Electronic Colloquium on Computational Complexity, Report No. 144 (2026)

A QUANTITATIVE CONTAINER CHARACTERIZATION OF ONE-SIDED TESTABILITY

GAIA CARENINI, CAMERON SETH, AND YUICHI YOSHIDA

ABSTRACT. We give a quantitative combinatorial characterization of size-oblivious one-sided testability in the dense graph model, resolving a question of Alon, Fischer, Newman, and Shapira. For hereditary graph properties, we prove that one-sided testability is quantitatively equivalent to the existence of suitable hypergraph containers, a central and widely used tool in modern combinatorics. Combining this equivalence with the Alon–Shapira notion of semi-hereditariness yields a quantitative characterization of arbitrary graph properties. The correspondence is effective in both directions and provides explicit translations between tester complexity and container parameters. Our proof is regularity-free and extends uniformly to every fixed finite relational signature of bounded arity, including digraphs, coloured graphs, and hypergraphs. As applications, we obtain quantitative closure results for partition properties and testers for properties defined by the existence of a linearly large induced substructure.

1. Introduction

1.1. **The dense graph model.** Throughout this paper, a *graph* is a finite simple labelled graph. Thus self-loops and multiple edges are not allowed. A *graph property* is a family of graphs closed under isomorphism. A graph property Π is *hereditary* if it is closed under taking induced subgraphs, and it is *extendable* if every graph $H \in \Pi$ occurs as an induced subgraph of a member of Π on $|V(H)| + m$ vertices for every $m \geq 0$.

We work in the *dense graph model*, also called the adjacency-matrix model, introduced in [GGR98] and surveyed in [Gol17; BY22]. An input $G = (V, E)$ is represented by its adjacency predicate $g : V \times V \to \{0, 1\}$, where $g(u, u) = 0$, $g(u, v) = g(v, u)$, and $g(u, v) = 1$ precisely when $\{u, v\} \in E$.

Definition 1.1 (Distance from a graph property). *If G and G' are graphs on the same n-element vertex set, define*

$$\operatorname{dist}(G, G') = \frac{|E(G) \triangle E(G')|}{n^2}, \qquad \operatorname{dist}(G, \Pi) = \min_{\substack{G' \in \Pi \\ V(G') = V(G)}} \operatorname{dist}(G, G').$$

We say that G is ε-far from Π if $\operatorname{dist}(G, \Pi) > \varepsilon$; otherwise it is ε-close to Π.

The normalization by n^2, rather than by $\binom{n}{2}$, is immaterial up to an absolute factor. We use this convention throughout Part I. In the relational setting of Part II, each relation is normalized by the number of tuples of its arity.

A tester has oracle access to the adjacency predicate and to independent uniform samples from the vertex set. In the canonical formulation below, a request for a sample larger than the input is interpreted as a request to inspect the whole graph. This convention allows a size-oblivious tester to decide membership exactly on sufficiently small inputs without being given their order.

Definition 1.2 (Size-oblivious tester). *Let Π be a graph property. A size-oblivious tester for Π is a randomized oracle algorithm which, given $\varepsilon > 0$ but not $|V(G)|$, satisfies the following conditions.*

(i) *If $G \in \Pi$, then the tester accepts G with probability at least $2/3$.*
(ii) *If G is ε-far from Π, then the tester rejects G with probability at least $2/3$.*

ISSN 1433-8092

Its query complexity is the maximum number of adjacency queries it makes, and its sample complexity is the maximum number of vertex labels it inspects. Both bounds are required to depend only on ε. The tester has one-sided error if it accepts every graph in Π with probability one.

We tacitly replace complexity bounds by their nonincreasing monotone closures in the proximity parameter. A *canonical tester* samples a uniformly random set S of a prescribed size, queries every pair in S, and decides from the isomorphism type of $G[S]$; if the prescribed sample size exceeds $|V(G)|$, it inspects the whole graph. By the canonicalization theorem of Goldreich and Trevisan [GT03], a size-oblivious tester making $Q(\varepsilon)$ adjacency queries can, after changing ε by an absolute constant factor, be converted into a canonical tester sampling $O(Q(\varepsilon))$ vertices. Conversely, revealing an induced subgraph on s vertices uses $O(s^2)$ adjacency queries.

For a hereditary property, a canonical one-sided tester may reject exactly when the sampled induced graph does not belong to the property. Indeed, every induced subgraph of a graph in Π again lies in Π, while enlarging the collection of rejection types outside Π can only increase the rejection probability.

The guiding question of this paper is whether graph properties admitting size-oblivious one-sided-error testers with prescribed query complexity can be characterized in combinatorial and quantitative terms.

The qualitative theory is due to Alon and Shapira [AS08a]. They proved that every hereditary graph property admits a size-oblivious one-sided-error tester, through an induced removal lemma for possibly infinite families of forbidden induced subgraphs. They also characterized the arbitrary graph properties admitting such testers through the notion of semi-hereditariness. Informally, a property is semi-hereditary if it is contained in a hereditary property $\mathcal{H}$, and every sufficiently large graph in $\mathcal{H}$ is already close to the original property.

Thus the qualitative boundary of one-sided testability is well understood. What remains unclear is the quantitative structure controlling the number of queries. Alon, Fischer, Newman, and Shapira asked in Section 8 of [Alo+09] for a quantitative version of the Alon–Shapira characterization. More recently, Goldreich [Gol21] highlighted the broader problem of classifying graph properties according to their query complexity, pointing to the size-oblivious model as a particularly natural regime in which such a characterization might be possible.

Our main result gives such a characterization. The answer is expressed through suitable hypergraph containers.

1.2. **From regularity to containers.** The proof of the qualitative theorem of Alon and Shapira is based on a strengthened form of Szemerédi's regularity lemma due to Alon, Fischer, Krivelevich, and Szegedy [Alo+00]. This framework is well suited to proving that a tester with complexity independent of the size of the graph exists. It is, however, poorly adapted to the quantitative question considered here.

Regularity-based proofs generally lead to very large quantitative losses. Even when a property is described by finitely many forbidden induced subgraphs, the resulting bounds are typically of tower type. For hypergraph properties, regularity arguments may produce still larger, Ackermann-type dependencies. When a hereditary property is described by an arbitrary, possibly infinite, collection of forbidden induced subgraphs, the compactness step in the proof need not yield an effective bound at all. In particular, regularity does not retain the dependence on an arbitrary prescribed tester complexity.

Analytic and limit approaches give a complementary explanation of the same qualitative phenomenon. The theory of convergent dense graph sequences and testable graph parameters was developed by Borgs, Chayes, Lovász, Sós, Szegedy, and Vesztergombi [Bor+08]. Lovász and Szegedy [LS10] placed hereditary testing in the setting of sampling-continuous graphon properties, while Elek and

Szegedy [ES12] developed measure-theoretic and ultraproduct methods for hypergraphs. These approaches explain the generality of hereditary testability, but they are likewise not designed to retain effective dependence on tester complexity.

It is therefore natural to search for a more effective combinatorial mechanism. In extremal combinatorics, hypergraph containers have repeatedly played precisely this role: they replace a global regularity decomposition by a direct description of the entire family of objects avoiding a collection of forbidden configurations.

The hypergraph container method was introduced independently by Balogh, Morris, and Samotij [BMS15] and by Saxton and Thomason [ST15], building on earlier work of Kleitman and Winston [KW82]; see [BMS18] for a survey. Over the past decade, it has become one of the main tools of extremal and probabilistic combinatorics.

Informally, a hypergraph container theorem gives a small family of subsets, called containers, such that every independent set of the hypergraph is contained in one of them, while each container spans only few hyperedges. Thus, rather than studying all independent sets individually, one can cover them by a controlled collection of larger sets that are still almost independent.

A standard example is the enumeration of triangle-free graphs. Consider the 3-uniform hypergraph whose vertex set is $E(K_n)$ and whose hyperedges are the triples of edges forming triangles. Its independent sets are precisely the triangle-free graphs on $[n]$. Applying the container method gives a family of subsets of $E(K_n)$, which we identify with graphs on $[n]$, such that every triangle-free graph is contained in one of these graphs and each container contains few triangles. Moreover, there are only $2^{o(n^2)}$ containers. Supersaturation then implies that every container has at most $n^2/4 + o(n^2)$ edges. Since each such container has at most $2^{n^2/4+o(n^2)}$ subgraphs, the number of labelled triangle-free graphs is at most $2^{n^2/4+o(n^2)}$. The matching lower bound follows by taking arbitrary subgraphs of a balanced complete bipartite graph.

The important point for us is that containers turn a quantitative abundance of forbidden configurations into an effective structural description of all objects avoiding them. Our approach follows the same philosophy. If a graph is far from a hereditary property, then it has many small induced subgraphs violating the property. These bad vertex sets form a dense uniform hypergraph, and the induced subgraphs satisfying the property correspond to independent sets in this hypergraph.

Containers have recently also appeared in property testing, where they were used to obtain improved quantitative bounds for particular classes of properties [BS24; BS25; Set25; GNS26]. Their role in the present paper is more structural: rather than merely providing improved testers for particular properties, suitable containers characterize the quantitative structure underlying one-sided testability itself.

1.3. **Ordered containers.**

The main idea underlying our characterization is that, inside any graph that is far from a hereditary property, every induced subgraph satisfying the property is forced to lie inside a substantially smaller ambient vertex set, and that this ambient set is determined by a small amount of information. We formalize this through the following notion.

For a set V, let $V^j_{\neq}$ denote the family of ordered j-tuples of distinct elements of V. We write $(x)_j = x(x-1)\cdots(x-j+1)$ for the falling factorial, with $(x)_0 = 1$, and use the convention that $\binom{a}{b} = 0$ when $b < 0$ or $b > a$.

Definition 1.3 (Ordered container lemma for a hereditary graph property)**.** *Let Π be a hereditary graph property. For $\varepsilon, \eta > 0$ and integers $N, q \geq 0$, we say that Π admits an $(\varepsilon, \eta, N, q)$-ordered container lemma if the following holds.*

For every $n \geq N$ and every n-vertex graph $G = (V, E)$ that is ε-far from Π, there exist maps $f : \{U \subseteq V : G[U] \in \Pi\} \longrightarrow \bigcup_{j=0}^{q} V^j_{\neq}$, $\qquad C : \mathrm{im}(f) \longrightarrow \binom{V}{\leq (1-\eta)n}$, such that every entry of $f(U)$ belongs to U and $U \subseteq C(f(U))$ whenever $G[U] \in \Pi$. We refer to $f(U)$ as the fingerprint of U, and to $C(f(U))$ as its container.

The container associated with an induced Π-subgraph is constructed by an iterative procedure. At each step, the procedure selects a vertex from the induced subgraph and uses it to discard part of the ambient vertex set. The selected vertices form the fingerprint: a short record of the choices made during the construction, from which the final container can be recovered.

The order on the fingerprint records the successive choices made in the construction of the container. It is not inherited from an ordering of the ambient graph. Thus an ordered container lemma asserts that every induced Π-subgraph of a graph far from Π is contained in a vertex set missing at least an η-fraction of the ambient vertices, and that this containing set is determined by at most q vertices of the induced subgraph itself.

With this notation in place, we may state our main theorem.

Theorem 1.4 (Characterization of hereditary properties having specified complexity)**.** *Let Π be a hereditary graph property, and let $Q : (0, 1) \to \mathbb{N}$. Then the following hold.*

(i) *If Π admits a size-oblivious one-sided-error tester with query complexity at most $Q(\varepsilon)$, then there is an absolute constant $c \in (0, 1)$ such that, for every $\varepsilon > 0$, the property Π admits an ordered container lemma with parameters satisfying*

$$q(\varepsilon) = O(Q(c\varepsilon)), \qquad \eta(\varepsilon) = \Omega\Big(Q(c\varepsilon)^{-1}\Big), \qquad N(\varepsilon) = O(Q(c\varepsilon)).$$

(ii) *Conversely, if Π admits ordered container lemmas with parameters $\eta(\varepsilon)$, $N(\varepsilon)$, and $q(\varepsilon)$, then Π admits a size-oblivious one-sided-error tester with query complexity $\widetilde{O}\Big(N(\varepsilon)^2 + \frac{(q(\varepsilon)+1)^2}{\eta(\varepsilon)^2}\Big)$.*

Here and throughout, $\widetilde{O}(\cdot)$ and $\widetilde{\Omega}(\cdot)$ hide polylogarithmic factors. The theorem shows that one-sided testability and ordered-container parameters are quantitatively equivalent up to explicit polynomial transformations.

