GRAPH-CERTIFIED ANYTIME-VALID TESTING: WHAT A DEPENDENCY GRAPH BUYS, AND WHAT CERTIFICATION COSTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Sequential monitoring of many machine-learning systems requires combining evidence from streams that share observations or sources of randomness. Multiplying evidence as if the streams were independent invalidates the test; dependence-agnostic aggregation preserves validity but discards structure. We study *graph-certified* anytime-valid testing, in which a conditional dependency graph D licenses exactly those products of local betting factors that are supported on independent sets, and a predictable learner allocates wealth across them. Our results are two-sided. On the positive side, the achievable log-growth multiplier is the *independence number* $\alpha(D)$, not the chromatic quantity $n/\chi_f(D)$; the two coincide on vertex-transitive graphs but the gap can be strict off that class, and n/χ_f understates the multiplier by up to $1.80\times$ on stars and grids in our sweep. On the negative side, a dependency graph is a $0/1$ notion and cannot express weak dependence, so certification has a price: on a corrupted-clique family we compute it in closed form and show it rises from 1 under exact duplication to the full clique size r as within-clique dependence vanishes. We give the domination result that actually justifies restricting to independent sets in the bounded-bet regime the algorithm uses, show that a ρ-approximate maximum-weight independent set inflates the expected rejection time by at most $1/\rho$, and introduce a balanced certified test that is within $\log 2$ of the better of the certified product and the cumulative mean, extending the sparse/dense trade-off known for independent streams to certified dependence. Across $10,000$ null trajectories per graph every valid method stays below α while the independence product inflates to 0.080–0.132. Finally, on growth rate a graph neural policy is beaten by a strong non-neural policy using the same certificate—on this problem the structure, not the representation learning, carries most of the gain—though the two are competitive on expected detection time, and we report both rather than the more flattering one.

1 INTRODUCTION

Machine-learning systems are increasingly monitored continuously rather than evaluated once: model endpoints are audited as traffic arrives, sensors are watched for drift, evaluation suites are re-run as models are updated. Such monitoring is naturally sequential, and the appropriate guarantee is *anytime validity*—error control that survives optional stopping (Ramdas et al., 2023; Grünwald et al., 2024). For a global null $\mathcal{P}_0$ asserting that every monitored stream is behaving, one wants a wealth process M_t with

$$\sup_{P\in\mathcal{P}_0} P\big(\exists\, t \geq 1 : M_t \geq 1/\alpha\big) \leq \alpha. \tag{1}$$

The streams, however, are rarely independent. Endpoints share evaluation units, sensors share weather, evaluation tasks share prompts. Multiplying local evidence as though the streams were independent breaks (1); in our experiments it raises the crossing probability from a nominal 0.05 to 0.080–0.132. Aggregation rules valid under arbitrary dependence—weighted averages of e-values—restore validity, but by construction they cannot exploit whatever independence is genuinely present.

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.

Between these extremes sits a natural object: a *dependency graph* D whose non-adjacent vertex sets index conditionally independent subvectors. Such a graph licenses the product of local betting factors over any independent set. This paper asks what that licence is worth, and what it costs.

Contributions.

1. **A construction and the domination result that justifies it** (§3). Products over independent sets, mixed with predictable weights, form a test supermartingale (Theorem 3). The usual justification—valid multilinear merging characterised with *unbounded* e-variables—does not apply to the bounded betting family the algorithm uses; we give the statement that does: every valid non-negative multilinear form is *pointwise* dominated by a mixture of independent-set products (Theorem 4).
2. **The right constant is $\alpha(D)$, not $n/\chi_f(D)$** (§4). The achievable multiplier is a maximum-weight independent set value, $\alpha(D)$ under homogeneous signal. Since $n/\chi_f(D) \le \alpha(D)$—equal on vertex-transitive graphs, strict off that class—the achievable growth rate is governed by a different, generally larger, parameter than Janson's sample-size re-weighting (Janson, 2004) (Theorem 7).
3. **The price of certification** (§5). A dependency graph cannot express weak dependence. On a corrupted-clique family, the closed-form ratio between the oracle joint likelihood ratio and the best certified product rises from 1 at exact duplication to the clique size r as dependence vanishes (Theorem 8): certification buys distribution-free validity under a composite null, not information.
4. **Computation transfers linearly** (§6). A ρ-approximate maximum-weight independent set yields growth rate at least ρ times optimal and inflates the first-order expected rejection time by at most $1/\rho$ (Theorem 10); exactly solvable on perfect graphs, inapproximable in general.
5. **A balanced certified test** (§3.3), within $\log(1/\pi)$ of both the certified product and the cumulative mean on every path, extending Bharti et al. (2026)'s sparse/dense trade-off from independent streams to a certified dependency graph.
6. **Experiments** (§8): null validity at $10,000$ trajectories per graph, the information rate, the price curve, approximation transfer, power without hiding non-detections, and a learned graph policy—mixed on the last: beaten by a strong non-neural policy on growth rate under the same certificate and allocation rule, but roughly tied on expected detection time.

2 SETUP

At round t we observe $Z_t = (Z_{1,t}, \dots, Z_{n,t}) \in [-1,1]^n$. Let $\mathcal{F}_t$ contain all observations through round t together with the algorithm's revealed randomness; all round-t decisions are made before Z_t is seen. For a one-sided global null,

$$H_0 : \quad \mathbb{E}_P[Z_{v,t} \mid \mathcal{F}_{t-1}] \le 0 \qquad \text{for every } v, t. \tag{2}$$

A predictable learner chooses $\lambda_{v,t} \in [0, \lambda_{\max}]$ with $\lambda_{\max} < 1$ and forms $b_{v,t} = 1 + \lambda_{v,t} Z_{v,t}$, so that $b_{v,t} \in [1 - \lambda_{\max}, 1 + \lambda_{\max}]$ and $\mathbb{E}_P[b_{v,t} \mid \mathcal{F}_{t-1}] \le 1$. We write $B := 1 + \lambda_{\max}$ for the upper end of this range; boundedness is not incidental, and Theorem 4 turns on it.

Dependency graphs, stated blockwise. A predictable graph D_t is a *conditional dependency graph* for Z_t if, under every $P \in \mathcal{P}_0$ and conditionally on $\mathcal{F}_{t-1}$, for all disjoint $A, B \subseteq V$ with no edge between them the subvectors $(Z_v)_{v \in A}$ and $(Z_v)_{v \in B}$ are independent. Two clarifications matter and are routinely elided.

Proposition 1 (A Markov graph is not a dependency graph)**.** *Let $X \sim N(0, \Sigma)$ with $\Sigma_{ij} = \rho^{|i-j|}$, $0 < \rho < 1$, whose Markov graph is the chain $1 - 2 - 3$. With $b_i = \exp(\lambda X_i - \lambda^2/2)$ each marginal expectation is 1, yet $\mathbb{E}[b_1 b_3] = \exp(\lambda^2 \rho^2) > 1$.*

Proposition 2 (Pairwise independence is not enough)**.** *Let X_1, X_2 be i.i.d. Rademacher and $X_3 = X_1 X_2$. All three pairs are independent, so the* empty *graph passes every pairwise check, yet with $b_i = 1 + \lambda X_i$,*

$$\mathbb{E}[b_1 b_2 b_3] = 1 + \lambda^3 \mathbb{E}[X_1 X_2 X_3] = 1 + \lambda^3 > 1.$$

Proposition 2 is the operationally dangerous one: pairwise tests are exactly what a practitioner runs to "validate" an estimated graph, and they do not certify what the construction needs. Proofs and Monte Carlo confirmations are in Appendix A.

Three graphs, three jobs. We keep separate the *dependence* graph D (which licenses products), the *representation* graph R (which need only be predictively useful), and any *signal* graph describing how alternatives co-occur. They may coincide but should not be equated by default.

3 GRAPH-CERTIFIED AGGREGATION

Let $\mathcal{I}(D_t)$ be the independent sets of D_t, including $\varnothing$. Before observing Z_t, choose a finite candidate family $\mathcal{C}_t \subseteq \mathcal{I}(D_t)$ and predictable weights $q_t(I) \geq 0$ with $\sum_{I \in \mathcal{C}_t} q_t(I) \leq 1$. Set

$$B_{I,t} = \prod_{v \in I} b_{v,t}, \qquad B_{\varnothing,t} = 1, \qquad R_t = \sum_{I \in \mathcal{C}_t} q_t(I) B_{I,t}, \qquad M_0 = 1, \quad M_t = M_{t-1} R_t.$$

Theorem 3 (Anytime validity)**.** *If D_t is a conditional dependency graph for Z_t under every $P \in \mathcal{P}_0$, and $D_t, \mathcal{C}_t, q_t, \lambda_{\cdot,t}$ are all $\mathcal{F}_{t-1}$-measurable, then (M_t) is a non-negative test supermartingale and* (1) *holds.*

The proof combines two standard facts—a product of conditionally independent e-values is an e-value, and a predictable convex mixture of valid wealth processes is valid—applied to the new object of independent-set products; everything downstream relies on it, though the combination itself is routine. The learner may be arbitrarily inaccurate without affecting Theorem 3; what it may not be is *unpredictable*. This robustness concerns prediction quality only—not an incorrect graph, on which see §10. Experiments use a fixed, non-adaptive D, so its predictability is never separately exercised.

3.1 WHY INDEPENDENT SETS: THE BOUNDED REGIME

A natural justification restricts attention to non-negative multilinear forms $F(e) = \sum_{A \subseteq V} c_A \prod_{v \in A} e_v$ and asks which are valid for *every* vector of non-negative marginal e-variables with dependency graph D. The answer is that $\sum_A c_A \leq 1$ and every supported A must be independent (Proposition 11, Appendix A); the converse direction is proved by inflating two adjacent e-variables to K with probability $1/K$ and letting $K \to \infty$.

That argument requires unbounded e-variables, and §2 fixes $b_{v,t} \in [0, B]$ with $B < 2$. In the bounded regime the characterisation is simply false: for any set A and *arbitrary* dependence, $\mathbb{E}[\prod_{v \in A} b_v] \leq B^{|A|-1}$, so $B^{-(|A|-1)} \prod_{v \in A} b_v$ is a valid e-value across a clique with no graph assumption at all. The restriction to independent sets therefore needs a different argument, and a stronger one is available.

Theorem 4 (Bounded-regime domination)**.** *Let $b_v \in [0, B]$ with $\mathbb{E}[b_v] \leq 1$, under arbitrary dependence. Then for every $A \subseteq V$ with $|A| = k \geq 1$,*

$$B^{-(k-1)} \prod_{v \in A} b_v \;\leq\; \frac{1}{k} \sum_{v \in A} b_v \qquad \textit{pointwise.} \tag{3}$$

Consequently $\mathbb{E}[\prod_{v \in A} b_v] \leq B^{k-1}$, and every non-negative multilinear form that is valid for all such b is pointwise dominated by a sub-probability mixture of independent-set products—indeed of singletons.

Proof sketch. $\prod_{v \in A} b_v \leq B^{k-1} \min_{v \in A} b_v \leq B^{k-1} \frac{1}{k} \sum_{v \in A} b_v$ gives (3), and taking expectations gives the moment bound. For the second claim, couple all b_v to equal B with probability $1/B$ and 0 otherwise; validity then forces $c_\varnothing + \sum_{A \neq \varnothing} c_A B^{|A|-1} \leq 1$, and applying (3) termwise expresses F as a sub-probability mixture over $\varnothing$ and singletons, all of which are independent sets. Full proof in Appendix A. □

Remark 5. Theorem 4 governs what is achievable under *no* independence assumption, so it does not undercut the paper's own construction: a genuine certificate on D licenses undiscounted independence

within I, which the theorem's hypothesis excludes by design. Its role is to explain why D is worth having—escaping this domination is what a correct certificate buys, and Theorem 7 quantifies how much.

