ON TESTABILITY OF FIRST-ORDER PROPERTIES IN BOUNDED-DEGREE GRAPHS AND CONNECTIONS TO PROXIMITY-OBLIVIOUS TESTING *

ISOLDE ADLER†, NOLEEN KÖHLER‡, AND PAN PENG§

Abstract. We study property testing of properties that are definable in first-order logic (FO) in the bounded-degree graph and relational structure models. We show that any FO property that is defined by a formula with quantifier prefix $\exists^*\forall^*$ is testable (i.e., testable with constant query complexity), while there exists an FO property that is expressible by a formula with quantifier prefix $\forall^*\exists^*$ that is not testable. In the dense graph model, a similar picture is long known (Alon, Fischer, Krivelevich, Szegedy, Combinatorica 2000), despite the very different nature of the two models. In particular, we obtain our lower bound by an FO formula that defines a class of bounded-degree expanders, based on zig-zag products of graphs. We expect this to be of independent interest.

We then use our class of FO definable bounded-degree expanders to answer a long-standing open problem for *proximity-oblivious testers (POTs)*. POTs are a class of particularly simple testing algorithms, where a basic test is performed a number of times that may depend on the proximity parameter, but the basic test itself is independent of the proximity parameter.

In their seminal work, Goldreich and Ron [STOC 2009; SICOMP 2011] show that the graph properties that are constant-query proximity-oblivious testable in the bounded-degree model are precisely the properties that can be expressed as a *generalised subgraph freeness (GSF)* property that satisfies the *non-propagation* condition. It is left open whether the non-propagation condition is necessary. Indeed, calling properties expressible as a generalised subgraph freeness property *GSF-local properties*, they ask whether all GSF-local properties are non-propagating. We give a negative answer by showing that our FO definable property is GSF-local and propagating. Hence in particular, our property does not admit a POT, despite being GSF-local. For this result we establish a new connection between FO properties and GSF-local properties via neighbourhood profiles.

Key words. Graph property testing, first-order logic, proximity-oblivious testing, locality, lower bound

MSC codes. 68Q25, 68R10, 68W20, 03B70

1. Introduction.

Graph property testing is a framework for studying sampling-based algorithms that solve a relaxation of classical decision problems on graphs. Given a graph G and a property $\mathcal{P}$ (e. g. triangle-freeness), the goal of a property testing algorithm, called a *property tester*, is to distinguish if a graph satisfies $\mathcal{P}$ or is *far* from satisfying $\mathcal{P}$, where the definition of *far* depends on the model. The general notion of property testing was first proposed by Rubinfeld and Sudan [34], with the motivation for the study of program checking. Goldreich, Goldwasser and Ron [19] then introduced the property testing for combinatorial objects and graphs. They formalized the *dense graph model* for testing graph properties, in which the algorithm

*Submitted to the editors February 28, 2023. This manuscript combines and extends two conference papers: "On Testability of First-Order Properties in Bounded-Degree Graphs" appeared at the 32nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2021) [3], and "GSF-Locality Is Not Sufficient For Proximity-Oblivious Testing" appeared at Computational Complexity Conference (CCC) 2021 [2]. In this paper we modified the property for the lower bound in [3] slightly so that it is GSF-local, which allows us to use it both as a non-testable Π_2-property and as a GSF-local property that is propagating. We also give an improved reduction from relational structures to undirected graphs which reduces the original (large) degree bound to 3. Finally, we complete the picture by showing that all GSF-local properties of degree at most 2 are non-propagating.

Funding: The third author was supported in part by NSFC grant 62272431.

†University of Bamberg, Bamberg, Germany (isolde.adler@uni-bamberg.de).

‡University of Leeds, Leeds, UK (n.koehler@leeds.ac.uk).

§School of Computer Science and Technology, University of Science and Technology of China, Hefei, China (ppeng@ustc.edu.cn).

can query if any pair of vertices of the input graph G with n vertices are adjacent or not, and the goal is to distinguish, with probability at least 2/3, the case of G satisfying a property $\mathcal{P}$ from the case that one has to modify (delete or insert) more than εn^2 edges to make it satisfy $\mathcal{P}$, for any specified proximity parameter $\varepsilon \in (0, 1]$. A property $\mathcal{P}$ is called testable (in the dense graph model), if it can be tested with constant query complexity, i.e., the number of queries made by the tester is bounded by a function of ε and is independent of the size of the input graph. Since [19], much effort has been made on the testability of graph properties in this model, culminating in the work by Alon et al. [5], who showed that a property is testable if and only if it can be reduced to testing for a finite number of regular partitions.

Since Goldreich and Ron's seminal work [21] introducing property testing on bounded-degree graphs, much attention has been paid to property testing in sparse graphs. Nevertheless, our understanding of testability of properties in such graphs is still limited. In the *bounded-degree graph model* [21], the algorithm has oracle access to the input graph G with maximum degree d, which is assumed to be a constant, and is allowed to perform *neighbour queries* to the oracle. That is, for any specified vertex v and index $i \leq d$, the oracle returns the i-th neighbour of v if it exists or a special symbol $\perp$ otherwise in constant time. A graph G with n vertices is called *ε-far* from satisfying a property $\mathcal{P}$, if one needs to modify more than εdn edges to make it satisfy $\mathcal{P}$. The goal now becomes to distinguish, with probability at least 2/3, if G satisfies a property $\mathcal{P}$ or is ε-far from satisfying $\mathcal{P}$, for any specified proximity parameter $\varepsilon \in (0, 1]$. Again, a property $\mathcal{P}$ is testable in the bounded-degree model, if it can be tested with constant query complexity, where the constant can depend on ε, d while being independent of n. So far, it is known that some properties are testable, including subgraph-freeness, k-edge connectivity, cycle-freeness, being Eulerian, degree-regularity [21], minor-freeness [7, 25, 29], hyperfinite properties [31], k-vertex connectivity [35, 16], and subdivision-freeness [28]. We now discuss the contributions of this paper.

1.1. Our contributions.

1.1.1. Non-testability of first-order logic. We study the testability of properties definable in first-order logic (FO) in the bounded-degree graph model. Recall that formulas of first-order logic on graphs are built from predicates for the edge relation and equality, using Boolean connectives $\vee, \wedge, \neg$ and universal and existential quantifiers $\forall, \exists$, where the variables represent graph vertices. First-order logic can e. g. express subgraph-freeness (i. e., no isomorphic copy of some fixed graph H appears as a subgraph) and subgraph containment (i. e., an isomorphic copy of some fixed H appears as a subgraph). Note however, that there are constant-query testable properties, such as connectivity and cycle-freeness, that cannot be expressed in FO. We study the question of which first-order properties are testable in the bounded-degree graph model. Our study extends to the bounded-degree *relational structure* model [1], while we focus on the classes of relational structures with binary relations, i.e., edge-coloured directed graphs. In this model for relational structures, one can perform neighbour queries, querying for both in- and out-neighbours and the edge colour that connects them. This model is natural in the context of relational databases, where each (edge-)relation is given by a list of the tuples it contains.

We consider the testability of first-order properties in the bounded-degree model according to quantifier alternation, inspired by a similar study for dense graphs by Alon et al. [4]. On relational structures of bounded-degree over a fixed finite signature, we have the following simple observation: Any first-order property definable by a sentence *without* quantifier alternations is testable. This means the sentence either consists of a quantifier prefix of the form $\exists^*$ (any finite number of existential quantifications), followed by a quantifier-free formula, or it consists of a quantifier prefix of the form $\forall^*$ (any finite number of universal quantifications), followed by a quantifier-free formula. Basically, every property of the form $\exists^*$ is testable because the structure required by the quantifier-free part of the formula can be planted with a small number of tuple modifications if the input structure is large enough (depending on the formula), and we can use an exact algorithm to determine the answer in constant time otherwise. Every property of the form $\forall^*$ is testable because a formula of the form $\forall\bar{x}\varphi(\bar{x})$, where φ is quantifier-free, is logically equivalent to a formula of the form $\neg\exists\bar{x}\psi(\bar{x})$, where ψ is quantifier-free. Testing $\neg\exists\bar{x}\psi(\bar{x})$ then amounts to testing for the absence of a finite number of induced substructures, which can be done similarly to testing subgraph freeness [21]. The testability of a property becomes less clear if it is defined by a sentence *with* quantifier alternations. Formally, we let Π_2 (resp. Σ_2) denote the set of properties that can be expressed by a formula in the $\forall^*\exists^*$-prefix (resp. $\exists^*\forall^*$-prefix) class. We obtain the following.

Every first-order property in Σ_2 is testable in the bounded-degree model (Theorem 6.1). On the other hand, there is a first-order property in Π_2, that is not testable in the bounded-degree model (Theorem 4.7).

The theorems that we refer to in the above statement speak about relational structures, while we also give a lower bound on graphs (Theorem 5.1), so the statement also holds when restricted to FO on graphs. Interestingly, the above dividing line is the same as for FO properties in dense graph model [4], despite the very different nature of the two models. Our proof uses a number of new proof techniques, combining graph theory, combinatorics and logic.