The main ingredient in the proof is a removal–container equivalence. Given an n-vertex graph G and an integer $s \geq 1$, we form the s-uniform hypergraph $H_{G,\Pi}$ on $V(G)$ with

$$E(H_{G,\Pi}) = \left\{ S \in \binom{V(G)}{s} : G[S] \notin \Pi \right\}.$$

Since Π is hereditary, every set $U \subseteq V(G)$ satisfying $G[U] \in \Pi$ is an independent set in $H_{G,\Pi}$.

If a canonical tester sampling s vertices rejects every graph that is ε-far from Π with probability at least $2/3$, then $e(H_{G,\Pi}) \geq \frac{2}{3}\binom{n}{s}$. The required containers therefore follow from a purely hypergraph-theoretic statement.

Theorem 1.5 (Dense ordered containers)**.** *Let H be an s-uniform hypergraph on an n-element vertex set V, and suppose that $e(H) \geq \delta\binom{n}{s}$. Then there exist maps $f : \mathcal{I}(H) \longrightarrow \bigcup_{j=0}^{s-1} V^j_{\neq}$, $\quad C : \mathrm{im}(f) \longrightarrow \binom{V}{\leq (1-\delta/s)n}$, such that every entry of $f(I)$ belongs to I and $I \subseteq C(f(I))$ for every independent set $I \in \mathcal{I}(H)$.*

The proof is an elementary iterative pivot argument. Starting from an independent set I, we choose a vertex $v \in I$ of maximum degree and record it in the fingerprint. We then pass to the link of v, namely the hypergraph whose edges are the sets $e \setminus v$ over all hyperedges e containing v. This reduces the uniformity by one. The maximality of the degree of v allows us to exclude from the eventual container every vertex whose degree is larger than that of v, since none of these vertices can belong to I. We repeat the procedure in the successive links. If, at some stage, there are many vertices of degree larger than the chosen pivot, then a positive proportion of the ambient vertices is excluded and the desired shrinkage follows. Otherwise, the chosen pivot has degree comparable to the average degree, so a definite proportion of the edge density survives in the next link. Iterating this alternative, either sufficient shrinkage occurs at an intermediate stage, or enough density remains when the process reaches a 1-uniform hypergraph. In the latter case, its edges are simply vertices that cannot belong to I, and deleting them gives the required container.

The converse direction is a counting argument. Suppose that every independent set is contained in a set of size at most $(1-\eta)n$ determined by a fingerprint of length at most q. If T is a uniformly random t-subset of the vertex set and T is independent, then T contains its fingerprint and all its remaining vertices lie in the corresponding container. Consequently,

$$\Pr[T \text{ is independent}] \le \sum_{j=0}^{q} (t)_j \frac{\binom{\lfloor (1-\eta)n \rfloor - j}{t-j}}{\binom{n-j}{t-j}} \le \sum_{j=0}^{q} t^j (1-\eta)^{t-j}.$$

The last expression is at most $1/3$ once $t = O\left(\frac{q+1}{\eta} \log\left(\frac{q+1}{\eta} + 2\right)\right)$. Thus an ordered container lemma yields a removal statement and hence a one-sided tester.

In this way, our proof factors through the equivalence between one-sided testing, induced removal, and ordered containers.

1.4. **Arbitrary graph properties.** The passage from hereditary properties to arbitrary ones is achieved through the notion of semi-hereditariness introduced by Alon and Shapira.

Definition 1.6 (Semi-hereditary graph property)**.** *A graph property Π is semi-hereditary if there exist a hereditary graph property $\mathcal{H} \supseteq \Pi$ and a function $M : (0,1) \to \mathbb{N}$ such that, for every $\varepsilon > 0$, every graph $G \in \mathcal{H}$ with at least $M(\varepsilon)$ vertices is ε-close to Π. We call $(\mathcal{H}, M)$ a semi-hereditary witness for Π.*

Starting from a one-sided tester for Π, the argument of Alon and Shapira constructs a hereditary envelope $\mathcal{H}$ by forbidding the finite induced graphs on which the tester can reject. The same tester also tests $\mathcal{H}$, and every sufficiently large member of $\mathcal{H}$ must be close to Π. Conversely, if Π is semi-hereditary with witness $(\mathcal{H}, M)$, then testing Π reduces to testing $\mathcal{H}$, together with exact inspection below the threshold M.

Combining this reduction with Theorem 1.4 gives the following.

Corollary 1.7 (Quantitative extension to arbitrary graph properties)**.** *Let Π be a graph property, and let $Q : (0,1) \to \mathbb{N}$. Then the following hold.*

(i) *If Π admits a size-oblivious one-sided-error tester with query complexity at most $Q(\varepsilon)$, then Π has a semi-hereditary witness $(\mathcal{H}, M)$ satisfying $M(\varepsilon) = O(Q(c\varepsilon))$ for an absolute constant $c > 0$, and $\mathcal{H}$ admits ordered-container parameters satisfying*

$$q(\varepsilon) = O(Q(c\varepsilon)), \qquad \eta(\varepsilon) = \Omega(Q(c\varepsilon)^{-1}), \qquad N(\varepsilon) = O(Q(c\varepsilon)).$$

(ii) *Conversely, if Π has a semi-hereditary witness $(\mathcal{H}, M)$ and $\mathcal{H}$ admits ordered-container parameters (η, N, q), then Π admits a size-oblivious one-sided-error tester with query complexity $O\big(M(\varepsilon/2)^2\big) + \widetilde{O}\Big(N(\varepsilon/2)^2 + \frac{(q(\varepsilon/2)+1)^2}{\eta(\varepsilon/2)^2}\Big)$.*

Thus the quantitative part of the problem is carried by ordered containers for the hereditary envelope, while semi-hereditariness describes precisely how an arbitrary graph property reduces to that setting.

1.5. **Beyond graphs.** Although graphs are the main setting of the paper and the source of the motivating question, the combinatorial argument above is not graph-specific. Once local obstructions are encoded as the edges of a uniform hypergraph, the proof no longer uses symmetry, irreflexivity, or the fact that the ambient relation has arity two.

The qualitative theory beyond graphs is already broad. Rödl and Schacht [RS07] extended hereditary testability to uniform hypergraphs by means of hypergraph regularity and removal. Austin and Tao [AT10] developed a general framework for multiple directed, coloured, and non-uniform hypergraphs, allowing loops and relations of different arities, and studied the stronger notion of local repairability. Alon, Ben-Eliezer, and Fischer [ABF17] obtained related results for edge-coloured ordered graphs and for two-dimensional matrices over finite alphabets.

Our argument gives a common quantitative and regularity-free treatment of these settings. We formulate this in the language of finite relational structures.

Definition 1.8 (Finite relational signatures and structures)**.** *A finite relational signature is a finite list $\tau = \{R_1, \dots, R_m\}$ of relation symbols together with a positive integer $a(R_i)$, called the arity of R_i, for each i. A finite τ-structure $\mathcal{A}$ consists of a finite vertex set $V(\mathcal{A})$ and an interpretation $R_i^{\mathcal{A}} \subseteq V(\mathcal{A})^{a(R_i)}$ of each symbol R_i. We write $r(\tau) = \max_{R\in\tau} a(R)$ for the maximum arity.*

A signature specifies only the names of the relations and the number of inputs each relation takes. It does not impose any further conditions on those relations. Conditions such as symmetry, the absence of loops, or the requirement that every pair receive exactly one colour are instead included in the definition of the property under consideration. For example, a simple graph is represented by one binary relation E, with the additional requirements that E be symmetric and irreflexive. A directed graph is represented by one binary relation with no symmetry requirement. An edge-coloured graph is represented by one binary relation for each colour, together with the requirement that these relations partition the pairs of distinct vertices. Finally, a k-uniform hypergraph is represented by one k-ary relation, required to be invariant under permutations of its arguments and to contain only tuples of distinct vertices. In this way, the same formalism treats directed, coloured, uniform, and non-uniform structures within a single framework.

For $U \subseteq V(\mathcal{A})$, the *induced substructure* $\mathcal{A}[U]$ is obtained by restricting every relation to tuples all of whose entries belong to U.

Definition 1.9 (Relational properties)**.** *A τ-property is a family of finite τ-structures closed under isomorphism. It is hereditary if $\mathcal{A} \in \mathcal{P}$ implies $\mathcal{A}[U] \in \mathcal{P}$ for every $U \subseteq V(\mathcal{A})$. It is extendable if, for every $\mathcal{A} \in \mathcal{P}$ and every $m \geq 0$, there is a structure $\mathcal{B} \in \mathcal{P}$ on $|V(\mathcal{A})| + m$ vertices containing $\mathcal{A}$ as an induced substructure.*

If $\mathcal{A}$ and $\mathcal{B}$ are τ-structures on the same n-element vertex set, we define $d_\tau(\mathcal{A}, \mathcal{B}) = \frac{1}{|\tau|}\sum_{R\in\tau} \frac{|R^{\mathcal{A}} \triangle R^{\mathcal{B}}|}{n^{a(R)}}$. Distance from a property and the notions of being ε-far or ε-close are defined in the usual way.

The intrinsic testing notion is again induced sampling.

Definition 1.10 (Oblivious induced-sampling tester)**.** *Let $\mathcal{P}$ be a τ-property. An oblivious induced-sampling tester for $\mathcal{P}$ consists, for every $\varepsilon > 0$, of an integer $s(\varepsilon)$ and a collection $\mathcal{R}_\varepsilon$ of isomorphism types of τ-structures on $s(\varepsilon)$ vertices.*

On an input $\mathcal{A}$ with $|V(\mathcal{A})| \geq s(\varepsilon)$, the tester chooses a uniformly random set $S \in \binom{V(\mathcal{A})}{s(\varepsilon)}$, reveals the complete induced structure $\mathcal{A}[S]$, and rejects if and only if $\mathcal{A}[S] \in \mathcal{R}_\varepsilon$. If $|V(\mathcal{A})| < s(\varepsilon)$, it may inspect the whole input.

The tester has one-sided error if it always accepts members of $\mathcal{P}$, and is sound if it rejects every ε-far input with probability at least $2/3$. Its sample complexity is $s(\varepsilon)$.

For hereditary properties, the tester may reject exactly when the sampled induced structure does not belong to the property.

Definition 1.11 (Ordered containers for relational properties)**.** *Let $\mathcal{P}$ be a hereditary τ-property. For $\varepsilon, \eta > 0$ and integers $N, q \geq 0$, we say that $\mathcal{P}$ admits an $(\varepsilon, \eta, N, q)$-ordered container lemma if, for every $n \geq N$ and every n-vertex τ-structure $\mathcal{A}$ that is ε-far from $\mathcal{P}$, there exist maps*

$$f : \{U \subseteq V(\mathcal{A}) : \mathcal{A}[U] \in \mathcal{P}\} \longrightarrow \bigcup_{j=0}^{q} V(\mathcal{A})^j_{\neq}, \qquad C : \mathrm{im}(f) \longrightarrow \binom{V(\mathcal{A})}{\leq (1-\eta)n},$$

such that every entry of $f(U)$ belongs to U and $U \subseteq C(f(U))$.