Corollary 6 (Per-round accounting)**.** *Adding a vertex to a non-independent product contributes at most* $\log B$ *to the expected log-return and costs exactly* $\log B$ *in discount. Non-independent products therefore never pay asymptotically—which is the operative reason to restrict to* $\mathcal{I}(D)$*, and it holds in the regime the algorithm uses.*

3.2 The chromatic benchmark is dominated increment by increment

For an r-colouring with classes $C_1, \dots, C_r$, conditional Hölder gives $\mathbb{E}[(\prod_v b_{v,t})^{1/r} \mid \mathcal{F}_{t-1}] \le \prod_j \mathbb{E}[\prod_{v \in C_j} b_{v,t} \mid \mathcal{F}_{t-1}]^{1/r} \le 1$, so the geometric-mean process is valid. Since the arithmetic mean of the colour products is pointwise at least their geometric mean, choosing q_t uniform on the colour classes dominates that benchmark at every increment. The same argument runs for a fractional cover $\{w_I\}$ with $W = \sum_I w_I$: the valid object is the weighted geometric mean $\prod_I (\prod_{v \in I} b_v)^{w_I/W}$, which weighted AM–GM dominates. We state it that way deliberately: identifying that object with $(\prod_v b_v)^{1/\chi_f}$ requires $b_v \ge 1$ pointwise, which betting factors violate by design.

3.3 A balanced certified test

The certified product is the right instrument when many streams deviate; when a single stream deviates, the cumulative mean $\bar{M}_T = \frac{1}{n}\sum_v \prod_t b_{v,t}$ is better, because it averages *cumulative* wealths rather than per-round returns. A convex mixture of valid wealth processes is valid, so for $\pi \in (0,1)$ the process

$$M_T^{\mathrm{bal}} = \pi M_T + (1-\pi)\bar{M}_T$$

is valid, and on every path

$$\log M_T^{\mathrm{bal}} \;\ge\; \max\{\log M_T,\ \log \bar{M}_T\} - \log\big(1/\min(\pi, 1-\pi)\big). \tag{4}$$

This inherits the $\alpha(D)$ multiplier in the dense regime and the union-bound behaviour in the sparse regime, at an additive $\log 2$ at $\pi = 1/2$. It is the certified analogue of the balanced test that Bharti et al. (2026) construct for *independent* streams.

4 What the graph buys: $\alpha(D)$, not $n/\chi_f(D)$

Fix a conditionally i.i.d. alternative Q and let $g_v = \mathbb{E}_Q[\log b_v]$ denote per-node log-growth.

Theorem 7 (Achievable multiplier)**.** *The best fixed certified product has log-growth* $I_{\mathrm{prod}}(D, g) = \max_{I \in \mathcal{I}(D)} \sum_{v \in I} g_v$*, a maximum-weight independent set value; under homogeneous* $g_v \equiv g > 0$ *it equals* $\alpha(D)\, g$*. Moreover* $n/\chi_f(D) \le \alpha(D)$*, with equality when* D *is vertex-transitive, and the inequality is strict in general.*

The identity $\chi_f = n/\alpha$ for vertex-transitive graphs (Lovász, 1979) explains why the disjoint-clique example that motivates this literature cannot distinguish the two constants: disjoint cliques are vertex-transitive. Vertex-transitivity is sufficient for equality but not necessary—even paths and some non-vertex-transitive grids also satisfy $n/\chi_f(D) = \alpha(D)$ exactly, as Table 2 itself shows (rows "path (n=8)" and "grid 2x5")—so the two constants can agree well outside that class. What matters for the sequential problem is not vertex-transitivity as such but whether $\alpha(D)$ itself exceeds $n/\chi_f(D)$, and in our sweep over 12 graph families 3 do show a strict gap, with the measured growth multiplier tracking $\alpha(D)$ throughout: on the star with $n = 10$, $\alpha(D) = 9$ and $n/\chi_f(D) = 5.00$ while the measured multiplier is 8.89—an understatement by a factor 1.80.

A caveat on what this comparison establishes. Janson's χ_f governs concentration of a fixed statistic built from all n coordinates; it is not a claim about achievable sequential growth, and no result we know of asserts $n/\chi_f(D)$ as that rate. What our construction dominates increment-by-increment is instead the *chromatic benchmark* of §3, built from χ_f via fractional-cover Hölder precisely because χ_f is the concentration-relevant constant. Theorem 7's point is that $\alpha(D)$—strictly larger whenever D is not vertex-transitive—is the right target for *achievability*, while $\chi_f(D)$ remains right for *concentration*: different questions about the same graph, not competing answers to one.

5 WHAT CERTIFICATION COSTS

A dependency graph is a $0/1$ notion: it records whether variables are independent, not how strongly they depend. A certified procedure must therefore discard all but one observation inside a clique *however weak* the dependence is. We quantify this.

The corrupted-clique family. Take m disjoint cliques of size r. Each clique j has a latent $Y_j \sim \mathrm{Bern}(p)$ under H_0 and $\mathrm{Bern}(q)$ under Q, and member i observes $X_{j,i} = Y_j \oplus \xi_{j,i}$ with $\xi_{j,i} \sim \mathrm{Bern}(\theta)$ independent. At $\theta = 0$ the members are identical copies; as $\theta \to 1/2$ their dependence vanishes. The clique graph is a genuine dependency graph for every θ. Write $\tilde{\pi} = \pi(1-\theta)+(1-\pi)\theta$.

Theorem 8 (Price of certification)**.** *Per clique, the best certified product grows at* $\mathrm{KL}(\mathrm{Bern}(\tilde{q}) \,\|\, \mathrm{Bern}(\tilde{p}))$ *and the oracle joint likelihood ratio at* $\mathrm{KL}(Q_S \| P_S)$*, where* $S = \sum_i X_{j,i}$*. Their ratio* $\Pi(\theta, r)$ *satisfies*

$$\Pi(0, r) = 1, \qquad \lim_{\theta \uparrow 1/2} \Pi(\theta, r) = r.$$

The limit follows from a local expansion at $\theta = 1/2 - \epsilon$: the certified rate is $8\epsilon^2(q-p)^2 + o(\epsilon^2)$ while the joint rate is $8r\epsilon^2(q-p)^2 + o(\epsilon^2)$, the factor r coming from the Binomial Fisher information $4r$ at $\pi = 1/2$. Numerically Π at $\theta = 0.4999$ is r=2: 2.0000,r=3: 3.0000,r=4: 4.0000,r=6: 6.0000,r=8: 8.0000. Figure 3 shows the closed form against simulated e-processes (maximum discrepancy 0.077 over the sweep) and confirms that the ratio of expected rejection times tracks the same curve.

The interpretation matters for how this line of work should be framed. Under exact duplication the certified product recovers the full information of the genuinely independent observations, which is the reassuring case usually displayed. Away from it, certification is strictly lossy, and the loss grows to the full clique size exactly where dependence is weakest. The following corollary is an accounting identity restating Theorems 7 and 8 side by side on this family, not a new derivation; we state it because juxtaposing the two constants is what makes the trade-off legible.

Corollary 9 (The two-sided ledger)**.** *On the corrupted-clique family with* $n = mr$*, the best certified product grows at* $\alpha(D)$ *times the rate of cumulative node averaging, with* $\alpha(D) = m = n/r$*; and at* $1/\Pi(\theta, r)$ *times the rate of the oracle joint likelihood ratio, with* $\Pi(0, r) = 1$ *and* $\Pi(\theta, r) \to r$ *as* $\theta \uparrow 1/2$*. A dependency-graph certificate therefore buys a factor* $\alpha(D)$ *over dependence-agnostic aggregation and costs up to a factor* r *against a fully specified joint model.*

The certificate is worth having when the alternative is dependence-agnostic averaging and a joint model is unavailable or untrustworthy—which is the auditing setting—and is worth abandoning when the joint law is known well enough to write down. Because $\alpha(D)$ grows with the graph and r is a local clique size, the trade is favourable on sparse graphs and unfavourable on graphs built from a few large cliques.

6 COMPUTATION

Selecting the certified product is a maximum-weight independent set (MWIS) problem in the weights g_v.

Theorem 10 (Approximation transfers linearly)**.** *Suppose a solver returns* $\hat{I} \in \mathcal{I}(D)$ *with* $\sum_{v \in \hat{I}} g_v \geq \rho \max_{I \in \mathcal{I}(D)} \sum_{v \in I} g_v$ *for some* $\rho \in (0, 1]$*. Then the resulting log-growth is at least* $\rho\, G^\star$ *and, with bounded increments,* $\mathbb{E}_Q[\hat{\tau}] \leq \big(\log(1/\alpha) + c\big)/(\rho G^\star)$*, so* $\limsup_{\alpha \to 0} \mathbb{E}_Q[\hat{\tau}]/\mathbb{E}_Q[\tau^\star] \leq 1/\rho$*.*

This converts graph-algorithmic facts into statistical ones. MWIS is solvable exactly in polynomial time on perfect graphs—hence on bipartite, chordal, interval and comparability graphs—via the theta body (Grötschel et al., 1988), giving $\rho = 1$ in principle; bipartite and chordal graphs admit simpler combinatorial algorithms still (max-flow; elimination-ordering DP). Our released solver implements neither, only branch-and-bound with an exact cutoff at $n = 26$ (validated in Appendix A.11), so realising $\rho = 1$ on a large such graph needs a dedicated implementation. On graphs of maximum degree Δ the greedy rule that repeatedly selects the vertex maximising $g_v/(1 + d_v)$ attains weight at least $\sum_v g_v/(1 + d_v)$—the weighted Caro–Wei bound—and hence $\rho \geq 1/(\Delta + 1)$; we verify both inequalities numerically over 4000 random graphs in Appendix A.11. In general MWIS is

inapproximable within $n^{1-\epsilon}$ unless P = NP (Håstad, 1999), so a graph-specific guarantee must restrict the class; Theorem 10 says exactly what that costs. Note the statement is first order in $\log(1/\alpha)$: at $\alpha = 0.05$ the overshoot term is not negligible, and §8 tests it at $\alpha = 10^{-9}$. Every experiment in this paper uses $n \leq 24$; Appendix A.13 measures the released solver directly out to $n = 100$ and reports where it becomes impractical, rather than leaving the computational story as theory alone.

7 Learning, and the discipline it requires

Allocation is not the open problem. Given a predictable candidate family of size K, choosing q_t is online portfolio selection over K assets with returns $B_{I,t}$. Cover's universal portfolio (Cover, 1991) attains log-wealth within $O(K \log T)$ of the best constant-rebalanced portfolio—the richer mixture comparator $\sup_q \mathbb{E}_Q[\log \sum_I q(I) B_I]$, not merely the best single set—and is implementable as a Dirichlet$(1/2)$ sample of such portfolios, itself a mixture of valid wealth processes. Of the three terms in a natural error decomposition (prediction error, allocation regret, combinatorial approximation) only the third is graph-specific, which is why §6 is where we place the theory.