We remark that our lower bound, i.e., the existence of a property in Π_2 that is not testable, is somewhat astonishing (on an intuitive level) due to the following two reasons. Firstly, it is proven by constructing a first-order definable class of structures that encode a class of expander graphs, which highlights that FO is surprisingly expressive on bounded-degree graphs, despite its locality [24, 17, 32]. Secondly, it is known that property testing algorithms in the bounded-degree model proceed by sampling vertices from the input graph and exploring their local neighbourhoods, and FO can only express 'local' properties, while our lower bound shows that this is not sufficient for testability. We elaborate on this in the following. On one hand, Hanf's Theorem [24] gives insight into first-order logic on graphs of bounded-degree and implies a strong normal form, called *Hanf Normal Form* (HNF) in [9], which we briefly sketch. For a graph G of maximum degree d and a vertex x in G, the neighbourhood of fixed radius r around x in G can be described by a first-order formula $\tau_r(x)$, up to isomorphism. A *Hanf sentence* is a first-order sentence of the form 'there are at least ℓ vertices x of neighbourhood (isomorphism) type $\tau_r(x)$'. A first-order sentence is in HNF, if it is a Boolean combination of Hanf sentences. By Hanf's Theorem, every first-order sentence is equivalent to a sentence in HNF on bounded-degree graphs [24, 32, 14]. Note that Hanf sentences only speak about local neighbourhoods. Hence this theorem gives evidence that first-order logic can only express local properties. On the other hand, if a property is constant-query testable in the bounded-degree graph model, then it can be tested by approximating the distribution of local neighbourhoods (see [11] and [22]). That is, a constant-query tester can essentially only test properties that are close to being defined by a distribution of local neighbourhoods. For these reasons[1], a priori, it could be true that every property that can be expressed in first-order logic is testable in the bounded-degree model. Indeed, the validity of this statement was raised as an open question in [1]. However, our lower bound gives a negative answer to this question.

1.1.2. GSF-locality is not sufficient for proximity oblivious testing. Typical property testers make decisions regarding the global property of the graph based on local views only. In the extreme case, a tester could make the size of the local views independent of the distance ε to a predetermined set of graphs. Motivated by this, Goldreich and Ron [22] initiated the study of (one-sided error) *proximity-oblivious testers (POTs)* for graphs, where a tester simply repeats a basic test for a number of times that depends on the proximity parameter, while the basic tester is oblivious of the proximity parameter. They gave characterizations of graph properties that can be tested with constant query complexity by a POT in both dense graph model and the bounded-degree model. In each model, it is known that the class of properties that have constant-query POTs is a strict subset of the class of properties that are testable (by standard testers).

Informally, a (one-sided error) POT for a property $\mathcal{P}$ is a tester that always accepts a graph G if it satisfies $\mathcal{P}$, and rejects G with probability that is a monotonically increasing function of the distance of G from the property $\mathcal{P}$. We say $\mathcal{P}$ is *proximity-oblivious testable* if such a tester exists with constant query complexity. To characterise the class of proximity-oblivious testable properties in the bounded-degree model, Goldreich and Ron [22] introduced a notion of generalized subgraph freeness (GSF), that extends the notions of induced subgraph freeness and (non-induced) subgraph freeness. A graph property is called a *GSF-local* property if it is expressible as a GSF property. It was shown in [22] that a graph property is constant-query proximity-oblivious testable if and only if it is a GSF-local property that satisfies a so-called *non-propagation* condition. Informally, a GSF-local property $\mathcal{P}$ is non-propagating if repairing a graph G that does not satisfy $\mathcal{P}$ does not trigger a global "chain reaction" of necessary modifications (see Section 7.1 for the formal definitions).

A major question that is left open in [22] is whether every GSF-local property satisfies the non-propagation condition. By using the aforementioned non-testable FO property and establishing a new connection between FO properties and GSF-local properties, we resolve this question by showing the following negative result.

There exists a GSF-local property of graphs of degree at most 3 *that is not testable in the bounded-degree model. Thus, not all GSF-local properties are non-propagating (Theorem 7.5).*

We expect this result will shed some light on a full characterisation of testable properties in the bounded-degree model. Indeed, in a recent work by Ito, Khoury and Newman [27], the authors gave a characterization of testable *monotone* graph properties and testable *hereditary* graph properties with one-sided error in the bounded-degree graph model; and they asked the open question "*is every property that is defined by a set of forbidden configurations testable?*". Since their definition of a property defined by a set of "forbidden configuration" is equivalent to a GSF-local property, our result above also gives a negative answer to their question.

We complete the picture by showing the following.

[1]Furthermore, previously, typical FO properties were all known to be testable, including degree-regularity for a fixed given degree, containing a k-clique and a dominating set of size k for fixed k (which are trivially testable), and the aforementioned subgraph-freeness and subgraph containment (see e.g. [18]).

Every GSF-local property of graphs of degree at most 2 *is non-propagating (Theorem 7.16).*

1.2. Our techniques.

1.2.1. On the testability and non-testability of FO properties. For showing that every property $\mathcal{P}$ defined by a formula φ in Σ_2 (i.e. of the form $\exists^*\forall^*$) is testable, we show that $\mathcal{P}$ is equivalent to the union of properties $\mathcal{P}_i$, each of which is 'indistinguishable' from a property $\mathcal{Q}_i$ that is defined by a formula of form $\forall^*$. Here the indistinguishability means we can transform any structure satisfying $\mathcal{P}_i$, into a structure satisfying $\mathcal{Q}_i$ by modifying a small fraction of the tuples of the structure and vice versa. This allows us to reduce the problem of testing $\mathcal{P}$ to testing properties defined by $\forall^*$ formulas. Then the testability of $\mathcal{P}$ follows, as any property of the form $\forall^*$ is testable and testable properties are closed under union [18]. The main challenge here is to deal with the interactions between existentially quantified variables and universally quantified variables. Intuitively, the degree bound limits the structure that can be imposed by the universally quantified variables. Using this, we are able to deal with the existential variables together with these interactions by 'planting' a required constant size substructure in such a way, that we are only a constant number of modifications 'away' from a formula of the form $\forall^*$.

Complementing this, we use Hanf's theorem to observe that every FO property on degree-regular structures is in Π_2 (see Lemma 4.5). Thus to prove that there exists a property defined by a formula in Π_2 which is not testable, it suffices to show the existence of an FO property that is not testable and degree-regular. For the latter, we note that it suffices to construct a formula φ, that defines a class of relational structures with binary relations only (edge-coloured directed graphs) whose underlying undirected graphs are expander graphs. To see this, we use an earlier result that if a property is constant-query testable, then the distance between the local (constant-size) neighbourhood distributions of a relational structure A satisfying the property φ and a relational structure B that is ε-far from having the property must be relatively large (see [1] which in turn is built upon the so-called "canonical testers" for bounded-degree graphs in [11, 22]). We then exploit a result of Alon (see Proposition 19.10 in [30]), that the neighbourhood distribution of an arbitrarily large relational structure A can be approximated by the neighbourhood distribution of a structure H of small constant size. Thus, for any A in φ, by taking the union of "many" disjoint copies of the "small" structure H, we obtain another structure B such that the local neighbourhood distributions of A and B have small distance. If the underlying undirected graphs of the structures in φ are expander graphs, it immediately follows that B is far from the property defined by the formula φ, from which we can conclude that the property φ is not testable. We remark that for simple undirected graphs, it was known before that any property that only consists of expander graphs is not testable [15].

Now we construct a formula φ, that defines a class of relational structures with binary relations only whose underlying undirected graphs are expander graphs, arising from the zig-zag product by Reingold, Vadhan and Wigderson [33]. For expressibility in FO, we hybridise the zig-zag construction of expanders with a tree structure. Roughly speaking, we start with a small graph H, which is a good expander, and the formula φ expresses that each model[2] looks like a rooted k-ary tree (for a suitable

[2]When the context is clear, "model" refers to a structure that satisfies some formula. This should not be confused with the names for our computational models, e.g., the bounded-degree model.

fixed k), where level 0 consists of the root only, level 1 contains $G_1 := H^2$, and level i contains the zig-zag product of G_{i-1}^2 with H. The class of trees is not definable in FO. However, we achieve that every finite model of our formula is connected and looks like a k-ary tree with the desired graphs on the levels. This structure is obtained by a recursive 'copying-inflating' mechanism, to mimic the expander construction locally between consecutive levels. For this we use a constant number of edge-colours, one set of colours for the edges of the tree, and another for the edges of the 'level' graphs G_i. On the way, many technicalities need to be tackled, such as encoding the zig-zag construction into the local copying mechanism (and achieving the right degrees), and finally proving connectivity. We then show that the underlying undirected graphs of the models of φ are expander graphs. Using a hardness reduction which inserts carefully designed gadgets to encode the different edge-colours, we finally obtain a non-testable property of undirected 3-regular graphs.

1.2.2. On GSF-locality and POTs. We then proceed to showing that this property of 3-regular graphs is GSF-local. For this, we first study the relation between locality of first-order logic and GSF-locality. Hanf's Theorem [24] implies that we can understand locality of FO as prescribing upper and lower bounds for the number of occurrences of certain local neighbourhood (isomorphism) types. On the other hand, a GSF-local property as defined in [22] prescribes the absence of some constant-size *marked* graphs, where a marked graph F specifies an induced subgraph and how it 'interacts' with the rest of the graph (see Definition 7.1). Intuitively, such a property just specifies a condition that the local neighbourhoods of a graph G should satisfy, i.e., certain types of local neighbourhoods cannot occur in G, or equivalently, these types have 0 occurrences. However, it does not follow that every GSF-local property is FO definable, because the set of forbidden marked graphs depends on the size n of the graphs in the class. Indeed, it is not hard to come up with undecidable properties that are GSF-local.