Theorem 1.12 (Container characterization for hereditary relational structures)**.** *Let $\mathcal{P}$ be a hereditary property of finite τ-structures.*

(i) *If $\mathcal{P}$ has a one-sided induced-sampling tester with sample complexity $s(\varepsilon)$, then it admits ordered-container parameters $N(\varepsilon) = s(\varepsilon)$, $q(\varepsilon) = s(\varepsilon) - 1$, and $\eta(\varepsilon) = 2/(3s(\varepsilon))$.*

(ii) *Conversely, if $\mathcal{P}$ admits ordered-container parameters (η, N, q), then it has a one-sided induced-sampling tester with sample complexity*

$$O\left(\max\left\{N(\varepsilon), \frac{q(\varepsilon)+1}{\eta(\varepsilon)} \log\left(\frac{q(\varepsilon)+1}{\eta(\varepsilon)} + 2\right)\right\}\right).$$

Revealing the complete induced structure on t vertices requires at most $\sum_{R\in\tau} t^{a(R)} \le |\tau| t^r$ relation queries. Thus the theorem also gives a characterization in terms of relation-query complexity.

Definition 1.13 (Semi-hereditary relational property)**.** *A τ-property $\mathcal{P}$ is semi-hereditary if there exist a hereditary property $\mathcal{H} \supseteq \mathcal{P}$ and a function $M : (0, 1) \to \mathbb{N}$ such that every $\mathcal{A} \in \mathcal{H}$ with $|V(\mathcal{A})| \ge M(\varepsilon)$ is ε-close to $\mathcal{P}$. We call $(\mathcal{H}, M)$ a semi-hereditary witness.*

The same hereditary-envelope argument gives a quantitative characterization of arbitrary relational properties. For hereditary properties in a fixed bounded-arity signature, polynomial induced-sampling complexity, polynomial relation-query complexity, and polynomial ordered-container parameters are equivalent. For an arbitrary property, the corresponding statement also requires polynomial control of the semi-hereditary threshold.

1.6. Applications.

The characterization has three consequences that we develop in Part III.

Partition properties. Let $\mathcal{P}_1, \dots, \mathcal{P}_k$ be properties of structures in a common binary signature. We write $\mathrm{Part}(\mathcal{P}_1, \dots, \mathcal{P}_k)$ for the property of admitting a partition $V(\mathcal{A}) = V_1 \cup \cdots \cup V_k$ such that $\mathcal{A}[V_i] \in \mathcal{P}_i$ for every i; tuples meeting distinct parts are unrestricted. We prove that if the $\mathcal{P}_i$ are nonempty, hereditary, extendable, and one-sided testable, then their partition property is also one-sided testable, with an explicit quantitative bound.

Theorem 1.14 (Partition closure for binary signatures)**.** *Let every relation symbol in τ have arity two, and let $\mathcal{P}_1, \dots, \mathcal{P}_k$ be nonempty hereditary and extendable τ-properties. Suppose that $\mathcal{P}_i$ has one-sided induced-sampling complexity $S_i(\delta)$, and put $S(\delta) = \max_{i\in[k]} S_i(\delta)$. Then* $\mathrm{Part}(\mathcal{P}_1, \dots, \mathcal{P}_k)$ *has a one-sided induced-sampling tester with sample complexity $\widetilde{O}\Big(\frac{k}{\varepsilon} S\big(\frac{\varepsilon}{8k}\big)^2\Big)$.*

For graphs, this applies to (r, s)-colourability, where the vertex set is partitioned into r independent sets and s cliques. It includes ordinary colourability and split graphs, and it gives a tester for bounded cochromatic number, the minimum number of parts each inducing either a clique or an independent set. For digraphs it gives a tester for bounded dichromatic number, the minimum number of acyclic parts.

Linearly large induced substructures. For $\rho \in (0, 1]$ and a property $\mathcal{P}$, let

$$\mathrm{Large}_\rho(\mathcal{P}) = \left\{\mathcal{A} : \begin{array}{c} \text{there exists } U \subseteq V(\mathcal{A}) \text{ with } |U| \ge \rho|V(\mathcal{A})| \\ \text{and } \mathcal{A}[U] \in \mathcal{P} \end{array}\right\}.$$

This property is generally not hereditary and need not be a bounded partition property. Nevertheless, iterating the ordered-container construction gives a tester whenever $\mathcal{P}$ is hereditary, extendable, and one-sided testable.

Theorem 1.15 (Testing for a large induced member)**.** *Let τ be a finite binary signature, let $\mathcal{P}$ be a nonempty hereditary and extendable τ-property, and let $S_{\mathcal{P}}(\delta)$ denote its one-sided induced-sampling complexity. Fix $\rho \in (0,1]$ and $0 < \varepsilon \leq \rho^2/8$, and put $\delta = \varepsilon/(4\rho^2)$. Then* $\mathrm{Large}_\rho(\mathcal{P})$ *has a two-sided induced-sampling tester using $\widetilde{O}\left(\frac{\rho^2}{\varepsilon^2} S_{\mathcal{P}}(\delta)^2 \log \frac{2}{\rho}\right)$ vertices.*

In particular, we obtain a tester for the existence of an induced acyclic subdigraph containing a prescribed positive proportion of the vertices.

Counting inside far hosts. Ordered containers also control all induced members of a hereditary property inside a fixed far host. If every induced Π-subgraph of an n-vertex graph G has a fingerprint of length at most q and a container of size at most $(1-\eta)n$, then $\left|\{U \subseteq V(G) : G[U] \in \Pi\}\right| \leq (q+1)n^q 2^{(1-\eta)n}$. Thus the family of vertex sets inducing members of Π has a uniform entropy deficit. Combining this observation with known edit-distance results for hereditary properties yields structural statements in random hosts; for example, with high probability every perfect induced subgraph of $G(n, 1/2)$ is contained in one of polynomially many vertex sets of size at most $(1-\eta)n$, for a constant $\eta > 0$. The edit-distance background and the formal random-host consequences are deferred to Section 7.

1.7. **Organization of the paper.** The paper is divided into three parts. Part I proves the graph characterization: first for hereditary properties through the removal–container equivalence, and then for arbitrary properties through semi-hereditary envelopes. Part II develops the corresponding theory for finite relational structures of bounded arity, including the induced-sampling model and its canonicalization. Part III contains the applications: closure under vertex partitions, properties defined by a linearly large induced substructure, and counting and structural consequences inside far hosts.

Acknowledgements. This project was carried out while the first two authors were participating in the 2025 NII International Internship Program. The authors are grateful to the program for its support. The first author is supported by the CB European PhD Studentship funded by Trinity College, Cambridge.

Part I. Graph properties

All graph-theoretic and testing terminology in this part is as fixed in the introduction.

2. The graph characterization

We suppress floor and ceiling signs when they have no effect on the argument.

2.1. **Hereditary properties.** We begin by isolating the combinatorial heart of the argument, namely the equivalence between removal statements and ordered container statements. The forward implication constructs, from a graph that is far from the property, a shrinking container around every induced Π-subgraph by iteratively exposing high-degree vertices in an auxiliary hypergraph. The reverse implication is a counting argument: once every induced Π-subgraph is controlled by a short fingerprint and a substantially smaller container, a random induced subgraph has a positive probability of witnessing a violation of Π.

Definition 2.1 (Removal lemma for a hereditary graph property)**.** *Let Π be a hereditary graph property. For real numbers $\varepsilon, \delta > 0$ and integers $N, s \geq 1$, we say that Π admits an $(\varepsilon, \delta, N, s)$-removal lemma if the following holds. For every integer $n \geq N$ and every n-vertex graph $G = (V, E)$ that is ε-far from Π, there are at least $\delta\binom{n}{s}$ sets $S \in \binom{V}{s}$ such that $G[S] \notin \Pi$.*

Theorem 2.2 (Removal–container equivalence)**.** *Let Π be a hereditary graph property.*

(1) If Π *admits an* $(\varepsilon, \delta, N, s)$*-removal lemma, then it admits the following ordered container statement: for every integer* $n \geq \max\{N, s\}$ *and every* n*-vertex graph* $G = (V, E)$ *that is* ε*-far from* Π*, there are maps* $f : \{U \subseteq V : G[U] \in \Pi\} \longrightarrow \bigcup_{j=0}^{s-1} V_{\neq}^{j}$, $\quad C : \operatorname{im}(f) \longrightarrow \binom{V}{\leq (1-\delta/s)n}$*, such that every entry of* $f(U)$ *belongs to* U *and* $U \subseteq C(f(U))$ *whenever* $G[U] \in \Pi$*.*

(2) Conversely, suppose that Π *admits an ordered container statement with proximity parameter* ε*, shrinkage* η*, threshold* N*, and fingerprint bound* q*. Then, for every integer* $t \geq q$ *such that* $\sum_{j=0}^{q} t^j (1-\eta)^{t-j} < 1$*, the property* Π *admits an* $(\varepsilon, \delta_t, N, t)$*-removal lemma, where* $\delta_t = 1 - \sum_{j=0}^{q} t^j (1-\eta)^{t-j} > 0$*.*

Proof. We first prove the forward implication. Let $G = (V, E)$ be an n-vertex graph with $n \geq \max\{N, s\}$ that is ε-far from Π, and let H be the s-uniform hypergraph on V whose edges are the sets $S \in \binom{V}{s}$ for which $G[S] \notin \Pi$. By hypothesis, $e(H) \geq \delta \binom{n}{s}$. Every set $U \subseteq V$ with $G[U] \in \Pi$ is an independent set of H.

Fix a total order on V, used only to make all choices canonical. If $s = 1$, let every independent set have the empty fingerprint and put $C(()) = V \setminus E(H)$, where the edge set of the 1-uniform hypergraph is identified with a subset of V. Then $|C(())| \leq (1-\delta)n$, and the conclusion follows. We may therefore assume $s \geq 2$.

If $n = s$, then H is nonempty and its unique possible edge is V. Hence every independent set is a proper subset of V. Assign to an independent set its increasing ordering as fingerprint and use the set itself as its container. Since $\delta \leq 1$, every such container has size at most $n - 1 \leq (1-\delta/s)n$.

Assume henceforth that $n > s$. If an independent set U has $|U| \leq s-2$, let $f(U)$ be its increasing ordering and set $C(f(U)) = U$. These fingerprints have length at most $s-2$. Now let $|U| \geq s-1$. Starting from $H^{(s)} = H$, define hypergraphs $H^{(s)}, H^{(s-1)}, \ldots, H^{(1)}$ recursively. For $i = s, s-1, \ldots, 2$, choose a vertex $v_i \in U \cap V(H^{(i)})$ of maximum degree among the vertices of $U \cap V(H^{(i)})$, breaking ties by the fixed order, and put $X_i = \{u \in V(H^{(i)}) : \deg_{H^{(i)}}(u) > \deg_{H^{(i)}}(v_i)\}$. Let $H^{(i-1)}$ be the $(i-1)$-uniform link of v_i in $H^{(i)}$, on the vertex set $V(H^{(i)}) \setminus \{v_i\}$. Since U is independent in $H^{(i)}$, the set $U \cap V(H^{(i-1)})$ is independent in $H^{(i-1)}$; in particular, the construction is well defined through the last step.

Set $f(U) = (v_s, v_{s-1}, \ldots, v_2)$ and define $C(f(U)) = V \setminus \left(E(H^{(1)}) \cup \bigcup_{i=2}^{s} X_i\right)$. Here again $E(H^{(1)})$ is viewed as a set of vertices. This is a genuine function of the fingerprint: starting from H, the ordered tuple $(v_s, \ldots, v_2)$ uniquely reconstructs all links $H^{(i)}$, all sets X_i, and hence the displayed container. Thus two independent sets with the same fingerprint receive the same container. Moreover, fingerprints arising in this case have length exactly $s-1$, so they cannot coincide with the fingerprints used for sets of size at most $s-2$.

By maximality of v_i inside $U \cap V(H^{(i)})$, every X_i is disjoint from U. Also, $U \cap V(H^{(1)})$ avoids $E(H^{(1)})$. Consequently $U \subseteq C(f(U))$.

It remains to prove the shrinkage. Write $n_i = |V(H^{(i)})| = n - s + i$ and $m_i = e(H^{(i)})$. Then $m_{i-1} = \deg_{H^{(i)}}(v_i)$ for $2 \leq i \leq s$. Let k be the least index in $[s]$ such that either $k = 1$ or $m_{k-1} < \frac{k-1}{n_k - 1} m_k$. For every $j = k+1, \ldots, s$ we therefore have $m_{j-1} \geq \frac{j-1}{n_j - 1} m_j$, and hence

$$m_k \geq \left(\prod_{j=k+1}^{s} \frac{j-1}{n_j - 1} \right) m_s \geq \frac{\delta n}{s} \binom{n_k - 1}{k-1}.$$

If $k = 1$, then $|E(H^{(1)})| = m_1 \geq \delta n/s$, which gives the required bound on $|C(f(U))|$.