Asymptotic optimality is not the same as being the right choice at a monitoring horizon, and here the gap bites. The universal portfolio's regret term is $O(K \log T)$, which at $K \approx 20$ and $T \approx 200$ is tens of nats—larger than the wealth the process accumulates—so exponentiated gradient (Kivinen & Warmuth, 1997) outperforms it by a factor of 1.18–1.19 in growth rate in our setting (Table 10). We therefore use exponentiated gradient throughout and hold the allocation rule *fixed* across arms when comparing predictors: an earlier version of this comparison let learned and non-learned arms use different allocation rules, and the differences were attributable almost entirely to allocation rather than to the learner.

A graph policy. Our learned arm encodes per-node summaries of the past—running mean, dispersion, last score, positive fraction, normalised degree, time—passes messages over the representation graph R, and emits a betting fraction and a score used to generate candidate independent sets by a greedy pass. Feasibility is always checked against D and never depends on the network's accuracy, so Theorem 3 holds whatever the policy outputs. Using R this way is in the spirit of universal portfolios with side information (Cover & Ordentlich, 1996), though our side information is a graph passed through a GNN rather than a fixed feature map, and—unlike that setting—R carries no validity obligation at all, since only D does.

Predictability discipline. The natural objective $\max_\theta \sum_t \log[\sum_I q_{\theta,t}(I) \prod_{v \in I}(1 + \lambda_{\theta,v,t} Z_{v,t})]$ is retrospective. If θ is fitted on rounds $1..T$ and the wealth is then recomputed on those same rounds, M_t is *not* a supermartingale. This is the most likely way to publish an invalid method in this setting, and the non-neural arms cannot exhibit it, so they cannot detect it. We therefore fit θ_t on rounds $1..t-1$ only, and we measure the consequence of not doing so (§8).

8 Experiments

All scores are generated by independent node and edge shocks, each node's score being a bounded function of its incident shocks; non-adjacent vertex sets then use disjoint shocks, so the declared graph is a genuine conditional dependency graph and the null holds exactly. (A sparse Gaussian precision matrix would supply a Markov graph, which Proposition 1 shows is different.) Code, seeds and result files accompany the submission.

Validity comes first. Figure 1 reports the probability of ever crossing $1/\alpha$ under the null over $10,000$ trajectories per graph at horizon 300 and $\alpha = 0.05$ (Monte Carlo standard error 0.0022). Every valid method stays below α, ours peaking at 0.0064; the independence product, a deliberately invalid diagnostic, reaches 0.080–0.132 on graphs with edges and is valid, as it must be, on the empty graph. Finite simulation cannot prove infinite-horizon validity, and because a supermartingale with slack crosses far less often than α, rates well below the nominal level are expected rather than evidence of a bug.

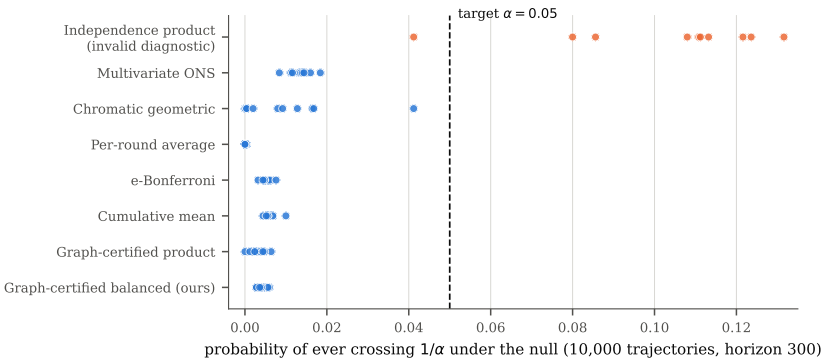


Figure 1: Null crossing probabilities, one point per graph ($10,000$ trajectories each). All valid methods fall left of the $\alpha = 0.05$ line; the independence product does not, except on the empty graph where it is valid.

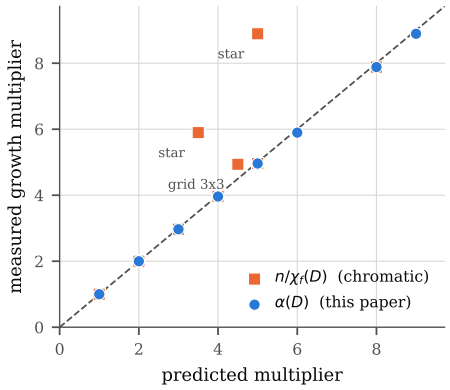


Figure 2: Measured growth multiplier against the two candidate predictions. Points for $\alpha(D)$ lie on the identity throughout; points for $n/\chi_f(D)$ coincide with it on some graphs (including the non-vertex-transitive path) and fall above it on others (e.g. the star), understating what is achievable whenever the two constants differ.

The theory, measured. Three checks, all tight. The measured growth multiplier lies on the identity against $\alpha(D)$ in every graph in the sweep; $n/\chi_f(D)$ agrees where the two constants coincide (e.g. vertex-transitive graphs, but also non-vertex-transitive ones such as even paths) and understates the multiplier elsewhere (Figure 2); the disjoint-clique rate reproduces at 7.80 against a predicted 8 for $m = 8, r = 4$ (Table 3). The price curve $\Pi(\theta, r)$ rises from 1.000 to 3.984 in closed form and 4.061 as simulated at $\theta = 0.48$, against $r = 4$ (Figure 3). Both halves of Theorem 10 hold on Erdős–Rényi and preferential-attachment graphs: the realised growth-rate ratio matches ρ to within 0.002 (smallest ρ observed 0.800), and at $\alpha = 10^{-9}$ the ratio of expected rejection times matches the reciprocal growth-rate ratio to within 0.0055 (Figure 4).

Power, with non-detections reported. Every aggregation rule receives the same local betting rule, so differences are attributable to aggregation. We report rejection probability by horizon and the censored mean stopping time $\mathbb{E}[\min(\tau, T)]$; reporting delay only among successful detections omits failures and can reverse rankings. Figure 5 shows the predicted sparse/dense split cleanly. Against the cumulative mean—the dependence-agnostic rule the certified product is meant to improve on—the product gains $+0.141$ in power on dense alternatives and loses -0.306 on single-node alternatives, and the balanced test of §3.3 is within 0.05 of the better of the two in 117 of 120 cells, as (4) requires.

We do not, however, claim dominance. Taking the maximum over all five baselines in each cell, the balanced test trails the best of them by 0.015–0.046 in power on average across regimes: in the dense regime multivariate ONS is the strongest single method, and in the sparse regime the cumulative mean is. What the certified construction delivers is the growth multiplier the theory predicts and robustness across regimes—it is strictly the worst method in 0 of 120 cells—not a uniformly best test. We think this is the right claim to make: the contribution here is the accounting in Corollary 9, and

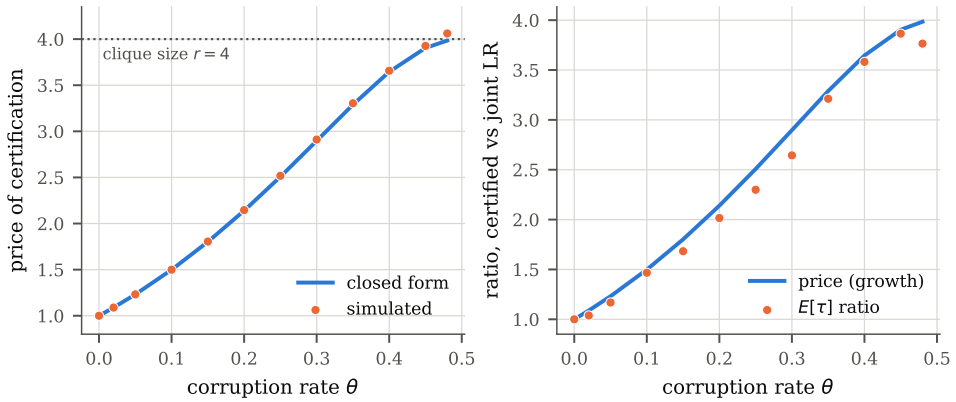


Figure 3: The price of certification on the corrupted-clique family ($r = 4$). Left: closed form (Theorem 8) against simulated e-processes. Right: the same curve against the measured ratio of expected rejection times. Certification is free only at exact duplication ($\theta = 0$) and costs the full clique size as dependence vanishes.

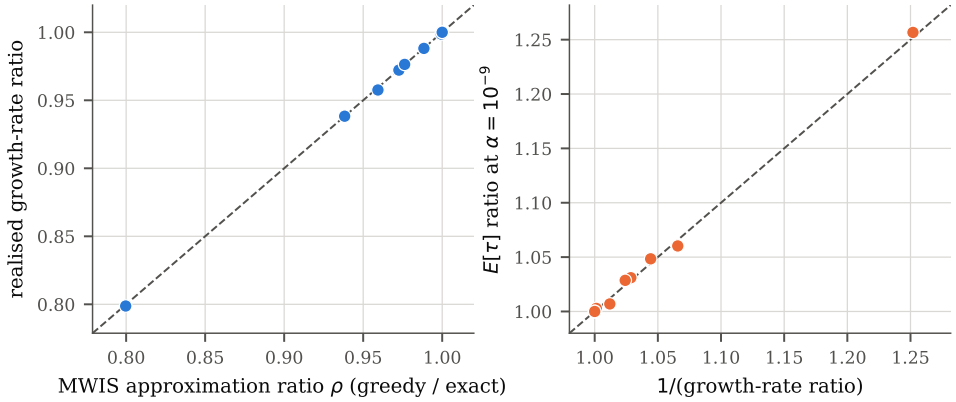


Figure 4: Approximation transfer. Left: realised growth-rate ratio against the MWIS approximation ratio ρ. Right: ratio of expected rejection times at $\alpha = 10^{-9}$ against the reciprocal growth-rate ratio. Dashed lines are the identity.

multivariate betting is a genuinely different and strong competitor that a dependency-graph certificate does not subsume.

Does the graph policy help? Figure 6 compares static, adaptive non-neural, online-fitted neural and transferred neural arms against an oracle, all sharing the certificate, betting family, and—critically—the allocation rule, isolating how bets and candidate sets are chosen. On *growth rate* the result is a clear negative: adaptive attains the highest rate in $8 - 1$ of 8 cells (up to 60% above the graph policy on the sparsest targets), and every arm reaches only 54–77% of the oracle rate. Expected stopping time tells a less one-sided story (power saturates at 1.0 throughout; Appendix Table 7): `gnn-xfer` is fastest to detect in 4 of 8 cells, tied with adaptive (4). A random representation graph does not degrade the neural arm either. The certificate carries the growth-rate gain; on stopping time the transferred policy is competitive—the more encouraging reading, and worth a follow-up. The claim is scoped to one architecture at one budget on solvable-exactly graphs, placing the burden of proof on the learned component rather than proving no architecture could help.