To establish a connection between FO properties and GSF-local properties, we first encode the bounds on the number of occurrences of local neighbourhood types into what we call *neighbourhood profiles*, and characterise FO definable properties of bounded-degree relational structures as finite unions of properties defined by neighbourhood profiles (Lemma 7.7). We then show that every FO formula defined by a non-trivial finite union of properties each of which is defined by a 0-*profile*, i. e. the prescribed lower bounds are all 0, is GSF-local (Theorem 7.9). Given the fundamental roles of local properties in graph theory, graph limits [30], we believe this new connection is of independent interest.

For technical reasons, we make use of the property defined by our formula φ above, which is a property of *relational structures* that is not testable in the bounded-degree model, instead of directly using our non-testable graph property of 3-regular graphs. We prove that the property defined by φ can actually be defined by 0-profiles (Lemma 7.12). We then derive that our non-testable graph property of 3-regular graphs is also GSF-local (Lemma 7.14), by showing that the reduction maintains definability by 0-profiles.

1.3. Other related work. Besides the aforementioned works on testing properties with constant query complexity in the bounded-degree graph model, Goldreich and Ron [22] have obtained a characterisation for a class of properties that are testable by a constant-query proximity-oblivious tester in bounded-degree graphs (and dense graphs). Such a class is a rather restricted subset of the class of all constant-query testable properties. Fichtenberger et al. [15] showed that every testable property is either finite or contains an infinite hyperfinite subproperty. Informally, a hyerfinite subproperty is a subset of graphs that can be partitioned into small connected components by removing a small fraction of the edges and is invariant under isomporphism. Ito et al. [27] gave characterisations of one-sided error (constant-query) testable monotone graph properties, and one-sided error testable hereditary graph properties in the bounded-degree (directed and undirected) graph model.

In the bounded-degree graph model, there exist properties (e.g. bipartiteness, expansion, k-clusterability) that need $\Omega(\sqrt{n})$ queries, and properties (e.g. 3-colorability) that need $\Omega(n)$ queries. We refer the reader to Goldreich's recent book [18].

Property testing on relational structures was recently motivated by the application in databases. Besides the aforementioned work [1], Chen and Yoshida [10] studied the testability of relational database queries for each relational structure in the framework of property testing.

The notion of POT was implicitly defined in [8]. Goldreich and Shinkar [23] studied two-sided error POTs for both dense graph and bounded-degree graph models. Goldreich and Kaufman [20] investigated the relation between local conditions that are invariant in an adequate sense and properties that have a constant-query proximity-oblivious testers.

1.4. Structure of the paper. Section 2 contains the preliminaries, including logic, property testing and the zig-zag construction of expander graphs. In Section 3 we construct the FO formula φ and prove properties of its models. In Section 4, we prove that there is a Π_2-property that is not testable, by proving that the property defined by φ on bounded-degree structures is not constant-query testable. Using a reduction, in Section 5 we then provide a Π_2-property of undirected graphs of degree at most 3 that is non-testable. In Section 6, we show that all Σ_2 properties are testable. In Section 7 we then turn to POTs, showing that our Π_2-property of undirected graphs of degree at most 3 is GSF-local and propagating. We then show that all GSF-local properties of degree at most 2 are non-propagating. We conclude in Section 8.

2. Preliminaries. We let $\mathbb{N}$ denote the set of natural numbers including 0, and $\mathbb{N}_{>0} := \mathbb{N} \setminus \{0\}$. For $n \in \mathbb{N}$ we let $[n] := \{0, 1, \dots, n-1\}$ denote the set of the first n natural numbers. For a set S and $k \in \mathbb{N}$ we denote the Cartesian product $S \times \cdots \times S$ of k copies of S by S^k. We use $\binom{S}{2}$ to denote the set of all two-element subsets of S, we denote the disjoint union of sets by $\sqcup$ and the symmetric difference by $\triangle$.

2.1. Undirected graphs. Unless otherwise specified we allow graphs to have self-loops and parallel edges. We represent an undirected graph G as a triple (V, E, f), where V is the set of vertices, E is the set of edges and $f : E \to V \cup \binom{V}{2}$ is the incidence map. An isomorphism from $G_1 = (V_1, E_1, f_1)$ to $G_2 = (V_2, E_2, f_2)$ is a pair of bijective maps (h_V, h_E), where $h_V : V_1 \to V_2$ and $h_E : E_1 \to E_2$, such that $h_V(f_1(e)) = f_2(h_E(e))$ for any $e \in E_1$, where $h_V(X) := \{h_V(x) \mid x \in X\}$ for any set $X \subseteq V_1$. Undirected graphs without self-loops and parallel edges are called *simple*. For a simple graph G, we also represent G as a tuple $G = (V(G), E(G))$, where $V(G)$ is the vertex set and $E(G) \subseteq \binom{V}{2}$. The *degree* $\deg_G(v)$ of a vertex v in a graph G is the number of edges to which v is incident. In particular, self-loops contribute one to the degree. We will say that a graph G is *d-regular* for some $d \in \mathbb{N}$ if every vertex in G has degree d. We specify paths in graphs by tuples of vertices. We further let all paths and cycles be simple, i. e. no vertex appears twice. The *length* of a path on n vertices is $n-1$. We define the distance between two vertices v and w in a graph G, denoted $\mathrm{dist}_G(v, w)$, as the length of a shortest path from v to w or ∞ if there is no path from v to w in G. Any subset $S \subseteq V$ of vertices *induces* a graph $G[S] := (S, \{e \in E \mid f(e) \in S \cup \binom{S}{2}\}, f|_S)$. A *connected component* of G is a graph induced by a maximal set S, such that each pair $v, w \in S$ has finite distance in G. A graph is connected if it has only one connected component. We refer the reader to [12] for the basic notions of graph theory.

We also consider rooted undirected trees. By specifying a root we can uniquely direct the edges away from the root. This allows us to use the terminology of *children* and *parents* for undirected rooted trees. We call a tree a *full k-ary tree* if every vertex has either none or exactly k children. If, in addition, for every $i \in \mathbb{N}$ there are either exactly k^i or no vertices of distance i to the root of the tree we call it a *balanced full k-ary tree.*

2.2. Relational structures and first-order logic. We will briefly introduce structures and first-order logic and point the reader to [14] for a more detailed introduction. A (relational) *signature* is a finite set $\sigma = \{R_1, \dots, R_\ell\}$ of relation symbols R_i. Every relation symbol R_i, $1 \leq i \leq \ell$ has an arity $\mathrm{ar}(R_i) \in \mathbb{N}_{>0}$. A *$\sigma$-structure* is a tuple $A = (U(A), R_1(A), \dots, R_\ell(A))$, where $U(A)$ is a *finite* set, called the *universe* of A and $R_i(A) \subseteq U(A)^{\mathrm{ar}(R_i)}$ is an $\mathrm{ar}(R_i)$-ary relation on $U(A)$. Note that if $\sigma = \{E_1, \dots, E_\ell\}$ is a signature where each E_i is a binary relation symbol, then σ-structures are directed simple graphs with ℓ edge-colours. Let $\sigma_{\mathrm{graph}} := \{E\}$ be a signature with one binary relation symbol E. Then we can understand undirected simple graphs as σ_{graph}-structures for which the relation E is symmetric (every undirected edge is represented by two tuples) and irreflexive. Using this we can transfer all notions defined below to simple graphs. Typically we name graphs G, H, F, we denote the set of vertices of a graph G by $V(G)$, the set of edges by $E(G)$ and vertices are typically named $u, v, w, u', v', w', \dots$. In contrast when we talk about a general relational structure we use A, B and $a, b, a', b', \dots$ to denote elements from the universe.

In the following we let σ be a relational signature. Two σ-structures A and B are *isomorphic* if there is a bijective map from $U(A)$ to $U(B)$ that preserves all relations. For a σ-structure A and a subset $S \subseteq U(A)$, we let $A[S]$ denote the *substructure* of A *induced* by S, i. e. $A[S]$ has universe S and $R(A[S]) := R(A) \cap S^{\mathrm{ar}(R)}$ for all $R \in \sigma$. The *degree* of an element $a \in U(A)$ denoted by $\deg_A(a)$ is defined to be the number of tuples in A containing a. We define the *degree* of A, denoted by $\deg(A)$, to be the maximum degree of its elements. A structure A is *d-regular* for some $d \in \mathbb{N}$ if every element $a \in U(A)$ has degree d. Given a signature σ and a constant d, we let $\mathcal{C}_{\sigma,d}$ be the class of all σ-structures of degree at most d, and let $\mathcal{C}_d$ the set of all simple graphs of degree at most d. Note that the degree of a graph differs by exactly a factor 2 from the degree of the corresponding σ_{graph}-structure. Let $\mathcal{C}$ be any class of σ-structures which is closed under isomorphism. A *property* $\mathcal{P}$ in $\mathcal{C}$ is a subset of $\mathcal{C}$ which is closed under isomorphism. We say that a structure A has property $\mathcal{P}$ if $A \in \mathcal{P}$.