Suppose that $k \geq 2$ and put

$$Y = \left\{ u \in V(H^{(k)}) : \deg_{H^{(k)}}(u) > \frac{k-1}{n_k - 1} m_k \right\}.$$

Since $n_k \geq k+1$, the average degree km_k/n_k is strictly larger than the threshold in this definition. Furthermore, $|Y| \geq \frac{m_k}{\binom{n_k-1}{k-1}}$. Indeed, otherwise the contribution to the degree sum from Y is less than m_k, while the remaining at most $n_k - 1$ vertices contribute at most $(k-1)m_k$, contradicting that the total degree sum is km_k. By the choice of k, $\deg_{H^{(k)}}(v_k) = m_{k-1} < \frac{k-1}{n_k-1}m_k$, so $Y \subseteq X_k$. It follows that

$$|X_k| \geq |Y| \geq \frac{m_k}{\binom{n_k-1}{k-1}} \geq \frac{\delta n}{s}.$$

Thus in every case $|C(f(U))| \leq (1-\delta/s)n$, proving the first implication.

For the converse, let $G = (V, E)$ be an n-vertex graph with $n \geq N$ that is ε-far from Π. The assertion is vacuous when $n < t$, so assume $n \geq t$ and choose $T \in \binom{V}{t}$ uniformly. If $G[T] \in \Pi$, then for some $j \leq q$ its fingerprint $F = f(T)$ is an ordered j-tuple contained in T, and $T \subseteq C(F)$, where $|C(F)| \leq (1-\eta)n$. For a fixed ordered tuple F of length j, $\Pr[F \subseteq T \subseteq C(F)] = \frac{\binom{|C(F)|-j}{t-j}}{\binom{n}{t}}$. There are at most $(n)_j$ possible ordered tuples of length j. Summing over all fingerprints and using $\frac{(n)_j}{\binom{n}{t}} = \frac{(t)_j}{\binom{n-j}{t-j}}$ gives

$$\begin{aligned}\Pr[G[T] \in \Pi] &\leq \sum_{j=0}^{q} (t)_j \frac{\binom{\lfloor(1-\eta)n\rfloor - j}{t-j}}{\binom{n-j}{t-j}} \\ &\leq \sum_{j=0}^{q} t^j (1-\eta)^{t-j}.\end{aligned}$$

Therefore at least a δ_t-fraction of the t-subsets induce graphs outside Π, as required. □

We now deduce the main characterization theorem from Theorem 2.2 together with the standard canonical reduction for one-sided testers in the dense graph model.

Proof of Theorem 1.4. Suppose first that Π has a size-oblivious one-sided tester of query complexity at most $Q(\varepsilon)$. By the canonicalization theorem, there are absolute constants $C > 0$ and $c \in (0,1)$ such that, at proximity ε, a canonical one-sided tester samples $s \leq CQ(c\varepsilon)$ vertices. Enlarge s to an integer if necessary. If an n-vertex graph G is ε-far from Π and $n \geq s$, then at least a $2/3$-fraction of its induced s-vertex subgraphs lie outside Π. Thus Π admits an $(\varepsilon, 2/3, s, s)$-removal lemma. The first part of Theorem 2.2 gives ordered containers with $q(\varepsilon) = s-1$, $\eta(\varepsilon) = \frac{2}{3s}$, and $N(\varepsilon) = s$, which proves the stated bounds.

Conversely, fix $\varepsilon > 0$ and abbreviate $q = q(\varepsilon)$, $\eta = \eta(\varepsilon)$, and $N = N(\varepsilon)$. By Lemma 4.2 below, there is an absolute constant C such that $t = \left\lceil \frac{C(q+1)}{\eta} \log\left(\frac{q+1}{\eta} + 2\right) \right\rceil$ satisfies $\sum_{j=0}^{q} t^j (1-\eta)^{t-j} \leq 1/3$. Set $T = \max\{N, t\}$. The tester requests a uniformly random set of T vertices; under our standing convention, if the input has fewer than T vertices, it receives and inspects the whole input. It rejects exactly when the revealed induced graph does not belong to Π.

The tester is one-sided by heredity. On an input with fewer than T vertices it decides membership exactly. If $n \geq T$ and the input is ε-far from Π, the second part of Theorem 2.2 shows that a uniformly random t-set is bad with probability at least $2/3$. A uniformly random T-set can be generated first and then a uniformly random t-subset chosen inside it. Since Π is hereditary, a good T-set contains only good t-sets; hence the probability that the sampled T-set is good is at most $1/3$. The query complexity is $O(T^2) = \widetilde{O}\left(N(\varepsilon)^2 + \frac{(q(\varepsilon)+1)^2}{\eta(\varepsilon)^2}\right)$. □

2.2. **Semi-hereditary envelopes.** We now pass from hereditary properties to arbitrary graph properties via the Alon–Shapira notion of semi-hereditariness.

Definition 2.3 (Semi-hereditary graph property)**.** *A graph property* Π *is said to be semi-hereditary if there exist a hereditary graph property* $H \supseteq \Pi$ *and a function* $M : (0,1) \to \mathbb{N}$ *such that, for every* $\varepsilon > 0$*, every graph* $G \in H$ *with at least* $M(\varepsilon)$ *vertices is* ε*-close to* Π*.*

The forward implication combines their hereditary-envelope construction with Theorem 1.4. For the converse, we test the hereditary envelope first and then use a brute-force routine below the semi-hereditary threshold.

Proof of Corollary 1.7. Suppose first that Π has a size-oblivious one-sided tester of query complexity at most $Q(\varepsilon)$. By canonicalization, there are absolute constants $C > 0$ and $c \in (0,1)$ and, for each $\varepsilon > 0$, a canonical one-sided tester T_ε sampling at most $s(\varepsilon) = CQ(c\varepsilon)$ vertices. Let $\mathcal{B}$ be the family of all finite graphs that occur as a rejection type of some T_ε, and let H be the hereditary property of containing no member of $\mathcal{B}$ as an induced subgraph.

One-sidedness gives $\Pi \subseteq H$. For each fixed ε, the same tester T_ε tests H: it never rejects an input in H, while an input that is ε-far from H is also ε-far from Π. Moreover, if $G \in H$, $|V(G)| \geq s(\varepsilon)$, and G were ε-far from Π, then T_ε would reject with positive probability and hence expose a member of $\mathcal{B}$ inside G, a contradiction. Thus (H, M) is a semi-hereditary witness with $M(\varepsilon) \leq s(\varepsilon)$. Applying Theorem 1.4 to H gives the asserted ordered-container parameters.

Conversely, suppose that (H, M) is a semi-hereditary witness for Π and that H has ordered-container parameter functions (η, N, q). Apply Theorem 1.4 to obtain a canonical one-sided tester for H at proximity $\varepsilon/2$, with sample size $S_H = \widetilde{O}\Big(N(\varepsilon/2) + \frac{q(\varepsilon/2)+1}{\eta(\varepsilon/2)}\Big)$. Set $T = \max\{M(\varepsilon/2), S_H\}$. Request a uniformly random set of T vertices. If the input has fewer than T vertices, inspect it completely and decide membership in Π exactly. Otherwise run the tester for H on the appropriate subset of the sampled vertices.

Inputs in Π are accepted because $\Pi \subseteq H$. Now let G be ε-far from Π and have at least T vertices. Then G is $\varepsilon/2$-far from H: otherwise there would be a graph $G' \in H$ on the same vertex set with $\operatorname{dist}(G, G') \leq \varepsilon/2$, while semi-hereditariness would make G' $\varepsilon/2$-close to Π, contrary to the triangle inequality. Hence the tester for H rejects with probability at least $2/3$. The resulting query complexity is $O\big(M(\varepsilon/2)^2\big) + \widetilde{O}\Big(N(\varepsilon/2)^2 + \frac{(q(\varepsilon/2)+1)^2}{\eta(\varepsilon/2)^2}\Big)$. □

Part II. Bounded-arity relational structures

The graph argument rests only on an ordered-container lemma for independent sets in a dense uniform hypergraph. It therefore does not depend on the ambient object having a single symmetric irreflexive binary relation. We now work over an arbitrary finite relational signature in the sense of Definition 1.8.

3. Finite relational structures and induced sampling

Fix a finite relational signature τ and write $r = r(\tau) = \max_{R \in \tau} a(R)$. A finite τ-structure $\mathcal{A}$ has vertex set $V(\mathcal{A})$ and relations $R^{\mathcal{A}} \subseteq V(\mathcal{A})^{a(R)}$ for $R \in \tau$.

Definition 3.1 (Relational properties)**.** *For* $U \subseteq V(\mathcal{A})$*, the induced substructure* $\mathcal{A}[U]$ *is obtained by restricting every relation to tuples all of whose entries lie in* U*. A* τ*-property is a family of finite* τ*-structures closed under isomorphism. It is hereditary if* $\mathcal{A} \in \mathcal{P}$ *implies* $\mathcal{A}[U] \in \mathcal{P}$ *for every* $U \subseteq V(\mathcal{A})$*, and it is extendable if, for every* $\mathcal{A} \in \mathcal{P}$ *and every* $m \geq 0$*, there is a structure* $\mathcal{B} \in \mathcal{P}$ *on* $|V(\mathcal{A})| + m$ *vertices containing* $\mathcal{A}$ *as an induced substructure.*

Definition 3.2 (Distance between relational structures)**.** *If* $\mathcal{A}$ *and* $\mathcal{B}$ *are* τ*-structures on the same* n*-element vertex set, define*

$$d_\tau(\mathcal{A}, \mathcal{B}) = \frac{1}{|\tau|} \sum_{R \in \tau} \frac{|R^{\mathcal{A}} \triangle R^{\mathcal{B}}|}{n^{a(R)}}. \tag{3.1}$$

We write $d_\tau(\mathcal{A},\mathcal{P})$ *for the minimum of* $d_\tau(\mathcal{A},\mathcal{B})$ *over all* $\mathcal{B}\in\mathcal{P}$ *on the same vertex set, and say that* $\mathcal{A}$ *is* ε*-far from* $\mathcal{P}$ *if this minimum is greater than* ε*; otherwise it is* ε*-close to* $\mathcal{P}$*. For a simple graph encoded as a symmetric irreflexive binary relation, this differs from the normalization in Part I only by an absolute constant.*

The intrinsic testing notion in this setting is induced sampling.

Definition 3.3 (Oblivious induced-sampling tester)**.** *Let* $\mathcal{P}$ *be a* τ*-property. An oblivious induced-sampling tester for* $\mathcal{P}$ *consists, for every* $\varepsilon>0$*, of an integer* $s(\varepsilon)$ *and a collection* $\mathcal{R}_\varepsilon$ *of isomorphism types of* τ*-structures on* $s(\varepsilon)$ *vertices. On an input* $\mathcal{A}$ *with* $n\ge s(\varepsilon)$*, the tester chooses a uniformly random* $s(\varepsilon)$*-subset* $S\subseteq V(\mathcal{A})$*, reveals the complete induced structure* $\mathcal{A}[S]$*, and rejects if and only if* $\mathcal{A}[S]\in\mathcal{R}_\varepsilon$*. When* $n<s(\varepsilon)$*, it may inspect the whole input and decide membership exactly.*

The tester has one-sided error if it always accepts structures in $\mathcal{P}$*, and is sound if it rejects every* ε*-far input with probability at least* $2/3$*. Its sample complexity is* $s(\varepsilon)$*.*

For hereditary properties, the rejection rule may always be taken to be the complement of the property on the sampled set.

