the independent WS-small predictor
`maxcut-decision-tradeoffs`	post-hoc mechanism audit	size controls, normalized structural factors, source hashes, and hard-gate checks without new solver calls
`maxcut-external-certificate-timing`	isolated efficiency	repeated cheap-certificate timing for pre-materialized commitments, excluding prediction, proposal generation, and completion
`revision-diagnostics`	post-hoc boundary audit	fixed-K and count-matched size controls, normalized additive slack, within-size correlations, and best-of composition
`mis-seed-stability`	main MIS	five-seed graph-cluster results and top-3 completion efficiency
`vertex-cover-rb-large-exploratory`	exploratory VC	RB-large structural-repair comparison
`vertex-cover-rb-giant-confirmation`	VC scale transfer	RB-giant structural-repair comparison
`vertex-cover-gate-diagnostic`	VC diagnostic	certified and unfiltered completion under gate attrition
`vertex-cover-absolute-budget-confirmation`	VC size control	fresh-holdout fixed-size and fractional-size comparisons
`paper-statistics`	reporting audit	aggregation units, confidence intervals, and endpoint reconciliation