The syntax and semantics of FO logic are defined in the usual way (see *e. g.* [14]). We use $\exists^{\geq m} x\, \varphi$ (and $\exists^{=m} x\, \varphi$, $\exists^{\leq m} x\, \varphi$, respectively) as a shortcut for the FO formula expressing that the number of witnesses x satisfying φ is at least m (exactly m, at most m, respectively). We say that a variable occurs *freely* in an FO formula if at least one of its occurrences is not bound by any quantifier. We use $\varphi(x_1, \dots, x_k)$ to express that the set of variables which occur freely in the FO formula φ is a subset of $\{x_1, \dots, x_k\}$. For a formula $\varphi(x_1, \dots, x_k)$, a σ-structure A and $a_1, \dots, a_k \in U(A)$ we write $A \models \varphi(a_1, \dots, a_k)$ if φ evaluates to true after assigning a_i to x_i, for $1 \leq i \leq k$. A *sentence* of FO is a formula with no free variables. For an FO sentence φ we say that A is a *model* of φ or A satisfies φ if $A \models \varphi$. Let $\mathcal{C}$ be a class of σ-structures closed under isomorphism. Every FO-sentence φ over σ defines a property $\mathcal{P}_\varphi \subseteq \mathcal{C}$ on $\mathcal{C}$, where $\mathcal{P}_\varphi := \{A \in \mathcal{C} \mid A \models \varphi\}$.

Hanf normal form. The *Gaifman graph* of a σ-structure A is the undirected graph $G(A) = (U(A), E)$, where $\{v, w\} \in E$, if $v \neq w$ and there is an $R \in \sigma$ and a tuple $\overline{a} = (a_1, \dots, a_{\mathrm{ar}(R)}) \in R(A)$, such that $v = a_j$ and $w = a_k$ for some $1 \leq k, j \leq \mathrm{ar}(R)$. We use $G(A)$ to apply graph theoretic notions to relational structures. Note that for any simple graph the Gaifman graph of the corresponding symmetric σ_{graph}-structure is the graph itself. We say that a σ-structure A is *connected* if its Gaifman graph $G(A)$ is connected. For two elements $a, b \in U(A)$, we define the *distance* between a and b in A, denoted by $\mathrm{dist}_A(a, b)$, as the length of a shortest path from a to b in $G(A)$, or ∞ if there is no such path. For $r \in \mathbb{N}$ and $a \in U(A)$, the *r-neighbourhood* of a is the set $N_r^A(a) := \{b \in U(A) : \mathrm{dist}_A(a, b) \leq r\}$. We define $\mathcal{N}_r^A(a) := A[N_r^A(a)]$ to be the substructure of A induced by the r-neighbourhood of a. For $r \in \mathbb{N}$ an *r-ball* is a tuple (B, b), where B is a σ-structure, $b \in U(B)$ and $U(B) = N_r^B(b)$, i.e. B has radius r and b is the centre. Note that by definition $(\mathcal{N}_r^A(a), a)$ is an r-ball for any σ-structure A and $a \in U(A)$. Two r-balls $(B, b), (B', b')$ are isomorphic if there is an isomorphism of σ-structure from B to B' that maps b to b'. We call the isomorphism classes of r-balls *r-types*. For an r-type τ and an element $a \in U(A)$ we say that *a has* (r-)type τ if $(\mathcal{N}_r^A(a), a) \in \tau$. Moreover, given such an r-type τ, there is a formula $\varphi_\tau(x)$ such that for every σ-structure A and for every $a \in U(A)$, $A \models \varphi_\tau(a)$ iff $(\mathcal{N}_r^A(a), a) \in \tau$. A *Hanf-sentence* is a sentence of the form $\exists^{\geq m} x \varphi_\tau(x)$, for some $m \in \mathbb{N}_{>0}$, where τ is an r-type. An FO sentence is in *Hanf normal form*, if it is a Boolean combination[3] of Hanf sentences. Two formulas $\varphi(x_1, \dots, x_k)$ and $\psi(x_1, \dots, x_k)$ of signature σ are called *d-equivalent*, denoted by $\varphi(x_1, \dots, x_k) \equiv_d \psi(x_1, \dots, x_k)$, if they are equivalent on $\mathcal{C}_{\sigma,d}$, i.e. for all $A \in \mathcal{C}_{\sigma,d}$ and all $(a_1, \dots, a_k) \in U(A)^k$ we have $A \models \varphi(a_1, \dots, a_k)$ iff $A \models \psi(a_1, \dots, a_k)$. Hanf's locality theorem for first-order logic [24] implies the following.

Theorem 2.1 (Hanf [24]). *Let $d \in \mathbb{N}$. Every sentence of first-order logic is d-equivalent to a sentence in Hanf normal form.*

Quantifier alternations of first-order formulas. Let σ be any relational signature. We use the following recursive definition, classifying first-order formulas according to the number of quantifier alterations in their quantifier prefix. Let $\Sigma_0 = \Pi_0$ be the class of all quantifier free first-order formulas over σ. Then for every $i \in \mathbb{N}_{>0}$ we let Σ_i be the set of all FO formulas $\varphi(y_1, \dots, y_\ell)$ for which there is $k \in \mathbb{N}$ and a formula $\psi(x_1, \dots, x_k, y_1, \dots, y_\ell) \in \Pi_{i-1}$ such that

$$\varphi \equiv \exists x_1 \dots \exists x_k \psi(x_1, \dots, x_k, y_1, \dots, y_\ell).$$

Analogously, Π_i consists of all FO formulas $\varphi(y_1, \dots, y_\ell)$ for which there is $k \in \mathbb{N}$ and a formula $\psi(x_1, \dots, x_k, y_1, \dots, y_\ell) \in \Sigma_{i-1}$ such that

$$\varphi \equiv \forall x_1 \dots \forall x_k \psi(x_1, \dots, x_k, y_1, \dots, y_\ell).$$

We further say that a property $\mathcal{P} \subseteq \mathcal{C}$ is in Σ_i or Π_i if there is an FO-sentence φ in Σ_i or Π_i, respectively, such that $\mathcal{P} = \mathcal{P}_\varphi$.

Example 1 (Substructure freeness). *Let B be a σ-structure, and let $d \in \mathbb{N}$. The property*

$$\mathcal{P} := \{A \in \mathcal{C}_{\sigma,d} \mid A \textit{ does not contain } B \textit{ as substructure}\}$$

is in Π_1.

[3] By Boolean combination we always mean *finite* Boolean combination.

2.3. Property testing. In the following, we give definitions of two models for property testing - the bounded-degree model for simple graphs introduced in [21] and a bounded-degree model for relational structures similar to the model introduced in [1]. The model for relational structures described here is chosen to simplify notation. It differs from the model in [1] in the way the query access is defined, however, they are equivalent in the sense that testability in either model implies testability in the other model. This can be easily seen using a local reduction as defined in Section 5.2. The bounded-degree model for relational structures extends the bounded-degree model for undirected graphs introduced in [21] and conforms with the bidirectional model of [11].

For notational convenience, $\mathcal{C}$ will either denote a class of graphs of bounded-degree d closed under isomorphism, or a class of σ-structures of bounded-degree d closed under isomorphism for some signature σ and some $d \in \mathbb{N}$. Let $\mathcal{P}$ be a property on $\mathcal{C}$. We will further refer to both graphs and σ-structures as structures. Let $\mathcal{P}_n$ be the subset of $\mathcal{P}$ with n vertices/elements. Thus $\mathcal{P} = \bigcup_{n\in\mathbb{N}} \mathcal{P}_n$. We define the distance of a structure A on n vertices/elements to a property $\mathcal{P} = \bigcup_{n\in\mathbb{N}} \mathcal{P}_n$ as

$$\mathrm{dist}(A, \mathcal{P}) := \min_{B\in\mathcal{P}_n} \frac{\sum_{R\in\sigma} |R(A) \triangle R(B)|}{dn}.$$

For $\epsilon \in (0, 1)$ we say that a structure A on n vertices/elements is *ϵ-close* to $\mathcal{P}$ if $\mathrm{dist}(A, \mathcal{P}) \leq \epsilon$, that is one can modify A into a structure in $\mathcal{P}$ by adding/deleting at most ϵdn tuples of A. We say that A is ϵ-far from $\mathcal{P}$ if A is not ϵ-close to $\mathcal{P}$.

An algorithm that processes a structure $A \in \mathcal{C}$ does not obtain an encoding of A as a bit string in the usual way. Instead, we assume that the algorithm receives the number n of elements/vertices of A, and that the elements/vertices of A are numbered $1, 2, \ldots, n$. In addition, the algorithm has direct access to A using an *oracle* which answers *neighbour queries* in A in constant time. A *query* to a σ-structure A of bounded-degree d has the form (a, i) for an element $a \in U(A)$, $i \in \{1, \ldots, d\}$ and is answered by $\mathrm{ans}(a, i) := (R, a_1, \ldots, a_{\mathrm{ar}(R)})$ where $(a_1, \ldots, a_{\mathrm{ar}(R)})$ is the i-th tuple (according to some fixed ordering) containing a and $(a_1, \ldots, a_{\mathrm{ar}(R)}) \in R(A)$, or a special symbol "$\perp$" if i is greater than the degree of a. A *query* to a graph G of bounded-degree d has the form (v, i) for $v \in V(G)$, $i \in \{1, \ldots, d\}$ and is answered by $\mathrm{ans}(v, i) := w$ where w is the i-th neighbour of v.

Now we give the formal definitions of standard property testing and proximity-oblivious testing.

DEFINITION 2.2 ((Standard) property testing). *Let $\mathcal{P} = \bigcup_{n\in\mathbb{N}} \mathcal{P}_n$ be a property. An ϵ-*tester *for $\mathcal{P}_n$ is a probabilistic algorithm which, given query access to a structure $A \in \mathcal{C}$ with n vertices/elements,*

- *accepts A with probability $2/3$, if $A \in \mathcal{P}_n$.*
- *rejects A with probability $2/3$, if A is ϵ-far from $\mathcal{P}_n$.*

We say that a property $\mathcal{P}$ is testable *if for every $n \in \mathbb{N}$ and $\epsilon \in (0, 1)$, there exists an ϵ-tester for $\mathcal{P}_n$ that makes at most $q = q(\epsilon, d)$ queries. We say the property $\mathcal{P}$ is testable with* one-sided error *if the ϵ-tester always accepts A if $A \in \mathcal{P}$.*

We introduce below the formal definition of proximity-oblivious testers.