Predictability. Table 8 contrasts one-step-ahead with retrospective fitting (all arms on one certified set, over 800 null trajectories): the exact retrospective maximiser raises the crossing probability to 0.150–0.519 ($10\times$ nominal), while predictable arms stay at ≤ 0.0362 (ONS) and ≤ 0.0000 (neural). Since a cold-started neural policy correctly learns to bet near zero under the null, that alone is not a stress test; a second arm, pretrained on separate source graphs and frozen before seeing the null

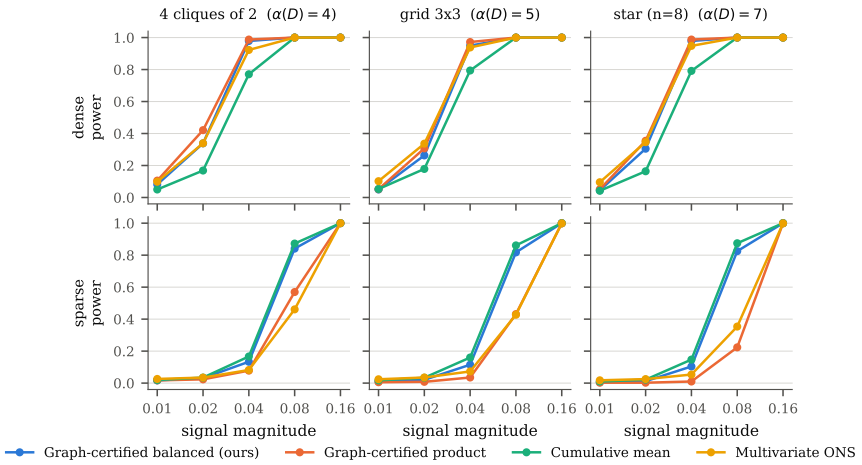


Figure 5: Power against signal magnitude (log scale), 2000 trajectories per cell, horizon 400, all methods sharing the same local betting rule. Top: a dense alternative, where the certified product leads and the cumulative mean trails. Bottom: a single-node alternative, where the ordering reverses. The balanced test of §3.3 tracks whichever is better in both rows, which is what (4) guarantees. Full results including non-detection rates and censored stopping times are in Table 14.

trajectory, bets 23–49× more (Table 9) yet still stays at ≤ 0.0125. The retrospective failure is a property of the objective, not of any optimiser, and only a learned arm can exhibit it.

9 Related work

E-values, test supermartingales, and Ville's inequality are the standard apparatus for anytime-valid inference (Ville, 1939; Vovk & Wang, 2021; Ramdas et al., 2023; Grünwald et al., 2024); admissible procedures rest on non-negative martingales (Ramdas et al., 2020), making ours the natural class a priori. Merging arbitrary e-values admits essentially only weighted averages (Wang, 2025; Vovk & Wang, 2020); we add a certified partial independence structure to that baseline. Products, averages, and calibration of e-values are due to Vovk & Wang (2021); false discovery control with e-values to Wang & Ramdas (2022); conformal prediction under non-exchangeable data (Barber et al., 2023) is the closest non-e-value treatment, targeting coverage rather than a sequential test. Betting-based tests with flexible predictors appear in Pandeva et al. (2024); Shaer et al. (2023); parameter-free coin-betting (Orabona & Pál, 2016) and regret-to-log-optimality arguments (Waudby-Smith et al., 2025; Waudby-Smith & Ramdas, 2024) underlie this machinery, which our learned arm inherits rather than extends. For the multi-stream global null, Bharti et al. (2026) give the sparse/dense trade-off between product, average and balanced merging—under *independence* across streams, which is precisely the assumption we relax; §3.3 is the certified analogue. Validity across filtrations, change-detection, and network applications of e-values are treated by Choe & Ramdas (2024); Dandapanthula & Ramdas (2025); Yanchenko et al. (2025). On the graph side, dependency graphs originate in normal-approximation error bounds (Baldi & Rinott, 1989); chromatic penalties go back to Janson (2004), whose χ_f re-weights sample size across coordinates; Theorem 7 shows the achievable growth rate is instead governed by $\alpha(D)$ (§4), the graph-dependent setting more broadly surveyed by Zhang & Amini (2022); Sun & Cai (2009).

10 Discussion and limitations

Discussion. $\mathcal{P}_0$ combines the mean condition (2) with D_t being a conditional dependency graph, so a rejection can reflect dependence misspecification, not stream misbehaviour: adding edges preserves validity and costs power, omitting one can break it, and estimating D on independent data does not establish correctness (diagnostic: a dependence-agnostic average alongside signals the certificate at work). Better designs know D by construction—clustered randomisation (Aronow & Samii, 2017) or shared templates—not by estimation. Separately, Theorem 4 gives an exponentially-conservative but certificate-free total-variation fallback (Appendix A.12); we treat only the global null.

REPRODUCIBILITY STATEMENT

Every claim in this paper is either proved in Appendix A or produced by code released with the submission. The theorems state their assumptions in full: Theorem 3 requires the blockwise conditional dependency-graph property of §2 together with predictability of $\mathcal{C}_t$, q_t and $\lambda_{\cdot,t}$; Theorem 4 requires only boundedness; Theorem 8 is stated for the corrupted-clique family defined in §5. Data generation is fully specified: the node/edge-shock construction and its certification argument are in Appendix A.10, and the corrupted-clique family has closed-form rates. Each experiment reports its number of trajectories, horizon, level, and the interval used. Every number quoted in the text is emitted by the analysis scripts into a macro file rather than transcribed by hand, so the text cannot drift from the results. A test suite checks the exact graph algorithms against brute-force enumeration, the supermartingale property of every wealth process under the null, the pointwise inequality of Theorem 4, and the limits of Theorem 8. The greedy-approximation numbers quoted in Appendix A.11 and the computational-scaling numbers in Appendix A.13 are each produced by a dedicated, released script (`exp08_greedy_validation.py`, `exp09_scaling.py`) rather than computed once and transcribed.

ETHICS STATEMENT

This work is methodological and uses only synthetic data generated by the processes described in §8; no human subjects, personal data, or released datasets are involved. The intended application—sequential auditing of deployed models—is a monitoring technology, and we note two attendant risks. First, a test that is valid only under a dependence certificate can be made to appear conclusive by asserting a certificate that does not hold; §10 explains why a rejection is evidence against a conjunction of the mean and structural conditions, and recommends running a dependence-agnostic procedure alongside as a diagnostic. Second, the anytime-valid guarantee concerns Type I error only, and a monitoring system that never rejects is not thereby trustworthy; we report non-detection rates throughout for this reason. We declare no conflicts of interest.

AI USE STATEMENT

Generative AI tools were used in this work to support writing and to implement some functions: drafting and editing prose, and assisting with parts of the code implementation. All AI-assisted output was reviewed, verified and edited by the authors, who take full responsibility for the final content of this work, including all text, claims and artifacts.

REFERENCES

Peter M. Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, 11(4):1912–1947, 2017.

Pierre Baldi and Yosef Rinott. On normal approximations of distributions in terms of dependency graphs. *The Annals of Probability*, 17(4):1646–1650, 1989.

Rina Foygel Barber, Emmanuel J. Candès, Aaditya Ramdas, and Ryan J. Tibshirani. Conformal prediction beyond exchangeability. *The Annals of Statistics*, 51(2):816–845, 2023.

Beepul Bharti, Ambar Pal, and Jeremias Sulam. Global sequential testing for multi-stream auditing. *arXiv preprint arXiv:2602.21479*, 2026.

Yo Joong Choe and Aaditya Ramdas. Combining evidence across filtrations. *Journal of the Royal Statistical Society: Series B*, 2024. arXiv:2402.09698.

Thomas M. Cover. Universal portfolios. *Mathematical Finance*, 1(1):1–29, 1991.

Thomas M. Cover and Erik Ordentlich. Universal portfolios with side information. *IEEE Transactions on Information Theory*, 42(2):348–363, 1996.

Sanjit Dandapanthula and Aaditya Ramdas. Multiple testing in multi-stream sequential change detection. *arXiv preprint arXiv:2501.04130*, 2025.

Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer, 1988.

Peter Grünwald, Rianne de Heide, and Wouter M. Koolen. Safe testing. *Journal of the Royal Statistical Society: Series B*, 86(5), 2024.

Johan Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Mathematica*, 182(1), 1999.

Svante Janson. Large deviations for sums of partly dependent random variables. *Random Structures & Algorithms*, 24(3):234–248, 2004. doi: 10.1002/rsa.20008.

Jyrki Kivinen and Manfred K. Warmuth. Exponentiated gradient versus gradient descent for linear predictors. *Information and Computation*, 132(1):1–63, 1997.

László Lovász. On the Shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25 (1):1–7, 1979.

Francesco Orabona and Dávid Pál. Coin betting and parameter-free online learning. *NeurIPS*, 2016.

Teodora Pandeva, Patrick Forré, Aaditya Ramdas, and Shubhanshu Shekhar. Deep anytime-valid hypothesis testing. In *AISTATS*, 2024.

Aaditya Ramdas, Johannes Ruf, Martin Larsson, and Wouter M. Koolen. Admissible anytime-valid sequential inference must rely on nonnegative martingales. *arXiv preprint arXiv:2009.03167*, 2020.

Aaditya Ramdas, Peter Grünwald, Vladimir Vovk, and Glenn Shafer. Game-theoretic statistics and safe anytime-valid inference. *Statistical Science*, 38(4), 2023.

Shalev Shaer, Gal Maman, and Yaniv Romano. Model-X sequential testing for conditional independence via testing by betting. In *AISTATS*, 2023.

Wenguang Sun and T. Tony Cai. Large-scale multiple testing under dependence. *Journal of the Royal Statistical Society: Series B*, 71(2):393–424, 2009.

Jean Ville. étude critique de la notion de collectif. *Gauthier-Villars, Paris*, 1939.

Vladimir Vovk and Ruodu Wang. Combining p-values via averaging. *Biometrika*, 107(4):791–808, 2020.

Vladimir Vovk and Ruodu Wang. E-values: Calibration, combination and applications. *The Annals of Statistics*, 49(3):1736–1754, 2021.

Ruodu Wang. The only admissible way of merging arbitrary e-values. *Biometrika*, 112(2), 2025. arXiv:2409.19888.

Ruodu Wang and Aaditya Ramdas. False discovery rate control with e-values. *Journal of the Royal Statistical Society: Series B*, 84(3):822–852, 2022.

Ian Waudby-Smith and Aaditya Ramdas. Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society: Series B*, 86(1), 2024.

Ian Waudby-Smith, Ricardo Sandoval, and Michael I. Jordan. Universal log-optimality for general classes of e-processes and sequential hypothesis tests. *arXiv preprint arXiv:2504.02818*, 2025.

Eric Yanchenko, Jonathan P. Williams, and Ryan Martin. Hypothesis testing for community structure in temporal networks using e-values. *arXiv preprint arXiv:2507.23034*, 2025.

Rui-Ray Zhang and Massih-Reza Amini. Generalization bounds for learning under graph-dependence: A survey. *arXiv preprint arXiv:2203.13534*, 2022.