Lemma 3.4. *Let* $\mathcal{P}$ *be hereditary and suppose that it has a one-sided induced-sampling tester with sample complexity* $s(\varepsilon)$*. Then the tester which samples the same number of vertices and rejects exactly when* $\mathcal{A}[S]\notin\mathcal{P}$ *is also one-sided and has at least the same rejection probability on every input.*

Proof. Every isomorphism type on which the original tester can reject lies outside $\mathcal{P}$. Indeed, if a structure $\mathcal{F}\in\mathcal{P}$ were a rejection type, then running the tester on $\mathcal{F}$ itself would contradict one-sidedness. Hence replacing the original rejection family by the whole complement of $\mathcal{P}$ can only increase the rejection probability. Heredity ensures that no induced substructure of an input in $\mathcal{P}$ lies outside $\mathcal{P}$. □

We shall also use the dense relation-query model corresponding to Part I. The tester is not given the order of the input; it obtains vertex labels from an independent uniform-sampling oracle, and a relation query specifies a relation symbol together with a tuple of previously sampled labels. Its complexity may depend on the proximity parameter and on the fixed signature, but not on the input order. The canonicalization theorem of Goldreich and Trevisan [GT03] extends without change from graphs to a fixed finite relational signature; we record the precise consequence used below.

Theorem 3.5 (Canonicalization for bounded-arity structures)**.** *Let* τ *be a fixed finite signature. There are constants* $c_\tau, C_\tau>0$ *such that the following holds. If a* τ*-property has a size-oblivious one-sided tester making at most* $Q(\varepsilon)$ *relation queries, then it has a one-sided induced-sampling tester sampling at most* $C_\tau Q(c_\tau\varepsilon)$ *vertices. If the property is hereditary, the induced sampler may reject exactly when the sampled induced structure does not satisfy the property.*

Proof. The random-relabeling proof of Goldreich and Trevisan uses only permutation invariance of the input and the fact that a query transcript mentions a bounded number of vertex labels. Both features remain valid for a fixed relational signature. Indeed, if r is the maximum arity, a transcript of Q relation queries involves at most rQ distinct labels. Conditional on its equality pattern, a uniformly random relabelling sends these labels to a uniformly random ordered tuple of distinct vertices. Hence revealing the complete induced structure on a uniformly random set of $O_\tau(Q)$ vertices contains all relation values needed to reproduce the transcript, including adaptive queries. The treatment of collisions and inputs whose order is comparable with Q, as well as the amplification and constant rescaling of the proximity parameter, is exactly the same as in the graph proof. This gives constants c_τ, C_τ depending only on the signature. Random relabeling preserves distance from an isomorphism-invariant property and preserves one-sidedness. The final assertion follows from Lemma 3.4. □

Revealing the complete induced structure on t vertices uses at most

$$\sum_{R\in\tau} t^{a(R)} \le |\tau| t^r \tag{3.2}$$

relation queries. Thus sample complexity is the natural arity-independent parameter, while relation-query complexity is recovered by taking the r-th power, up to a constant depending on the signature.

4. The relational characterization

We first isolate the pure hypergraph statement already proved inside Theorem 2.2. If H is a hypergraph, let $\mathcal{I}(H)$ denote its family of independent sets.

Lemma 4.1 (Dense ordered containers)**.** *Let H be an s-uniform hypergraph on an n-element vertex set V, where $1 \le s \le n$, and suppose that $e(H) \ge \delta\binom{n}{s}$. Then there are a family $\mathcal{F}$ of ordered tuples of distinct vertices, each of length at most $s-1$, and maps $f : \mathcal{I}(H) \to \mathcal{F}, C : \mathcal{F} \to \binom{V}{\le(1-\delta/s)n}$, such that every entry of $f(I)$ belongs to I and $I \subseteq C(f(I))$ for every $I \in \mathcal{I}(H)$.*

Proof. This is precisely the ordered pivot argument in the first implication of Theorem 2.2. That proof begins with an arbitrary dense s-uniform hypergraph, chooses successive maximum-degree pivots from a fixed independent set, passes to links, and obtains a container omitting at least $\delta n/s$ vertices. No property of the host graph is used at any point; the graph enters there only to define the obstruction hypergraph. Therefore the same construction, with H as the initial hypergraph, gives the asserted maps. $\square$

The reverse implication is also purely hypergraph-theoretic.

Lemma 4.2 (Counting from ordered containers)**.** *Suppose that an n-vertex hypergraph H admits maps $f : \mathcal{I}(H) \to \bigcup_{j=0}^{q} V^j_{\neq}, C : \operatorname{im}(f) \to \binom{V}{\le(1-\eta)n}$, such that every entry of $f(I)$ belongs to I and $I \subseteq C(f(I))$. Then, for every $q \le t \le n$,*

$$\frac{|\mathcal{I}(H) \cap \binom{V}{t}|}{\binom{n}{t}} \le \sum_{j=0}^{q} (t)_j \frac{\binom{\lfloor(1-\eta)n\rfloor-j}{t-j}}{\binom{n-j}{t-j}} \tag{4.1}$$

$$\le \sum_{j=0}^{q} t^j (1-\eta)^{t-j}. \tag{4.2}$$

Consequently, there is an absolute constant $C > 0$ such that the right-hand side of (4.2) is at most $1/3$ whenever

$$t \ge \frac{C(q+1)}{\eta} \log\left(\frac{q+1}{\eta} + 2\right). \tag{4.3}$$

Proof. Choose $T \in \binom{V}{t}$ uniformly. If T is independent and its fingerprint has length j, then $f(T) \subseteq T \subseteq C(f(T))$. For a fixed ordered tuple F of length j, $\Pr[F \subseteq T \subseteq C(F)] = \frac{\binom{|C(F)|-j}{t-j}}{\binom{n}{t}}$. There are at most $(n)_j$ possible ordered tuples of length j. Summing over j gives (4.1), using $\frac{(n)_j}{\binom{n}{t}} = \frac{(t)_j}{\binom{n-j}{t-j}}$. Moreover, $\frac{\binom{\lfloor(1-\eta)n\rfloor-j}{t-j}}{\binom{n-j}{t-j}} \le (1-\eta)^{t-j}, (t)_j \le t^j$, which proves (4.2). Finally, using $1-\eta \le e^{-\eta}$, each summand is at most $\exp(j\log t - \eta(t-j))$. If t satisfies (4.3) with a sufficiently large absolute constant, then the sum of the $q+1$ terms is at most $1/3$. $\square$

We now formulate the relational version of the ordered-container condition.

Definition 4.3 (Ordered container lemma for a hereditary relational property). *Let $\mathcal{P}$ be a hereditary τ-property. For $\varepsilon, \eta > 0$ and $N, q \in \mathbb{N}$, we say that $\mathcal{P}$ admits an $(\varepsilon, \eta, N, q)$-ordered container lemma if, for every $n \geq N$ and every n-vertex τ-structure $\mathcal{A}$ that is ε-far from $\mathcal{P}$, there are maps $f : \{U \subseteq V(\mathcal{A}) : \mathcal{A}[U] \in \mathcal{P}\} \longrightarrow \bigcup_{j=0}^{q} V(\mathcal{A})^{j}_{\neq}$ and $C : \mathrm{im}(f) \longrightarrow \binom{V(\mathcal{A})}{\leq (1-\eta)n}$ such that every entry of $f(U)$ belongs to U and $U \subseteq C(f(U))$.*

Theorem 4.4 (Container characterization for hereditary relational structures). *Let $\mathcal{P}$ be a hereditary property of finite τ-structures.*

(1) If $\mathcal{P}$ has a one-sided induced-sampling tester with sample complexity $s(\varepsilon)$, then, for every $\varepsilon > 0$, it admits an ordered container lemma with $N(\varepsilon) = s(\varepsilon)$, $q(\varepsilon) = s(\varepsilon) - 1$, and $\eta(\varepsilon) = 2/(3s(\varepsilon))$.

(2) Conversely, suppose that $\mathcal{P}$ admits ordered container lemmas with parameter functions η, N, and q. Then it has a one-sided induced-sampling tester with sample complexity at most

$$T_{\mathcal{P}}(\varepsilon) = C \max\left\{ N(\varepsilon), \frac{q(\varepsilon)+1}{\eta(\varepsilon)} \log\left(\frac{q(\varepsilon)+1}{\eta(\varepsilon)} + 2\right)\right\}, \tag{4.4}$$

where C is an absolute constant.

Proof. Fix $\varepsilon > 0$ and write $s = s(\varepsilon)$. Let $\mathcal{A}$ be an n-vertex structure, where $n \geq s$, that is ε-far from $\mathcal{P}$. By Lemma 3.4, at least a 2/3-fraction of the s-subsets $S \subseteq V(\mathcal{A})$ satisfy $\mathcal{A}[S] \notin \mathcal{P}$. Define an s-uniform obstruction hypergraph $H_{\mathcal{A}}$ on $V(\mathcal{A})$ by $E(H_{\mathcal{A}}) = \{S \in \binom{V(\mathcal{A})}{s} : \mathcal{A}[S] \notin \mathcal{P}\}$. Every $U \subseteq V(\mathcal{A})$ with $\mathcal{A}[U] \in \mathcal{P}$ is independent in $H_{\mathcal{A}}$, by heredity. Since $e(H_{\mathcal{A}}) \geq (2/3)\binom{n}{s}$, Lemma 4.1 gives fingerprints of length at most $s-1$ and containers of size at most $(1 - 2/(3s))n$. This proves the first assertion.

Conversely, fix $\varepsilon > 0$ and choose $T = T_{\mathcal{P}}(\varepsilon)$ with the constant large enough that Lemma 4.2 applies. If the input has fewer than T vertices, inspect it completely. Otherwise, choose a uniformly random T-subset S, reveal $\mathcal{A}[S]$, and reject exactly when $\mathcal{A}[S] \notin \mathcal{P}$. Heredity gives one-sidedness. If $\mathcal{A}$ is ε-far and has at least $T \geq N(\varepsilon)$ vertices, apply the assumed ordered container lemma to the family of sets U with $\mathcal{A}[U] \in \mathcal{P}$. Lemma 4.2 then shows that $\Pr[\mathcal{A}[S] \in \mathcal{P}] \leq \frac{1}{3}$. Thus the tester rejects every ε-far input with probability at least 2/3. ☐

Combining this theorem with canonicalization gives the query-complexity form.

Corollary 4.5 (Relation-query formulation). *Let τ have maximum arity r, and let $\mathcal{P}$ be hereditary.*

(1) If $\mathcal{P}$ has a size-oblivious one-sided tester making at most $Q(\varepsilon)$ relation queries, then, for constants $c_\tau, C_\tau > 0$, it admits ordered-container parameters satisfying $N(\varepsilon), q(\varepsilon) \leq C_\tau Q(c_\tau \varepsilon)$, $\eta(\varepsilon) \geq \frac{1}{C_\tau Q(c_\tau \varepsilon)}$.

(2) Conversely, ordered-container parameters (η, N, q) give a size-oblivious one-sided tester with relation-query complexity

$$O_\tau\left(\max\left\{N(\varepsilon)^r, \left(\frac{q(\varepsilon)+1}{\eta(\varepsilon)} \log\left(\frac{q(\varepsilon)+1}{\eta(\varepsilon)} + 2\right)\right)^r\right\}\right).$$

Proof. The first implication follows from Theorem 3.5 and Theorem 4.4(1). For the reverse implication, use the induced-sampling tester from Theorem 4.4(2) and reveal every relation on the sampled set. The query bound follows from (3.2). ☐

In particular, for every fixed bounded-arity signature, polynomial one-sided induced-sampling complexity, polynomial relation-query complexity, and polynomial ordered-container parameters are equivalent.

4.1. **Arbitrary properties and semi-hereditary envelopes.** The passage from hereditary to arbitrary properties is also independent of the signature.