DEFINITION 2.3 (Proximity-oblivious testing (with one-sided error)). *Let $\mathcal{P} = \bigcup_{n\in\mathbb{N}} \mathcal{P}_n$ be a property. Let $\eta : (0, 1] \to (0, 1]$ be a monotonically non-decreasing function. A* proximity-oblivious tester (POT) *with detection probability η for $\mathcal{P}_n$ is a probabilistic algorithm which, given query access to a structure $A \in \mathcal{C}$ with n vertices/elements,*

- *accepts A with probability 1, if $A \in \mathcal{P}_n$.*
- *rejects A with probability at least $\eta(\mathrm{dist}(A, \mathcal{P}_n))$, if $A \notin \mathcal{P}_n$. Equivalently, for any A that is not in $\mathcal{P}_n$, the algorithm accepts A with probability at most $1 - \eta(\mathrm{dist}(A, \mathcal{P}_n))$.*

We say that a property $\mathcal{P}$ is proximity-oblivious testable *if for every $n \in \mathbb{N}$, there exists a POT for $\mathcal{P}_n$ of constant query complexity with detection probability η.*

We remind the reader of the following which we argued in the introduction.

Remark 2.4. Let $d \in \mathbb{N}$. Every propery definable in Σ_1 is testable on C_d, and every property definable in Π_1 is testable on C_d.

2.4. Expansion and the zig-zag product. In this section we recall a construction of a class of expanders introduced in [33]. This construction uses undirected graphs with parallel edges and self-loops.

Let $G = (V, E, f)$ be an undirected D-regular graph on N vertices. We follow the convention that each self-loop counts 1 towards the degree. Let I be a set of size D. Then a *rotation map* of G is a function $\mathrm{ROT}_G : V \times I \to V \times I$ such that for every two not necessarily different vertices $u, v \in V$

$$|\{(i,j) \in I \times I \mid \mathrm{ROT}_G(u,i) = (v,j)\}| = 2|\{e \in E \mid f(e) = \{u,v\}\}|$$

and ROT_G is self inverse, i.e. $\mathrm{ROT}_G(\mathrm{ROT}_G(v,i)) = (v,i)$ for all $v \in V, i \in I$. A rotation map is a representation of a graph that additionally fixes for every vertex v an order on all edges incident to v. We let the normalised adjacency matrix M of G be defined by

$$M_{u,v} := \frac{1}{D} \cdot |\{e \mid f(e) = \{u,v\}\}|.$$

Since M is real, symmetric, contains no negative entries and all columns sum up to 1, all its eigenvalues are in the real interval $[-1, 1]$. Let $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq -1$ denote the eigenvalues of M. We let $\lambda(G) := \max\{|\lambda_2|, |\lambda_N|\}$. Note that these notions do not depend on the rotation map. We say that a graph is an (N, D, λ)-graph, if G has N vertices, is D-regular and $\lambda(G) \leq \lambda$. We will use the following lemma.

LEMMA 2.5 ([26]). *The graph G is connected if and only if $\lambda_2 < 1$. Furthermore, if G is connected, then G is bipartite if and only if $\lambda_N = -1$.*

For any subsets $S, T \subseteq V$ let $\langle S, T \rangle_G := \{e \in E \mid f(e) \cap S \neq \emptyset, f(e) \cap T \neq \emptyset\}$ be the set of edges *crossing* between S and T.

DEFINITION 2.6. *For any set $S \subseteq V$, we let $h(S) := \frac{|\langle S, \overline{S} \rangle_G|}{|S|}$ be the* expansion *of S. We let $h(G)$ be the* expansion ratio *of G defined by $h(G) := \min_{\{S \subset V \mid |S| \leq N/2\}} h(S)$.*

For any constant $\epsilon > 0$ we call a sequence $\{G_m\}_{m \in \mathbb{N}_{>0}}$ of graphs of increasing number of vertices a *family of ϵ-expanders*, if $h(G_m) \geq \epsilon$ for all $m \in \mathbb{N}_{>0}$. We say that a family of graphs is a family of expanders if it is a family of ϵ-expanders for some constant $\epsilon > 0$. We further often call a graph from a family of expanders an expander. There exists the following connection between $h(G)$ and $\lambda(G)$.

THEOREM 2.7 ([13, 6]). *Let G be a D-regular graph. Then it holds that $h(G) \geq D(1 - \lambda(G))/2$.*

This implies that for a sequence of graphs $\{G_m\}_{m \in \mathbb{N}_{>0}}$ of increasing number of vertices, if there is a constant $\epsilon < 1$ such that $\lambda(G_m) \leq \epsilon$ for all $m \in \mathbb{N}_{>0}$, then the sequence $\{G_m\}_{m \in \mathbb{N}_{>0}}$ is a family of $D(1 - \varepsilon)/2$-expanders.

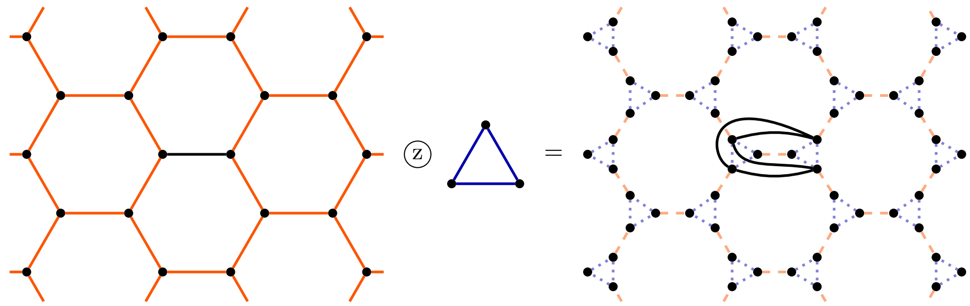


Fig. 1: Zig-zag product of a 3-regular grid with a triangle

Definition 2.8. *Let G be a D-regular graph on N vertices with rotation map $\mathrm{ROT}_G : V \times I \to V \times I$ and I a set of size D. Then the* square *of G, denoted by G^2, is a D^2-regular graph on N vertices with rotation map $\mathrm{ROT}_{G^2}(u,(k_1,k_2)) := (w,(\ell_2,\ell_1))$, where*

$$\mathrm{ROT}_G(u,k_1) = (v,\ell_1) \text{ and } \mathrm{ROT}_G(v,k_2) = (w,\ell_2),$$

and $u,v,w \in V$, $k_1,k_2,\ell_1,\ell_2 \in I$.

Note that the edges of G^2 correspond to walks of length 2 in G and the adjacency matrix of G^2 is the square of the adjacency matrix of G. Note here that if G is bipartite then G^2 is not connected, which can be easily seen by using Lemma 2.5.

Lemma 2.9 ([33]). *If G is a (N,D,λ)-graph then G^2 is a (N,D^2,λ^2)-graph.*

Definition 2.10. *Let $G_1 = (V_1,E_1,f_1)$ be a D_1-regular graph on N_1 vertices, I_1 a set of size D_1 and $\mathrm{ROT}_{G_1} : V_1 \times I_1 \to V_1 \times I_1$ a rotation map of G_1. Let $G_2 = (I_1,E_2,f_2)$ be a D_2-regular graph, let I_2 be a set of size D_2 and $\mathrm{ROT}_{G_2} : I_1 \times I_2 \to I_1 \times I_2$ be a rotation map of G_2. Then the* zig-zag product of G_1 and G_2, *denoted by $G_1 \textcircled{z} G_2$, is the D_2^2-regular graph on vertex set $V_1 \times I_1$ with rotation map given by $\mathrm{ROT}_{G_1 \textcircled{z} G_2}((v,k),(i,j)) := ((w,\ell),(j',i'))$, where*

$$\mathrm{ROT}_{G_2}(k,i) = (k',i'), \mathrm{ROT}_{G_1}(v,k') = (w,\ell'), \text{ and } \mathrm{ROT}_{G_2}(\ell',j) = (\ell,j'),$$

and $v,w \in V_1$, $k,k',\ell,\ell' \in I_1$, $i,i',j,j' \in I_2$.

The zig-zag product $G_1 \textcircled{z} G_2$ can be seen as the result of the following construction. First pick some numbering of the vertices of G_2. Then replace every vertex in G_1 by a copy of G_2 where we colour edges from G_1, say, red, and edges from G_2 blue. We do this in such a way that the i-th edge of a vertex v in G_1 will be connected to vertex i of the replica of G_2, which replaces the vertex v in the preceding step. Then for every red edge (v,w) and for every tuple $(i,j) \in I_2 \times I_2$ we add an edge to the zig-zag product $G_1 \textcircled{z} G_2$ connecting v' and w' where v' is the vertex reached from v by taking its i-th blue edge and w' can be reached from w by taking its j-th blue edge. Figure 1 shows an example, where in the graph on the right hand side we show the 4 edges that are added to the zig-zag product for the highlighted edge of the graph on the left hand side.