A PROOFS

A.1 PROOF OF THEOREM 3

Fix t and work conditionally on $\mathcal{F}_{t-1}$. Let $I \in \mathcal{C}_t \subseteq \mathcal{I}(D_t)$ with $I = \{v_1, \dots, v_k\}$. Because I contains no edge of D_t, the sets $\{v_k\}$ and $\{v_1, \dots, v_{k-1}\}$ are disjoint with no edge between them, so the blockwise dependency-graph property gives $Z_{v_k} \perp\!\!\!\perp (Z_{v_1}, \dots, Z_{v_{k-1}})$ conditionally on $\mathcal{F}_{t-1}$. Iterating downwards yields conditional mutual independence of $(Z_v)_{v \in I}$, hence

$$\mathbb{E}_P\Big[\prod_{v \in I} b_{v,t} \;\Big|\; \mathcal{F}_{t-1}\Big] = \prod_{v \in I} \mathbb{E}_P[b_{v,t} \mid \mathcal{F}_{t-1}] \le 1,$$

since each factor has conditional expectation $1 + \lambda_{v,t}\mathbb{E}_P[Z_{v,t} \mid \mathcal{F}_{t-1}] \le 1$ by predictability of $\lambda_{v,t}$, non-negativity of $\lambda_{v,t}$ and (2). The empty set contributes $B_{\varnothing,t} = 1$. As $\mathcal{C}_t$ and q_t are $\mathcal{F}_{t-1}$-measurable and $\sum_I q_t(I) \le 1$,

$$\mathbb{E}_P[R_t \mid \mathcal{F}_{t-1}] = \sum_{I \in \mathcal{C}_t} q_t(I)\, \mathbb{E}_P[B_{I,t} \mid \mathcal{F}_{t-1}] \le \sum_{I \in \mathcal{C}_t} q_t(I) \le 1.$$

Therefore $\mathbb{E}_P[M_t \mid \mathcal{F}_{t-1}] = M_{t-1}\mathbb{E}_P[R_t \mid \mathcal{F}_{t-1}] \le M_{t-1}$, and $M_t \ge 0$ with $M_0 = 1$. Ville's inequality for non-negative supermartingales gives $P(\exists t \ge 1 : M_t \ge 1/\alpha) \le \alpha$ for every $P \in \mathcal{P}_0$, which is (1). □

Remark. Mutual independence is what the proof needs, and it follows from the blockwise statement only. Proposition 2 shows that the pairwise reading is insufficient.

A.2 PROOF OF PROPOSITION 1

$X_1 + X_3 \sim N(0, 2 + 2\rho^2)$, so $\mathbb{E}[\exp(\lambda(X_1 + X_3))] = \exp(\lambda^2(1 + \rho^2))$ and $\mathbb{E}[b_1 b_3] = \exp(\lambda^2(1 + \rho^2) - \lambda^2) = \exp(\lambda^2\rho^2) > 1$, while $\mathbb{E}[b_i] = 1$ for each i because $X_i \sim N(0, 1)$. The Markov graph of Σ is the chain $1 - 2 - 3$ since the precision matrix of an AR(1) covariance is tridiagonal, yet $\mathrm{Cov}(X_1, X_3) = \rho^2 \ne 0$. Monte Carlo with 4×10^6 draws at $\rho = 0.7, \lambda = 0.5$ gives 1.1299 against the exact 1.1303. □

A.3 PROOF OF PROPOSITION 2

X_1, X_2 are i.i.d. Rademacher and $X_3 = X_1X_2$. Each pair is independent: $X_1X_3 = X_1^2X_2 = X_2$, which is independent of X_1, and symmetrically. But $X_1X_2X_3 = X_1^2X_2^2 = 1$, so expanding $\mathbb{E}[\prod_i(1 + \lambda X_i)] = 1 + \lambda\sum_i \mathbb{E}X_i + \lambda^2\sum_{i<j}\mathbb{E}X_iX_j + \lambda^3\mathbb{E}X_1X_2X_3 = 1 + \lambda^3$. Monte Carlo with 2×10^6 draws at $\lambda = 1/2$ gives 1.1258 against the exact 1.125, with all three sample covariances below 6×10^{-4}. □

A.4 THE UNBOUNDED CHARACTERISATION

Proposition 11 (Unbounded regime)**.** *Let $F(e) = \sum_{A \subseteq V} c_A \prod_{v \in A} e_v$ with $c_A \ge 0$. Then $\mathbb{E}[F(E)] \le 1$ for every vector of non-negative e-variables with dependency graph D, without boundedness restrictions, if and only if $\sum_A c_A \le 1$ and every A with $c_A > 0$ is an independent set of D.*

Proof. Sufficiency. Each supported A is independent, so $\mathbb{E}[\prod_{v \in A} E_v] = \prod_{v \in A}\mathbb{E}[E_v] \le 1$ as in Theorem 3, whence $\mathbb{E}[F(E)] \le \sum_A c_A \le 1$.

Necessity of $\sum_A c_A \le 1$. Take $E_v \equiv 1$; constants are independent of everything, so any D is a dependency graph, and $F(E) = \sum_A c_A$.

Necessity of independence. Suppose $c_{A_0} > 0$ and $(u, v) \in A_0$ is an edge. Let $E_u = E_v = K$ on an event of probability $1/K$ and 0 otherwise (the *same* event for both), and $E_w \equiv 1$ for $w \notin \{u, v\}$. Then $\mathbb{E}[E_u] = \mathbb{E}[E_v] = 1$; E_u, E_v are dependent but adjacent, so this law is compatible with D. All terms are non-negative and the A_0 term contributes $c_{A_0}K^{|A_0|} \cdot (1/K) \ge c_{A_0}K$ (using $|A_0| \ge 2$), so $\mathbb{E}[F(E)] \to \infty$. □

The construction is the reason the characterisation does not survive boundedness: an e-variable confined to $[0, B]$ cannot be inflated.

A.5 Proof of Theorem 4

Let $|A| = k \geq 1$ and $b_v \in [0, B]$. Each factor is at most B, so for any $u \in A$, $\prod_{v\in A} b_v \leq B^{k-1} b_u$; taking u to be a minimiser gives $\prod_{v\in A} b_v \leq B^{k-1} \min_{v\in A} b_v$. Since the minimum is at most the mean, $\prod_{v\in A} b_v \leq B^{k-1} \frac{1}{k} \sum_{v\in A} b_v$, which is (3). Taking expectations and using $\mathbb{E}[b_v] \leq 1$ gives $\mathbb{E}[\prod_{v\in A} b_v] \leq B^{k-1}$.

For the domination claim, let $F(b) = \sum_A c_A \prod_{v\in A} b_v$ with $c_A \geq 0$ be valid for all such b. Couple the coordinates: let $b_v = B$ for every v on an event of probability $1/B$ and $b_v = 0$ for every v otherwise. Then $\mathbb{E}[b_v] = 1$, and $\mathbb{E}[\prod_{v\in A} b_v] = B^{|A|}/B = B^{|A|-1}$ simultaneously for all non-empty A, so validity forces

$$c_\varnothing + \sum_{A\neq\varnothing} c_A B^{|A|-1} \leq 1. \tag{5}$$

Applying (3) to each non-empty term,

$$F(b) \;\leq\; c_\varnothing + \sum_{A\neq\varnothing} c_A B^{|A|-1} \frac{1}{|A|} \sum_{v\in A} b_v \;=\; c_\varnothing + \sum_{v\in V} \underbrace{\Big(\sum_{A\ni v} \frac{c_A B^{|A|-1}}{|A|}\Big)}_{=:w_v} b_v,$$

and $c_\varnothing + \sum_v w_v = c_\varnothing + \sum_{A\neq\varnothing} c_A B^{|A|-1} \leq 1$ by (5). The right-hand side is thus a sub-probability mixture over $\varnothing$ and the singletons $\{v\}$, every one of which is an independent set of any graph. □

Proof of Corollary 6. Using the sharp discount $B^{-(k-1)}$, the expected log-return of a k-set is $\sum_{v\in A} \mathbb{E}_Q[\log b_v] - (k-1)\log B$. Since $b_v \leq B$ we have $\mathbb{E}_Q[\log b_v] \leq \log B$, so appending a vertex changes the objective by at most $\log B - \log B = 0$. Hence no non-independent product strictly improves on the best independent one asymptotically. □

A.6 Proof of Theorem 7

Under a conditionally i.i.d. alternative the log-wealth of the fixed product over I is $\sum_{t\leq T} \sum_{v\in I} \log b_{v,t}$, whose almost-sure growth rate is $\sum_{v\in I} g_v$ by the strong law; maximising over $\mathcal{I}(D)$ gives I_{prod}, and with $g_v \equiv g > 0$ the maximiser is a maximum independent set, of size $\alpha(D)$. For the graph inequality, let $\{w_I\}$ be an optimal fractional colouring, so $\sum_{I\ni v} w_I \geq 1$ for all v and $\chi_f = \sum_I w_I$. Summing over v, $n \leq \sum_v \sum_{I\ni v} w_I = \sum_I w_I |I| \leq \alpha(D)\chi_f(D)$, i.e. $n/\chi_f(D) \leq \alpha(D)$. For vertex-transitive D, $\chi_f(D) = n/\alpha(D)$ (Lovász, 1979), giving equality. Strictness in general is witnessed by the star $K_{1,n-1}$, where $\alpha = n-1$ and $\chi_f = 2$, so $n/\chi_f = n/2 < n-1$ for $n \geq 3$. □

A.7 Proof of Theorem 8

Write $\tilde{\pi} = \pi(1-\theta) + (1-\pi)\theta$, so an observed copy is $\mathrm{Bern}(\tilde{p})$ under H_0 and $\mathrm{Bern}(\tilde{q})$ under Q. An independent set of the clique graph contains at most one member of each clique, so the best certified product per clique is that member's marginal likelihood ratio, with growth $\mathrm{KL}(\mathrm{Bern}(\tilde{q})\|\mathrm{Bern}(\tilde{p}))$. The clique's law depends on $(X_{j,1}, \ldots, X_{j,r})$ only through $S = \sum_i X_{j,i}$, which is a two-component mixture of Binomials, $P_S(s) = p\binom{r}{s}(1-\theta)^s\theta^{r-s} + (1-p)\binom{r}{s}\theta^s(1-\theta)^{r-s}$ and likewise Q_S with q in place of p; the oracle joint growth is $\mathrm{KL}(Q_S\|P_S)$.

At $\theta = 0$ every member equals Y_j, so $S \in \{0, r\}$ carries exactly the information in Y_j and both rates equal $\mathrm{KL}(\mathrm{Bern}(q)\|\mathrm{Bern}(p))$, giving $\Pi(0, r) = 1$.

For the limit set $\theta = 1/2 - \epsilon$. Then $\tilde{q} - \tilde{p} = (q-p)(1-2\theta) = 2\epsilon(q-p)$ and $\tilde{p} \to 1/2$, so

$$\mathrm{KL}(\mathrm{Bern}(\tilde{q})\|\mathrm{Bern}(\tilde{p})) = \frac{(\tilde{q}-\tilde{p})^2}{2\tilde{p}(1-\tilde{p})} + o(\epsilon^2) = 8\epsilon^2(q-p)^2 + o(\epsilon^2).$$

For the joint rate, write $f_1 = \mathrm{Bin}(r, 1/2+\epsilon)$ and $f_0 = \mathrm{Bin}(r, 1/2-\epsilon)$, so $Q_S - P_S = (q-p)(f_1 - f_0)$ and $P_S \to f := \mathrm{Bin}(r, 1/2)$. Using $\mathrm{KL}(Q_S \| P_S) = \frac{1}{2}\sum_s (Q_S - P_S)^2/P_S + o(\epsilon^2)$ and $f_1 - f_0 = 2\epsilon\, \partial_\pi f|_{\pi=1/2} + O(\epsilon^3)$ together with the Binomial Fisher information $\sum_s (\partial_\pi f)^2/f = r/(\pi(1-\pi)) = 4r$ at $\pi = 1/2$,

$$\mathrm{KL}(Q_S \| P_S) = \tfrac{1}{2}(q-p)^2 \cdot 4\epsilon^2 \cdot 4r + o(\epsilon^2) = 8r\epsilon^2(q-p)^2 + o(\epsilon^2).$$

Dividing, $\Pi(\theta, r) \to r$. □

Numerically, evaluating the exact expressions at $\theta = 0.4999$ gives r=2: 2.0000,r=3: 3.0000,r=4: 4.0000,r=6: 6.0000,r=8: 8.0000, and the closed form agrees with simulated e-processes to within 0.077 across the sweep in Table 4.