Definition 4.6 (Semi-hereditary relational property)**.** *A τ-property $\mathcal{P}$ is semi-hereditary if there are a hereditary property $\mathcal{H} \supseteq \mathcal{P}$ and a function $M : (0,1) \to \mathbb{N}$ such that, for every $\varepsilon > 0$, every $\mathcal{A} \in \mathcal{H}$ with $|V(\mathcal{A})| \geq M(\varepsilon)$ is ε-close to $\mathcal{P}$. We call $(\mathcal{H}, M)$ a semi-hereditary witness.*

Theorem 4.7 (Container characterization for arbitrary relational properties)**.** *Let $\mathcal{P}$ be a property of finite τ-structures.*

(1) Suppose that $\mathcal{P}$ has an oblivious one-sided induced-sampling tester with sample complexity $s(\varepsilon)$. Then $\mathcal{P}$ has a semi-hereditary witness $(\mathcal{H}, M)$ satisfying $M(\varepsilon) \leq s(\varepsilon)$. The same tester tests $\mathcal{H}$, and $\mathcal{H}$ admits ordered container parameters $N(\varepsilon) = s(\varepsilon)$, $q(\varepsilon) = s(\varepsilon) - 1$, and $\eta(\varepsilon) = 2/(3s(\varepsilon))$.

(2) Conversely, suppose that $\mathcal{P}$ has a semi-hereditary witness $(\mathcal{H}, M)$ and that $\mathcal{H}$ admits ordered container parameters (η, N, q). Then $\mathcal{P}$ has an oblivious one-sided induced-sampling tester with sample complexity

$$O\left(\max\left\{M(\varepsilon/2), N(\varepsilon/2), \frac{q(\varepsilon/2)+1}{\eta(\varepsilon/2)} \log\left(\frac{q(\varepsilon/2)+1}{\eta(\varepsilon/2)} + 2\right)\right\}\right). \tag{4.5}$$

Proof. Fix the given one-sided induced-sampling tester. Let $\mathcal{B}$ be the family of all finite τ-structures that can occur as a rejection pattern for this tester at some proximity parameter, and let $\mathcal{H}$ be the property of containing no member of $\mathcal{B}$ as an induced substructure. Then $\mathcal{H}$ is hereditary. One-sidedness implies $\mathcal{P} \subseteq \mathcal{H}$: otherwise, an input in $\mathcal{P}$ containing a rejection pattern would be rejected with positive probability.

For each fixed ε, the original tester is one-sided for $\mathcal{H}$, since no induced substructure of an input in $\mathcal{H}$ is a rejection pattern. If an input is ε-far from $\mathcal{H}$, then it is also ε-far from $\mathcal{P}$, because $\mathcal{P} \subseteq \mathcal{H}$, and hence the tester rejects with probability at least $2/3$. Thus the same tester tests $\mathcal{H}$.

We next verify semi-hereditariness. Suppose that $\mathcal{A} \in \mathcal{H}$, $|V(\mathcal{A})| \geq s(\varepsilon)$, and $\mathcal{A}$ is ε-far from $\mathcal{P}$. The tester must reject with positive probability, so some sampled $s(\varepsilon)$-vertex induced substructure belongs to $\mathcal{B}$, contradicting $\mathcal{A} \in \mathcal{H}$. Hence one may take $M(\varepsilon) = s(\varepsilon)$. Applying Theorem 4.4(1) to $\mathcal{H}$ gives the container parameters in the first assertion.

Conversely, fix $\varepsilon > 0$. If the input has fewer than $M(\varepsilon/2)$ vertices, inspect it completely. Otherwise run the tester for $\mathcal{H}$ supplied by Theorem 4.4(2), at proximity parameter $\varepsilon/2$. Inputs in $\mathcal{P}$ are accepted because $\mathcal{P} \subseteq \mathcal{H}$. If $\mathcal{A}$ is ε-far from $\mathcal{P}$ and has at least $M(\varepsilon/2)$ vertices, then it is $\varepsilon/2$-far from $\mathcal{H}$. Indeed, otherwise there would be a structure $\mathcal{A}' \in \mathcal{H}$ on the same vertex set with $d_\tau(\mathcal{A}, \mathcal{A}') \leq \varepsilon/2$; semi-hereditariness would make $\mathcal{A}'$ $\varepsilon/2$-close to $\mathcal{P}$, and the triangle inequality would contradict the assumption on $\mathcal{A}$. Therefore the tester for $\mathcal{H}$ rejects with probability at least $2/3$, and the sample bound is (4.5). □

Together with Theorem 3.5 and (3.2), this gives a query-complexity characterization of arbitrary bounded-arity properties. The quantitative information is carried jointly by the ordered-container parameters of the hereditary envelope and by the threshold M, which measures the difference between the original property and that envelope.

Part III. Applications

The preceding parts identify ordered containers as the common quantitative structure behind one-sided testability. We now use that structure in three directions. The first two applications depend only on the binary nature of the ambient relations, which makes the edit cost of changing a small exceptional vertex set quadratic. The third applies directly to every graph property covered by the characterization.

Remark 4.8. One may ask whether the applications below can be deduced directly from the container characterization. This is possible, but it is convenient to use a slightly stronger version of the characterization in which, in addition to containing every induced Π-subgraph, each container itself induces a graph that is close to Π. Such containers can be obtained by iterating the container construction until the auxiliary hypergraph of local obstructions becomes sparse. The resulting statement has somewhat weaker quantitative bounds, since the longer procedure requires larger fingerprints, but it makes the applications below essentially formal.

5. Partition properties

Given properties $\mathcal{P}_1, \dots, \mathcal{P}_k$ of structures in a common binary signature, we write $\mathrm{Part}(\mathcal{P}_1, \dots, \mathcal{P}_k)$ for the property of admitting a vertex partition whose ith part induces a member of $\mathcal{P}_i$. The next theorem gives the signature-independent closure principle.

5.1. **Binary relational structures.** For τ-properties $\mathcal{P}_1, \dots, \mathcal{P}_k$, let $\mathrm{Part}(\mathcal{P}_1, \dots, \mathcal{P}_k)$ consist of the structures $\mathcal{A}$ admitting a partition $V(\mathcal{A}) = V_1 \cup \cdots \cup V_k$ with $\mathcal{A}[V_i] \in \mathcal{P}_i$ for every i. Relation tuples meeting two different parts are unrestricted.

Theorem 5.1 (Partition closure for binary signatures)**.** *Let every symbol of τ have arity two, and let $\mathcal{P}_1, \dots, \mathcal{P}_k$ be nonempty hereditary and extendable τ-properties. Suppose that $\mathcal{P}_i$ has one-sided induced-sampling complexity $S_i(\delta)$, and put $S(\delta) = \max_i S_i(\delta)$. Then* $\mathrm{Part}(\mathcal{P}_1, \dots, \mathcal{P}_k)$ *has a one-sided induced-sampling tester with sample complexity $\widetilde{O}\Big(\frac{k}{\varepsilon} S\big(\frac{\varepsilon}{8k}\big)^2\Big)$. Consequently, its relation-query complexity is the square of this quantity, up to a constant depending on τ.*

Proof. Set $\gamma = \frac{\varepsilon}{8k}$, $\alpha = \sqrt{\frac{\varepsilon}{8k}}$, $\lambda = \frac{\varepsilon}{12}$, and $s = S(\gamma)$.

By increasing s if necessary, we may assume that, for every i, every structure that is γ-far from $\mathcal{P}_i$ has at least a $2/3$-fraction of bad induced s-substructures whenever its order is at least s. Indeed, if a larger sampled structure belonged to $\mathcal{P}_i$, then all of its induced substructures of the original sample size would belong to $\mathcal{P}_i$ by heredity. Inputs below any fixed threshold depending on k, ε, and s will be inspected completely, so throughout the structural argument we assume $n \geq s/\alpha$.

Let $\mathcal{A}$ be ε-far from $\mathrm{Part}(\mathcal{P}_1, \dots, \mathcal{P}_k)$. We first claim that whenever $U_1, \dots, U_k \subseteq V(\mathcal{A})$ satisfy $|\bigcup_i U_i| > (1-\lambda)n$, some i satisfies

$$|U_i| \geq \alpha n \qquad \text{and} \qquad \mathcal{A}[U_i] \text{ is } \gamma\text{-far from } \mathcal{P}_i. \tag{5.1.1}$$

Suppose not. Choose disjoint sets $V_i \subseteq U_i$ that partition $U = \bigcup_i U_i$. If $|U_i| \geq \alpha n$, edit $\mathcal{A}[U_i]$ into a member of $\mathcal{P}_i$ and restrict the resulting structure to V_i; heredity shows that the restriction still belongs to $\mathcal{P}_i$, and the total normalized cost over all such indices is at most $k\gamma = \varepsilon/8$. If $|U_i| < \alpha n$, replace the complete structure on V_i by a member of $\mathcal{P}_i$ of the same order. Such a member exists because $\mathcal{P}_i$ is nonempty, hereditary, and extendable, and the total cost is at most $k\alpha^2 = \varepsilon/8$. Finally, assign the leftover set $L = V(\mathcal{A}) \setminus U$ to one part and use extendability. Only relation tuples meeting L need be changed, at normalized cost at most $2|L|/n + (|L|/n)^2 < 3\lambda = \varepsilon/4$. The total cost is less than ε, a contradiction. This proves the claim.

For each i, let H_i be the s-uniform hypergraph on $V(\mathcal{A})$ whose edges are the s-sets inducing structures outside $\mathcal{P}_i$. Fix an arbitrary total order of the vertex set and use the canonical maps furnished by the proof of Lemma 4.1. Let W satisfy the partition property and fix a witnessing partition $W = W_1 \cup \cdots \cup W_k$. Initialize $C_i^{(0)} = V(\mathcal{A})$ for every i. While the union of the current containers has size greater than $(1-\lambda)n$, the claim gives an index i for which $|C_i| \geq \alpha n$ and $\mathcal{A}[C_i]$ is γ-far from $\mathcal{P}_i$. Choose the least such index. Then $H_i[C_i]$ has density at least $2/3$, and W_i is an independent set in this hypergraph. Applying Lemma 4.1 inside C_i replaces C_i by a set containing W_i and of size at most $(1 - 2/(3s))|C_i|$, while recording an ordered block of at most $s-1$ vertices of W_i.

An index can be selected at most $R_0 = 1 + \left\lceil \frac{\log(1/\alpha)}{-\log(1-2/(3s))} \right\rceil = O\left(s \log \frac{1}{\alpha}\right)$ times. Thus the procedure has at most $R = kR_0$ rounds and records at most $q = R(s-1) = O\left(ks^2 \log \frac{1}{\alpha}\right)$ vertex occurrences. Its terminal union C contains W and satisfies $|C| \le (1-\lambda)n$.

The complete run is determined by its at most q recorded vertex occurrences, the round boundaries, and the selected indices. After identifying repeated vertices, the number of possible auxiliary transcripts is $\exp(\widetilde{O}(q))$. Summing over these transcripts and over the ordered fingerprints gives, exactly as in the standard container-counting argument, that a uniformly random t-set satisfies the partition property with probability at most $1/3$ for $t = \widetilde{O}\left(\frac{q+1}{\lambda}\right) = \widetilde{O}\left(\frac{k}{\varepsilon} s^2\right)$. We take t also larger than the fixed small-input threshold; this does not change the displayed bound. The tester samples t vertices, inspects the whole induced structure, and rejects exactly when it fails the partition property. It is one-sided, and the preceding bound gives soundness. The relation-query bound follows from (3.2) with $r = 2$. ☐

5.2. **Graphs and generalized colouring parameters.** Specializing the preceding theorem to the signature of simple graphs and using canonicalization gives the following query-complexity formulation.