Theorem 2.11 ([33]). *If G_1 is an (N_1,D_1,λ_1)-graph and G_2 is a (D_1,D_2,λ_2)-*

graph then $G_1 ⓩ G_2$ is an $(N_1 \cdot D_1, D_2^2, g(\lambda_1, \lambda_2))$-graph, where

$$g(\lambda_1, \lambda_2) = \frac{1}{2}(1-\lambda_2^2)\lambda_1 + \frac{1}{2}\sqrt{(1-\lambda_2^2)^2\lambda_1 + 4\lambda_2^2}.$$

This function has the following properties.

1. *If both $\lambda_1 < 1$ and $\lambda_2 < 1$ then $g(\lambda_1, \lambda_2) < 1$.*
2. *$g(\lambda_1, \lambda_2) < \lambda_1 + \lambda_2$.*

DEFINITION 2.12 ([26]). *Let D be a sufficiently large prime power (e.g. $D = 2^{16}$). Let H be a $(D^4, D, 1/4)$ expander (an explicit constructions for H exist, cf. [33].) We define $\{G_m\}_{m\in\mathbb{N}_{>0}}$ by*

$$G_1 := H^2, \qquad G_m := G_{m-1}^2 ⓩ H \text{ for } m > 1. \tag{2.1}$$

PROPOSITION 2.13 ([26]). *For any $m \in \mathbb{N}_{>0}$, the graph G_m is a $(D^{4m}, D^2, 1/2)$-graph.*

In the next section we will use the following lemma.

LEMMA 2.14. *Let G be a D-regular graph and S be the set of vertices of a connected component of G^2. Then $\lambda(G^2[S]) < 1$.*

Proof. Let $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ be the eigenvalues of $G^2[S]$. Since $G^2[S]$ is connected, Lemma 2.5 implies that $\lambda_1 > \lambda_2$. Now assume that -1 is an eigenvalue of $G^2[S]$ with eigenvector $\overline{v}$. Then the vector $\overline{v}'$ defined by $\overline{v}'_v = \overline{v}_v$ for all $v \in S$ and $\overline{v}'_v = 0$ otherwise is the eigenvector for eigenvalue -1 of the graph G^2. But G^2 can not have a negative eigenvalue as every eigenvalue of G^2 is a square of a real number. Therefore $\lambda_1 \neq \lambda_N$ and $\lambda(G^2[S]) < 1$ as claimed. ∎

3. A class of expanders definable in FO. In this section we define a formula such that the underlying graphs of its models are expanders. We start with a high-level description of the formula. Let $\{G_m\}_{m\in\mathbb{N}_{>0}}$ be as in Definition 2.12. Loosely speaking, each model of our formula is a structure which consists of the disjoint union of $G_1, \ldots, G_n$ for some $n \in \mathbb{N}_{>0}$ with some underlying tree structure connecting G_{m-1} to G_m for all $m \in \{2, \ldots, n\}$. For illustration see Figure 2. The tree structure enables us to provide an FO-checkable certificate for the construction of expanders. The tree structure is a D^4-ary tree, that is used to connect a vertex v of G_{m-1} to every vertex of the copy of H which will replace v in G_m. We use D^4 relations $\{F_k\}_{k\in([D]^2)^2}$ to enforce an ordering on the D^4 children of each vertex. We use additional relations to encode rotation maps. For $i, j \in [D]^2$ let $E_{i,j}$ be a binary relation. For every pair $i, j \in [D]^2$ we represent an edge $\{v, w\}$ in G_m by the two tuples $(v, w) \in E_{i,j}(A)$ and $(w, v) \in E_{j,i}(A)$. This allows us to encode the relationship $\text{ROT}_{G_m}(v, i) = (w, j)$ in first-order logic using the formula '$E_{i,j}(v, w)$'.

We use auxiliary relations R and L_k for $k \in ([D]^2)^2$, to force the models to be degree regular. The relation R contains the tuple (r, r) for the root r of the tree, and L_k will contain the tuple (v, v) for every leaf v of the tree.

We now give the precise definition of the formula. We use $[n] := \{0, 1, \ldots, n-1\}$ for $n \in \mathbb{N}$. Let

$$\sigma := \big\{\{E_{i,j}\}_{i,j\in[D]^2}, \{F_k\}_{k\in([D]^2)^2}, R, \{L_k\}_{k\in([D]^2)^2}\big\},$$

where $E_{i,j}$, F_k, R and L_k are binary relation symbols for $i, j \in [D]^2$ and $k \in ([D]^2)^2$. For convenience we introduce auxiliary relations E and F with the property that for every σ-structure A we have $E(A) := \bigcup_{i,j\in[D]^2} E_{i,j}(A)$ and $F(A) := \bigcup_{k\in([D]^2)^2} F_k(A)$.

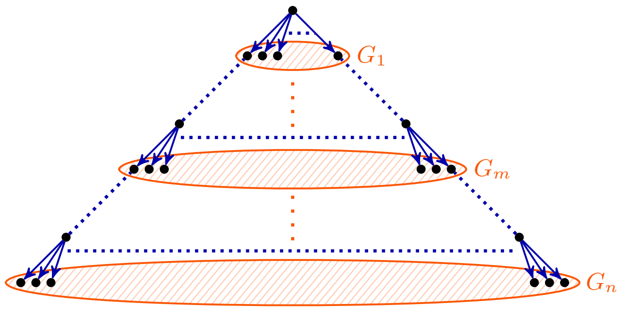


Fig. 2: Schematic representation of a model of $\varphi_{\textcircled{z}}$, where the parts in red (grey) only contain relations from E and relations in F are blue (black). Relation R and L are omitted.

In any formula we can reverse using these auxiliary relations by replacing formulas of the form "$E(x,y)$" by "$\bigvee_{i,j\in[D]^2} E_{i,j}(x,y)$" and formulas of the form "$F(x,y)$" by "$\bigvee_{k\in([D]^2)^2} F_k(x,y)$" below.

We use the following formula $\varphi_{\text{root}}(x) := \forall y \neg F(y,x)$ and we say that an element $a \in U(A)$ is a root of a structure A if $A \models \varphi_{\text{root}}(a)$.

We now define a formula φ_{tree}, which expresses that any model restricted to the relation F locally looks like a D^4-ary tree. More precisely, the formula defines that the structure has no more than one root, that every other vertex has exactly one parent and every vertex has either no children or exactly one child for each of the D^4 relations F_k. It also defines the self-loops used to make the structure degree regular.

$$
\begin{aligned}
\varphi_{\text{tree}} := \exists^{\leq 1} x \varphi_{\text{root}}(x) \wedge \forall x \Big(\big(\varphi_{\text{root}}(x) \wedge R(x,x)\big) \vee \\
\big(\exists^{=1} y F(y,x) \wedge \neg\exists y R(x,y) \wedge \neg\exists y R(y,x)\big)\Big) \wedge \\
\forall x \Bigg(\Big[\neg\exists y F(x,y) \wedge \bigwedge_{k\in([D]^2)^2} L_k(x,x) \wedge \\
\forall y \big(y \neq x \rightarrow \bigwedge_{k\in([D]^2)^2} \neg L_k(x,y) \wedge \bigwedge_{k\in([D]^2)^2} \neg L_k(y,x)\big)\Big] \\
\vee \Big[\neg\exists y \bigvee_{k\in([D]^2)^2} \big(L_k(x,y) \vee L_k(y,x)\big) \wedge \bigwedge_{k\in([D]^2)^2} \exists y_k \Big(x \neq y_k \wedge F_k(x,y_k) \wedge \\
\big(\bigwedge_{k'\in([D]^2)^2, k'\neq k} \neg F_{k'}(x,y_k)\big) \wedge \forall y (y \neq y_k \rightarrow \neg F_k(x,y))\Big)\Big]\Bigg).
\end{aligned}
$$

The formula $\varphi_{\text{rotationMap}}$ will define the properties the relations in E need to have in order to encode rotation maps of D^2-regular graphs. For this we make sure that the edge colours encode a map, i.e. for any pair of a vertex x and index $i \in [D]^2$ there is only one pair of vertex y and index $j \in [D]^2$ such that $E_{i,j}(x,y)$ holds and that the map is self inverse, i.e. if $E_{i,j}(x,y)$ then $E_{j,i}(y,x)$.

$$
\varphi_{\text{rotationMap}} := \forall x \forall y \Big(\bigwedge_{i,j\in[D]^2} (E_{i,j}(x,y) \rightarrow E_{j,i}(y,x)) \Big) \wedge
$$

$$\forall x\Big(\bigwedge_{i\in[D]^2}\Big(\bigvee_{j\in[D]^2}\big(\exists^{=1}yE_{i,j}(x,y)\wedge\bigwedge_{\substack{j'\in[D]^2\\ j'\neq j}}\neg\exists yE_{i,j'}(x,y)\big)\Big)\Big)$$

We now define a formula φ_{base} which expresses that every root x of a structure has a self-loop (x,x) in each relation $E_{i,j}$ and that the D^4 children of a root form G_1. Let H be the $(D^4,D,1/4)$-graph from Definition 2.12. We assume that H has vertex set $([D]^2)^2$. We then identify vertex $k\in([D]^2)^2$ with the element y such that $(x,y)\in F_k(A)$ for each root x. Let $\text{ROT}_H:([D]^2)^2\times[D]\to([D]^2)^2\times[D]$ be any rotation map of H. Fixing a rotation map for H fixes the rotation map for H^2. Recall that $G_1:=H^2$. We can define G_1 by a conjunction over all edges of G_1.