A.8 Proof of Theorem 10

Let $G^\star = \max_I \sum_{v\in I} g_v$ and $\hat{G} = \sum_{v\in \hat{I}} g_v \geq \rho G^\star$. The wealth of the fixed product over $\hat{I}$ has i.i.d. log increments with mean $\hat{G}$, bounded by $|I| \log B$ in absolute value. By Wald's identity applied at the stopping time $\hat{\tau} = \inf\{t : \log M_t \geq a\}$ with $a = \log(1/\alpha)$, and bounded overshoot $c \leq |\hat{I}| \log B$,

$$\frac{a}{\hat{G}} \leq \mathbb{E}_Q[\hat{\tau}] \leq \frac{a+c}{\hat{G}} \leq \frac{a+c}{\rho G^\star},$$

and similarly $\mathbb{E}_Q[\tau^\star] \geq a/G^\star$. Hence $\mathbb{E}_Q[\hat{\tau}]/\mathbb{E}_Q[\tau^\star] \leq (a+c)/(\rho a) \to 1/\rho$ as $\alpha \to 0$. □

The bound is first order in a: at moderate α the overshoot c is not negligible relative to a, and the finite-α ratio can exceed $1/\rho$. Our measurements confirm the sharper statement that the ratio of expected rejection times converges to the reciprocal of the *realised* growth-rate ratio, matching to within 0.0055 at $\alpha = 10^{-9}$ (Table 5).

A.9 The balanced bound (4)

$M_T^{\mathrm{bal}} = \pi M_T + (1-\pi)\bar{M}_T \geq \max\{\pi M_T, (1-\pi)\bar{M}_T\}$, so $\log M_T^{\mathrm{bal}} \geq \max\{\log M_T + \log \pi,\ \log \bar{M}_T + \log(1-\pi)\}$, which is (4). Validity is immediate: a convex combination of non-negative supermartingales with initial value 1 is a non-negative supermartingale with initial value 1. □

A.10 The shock generator certifies its graph

Let u_v be independent node shocks and w_e independent edge shocks, and let Z_v be a bounded measurable function of u_v and $\{w_e : v \in e\}$. Let $A, B \subseteq V$ be disjoint with no edge between them. A shock read by both groups would have to be an edge shock w_e with e incident to a vertex of A and to a vertex of B; since e has two endpoints, this forces $e = \{a, b\}$ with $a \in A, b \in B$, an A–B edge, which is excluded. Hence $(Z_v)_{v\in A}$ and $(Z_v)_{v\in B}$ are functions of disjoint independent collections and are independent, so D is a dependency graph. With Z_v an odd function of its shocks plus a shift δ_v, the null $\delta \equiv 0$ makes each Z_v symmetric and hence exactly mean zero; clipping to $[-1, 1]$ preserves symmetry.

A.11 The greedy surrogate and its approximation ratio

The greedy rule used throughout repeatedly selects the remaining vertex maximising $g_v/(1+d_v)$, where d_v counts remaining neighbours, and deletes its closed neighbourhood. Its output weight is at least the weighted Caro–Wei quantity $\sum_v g_v/(1+d_v)$, where here d_v is the degree in the original graph. Since $\max_I \sum_{v\in I} g_v \leq \sum_v g_v$ and $d_v \leq \Delta$, this yields $\rho \geq 1/(\Delta+1)$.

We checked both statements numerically. Over 4000 Erdős–Rényi graphs with $n \in [5, 30)$ and edge probability drawn uniformly from $[0.05, 0.6]$, the greedy weight fell below the weighted Caro–Wei bound in 0 instances (minimum slack 0, i.e. the bound is attained but never violated). Over 600 graphs with $n \in [5, 20)$ compared against exact branch-and-bound maximum-weight independent sets, the minimum of $\rho \cdot (\Delta+1)$ was 1.0, confirming $\rho \geq 1/(\Delta+1)$ and showing the bound is tight.

Our exact solver is branch and bound with a degree-based branching vertex and a positive-weight upper bound; it was validated against exhaustive enumeration on 300 random weighted graphs with

$n \leq 10$ (zero discrepancies), and the exact independence numbers and fractional chromatic numbers reported in Table 2 were likewise cross-checked against enumeration. All three checks in this section are reproduced exactly by `experiments/exp08_greedy_validation.py`, released with the submission, and are additionally covered by `tests/test_core.py`.

A.12 A ROBUSTNESS CORRECTION UNDER BOUNDED MISSPECIFICATION

Suppose $0 \leq B_{I,t} \leq L_{I,t}$ with $L_{I,t}$ predictable, and a reference conditional law Q_t satisfies $\mathbb{E}_{Q_t}[B_{I,t} \mid \mathcal{F}_{t-1}] \leq 1$. If the true conditional law P_t is within total variation η_t of Q_t, where $\mathrm{TV} = \sup_A |P(A) - Q(A)|$, then $|\mathbb{E}_{P_t}[B_{I,t} \mid \mathcal{F}_{t-1}] - \mathbb{E}_{Q_t}[B_{I,t} \mid \mathcal{F}_{t-1}]| \leq L_{I,t}\eta_t$, so $\mathbb{E}_{P_t}[B_{I,t} \mid \mathcal{F}_{t-1}] \leq 1 + L_{I,t}\eta_t$ and $B_{I,t}/(1 + L_{I,t}\eta_t)$ is valid under that bound. (With the convention $\mathrm{TV} = \int |p - q|$ the bound is $L_{I,t}\eta_t/2$.)

This is exponentially conservative: $L_{I,t} \approx B^{|I|}$, so η_t must be of order $B^{-|I|}$ for the divisor to be non-trivial, which is a stronger requirement than certifying the edges directly. Theorem 4 supplies a fallback. Since $\mathbb{E}[\prod_{v \in I} b_v] \leq B^{|I|-1}$ holds under *arbitrary* dependence, the divisor

$$\min\{1 + L_{I,t}\eta_t,\ B^{|I|-1}\}$$

is valid, and its second branch requires no certificate and no bound on η_t. The corrected process is therefore never vacuous, whatever the misspecification. Note also that an unverified estimate of η_t does not support exact validity; the guarantee is conditional on the stated bound holding.

A.13 COMPUTATIONAL SCALING

§6 argues that MWIS approximation transfers linearly to statistical loss, but reports no measurement of wall-clock cost. We supply one here, using the released branch-and-bound exact solver and greedy surrogate directly, at n well beyond every experiment elsewhere in the paper.

Table 11 shows exact-solver time against n at three edge densities (a 20-second budget was set in advance per density: once a completed call exceeds it, larger n are not attempted at that density — the check cannot abort a call already in progress). At the sparsest density the budget is exhausted before $n = 100$: the solver took 65s at $n = 90$ and is not attempted at $n = 100$. At a denser setting the $n = 100$ call was already in flight when this rule would have skipped it, and ran to completion in 91.8 seconds — itself well over the budget. Either way the exact solver is impractical at this scale. This confirms, rather than merely asserts, that a real deployment beyond a few dozen nodes needs the greedy fallback — or, on the graph classes named in §6, a dedicated polynomial algorithm neither implemented here nor commonly available off the shelf. Greedy itself scales comfortably: Table 12 shows it remains under a second per call up to $n = 10,000$ nodes (11.7s at that size).

Table 13 shows a second, independent scaling concern: the number of independent sets a graph contains, which upper-bounds the size of the fractional-cover component of `static_candidates` (§3). At $n = 307$ the mean count is 307; by $n = 14,720$, the exact cutoff at which `static_candidates` disables the fractional cover, the mean count is already $14,720$ and continues to grow (to $81,077$ by $n = 22$). The candidate family size K itself, however, stays small and roughly stable across the cutoff — because the fractional cover is only one of several components of `static_candidates`, and disabling it past $n = 18$ removes a component whose own size was already growing quickly, not because the underlying combinatorics become any less severe. A deployment at graph sizes materially beyond those tested here should not assume the candidate-family strategy used throughout this paper continues to track $I_{\mathrm{prod}}(D, g)$ well; this is a genuine open question our experiments do not resolve.

All numbers in this section are produced by `experiments/exp09_scaling.py`, released with the submission.

B ADDITIONAL FIGURES AND TABLES

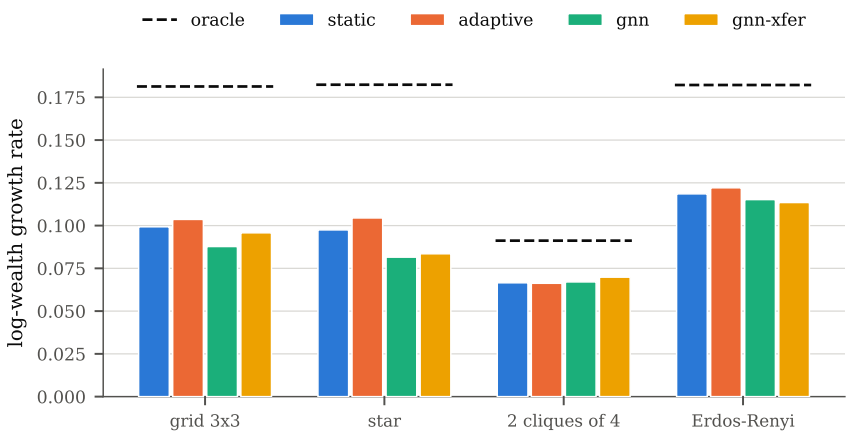


Figure 6: Log-wealth growth rate on target graphs unseen during pretraining. All arms share the dependence certificate, the betting family and the allocation rule, so the comparison isolates how bets and candidate sets are chosen. "static" uses a graph-only candidate family; "adaptive" regenerates candidates from running node-value estimates; "gnn" fits a graph policy online on past rounds; "gnn-xfer" is pretrained on smaller source graphs and frozen. The dashed line is the oracle. The non-neural adaptive policy is the strongest learned arm, and all arms fall well short of the oracle.

Table 1: Null crossing probabilities at $\alpha = 0.05$, $10,000$ trajectories per graph, horizon 300. Monte Carlo standard error near 0.05 is 0.0022. †deliberately invalid diagnostic.

Method	empty (n=8)	complete (n=8)	4 cliques of 2	2 cliques of 4	path (n=8)	cycle (n=8)	grid 3x3	star (n=8)	Erdos-Renyi (n=12, p=.3)	power law (n=12, m=2)
Graph-certified balanced (ours)	0.0052	0.0028	0.0060	0.0044	0.0044	0.0056	0.0028	0.0040	0.0040	0.0036
Graph-certified product	0.0028	0.0000	0.0064	0.0020	0.0024	0.0036	0.0012	0.0024	0.0048	0.0044
Cumulative mean	0.0068	0.0052	0.0064	0.0060	0.0068	0.0100	0.0044	0.0056	0.0056	0.0052
e-Bonferroni	0.0064	0.0052	0.0056	0.0056	0.0060	0.0076	0.0032	0.0048	0.0048	0.0044
Per-round average	0.0000	0.0000	0.0000	0.0004	0.0000	0.0004	0.0004	0.0000	0.0000	0.0000
Chromatic geometric	0.0412	0.0000	0.0080	0.0004	0.0128	0.0164	0.0168	0.0092	0.0004	0.0020
Multivariate ONS	0.0132	0.0184	0.0148	0.0136	0.0140	0.0160	0.0112	0.0084	0.0116	0.0144
Independence product†	0.0412	0.1316	0.0856	0.1216	0.1108	0.1132	0.1080	0.0800	0.1236	0.1112

Table 2: Independence number against the chromatic constant, with the measured growth multiplier under a homogeneous alternative. "gap" marks $n/\chi_f(D) < \alpha(D)$.