Theorem 5.2 (Testing partitions of hereditary graph properties)**.** *Let $k \ge 1$, and let $\Pi_1, \dots, \Pi_k$ be hereditary and extendable graph properties. Suppose that Π_i admits a size-oblivious one-sided error tester of query complexity at most $Q_i(\varepsilon)$, and let $Q(\varepsilon) = \max_{i \in [k]} Q_i(\varepsilon)$. Then* $\mathrm{Part}(\Pi_1, \dots, \Pi_k)$ *admits a size-oblivious one-sided error tester with query complexity $\widetilde{O}\left(\frac{k^2}{\varepsilon^2} Q^4\left(\frac{\varepsilon}{Ck}\right)\right)$, where $C > 0$ is an absolute constant.*

Proof. By canonicalization, each Π_i has one-sided induced-sampling complexity $S_i(\delta) = O(Q_i(c\delta))$ for an absolute constant $c > 0$. Apply Theorem 5.1 to the symmetric irreflexive binary signature. The resulting tester samples $\widetilde{O}\left(\frac{k}{\varepsilon} Q^2\left(\frac{\varepsilon}{Ck}\right)\right)$ vertices and queries every pair among them. Squaring the sample bound proves the claim. ☐

For integers $r, s \ge 0$ with $r + s \ge 1$, a graph is (r, s)-*colourable* if its vertex set can be partitioned into r independent sets and s cliques; empty parts are allowed. We denote this property by $\Pi_{r,s}$. Thus $\Pi_{k,0}$ is ordinary k-colourability. A *split graph* is a $(1, 1)$-colourable graph, equivalently a graph whose vertex set is the union of one independent set and one clique. The *cochromatic number* $z(G)$ is the least k for which $V(G)$ can be partitioned into k parts, each inducing either an independent set or a clique. Equivalently, $z(G) = \min\{r + s : G \in \Pi_{r,s}\}$.

Corollary 5.3 ((r, s)-colourability)**.** *For fixed $r, s \ge 0$ with $r + s \ge 1$, the property $\Pi_{r,s}$ admits a size-oblivious one-sided error tester with query complexity $\widetilde{O}\left((r+s)^6/\varepsilon^6\right)$.*

Proof. The properties of being edgeless and complete are hereditary and extendable, and each has a one-sided tester using $O(1/\varepsilon)$ queries. Apply Theorem 5.2 with $k = r + s$. ☐

Two familiar cases are worth recording separately.

Corollary 5.4. *For every $k \ge 1$, k-colourability admits a size-oblivious one-sided error tester with query complexity $\widetilde{O}(k^6/\varepsilon^6)$.*

Corollary 5.5. *The property of being a split graph admits a size-oblivious one-sided error tester with query complexity $\widetilde{O}(1/\varepsilon^6)$.*

For a digraph D, the *dichromatic number* $\vec{\chi}(D)$ is the least integer k for which $V(D)$ can be partitioned into k sets, each inducing an acyclic digraph.

Corollary 5.6 (Dichromatic number). *For every $k \geq 1$, the property $\vec{\chi}(D) \leq k$ admits a one-sided induced-sampling tester with sample complexity $\widetilde{O}\left(\frac{k^3}{\varepsilon^3}\right)$. Consequently, it admits a size-oblivious one-sided-error tester with relation-query complexity $\widetilde{O}\left(\frac{k^6}{\varepsilon^6}\right)$.*

Proof. Let $\mathcal{P}_{\text{acyc}}$ be the property of being acyclic. A digraph has dichromatic number at most k exactly when it belongs to $\text{Part}(\mathcal{P}_{\text{acyc}}, \dots, \mathcal{P}_{\text{acyc}})$ with k factors. Bender and Ron [BR02] prove that a uniformly random set of $\widetilde{O}(1/\delta)$ vertices detects a directed cycle with probability at least $2/3$ in every digraph that is δ-far from acyclic. Thus $S_{\text{acyc}}(\delta) = \widetilde{O}(1/\delta)$. Applying Theorem 5.1 gives sample complexity

$$\widetilde{O}\left(\frac{k}{\varepsilon} S_{\text{acyc}}\left(\frac{\varepsilon}{8k}\right)^2\right) = \widetilde{O}\left(\frac{k^3}{\varepsilon^3}\right).$$

Revealing all ordered pairs in the sample gives the stated relation-query bound. □

Remark 5.7. The sample bound in the Bender–Ron tester is important here. Their adjacency-matrix query bound is $\widetilde{O}(\delta^{-2})$; applying the graph query-complexity closure theorem directly to that estimate would give only $\widetilde{O}(k^{10}\varepsilon^{-10})$.

Corollary 5.8 (Cochromatic number). *For every $k \geq 1$, the property of having cochromatic number at most k admits a size-oblivious one-sided error tester with query complexity $\widetilde{O}\left(k^7/\varepsilon^6\right)$.*

Proof. Because empty parts are allowed, $z(G) \leq k$ if and only if $G \in \bigcup_{r=0}^{k} \Pi_{r,k-r}$. Run independent amplified one-sided testers for these $k+1$ properties and reject precisely when all of them reject. Completeness is one-sided. If the input is ε-far from the union, then it is ε-far from every constituent property; amplifying each tester to error probability at most $1/(3(k+1))$ and applying the union bound gives soundness at least $2/3$. The factor $k+1$ from running all the testers gives the stated bound, with amplification absorbed by the polylogarithmic factor. □

6. Linearly large induced substructures

The next application concerns properties asserting that there is a linearly large induced substructure satisfying $\mathcal{P}$. Such properties are generally not hereditary and need not be expressible through a bounded vertex partition. Nevertheless, iterating the ordered-container construction forces every induced member of $\mathcal{P}$ into a container whose density lies below the target density.

6.1. A general large-induced-substructure theorem.

For $\rho \in (0,1]$ and a property $\mathcal{P}$, write $\text{Large}_\rho(\mathcal{P}) = \{\mathcal{A} : \text{there is } U \subseteq V(\mathcal{A}),\ |U| \geq \rho|V(\mathcal{A})|,\ \mathcal{A}[U] \in \mathcal{P}\}$.

Theorem 6.1 (Testing for a large induced member). *Let τ be a finite binary signature, let $\mathcal{P}$ be a nonempty hereditary and extendable τ-property, and let $S_{\mathcal{P}}(\delta)$ be its one-sided induced-sampling complexity. Fix $\rho \in (0,1]$ and $0 < \varepsilon \leq \rho^2/8$, and let $\delta = \varepsilon/4\rho^2$, $\gamma = \varepsilon/16\rho$, and $s = S_{\mathcal{P}}(\delta)$. Then $\text{Large}_\rho(\mathcal{P})$ has a two-sided induced-sampling tester using $\widetilde{O}\left(\frac{\rho^2}{\varepsilon^2} s^2 \log \frac{2}{\rho}\right)$ vertices. Its relation-query complexity is the square of this quantity, up to a constant depending on τ.*

Proof. Put $n = |V(\mathcal{A})|$. We first establish the local-farness statement used by the container iteration. There is a threshold $n_0 = O_\tau\left(\frac{s}{\rho} + \frac{1}{\varepsilon} + 1\right)$, such that if $n \geq n_0$ and $\mathcal{A}$ is ε-far from $\text{Large}_\rho(\mathcal{P})$, then every $U \subseteq V(\mathcal{A})$ with $|U| \geq (\rho - \gamma)n$ is δ-far from $\mathcal{P}$.

Indeed, suppose that $\mathcal{A}[U]$ can be changed into a structure $\mathcal{B} \in \mathcal{P}$ at cost at most δ relative to $|U|^2$. If $|U| \geq \rho n$, choose a uniformly random set $W \subseteq U$ of size $\lceil \rho n \rceil$. Restricting the edit set to W and averaging shows that some such W requires at most $\delta\rho^2 n^2 + O_\tau(n)$ relation changes; the $O_\tau(n)$ term accounts for tuples with repeated coordinates. Since $\mathcal{B}[W] \in \mathcal{P}$, this makes $\mathcal{A}$ closer than ε to

$\mathrm{Large}_\rho(\mathcal{P})$ for sufficiently large n. If instead $|U| < \rho n$, extend $\mathcal{B}$ to a member of $\mathcal{P}$ on $\lceil \rho n \rceil$ vertices. Besides the edits inside U, only binary relation tuples meeting at most $\gamma n + 1$ new vertices need be changed, at normalized cost at most $3\gamma\rho + o(1)$. Since $\delta\rho^2 + 3\gamma\rho = \frac{\varepsilon}{4} + \frac{3\varepsilon}{16} < \varepsilon$, we again obtain a contradiction. Increase n_0 also so that $(\rho - \gamma)n_0 \geq s$.

Let H be the s-uniform obstruction hypergraph on $V(\mathcal{A})$, $E(H) = \{S \in \binom{V(\mathcal{A})}{s} : \mathcal{A}[S] \notin \mathcal{P}\}$. For every $C \subseteq V(\mathcal{A})$ with $|C| > (\rho - \gamma)n$, the local-farness statement and the definition of s imply that $H[C]$ has density at least $2/3$. Fix a total order of the vertices and use the canonical maps from Lemma 4.1. Given a set I with $\mathcal{A}[I] \in \mathcal{P}$, start with $C_0 = V(\mathcal{A})$ and, while $|C_a| > (\rho - \gamma)n$, apply the dense ordered-container lemma to the independent set $I \subseteq C_a$ in $H[C_a]$. Each round records at most $s - 1$ vertices of I and replaces the current container by one of size at most $(1 - 2/(3s))|C_a|$.

The number of rounds is at most $R = 1 + \left\lceil \frac{\log(1/(\rho-\gamma))}{-\log(1-2/(3s))} \right\rceil = O\left(s \log \frac{2}{\rho}\right)$, where we used $\gamma \leq \rho/128$. Thus the concatenated transcript has length at most $q = R(s-1) = O\left(s^2 \log \frac{2}{\rho}\right)$. The complete run is determined by its at most q recorded vertex occurrences and its round boundaries. Identifying repeated vertices leaves $\exp(\widetilde{O}(q))$ possible auxiliary transcripts, each of which, together with an ordered fingerprint of length at most q, determines a container of size at most $(\rho - \gamma)n$. A union bound over these choices, followed by the hypergeometric tail bound, shows that it is enough to take $t = \widetilde{O}\left(\frac{q+1}{\gamma^2}\right) = \widetilde{O}\left(\frac{\rho^2}{\varepsilon^2} s^2 \log \frac{2}{\rho}\right)$. Choose t also larger than n_0. The tester samples a uniformly random t-set T and accepts precisely when there is $I \subseteq T$ such that $|I| \geq (\rho - \gamma/2)t \quad$ and $\quad \mathcal{A}[I] \in \mathcal{P}$. If the input has fewer than t vertices, it is inspected completely.

If $\mathcal{A}$ contains an induced member of $\mathcal{P}$ on at least ρn vertices, Hoeffding's inequality for the hypergeometric distribution gives $\Pr\left[|T \cap I| < (\rho - \gamma/2)t\right] \leq e^{-\gamma^2 t/2} < \frac{1}{3}$ for a sufficiently large implicit constant; heredity then gives completeness. If $\mathcal{A}$ is ε-far, the preceding union bound shows that the acceptance probability is at most $1/3$. This proves the sample bound. The relation-query bound follows from (3.2) with $r = 2$. □

6.2. **Large acyclic subdigraphs.** For $\rho \in (0, 1]$, we call a digraph D on n vertices a ρ-DAG if it contains an induced acyclic subdigraph on at least ρn vertices. Since acyclicity is monotone under deletion of arcs, D is ε-far from this property exactly when, for every $S \subseteq V(D)$ with $|S| = \lceil \rho n \rceil$, more than εn^2 arcs must be deleted from $D[S]$ to make it acyclic.

Theorem 6.2. *Let $\rho \in (0, 1]$. The property of being a ρ-DAG admits a size-oblivious two-sided error tester with query complexity $\widetilde{O}\left(\rho^{12}\varepsilon^{-8}\right)$ whenever $0 < \varepsilon \leq \rho^2/8$. For arbitrary $\varepsilon > 0$, the query complexity is $\widetilde{O}\left(\rho^{12} \min\{\varepsilon, \rho^2\}^{-8}\right)$.*

Proof. Apply Theorem 6.1 to the binary directed signature and to the hereditary, extendable property of acyclicity. Bender and Ron [BR02] show that a uniformly random set of $O(\delta^{-1} \log(1/\delta))$ vertices detects a directed cycle with probability at least $2/3$ in every digraph that is δ-far from acyclic. Thus $S_{\mathrm{acyc}}(\delta) = \widetilde{O}(1/\delta)$. With $\delta = \varepsilon/(4\rho^2)$, Theorem 6.1 gives sample complexity $\widetilde{O}(\rho^6\varepsilon^{-4})$. Revealing all ordered pairs in the sample gives the asserted query bound. The second statement follows by monotonicity in ε. □

Corollary 6.3 (Polynomial preservation)**.** *Let τ be a fixed binary signature and let $\mathcal{P}$ be hereditary and extendable. If $\mathcal{P}$ has polynomial one-sided induced-sampling complexity, then* $\mathrm{Large}_\rho(\mathcal{P})$ *has polynomial two-sided induced-sampling complexity for every fixed $\rho > 0$.*

Proof. This is immediate from Theorem 6.1. □

7. Counting and structure inside far hosts

The characterization contains information beyond testing. A short fingerprint and a smaller container give an immediate entropy deficit for the family of all induced substructures satisfying the property.