$$\begin{aligned}\varphi_{\text{base}}:=&\forall x\Big(\varphi_{\text{root}}(x)\to\Big[\bigwedge_{i,j\in[D]^2}\Big(E_{i,j}(x,x)\wedge\\ &\forall y\Big(x\neq y\to\big(\neg E_{i,j}(x,y)\wedge\neg E_{i,j}(y,x)\big)\Big)\Big)\wedge\\ &\bigwedge_{\substack{\text{ROT}_{H^2}(k,i)=(k',i')\\ k,k'\in([D]^2)^2\\ i,i'\in[D]^2}}\exists y\exists y'\big(F_k(x,y)\wedge F_{k'}(x,y')\wedge E_{i,i'}(y,y')\big)\Big]\Big)\end{aligned}$$

We will now define a formula $\varphi_{\text{recursion}}$ which will ensure that level m of the tree contains G_m. Recall that $G_m:=G_{m-1}^2\textcircled{z}H$. We therefore express that if there is a path of length two between two vertices x,z then for every pair $i,j\in[D]$ there is an edge connecting the corresponding children of x and z according to the definition of the zig-zag product. Here it is important that x and z either both have no children in the underlying tree structure or they both have children. This will also be encoded in the formula.

$$\begin{aligned}\varphi_{\text{recursion}}:=&\forall x\forall z\Big[\Big(\neg\exists yF(x,y)\wedge\neg\exists yF(z,y)\Big)\vee\\ &\bigwedge_{\substack{k_1',k_2'\in[D]^2\\ \ell_1',\ell_2'\in[D]^2}}\Big(\exists y\big[E_{k_1',\ell_1'}(x,y)\wedge E_{k_2',\ell_2'}(y,z)\big]\to\\ &\bigwedge_{\substack{i,j,i',j'\in[D],k,\ell\in([D]^2)^2\\ \text{ROT}_H(k,i)=((k_1',k_2'),i')\\ \text{ROT}_H((\ell_2',\ell_1'),j)=(\ell,j')}}\exists x'\exists z'\big[F_k(x,x')\wedge F_\ell(z,z')\wedge E_{(i,j),(j',i')}(x',z')\big]\Big)\Big]\end{aligned}$$

We finally let $\varphi_{\textcircled{z}}:=\varphi_{\text{tree}}\wedge\varphi_{\text{rotationMap}}\wedge\varphi_{\text{base}}\wedge\varphi_{\text{recursion}}$. This concludes defining the formula.

3.1. Proving expansion. In this section we prove that the formula $\varphi_{\textcircled{z}}$ defines a property of expanders on bounded-degree relational structures.

Let $d:=2D^2+D^4+1$, which is chosen in such a way to allow for any element of a σ-structure in $\mathcal{C}_{\sigma,d}$ to be in $2D^2$ E-relations (G_m is D^2 regular and every edge of G_m is modelled by two tuples), to have either D^4 F-children or D^4 L-self-loops and to either have one F-parent or be in one R-self-loop.

To each model A of $\varphi_{\textcircled{z}}$ we will associate an undirected (with parallel edges and self loops) graph $\underline{G}(A)$ with vertex set $U(A)$. For every tuple in each of the relations of A, the graph $\underline{G}(A)$ will have an edge. We will define $\underline{G}(A)$ by a rotation map, which extends the rotation map encoded by the relation E. For this let $I := \{0\} \sqcup ([D]^2)^2 \sqcup [D]^2$ be an index set. Formally, we define the *underlying graph* $\underline{G}(A)$ of a model A of $\varphi_{\textcircled{z}}$ to be the undirected graph with vertex set $U(A)$ given by the rotation map $\mathrm{ROT}_{\underline{G}(A)} : A \times I \to A \times I$ defined by

$$\mathrm{ROT}_{\underline{G}(A)}(v, i) := \begin{cases} (v, 0) & \text{if } i = 0 \text{ and } (v, v) \in R(A) \\ (w, j) & \text{if } i = 0 \text{ and } (w, v) \in F_j(A) \\ (w, 0) & \text{if } i \in ([D]^2)^2 \text{ and } (v, w) \in F_i(A) \\ (v, i) & \text{if } i \in ([D]^2)^2 \text{ and } (v, v) \in L_i(A) \\ (w, j) & \text{if } i \in [D]^2 \text{ and } (v, w) \in E_{i,j}(A). \end{cases}$$

We can understand this rotation map as labelling the tuples containing an element v as follows: $(v, v) \in R(A)$ or $(w, v) \in F_k(A)$ respectively is labelled by 0, $(v, w) \in F_k(A)$ or $(v, v) \in L_k(A)$ respectively is labelled by k and $(v, w) \in E_{i,j}(A)$ is labelled by i. Note that $\underline{G}(A)$ is $(D^2 + D^4 + 1)$-regular. We chose the notion of an underlying graph here instead of the Gaifman graph, and it is more convenient in particular for using results from [33]. However the Gaifman graph can be obtained from the underlying graph by ignoring self-loops and multiple edges.

Theorem 3.1. *There is an $\epsilon > 0$ such that the class $\{\underline{G}(A) \mid A \models \varphi_{\textcircled{z}}\}$ is a family of ϵ-expanders.*

In the rest of this section, we give the proof of Theorem 3.1. Let A be a model of $\varphi_{\textcircled{z}}$. Let $A|_F$ be the $\{(F_k)_{k\in([D]^2)^2}\}$-structure $(U(A), (F_k(A))_{k\in([D]^2)^2})$. Recall that we denote the Gaifman graph of $A|_F$ by $G(A|_F)$. Let $A|_E$ be the $\{(E_{i,j})_{i,j\in[D]^2}\}$-structure $(U(A), (E_{i,j}(A))_{i,j\in[D]^2})$. We further define the *underlying graph* $\underline{G}(A|_E)$ of $A|_E$ as the undirected graph specified by the rotation map $\mathrm{ROT}_{\underline{G}(A|_E)}$ which is defined by $\mathrm{ROT}_{\underline{G}(A|_E)}(v, i) := (w, j)$ if $(v, w) \in E_{i,j}(A)$. This is well defined as $A \models \varphi_{\mathrm{rotationMap}}$. We use the substructures $G(A|_F)$ and $\underline{G}(A|_E)$ to express the structural properties of models of $\varphi_{\textcircled{z}}$. More precisely, we want to prove that $G(A|_F)$ is a rooted balanced full tree and $\underline{G}(A|_E)$ is the disjoint union of the expanders $G_1, \ldots, G_n$ for some $n \in \mathbb{N}$ (Lemma 3.10). To prove this we use two technical lemmas (Lemma 3.2 and Lemma 3.5). Lemma 3.2 intuitively shows that the children in $G(A|_F)$ of each connected part of $\underline{G}(A|_E)$ form the zig-zag product with H of the square of the connected part. Lemma 3.5 shows that $G(A|_F)$ is connected. To prove Theorem 3.1 we use that a tree with an expander on each level has good expansion. Loosely speaking, this is true because cutting the tree 'horizontally' takes many edge deletions and for cutting the tree 'vertically' we cut many expanders.

Lemma 3.2. *Let A be a model of $\varphi_{\textcircled{z}}$ and assume S is the set of all vertices belonging to a connected component of $(\underline{G}(A|_E))^2$ not containing a root and let $S' := \{w \in U(A) \mid (v, w) \in F(A), v \in S\}$. If $S' \neq \emptyset$ then $\underline{G}(A|_E)[S']$ is a connected component of $\underline{G}(A|_E)$ and $\underline{G}(A|_E)[S'] \cong ((\underline{G}(A|_E))^2[S]) \textcircled{z} H$.*

We use connected components of $(\underline{G}(A|_E))^2$ as the square of a connected component of $\underline{G}(A|_E)$ may not be connected, in which case the zig-zag product with H of the square of the connected component cannot be connected.

Proof of Lemma 3.2. Assume that $S' \neq \emptyset$. We first show that $\underline{G}(A|_E)[S'] \cong ((\underline{G}(A|_E))^2[S]) \textcircled{z} H$. For this we use the following two claims.

CLAIM 3.3. *If*

$$\mathrm{ROT}_{(\underline{G}(A|_E))^2[S] \textcircled{z} H}((u,k),(i,j)) = ((w,\ell),(j',i'))$$

for some $u, w \in S$, $k, \ell \in ([D]^2)^2$, $i, j, i', j' \in [D]$ *then there is* $v \in S$ *such that* $(u,v) \in E_{k'_1,\ell'_1}(A)$ *and* $(v,w) \in E_{k'_2,\ell'_2}(A)$ *where* $\mathrm{ROT}_H(k,i) = ((k'_1,k'_2),i')$ *and* $\mathrm{ROT}_H((\ell'_2,\ell'_1),j) = (\ell,j')$.

Proof. By the precondition of the Claim and the definition of the zig-zag product, we have that $\mathrm{ROT}_{(\underline{G}(A|_E))^2[S]}(u,(k'_1,k'_2)) = (w,(\ell'_2,\ell'_1))$ for $\mathrm{ROT}_H(k,i) = ((k'_1,k'_2),i')$ and $\mathrm{ROT}_H((\ell'_2,\ell'_1),j) = (\ell,j')$.