Graph D	n	$\alpha(D)$	$\chi_f(D)$	$n/\chi_f(D)$	measured	gap
empty (n=8)	8	8	1.00	8.00	7.88	–
complete (n=8)	8	1	8.00	1.00	1.00	–
path (n=8)	8	4	2.00	4.00	3.95	–
cycle (n=8)	8	4	2.00	4.00	3.96	–
cycle (n=5)	5	2	2.50	2.00	2.00	–
star (n=7)	7	6	2.00	3.50	5.90	yes
star (n=10)	10	9	2.00	5.00	8.89	yes
grid 3x3	9	5	2.00	4.50	4.94	yes
grid 2x5	10	5	2.00	5.00	4.96	–
2 cliques of 4	8	2	4.00	2.00	2.00	–
3 cliques of 3	9	3	3.00	3.00	2.97	–
4 cliques of 2	8	4	2.00	4.00	3.96	–

Table 3: Disjoint-clique log-growth: certified product versus cumulative mean. Predicted ratio $n/r = m$.

m	r	n	certified rate		cumulative mean		ratio	
			pred.	meas.	pred.	meas.	pred.	meas.
8	4	32	1.5363	1.5232	0.1920	0.1954	8	7.80
6	3	18	1.1523	1.1448	0.1920	0.1942	6	5.89
4	5	20	0.7682	0.7626	0.1920	0.1952	4	3.91
12	2	24	2.3045	2.2899	0.1920	0.1974	12	11.60

Table 4: Price of certification $\Pi(\theta, r)$ at $m = 6$, $r = 4$, $p = 0.3$, $q = 0.6$: closed form, simulation, and the corresponding expected rejection times.

θ	certified	joint LR	Π (theory)	Π (meas.)	$E[\tau]$ cert.	$E[\tau]$ joint	τ ratio
0.00	0.19204	0.19204	1.000	1.000	3.7	3.7	1.000
0.02	0.17602	0.19157	1.088	1.089	3.9	3.8	1.039
0.05	0.15353	0.18905	1.231	1.232	4.4	3.8	1.169
0.10	0.11993	0.17971	1.498	1.500	5.8	3.9	1.467
0.15	0.09093	0.16383	1.802	1.805	7.1	4.2	1.683
0.20	0.06625	0.14174	2.139	2.146	9.5	4.7	2.016
0.25	0.04569	0.11457	2.507	2.518	13.0	5.7	2.299
0.30	0.02908	0.08424	2.897	2.911	19.6	7.4	2.644
0.35	0.01629	0.05357	3.289	3.304	34.3	10.7	3.212
0.40	0.00722	0.02631	3.645	3.657	73.6	20.6	3.582
0.45	0.00180	0.00703	3.904	3.926	291.6	75.4	3.865
0.48	0.00029	0.00115	3.984	4.061	1689.6	448.8	3.765

Table 5: Approximation transfer. ρ is the greedy/exact MWIS ratio; $E[\tau]$ ratio is measured at $\alpha = 10^{-9}$.

Graph	$\lvert I^\star \rvert$	$\lvert \hat{I} \rvert$	ρ	growth ratio	$E[\tau]$ ratio	$1/\text{growth}$
ER ($n = 16$, 0.15)	9	9	1.000	1.000	1.000	1.000
ER ($n = 16$, 0.3)	7	7	1.000	1.000	1.000	1.000
ER ($n = 16$, 0.5)	5	5	1.000	1.000	1.000	1.000
ER ($n = 20$, 0.15)	9	8	0.938	0.938	1.060	1.066
ER ($n = 20$, 0.3)	6	6	0.999	0.999	1.003	1.001
ER ($n = 20$, 0.5)	6	4	0.800	0.799	1.257	1.252
ER ($n = 24$, 0.15)	10	11	0.973	0.972	1.031	1.029
ER ($n = 24$, 0.3)	8	7	0.976	0.976	1.029	1.024
ER ($n = 24$, 0.5)	5	5	1.000	1.000	1.000	1.000
PL ($n = 16$, 2)	9	9	0.959	0.958	1.048	1.044
PL ($n = 20$, 2)	12	12	1.000	1.000	1.000	1.000
PL ($n = 24$, 2)	12	12	0.988	0.988	1.007	1.012

Table 6: Log-wealth growth rate by arm on unseen target graphs. All arms share the dependence certificate; "oracle" uses bets tuned to the true alternative on an exact maximum-weight independent set.

Target graph	signal	static	adaptive	gnn	gnn-xfer	oracle
grid 3x3 (unseen)	homophilous	0.0993	0.1036	0.0877	0.0958	0.1813
grid 3x3 (unseen)	scattered	0.0996	0.1498	0.0920	0.0943	0.2754
star (n=8, unseen)	homophilous	0.0975	0.1045	0.0815	0.0835	0.1824
star (n=8, unseen)	scattered	0.0981	0.1048	0.0815	0.0834	0.1826
2 cliques of 4 (unseen)	homophilous	0.0665	0.0663	0.0671	0.0698	0.0912
2 cliques of 4 (unseen)	scattered	0.1012	0.1069	0.1010	0.1058	0.1838
Erdos-Renyi (n=12, unseen)	homophilous	0.1185	0.1221	0.1152	0.1135	0.1822
Erdos-Renyi (n=12, unseen)	scattered	0.1611	0.2283	0.1636	0.1780	0.3654

Table 7: The other axis of the same comparison: expected censored stopping time $E[\min(\tau, T)]$ (lower is better), same cells as Table 6. Power is saturated at 1.0 throughout, so growth rate and stopping time are the two informative summaries here, and they do not always rank the arms the same way: the transferred policy `gnn-xfer` is fastest to detect in 4 of 8 cells, tied with `adaptive` (4 of 8) rather than trailing it as the growth-rate comparison alone would suggest.

Target graph	signal	static	adaptive	gnn	gnn-xfer	fastest
grid 3x3 (unseen)	homophilous	56.8	53.3	65.0	53.3	**adaptive**
grid 3x3 (unseen)	scattered	55.8	50.9	59.8	53.2	**adaptive**
star (n=8, unseen)	homophilous	56.3	54.3	59.1	57.3	**adaptive**
star (n=8, unseen)	scattered	56.3	54.1	59.3	57.5	**adaptive**
2 cliques of 4 (unseen)	homophilous	65.4	67.1	66.3	58.9	**gnn-xfer**
2 cliques of 4 (unseen)	scattered	54.7	55.6	55.3	50.1	**gnn-xfer**
Erdos-Renyi (n=12, unseen)	homophilous	42.8	42.6	42.6	39.4	**gnn-xfer**
Erdos-Renyi (n=12, unseen)	scattered	49.0	43.5	48.4	39.0	**gnn-xfer**

Table 8: Null crossing probability at $\alpha = 0.05$ under predictable and retrospective fitting. "neural, online-fit" starts from random weights on a null trajectory and correctly learns to bet near zero; "neural, pretrained" is instead pretrained to expect a real signal on separate source graphs, then frozen, so it enters the null trajectory betting substantially and its respecting the level is not merely a consequence of not betting (see Table 9). †the retrospective fit is the natural but invalid reading of the training objective in §7.

Graph	ONS	neural, online-fit	neural, pretrained	retrospective†
grid 4x4	0.0275	0.0000	0.0037	0.3100
star (n=12)	0.0288	0.0000	0.0125	0.5188
6 cliques of 2	0.0338	0.0000	0.0013	0.2137
grid 3x3	0.0362	0.0000	0.0013	0.1500

Table 9: Mean log-wealth at the horizon for the same arms and graphs as Table 8, confirming the pretrained arm is actually betting: its magnitude is 23–$49\times$ larger than the online-fit arm's, which stays within Monte Carlo noise of zero.

Graph	ONS	neural, online-fit	neural, pretrained	retrospective†
grid 4x4	-4.912	-0.008	-0.312	+1.968
star (n=12)	-6.810	-0.008	-0.394	+2.796
6 cliques of 2	-3.669	-0.010	-0.242	+1.475
grid 3x3	-3.106	-0.009	-0.199	+1.257

Table 10: Allocation rule at a monitoring horizon ($T = 200$, $K \approx 20$), with candidates and bets held fixed. Cover's universal portfolio is asymptotically stronger but its $O(K \log T)$ regret term dominates at this horizon.

Graph	EG	universal portfolio	ratio
grid 3x3 (unseen)	0.0993	0.0845	1.18
star (n=8, unseen)	0.0975	0.0823	1.18
2 cliques of 4 (unseen)	0.0665	0.0560	1.19

Table 11: Computational scaling: exact branch-and-bound vs. the greedy surrogate, edge probability 0.1. "timeout" means the per-call ceiling of 20s was reached; the solver is not attempted at larger n once this happens.

n	exact (s)	greedy (ms)
10	0.000	0.03
20	0.001	0.06
30	0.005	0.12
40	0.014	0.18
50	0.046	0.31
60	0.179	0.44
70	0.843	0.51
80	5.550	0.68
90	11.949	0.75
100	91.775	0.98

Table 12: Greedy alone at deployment-relevant scale, same edge density scaled to keep mean degree roughly constant.

n	greedy (ms)
10	0.03
50	0.25
100	0.89
300	6.88
1000	93.59
3000	956.89
10000	11693.70

Table 13: Independent-set count $|\mathcal{I}(D)|$ and the resulting `static_candidates` family size K, mean of 5 random graphs per n at edge probability 0.1. The fractional-cover component of K is disabled past $n = 18$ (§3); $|\mathcal{I}(D)|$ is already large at that cutoff and keeps growing.

n	$\lvert\mathcal{I}(D)\rvert$ (mean of 5)	candidate family size K (mean of 5)
10	307	16.6 (fractional cover)
12	1509	–
14	2342	22.2 (fractional cover)
15	4580	–
16	7596	30.8 (fractional cover)
17	12590	–
18	14720	35.2 (fractional cover)
19	17950	–
20	50048	24.6
21	76656	–
22	81077	–
24	–	29.0
30	–	35.0
50	–	55.8

Table 14: Power by graph, signal pattern and magnitude (2000 trajectories per cell, horizon 400, $\alpha = 0.05$). "balanced" and "product" are the two certified variants; "best baseline" is the strongest non-certified method in that cell.