Corollary 7.1 (Counting induced members)**.** *Let Π be a hereditary graph property, and suppose that for an n-vertex graph G every set $U \subseteq V(G)$ with $G[U] \in \Pi$ has an ordered fingerprint of length at most q whose associated container has size at most $(1-\eta)n$. Then $\left|\{U \subseteq V(G) : G[U] \in \Pi\}\right| \leq (q+1)n^q 2^{(1-\eta)n}$. Moreover, for every $t \geq q$, $\frac{|\{U \in \binom{V(G)}{t} : G[U] \in \Pi\}|}{\binom{n}{t}} \leq \sum_{j=0}^{q} t^j (1-\eta)^{t-j}$.*

Proof. There are at most $\sum_{j=0}^{q} n^j \leq (q+1)n^q$ ordered fingerprints. Every induced member assigned to a fixed fingerprint is a subset of its container, which has at most $(1-\eta)n$ vertices. This proves the first bound. The second is the counting argument of Lemma 4.2, specialized to the obstruction hypergraph associated with G and Π. ☐

7.1. **Random hosts and edit distance.** The study of hereditary graph properties has largely focused on the structure and typical form of graphs belonging to the property; see, for example, [Alo+11] and the references therein. Here the host graph lies far outside the property, and the question is how the induced members of the property can be distributed inside it.

For a graph property Π, let $d_\Pi(G) = \min\{|E(G) \triangle E(G')| : G' \in \Pi, V(G') = V(G)\}$ and let $\mathrm{ed}(n, \Pi) = \max_{|V(G)|=n} d_\Pi(G)$. Alon and Stav [AS08c] proved that for every hereditary property Π there is a density $p = p(\Pi)$ such that, with high probability, $\mathrm{ed}(n, \Pi) = d_\Pi(G(n,p)) + o(n^2)$. Their subsequent work [AS08b] determines the extremal density and the limiting distance for several natural classes and, in particular, determines the distance of $G(n, 1/2)$ from every hereditary property.

Our characterization refines this edit-distance statement at the level of all induced members of Π. Fix $\varepsilon > 0$. If Π has a one-sided tester of query complexity Q, then every sufficiently large graph that is ε-far from Π has ordered-container parameters $q = O(Q(c\varepsilon))$ and $\eta = \Omega(Q(c\varepsilon)^{-1})$. Hence every induced Π-subgraph lies in a set of size at most $(1-\eta)n$, and that set is determined by an ordered tuple of at most q vertices. In particular, the possible locations of all induced members are controlled by only polynomially many containers. The following random-host consequence isolates precisely what is needed later.

Corollary 7.2. *Let Π be a hereditary graph property, and suppose that there exist $p \in [0,1]$ and $\varepsilon_0 > 0$ such that $G(n,p)$ is ε_0-far from Π with probability that tends to 1 as $n \to \infty$. Then there exist constants $\eta_0 > 0$ and $q_0 \in \mathbb{N}$, such that, with probability that tends to 1 as $n \to \infty$, every set $U \subseteq V(G(n,p))$ with $G(n,p)[U] \in \Pi$ is contained in one of at most n^{q_0} vertex sets of size at most $(1-\eta_0)n$.*

Proof. By the hereditary testing theorem of Alon and Shapira, Π has a size-oblivious one-sided tester. Apply Theorem 1.4 at the fixed proximity parameter ε_0. This gives constants $\eta > 0$, $q \in \mathbb{N}$, and $N \in \mathbb{N}$ such that every n-vertex graph with $n \geq N$ that is ε_0-far from Π has an ordered-container lemma with fingerprint length at most q and container size at most $(1-\eta)n$.

With high probability, the realization of $G(n,p)$ is ε_0-far from Π, so the preceding conclusion applies. There are at most $\sum_{j=0}^{q} (n)_j \leq (q+1)n^q$ possible ordered fingerprints. Set $q_0 = q+1$. After increasing the fixed threshold N if necessary, $(q+1)n^q \leq n^{q_0}$. Taking $\eta_0 = \eta$, the corresponding containers have the required size and cover every set inducing a member of Π. ☐

To obtain a concrete example, we use the exact asymptotic distance of $G(n, 1/2)$ from a hereditary property established by Alon and Stav [AS08b]. Define the *binary chromatic number* of Π by $\chi_B(\Pi) = 1 + \max\{r + s : \Pi_{r,s} \subseteq \Pi\}$, where $\Pi_{r,s}$ denotes the property of being (r,s)-colorable.

Theorem 7.3 (Alon–Stav [AS08b])**.** *Let Π be a hereditary graph property for which $2 \leq \chi_B(\Pi) < \infty$. Then, with high probability, $d_\Pi(G(n,1/2)) = \left(\frac{1}{2(\chi_B(\Pi)-1)} \pm o(1)\right)\binom{n}{2}$.*

Under the hypotheses of the theorem, this quantity is a positive constant multiple of n^2, so Corollary 7.2 applies to $p = 1/2$.

Let Π_{perf} be the property of being perfect. Every bipartite graph is perfect, so $\Pi_{2,0} \subseteq \Pi_{\mathrm{perf}}$ and hence $\chi_B(\Pi_{\mathrm{perf}}) \geq 3$. Conversely, C_5 is not perfect and is (r,s)-colourable for every $r, s \geq 0$ with $r + s = 3$. Therefore no $\Pi_{r,s}$ with $r + s = 3$ is contained in Π_{perf}, and $\chi_B(\Pi_{\mathrm{perf}}) = 3$. The Alon–Stav theorem gives, with high probability, $d_{\Pi_{\mathrm{perf}}}(G(n,1/2)) = \left(\frac{1}{4} \pm o(1)\right)\binom{n}{2}$. Combining this with Corollary 7.2 yields the following.

Corollary 7.4. *There exist constants $\eta > 0$ and $q \in \mathbb{N}$ such that, with probability tending to 1 as $n \to \infty$, every perfect induced subgraph of $G(n,1/2)$ is contained in one of at most n^q vertex sets of size at most $(1-\eta)n$.*

References

[ABF17] Noga Alon, Omri Ben-Eliezer, and Eldar Fischer. "Testing Hereditary Properties of Ordered Graphs and Matrices". In: Proceedings of the 58th IEEE Symposium on Foundations of Computer Science. 2017, pp. 848–858.

[Alo+00] Noga Alon, Eldar Fischer, Michael Krivelevich, and Mario Szegedy. "Efficient Testing of Large Graphs". In: Combinatorica 20.4 (2000), pp. 451–476.

[Alo+09] Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira. "A Combinatorial Characterization of the Testable Graph Properties: It's All About Regularity". In: SIAM Journal on Computing 39.1 (2009), pp. 143–167.

[Alo+11] Noga Alon, József Balogh, Béla Bollobás, and Robert Morris. "The Structure of Almost All Graphs in a Hereditary Property". In: Journal of Combinatorial Theory, Series B 101.2 (2011), pp. 85–110.

[AS08a] Noga Alon and Asaf Shapira. "A Characterization of the (Natural) Graph Properties Testable with One-Sided Error". In: SIAM Journal on Computing 37.6 (2008), pp. 1703–1727.

[AS08b] Noga Alon and Uri Stav. "The Maximum Edit Distance from Hereditary Graph Properties". In: Journal of Combinatorial Theory, Series B 98.4 (2008), pp. 672–697.

[AS08c] Noga Alon and Uri Stav. "What Is the Furthest Graph from a Hereditary Property?" In: Random Structures & Algorithms 33.1 (2008), pp. 87–104.

[AT10] Tim Austin and Terence Tao. "On the Testability and Repair of Hereditary Hypergraph Properties". In: Random Structures & Algorithms 36.4 (2010), pp. 373–463.

[BMS15] József Balogh, Robert Morris, and Wojciech Samotij. "Independent Sets in Hypergraphs". In: Journal of the American Mathematical Society 28.3 (2015), pp. 669–709.

[BMS18] József Balogh, Robert Morris, and Wojciech Samotij. "The Method of Hypergraph Containers". In: Proceedings of the International Congress of Mathematicians, Rio de Janeiro 2018. Vol. 4. World Scientific, 2018, pp. 3077–3110.

[Bor+08] Christian Borgs, Jennifer T. Chayes, László Lovász, Vera T. Sós, and Katalin Vesztergombi. "Convergent Sequences of Dense Graphs I: Subgraph Frequencies, Metric Properties and Testing". In: Advances in Mathematics 219.6 (2008), pp. 1801–1851.

[BR02] Michael A. Bender and Dana Ron. "Testing Properties of Directed Graphs: Acyclicity and Connectivity". In: Random Structures & Algorithms 20.2 (2002), pp. 184–205.

[BS24] Eric Blais and Cameron Seth. "New Graph and Hypergraph Container Lemmas with Applications in Property Testing". In: Proceedings of the 56th Annual ACM Symposium on Theory of Computing. 2024, pp. 1793–1804.

[BS25] Eric Blais and Cameron Seth. "Testing Graph Properties with the Container Method". In: SIAM Journal on Computing 54.6 (2025), pp. 1626–1642.

[BY22] Arnab Bhattacharyya and Yuichi Yoshida. Property Testing: Problems and Techniques. Singapore: Springer, 2022.

[ES12] Gábor Elek and Balázs Szegedy. "A Measure-Theoretic Approach to the Theory of Dense Hypergraphs". In: Advances in Mathematics 231.3–4 (2012), pp. 1731–1772.

[GGR98] Oded Goldreich, Shafi Goldwasser, and Dana Ron. "Property Testing and Its Connection to Learning and Approximation". In: Journal of the ACM 45.4 (1998), pp. 653–750.

[GNS26] Elena Grigorescu, Shreya Nasa, and Cameron Seth. "Testing the Independent Set Property in Hypergraphs". In: arXiv preprint arXiv:2607.13011 (2026).

[Gol17] Oded Goldreich. Introduction to Property Testing. Cambridge: Cambridge University Press, 2017.

[Gol21] Oded Goldreich. Open Problems in Property Testing of Graphs. Tech. rep. TR21-088. Electronic Colloquium on Computational Complexity, 2021.

[GT03] Oded Goldreich and Luca Trevisan. "Three Theorems Regarding Testing Graph Properties". In: Random Structures & Algorithms 23.1 (2003), pp. 23–57.

[KW82] Daniel J. Kleitman and Kenneth J. Winston. "On the Number of Graphs Without 4-Cycles". In: Discrete Mathematics 41.2 (1982), pp. 167–172.

[LS10] László Lovász and Balázs Szegedy. "Testing Properties of Graphs and Functions". In: Israel Journal of Mathematics 178 (2010), pp. 113–156.

[RS07] Vojtěch Rödl and Mathias Schacht. "Property Testing in Hypergraphs and the Removal Lemma". In: Proceedings of the 39th Annual ACM Symposium on Theory of Computing. 2007, pp. 488–495.

[Set25] Cameron Seth. "A Tolerant Independent Set Tester". In: Proceedings of the 57th Annual ACM Symposium on Theory of Computing. 2025, pp. 1031–1042.

[ST15] David Saxton and Andrew Thomason. "Hypergraph Containers". In: Inventiones Mathematicae 201.3 (2015), pp. 925–992.

Trinity College Cambridge, Department of Pure Mathematics and Mathematical Statistics, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, United Kingdom.
Email address: gc645@cam.ac.uk

Cheriton School of Computer Science, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 5Z5, Canada
Email address: cjmpseth@uwaterloo.ca

National Institute of Informatics (NII), 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan.
Email address: yyoshida@nii.ac.jp

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il