Graph	$\alpha(D)$	pattern	mag.	balanced	product	best baseline	which
4 cliques of 2	4	dense	0.01	0.082	0.105	0.099	Multivariate ONS
4 cliques of 2	4	dense	0.02	0.339	0.421	0.339	Multivariate ONS
4 cliques of 2	4	dense	0.04	0.979	0.988	0.958	Chromatic geometric
4 cliques of 2	4	dense	0.08	1.000	1.000	1.000	Cumulative mean
4 cliques of 2	4	dense	0.16	1.000	1.000	1.000	Cumulative mean
4 cliques of 2	4	sparse	0.01	0.017	0.017	0.026	Multivariate ONS
4 cliques of 2	4	sparse	0.02	0.031	0.024	0.035	Cumulative mean
4 cliques of 2	4	sparse	0.04	0.134	0.078	0.167	Cumulative mean
4 cliques of 2	4	sparse	0.08	0.841	0.570	0.873	Cumulative mean
4 cliques of 2	4	sparse	0.16	1.000	1.000	1.000	Cumulative mean
4 cliques of 2	4	clustered	0.01	0.021	0.021	0.029	Multivariate ONS
4 cliques of 2	4	clustered	0.02	0.053	0.048	0.056	Multivariate ONS
4 cliques of 2	4	clustered	0.04	0.261	0.200	0.275	Cumulative mean
4 cliques of 2	4	clustered	0.08	0.960	0.906	0.965	Cumulative mean
4 cliques of 2	4	clustered	0.16	1.000	1.000	1.000	Cumulative mean
4 cliques of 2	4	scattered	0.01	0.024	0.022	0.032	Multivariate ONS
4 cliques of 2	4	scattered	0.02	0.056	0.051	0.058	Cumulative mean
4 cliques of 2	4	scattered	0.04	0.301	0.259	0.304	Cumulative mean
4 cliques of 2	4	scattered	0.08	0.983	0.971	0.984	Cumulative mean
4 cliques of 2	4	scattered	0.16	1.000	1.000	1.000	Cumulative mean
2 cliques of 4	2	dense	0.01	0.041	0.036	0.094	Multivariate ONS
2 cliques of 4	2	dense	0.02	0.196	0.211	0.327	Multivariate ONS
2 cliques of 4	2	dense	0.04	0.887	0.922	0.911	Multivariate ONS
2 cliques of 4	2	dense	0.08	1.000	1.000	1.000	Cumulative mean
2 cliques of 4	2	dense	0.16	1.000	1.000	1.000	Cumulative mean
2 cliques of 4	2	sparse	0.01	0.010	0.004	0.024	Multivariate ONS
2 cliques of 4	2	sparse	0.02	0.018	0.006	0.034	Cumulative mean
2 cliques of 4	2	sparse	0.04	0.126	0.024	0.170	Cumulative mean
2 cliques of 4	2	sparse	0.08	0.829	0.391	0.874	Cumulative mean
2 cliques of 4	2	sparse	0.16	1.000	1.000	1.000	Cumulative mean
2 cliques of 4	2	clustered	0.01	0.015	0.009	0.040	Multivariate ONS
2 cliques of 4	2	clustered	0.02	0.049	0.026	0.079	Multivariate ONS
2 cliques of 4	2	clustered	0.04	0.365	0.233	0.414	Cumulative mean
2 cliques of 4	2	clustered	0.08	0.991	0.977	0.995	Cumulative mean
2 cliques of 4	2	clustered	0.16	1.000	1.000	1.000	Cumulative mean
2 cliques of 4	2	scattered	0.01	0.010	0.004	0.024	Multivariate ONS
2 cliques of 4	2	scattered	0.02	0.018	0.006	0.034	Cumulative mean
2 cliques of 4	2	scattered	0.04	0.126	0.024	0.170	Cumulative mean
2 cliques of 4	2	scattered	0.08	0.829	0.391	0.874	Cumulative mean
2 cliques of 4	2	scattered	0.16	1.000	1.000	1.000	Cumulative mean
grid 3x3	5	dense	0.01	0.050	0.052	0.102	Multivariate ONS
grid 3x3	5	dense	0.02	0.263	0.307	0.337	Multivariate ONS
grid 3x3	5	dense	0.04	0.950	0.972	0.964	Chromatic geometric
grid 3x3	5	dense	0.08	1.000	1.000	1.000	Cumulative mean
grid 3x3	5	dense	0.16	1.000	1.000	1.000	Cumulative mean
grid 3x3	5	sparse	0.01	0.010	0.005	0.024	Multivariate ONS
grid 3x3	5	sparse	0.02	0.021	0.008	0.035	Multivariate ONS
grid 3x3	5	sparse	0.04	0.115	0.035	0.160	Cumulative mean
grid 3x3	5	sparse	0.08	0.818	0.431	0.861	Cumulative mean
grid 3x3	5	sparse	0.16	1.000	0.999	1.000	Cumulative mean
grid 3x3	5	clustered	0.01	0.016	0.011	0.037	Multivariate ONS
grid 3x3	5	clustered	0.02	0.050	0.032	0.071	Multivariate ONS
grid 3x3	5	clustered	0.04	0.360	0.270	0.397	Cumulative mean
grid 3x3	5	clustered	0.08	0.995	0.988	0.996	Cumulative mean
grid 3x3	5	clustered	0.16	1.000	1.000	1.000	Cumulative mean
grid 3x3	5	scattered	0.01	0.012	0.008	0.032	Multivariate ONS
grid 3x3	5	scattered	0.02	0.040	0.018	0.059	Cumulative mean
grid 3x3	5	scattered	0.04	0.256	0.158	0.297	Cumulative mean
grid 3x3	5	scattered	0.08	0.981	0.953	0.986	Cumulative mean
grid 3x3	5	scattered	0.16	1.000	1.000	1.000	Cumulative mean
star (n=8)	7	dense	0.01	0.045	0.051	0.096	Multivariate ONS
star (n=8)	7	dense	0.02	0.305	0.354	0.346	Multivariate ONS
star (n=8)	7	dense	0.04	0.979	0.988	0.958	Chromatic geometric
star (n=8)	7	dense	0.08	1.000	1.000	1.000	Cumulative mean
star (n=8)	7	dense	0.16	1.000	1.000	1.000	Cumulative mean
star (n=8)	7	sparse	0.01	0.006	0.003	0.018	Multivariate ONS
star (n=8)	7	sparse	0.02	0.012	0.003	0.025	Multivariate ONS
star (n=8)	7	sparse	0.04	0.104	0.010	0.147	Cumulative mean
star (n=8)	7	sparse	0.08	0.825	0.224	0.875	Cumulative mean
star (n=8)	7	sparse	0.16	1.000	1.000	1.000	Cumulative mean
star (n=8)	7	clustered	0.01	0.011	0.007	0.031	Multivariate ONS
star (n=8)	7	clustered	0.02	0.045	0.029	0.069	Multivariate ONS
star (n=8)	7	clustered	0.04	0.367	0.265	0.406	Cumulative mean
star (n=8)	7	clustered	0.08	0.998	0.990	0.999	Cumulative mean
star (n=8)	7	clustered	0.16	1.000	1.000	1.000	Cumulative mean
star (n=8)	7	scattered	0.01	0.013	0.011	0.032	Multivariate ONS
star (n=8)	7	scattered	0.02	0.057	0.048	0.083	Multivariate ONS
star (n=8)	7	scattered	0.04	0.464	0.441	0.431	Cumulative mean
star (n=8)	7	scattered	0.08	1.000	1.000	0.999	Cumulative mean
star (n=8)	7	scattered	0.16	1.000	1.000	1.000	Cumulative mean
path (n=8)	4	dense	0.01	0.036	0.042	0.095	Multivariate ONS
path (n=8)	4	dense	0.02	0.224	0.272	0.319	Multivariate ONS
path (n=8)	4	dense	0.04	0.931	0.956	0.947	Chromatic geometric
path (n=8)	4	dense	0.08	1.000	1.000	1.000	Cumulative mean
path (n=8)	4	dense	0.16	1.000	1.000	1.000	Cumulative mean
path (n=8)	4	sparse	0.01	0.007	0.004	0.021	Multivariate ONS
path (n=8)	4	sparse	0.02	0.015	0.008	0.034	Multivariate ONS

Graph	$\alpha(D)$	pattern	mag.	balanced	product	best baseline	which
path (n=8)	4	sparse	0.04	0.122	0.032	0.160	Cumulative mean
path (n=8)	4	sparse	0.08	0.846	0.487	0.878	Cumulative mean
path (n=8)	4	sparse	0.16	1.000	1.000	1.000	Cumulative mean
path (n=8)	4	clustered	0.01	0.009	0.006	0.029	Multivariate ONS
path (n=8)	4	clustered	0.02	0.030	0.018	0.046	Multivariate ONS
path (n=8)	4	clustered	0.04	0.222	0.127	0.261	Cumulative mean
path (n=8)	4	clustered	0.08	0.963	0.875	0.972	Cumulative mean
path (n=8)	4	clustered	0.16	1.000	1.000	1.000	Cumulative mean
path (n=8)	4	scattered	0.01	0.009	0.005	0.030	Multivariate ONS
path (n=8)	4	scattered	0.02	0.034	0.020	0.046	Cumulative mean
path (n=8)	4	scattered	0.04	0.249	0.166	0.294	Cumulative mean
path (n=8)	4	scattered	0.08	0.982	0.958	0.986	Cumulative mean
path (n=8)	4	scattered	0.16	1.000	1.000	1.000	Cumulative mean
Erdos-Renyi (n=12, p=.3)	5	dense	0.01	0.077	0.106	0.108	Multivariate ONS
Erdos-Renyi (n=12, p=.3)	5	dense	0.02	0.400	0.494	0.400	Multivariate ONS
Erdos-Renyi (n=12, p=.3)	5	dense	0.04	0.988	0.996	0.968	Multivariate ONS
Erdos-Renyi (n=12, p=.3)	5	dense	0.08	1.000	1.000	1.000	Cumulative mean
Erdos-Renyi (n=12, p=.3)	5	dense	0.16	1.000	1.000	1.000	Cumulative mean
Erdos-Renyi (n=12, p=.3)	5	sparse	0.01	0.007	0.005	0.018	Multivariate ONS
Erdos-Renyi (n=12, p=.3)	5	sparse	0.02	0.014	0.006	0.025	Multivariate ONS
Erdos-Renyi (n=12, p=.3)	5	sparse	0.04	0.102	0.015	0.143	Cumulative mean
Erdos-Renyi (n=12, p=.3)	5	sparse	0.08	0.785	0.189	0.845	Cumulative mean
Erdos-Renyi (n=12, p=.3)	5	sparse	0.16	1.000	0.996	1.000	Cumulative mean
Erdos-Renyi (n=12, p=.3)	5	clustered	0.01	0.016	0.012	0.027	Multivariate ONS
Erdos-Renyi (n=12, p=.3)	5	clustered	0.02	0.044	0.034	0.056	Multivariate ONS
Erdos-Renyi (n=12, p=.3)	5	clustered	0.04	0.314	0.206	0.352	Cumulative mean
Erdos-Renyi (n=12, p=.3)	5	clustered	0.08	0.987	0.948	0.992	Cumulative mean
Erdos-Renyi (n=12, p=.3)	5	clustered	0.16	1.000	1.000	1.000	Cumulative mean
Erdos-Renyi (n=12, p=.3)	5	scattered	0.01	0.014	0.017	0.022	Multivariate ONS
Erdos-Renyi (n=12, p=.3)	5	scattered	0.02	0.045	0.050	0.039	Cumulative mean
Erdos-Renyi (n=12, p=.3)	5	scattered	0.04	0.301	0.297	0.257	Cumulative mean
Erdos-Renyi (n=12, p=.3)	5	scattered	0.08	0.987	0.984	0.982	Cumulative mean
Erdos-Renyi (n=12, p=.3)	5	scattered	0.16	1.000	1.000	1.000	Cumulative mean