Since $\mathrm{ROT}_{(\underline{G}(A|_E))^2[S]}$ is equal to $\mathrm{ROT}_{(\underline{G}(A|_E))^2}$ restricted to elements of the set S, we have that $\mathrm{ROT}_{(\underline{G}(A|_E))^2}(u,(k'_1,k'_2)) = (w,(\ell'_2,\ell'_1))$. Consequently, by the definition of the square of a graph $\mathrm{ROT}_{(\underline{G}(A|_E))^2}(u,(k'_1,k'_2)) = (w,(\ell'_2,\ell'_1))$ implies that there is v such that $\mathrm{ROT}_{\underline{G}(A|_E)}(u,k'_1) = (v,\ell'_1)$ and $\mathrm{ROT}_{\underline{G}(A|_E)}(v,k'_2) = (w,\ell'_2)$. ∎

CLAIM 3.4. *If* $(u,v) \in E_{k'_1,\ell'_1}(A)$ *and* $(v,w) \in E_{k'_2,\ell'_2}(A)$ *for* $u, v, w \in U(A)$, $k'_1, k'_2, \ell'_1, \ell'_2 \in ([D]^2)^2$ *and there is* $u' \in U(A)$ *with* $(u,u') \in F(A)$ *then there is* $w' \in U(A)$ *such that* $(w,w') \in F(A)$. *Furthermore for any* $i, i', j, j' \in [D]$ *there are* $\tilde{u}, \tilde{w} \in U(A)$, $k, \ell \in ([D]^2)^2$ *such that* $(\tilde{u},\tilde{w}) \in E_{(i,j),(j'i')}(A)$ *for* $(u,\tilde{u}) \in F_k(A)$ *and* $(w,\tilde{w}) \in F_\ell(A)$ *where* $\mathrm{ROT}_H(k,i) = ((k'_1,k'_2),i')$ *and* $\mathrm{ROT}_H((\ell'_2,\ell'_1),j) = (\ell,j')$.

Proof. We only use that $A \models \varphi_{\mathrm{recursion}}$. Since $\varphi_{\mathrm{recursion}}$ has the form $\forall x \forall z \psi(x,z)$ for some formula $\psi(x,z)$ we know that $A \models \psi(u,w)$. Since $(u,u') \in F(A)$ we have $A \not\models \neg\exists y F(u,y) \wedge \neg\exists y F(w,y)$. Since $A \models \exists y \big[E_{k'_1,\ell'_1}(u,y) \wedge E_{k'_2,\ell'_2}(w,z)\big]$

$$A \models \bigwedge_{\substack{i,j,i',j' \in [D], k,\ell \in ([D]^2)^2 \\ \mathrm{ROT}_H(k,i) = ((k'_1,k'_2),i') \\ \mathrm{ROT}_H((\ell'_2,\ell'_1),j) = (\ell,j')}} \exists x' \exists z' \big[F_k(u,x') \wedge F_\ell(w,z') \wedge E_{(i,j),(j',i')}(x',z')\big]$$

Since H is D-regular, for every $k'_1, k'_2 \in [D]^2$ and $i, i' \in [D]$, there is $k \in ([D]^2)^2$ such that $\mathrm{ROT}_H(k,i) = ((k'_1,k'_2,i')$ (and the same for ℓ'_1, ℓ'_2, j, j'). Thus, the above conjunction is not empty. This further implies that for any $i, i', j, j' \in [D]$ there are $\tilde{u}, \tilde{w} \in U(A)$, $k, \ell \in ([D]^2)^2$ as claimed. In particular there is $w' \in U(A)$ such that $(w,w') \in F(A)$. ∎

We will argue that for every element $w \in S$ there is a $w' \in S'$ such that $(w,w') \in F(A)$. For this pick any $u' \in S'$. Let $u \in S$ be the element such that $(u,u') \in F(A)$. By combining Lemma 2.14, Theorem 2.11 and Lemma 2.5 it follows that $((\underline{G}(A|_E))^2[S]) \textcircled{z} H$ is connected. Therefore, there exists a path $(u'_0, \dots, u'_m)$ in $((\underline{G}(A|_E))^2[S]) \textcircled{z} H$ from $u'_0 = (u,(k_1,k_2))$ to $u'_m = (w,(\ell_1,\ell_2))$ for some $k_1, k_2, \ell_1, \ell_2 \in [D]^2$. By Claim 3.3 there is a path $(u_0, v_0, u_1, v_1, \dots u_{m-1}, v_{m-1}, u_m)$ in $\underline{G}(A|_E)$ from $u_0 = u$ to $u_m = w$. By inductively using Claim 3.4 on the path we find w' such that $(w,w') \in F(A)$.

Combining this with $A \models \varphi_{\mathrm{tree}}$ implies that the map $f : S \times ([D]^2)^2 \to S'$, given by $f(v,k) = u$ if $(v,u) \in F_k(A)$, is well-defined. Furthermore, by Claim 3.3 and 3.4, we have that if it holds that $\mathrm{ROT}_{(\underline{G}(A|_E))^2[S] \textcircled{z} H}((u,k),(i,j)) = ((w,\ell),(j',i'))$ then

$$\mathrm{ROT}_{(\underline{G}(A|_E))[S']}(f((u,k)),(i,j)) = (f((w,\ell)),(j',i')).$$

This proves that f maps each edge in $((\underline{G}(A|_E))^2[S]) \textcircled{z} H$ injectively to an edge in $\underline{G}(A|_E)[S']$. Then the map f together with the corresponding edge map is an isomorphism from $((\underline{G}(A|_E))^2[S]) \textcircled{z} H$ to $\underline{G}(A|_E)$ as both are D^2-regular.

Moreover, $\underline{G}(A|_E)[S'] \cong ((\underline{G}(A|_E))^2[S]) \textcircled{z}\, H$ implies that $\underline{G}(A|_E)[S']$ is connected and D^2-regular. Since $A \models \varphi_{\text{rotationMap}}$ enforces that $\underline{G}(A|_E)$ is D^2-regular, no vertex $v \in S'$ can have neighbours which are not in S' and therefore $\underline{G}(A|_E)[S']$ is a connected component of $\underline{G}(A|_E)$. ∎

Lemma 3.5. *Let $A \in \mathcal{C}_{\sigma,d}$ be a model of $\varphi_{\textcircled{z}}$. Then every connected component of $G(A|_F)$ contains a root of A. In particular for every model $A \in \mathcal{C}_{\sigma,d}$ of $\varphi_{\textcircled{z}}$ the graph $G(A|_F)$ is connected.*

Note that the connectivity of $G(A|_F)$ for a model $A \in \mathcal{C}_{\sigma,d}$ of $\varphi_{\textcircled{z}}$ implies that A is connected as $G(A|_F)$ is a subgraph of the Gaifman graph of A containing the same set of vertices. Hence the following corollary follows immediately from Lemma 3.5.

Corollary 3.6. *Any model $A \in \mathcal{C}_{\sigma,d}$ of $\varphi_{\textcircled{z}}$ is connected.*

Proof of Lemma 3.5. Assume that there is a connected component of $G(A|_F)$ which contains no root of A and let G' to be a connected component of $G(A|_F)$ with vertex set $V \subseteq U(A)$ such that $A \not\models \varphi_{\text{root}}(v)$ for every $v \in V$. For the next claim we should have in mind that $(A|_F)[V]$ can be understood as a directed graph in which every vertex has in-degree 1 and the corresponding undirected graph G' is connected. Hence $(A|_F)[V]$ must consist of a set of disjoint directed trees whose roots form a directed cycle. Consequently G' has the structure as given in the following claim.

Claim 3.7. *G' contains a tuple of vertices $(c_0, \dots, c_{\ell-1})$ such that either $\ell = 2$ and $(c_0, c_1), (c_1, c_0) \in F(A)$ or $(c_0, \dots, c_{\ell-1})$ is a cycle. Furthermore, for every vertex v of G' there is exactly one path $(p_0, \dots, p_m)$ in G' with $p_0 = v$, $p_m \in \{c_0, \dots, c_{\ell-1}\}$ and $p_i \notin \{c_0, \dots, c_{\ell-1}\}$ for all $i \in [m-1]$.*

Proof. We first identify a cycle (or a pair as mentioned in the statement) by traversing along the path of incoming edges until encountering a repeated vertex. Let v_0 be any vertex in G' and let $S_0 = \{v_0\}$. We will now recursively define v_i to be the vertex of G' such that $(v_i, v_{i-1}) \in F(A)$. Such a vertex always exists by the choice of G' (i.e. that no root is in G') and the fact that $A \models \varphi_{\text{tree}}$. Furthermore, such a vertex is unique as $A \models \varphi_{\text{tree}}$. We let $S_i := S_{i-1} \cup \{v_i\}$. Since $U(A)$ is finite, the chain $S_0 \subseteq S_1 \subseteq \dots \subseteq S_i \subseteq \dots$ must become stationary at some point. Let $i \in \mathbb{N}$ be the minimum index such that $S_{i-1} = S_i$ and let $j < i$ be such that $v_i = v_j$. Then $(v_j, v_{j+1}, \dots, v_{i-1}, v_i)$ is either a cycle in G' or (in case $j = i-1$) $(v_j, v_i), (v_i, v_j) \in F(A)$. Let $C = \{c_0, \dots, c_{\ell-1}\}$ be the vertices of the cycle or pair of vertices.

We now show that for every vertex v in G' there exists a unique path from v to C. We first note that since G' is connected, for every v, a path that satisfies the property as described in the assertion of the claim always exists. Assume that there exists one vertex v, from which there are two different paths to C, denoted by $(p_0 = v, \dots, p_m)$ and $(p'_0 = v, \dots, p'_{m'})$, respectively. We let $p_m = c_i$ and $p'_{m'} = c_j$. Let $k \le \min\{m, m'\}$ be the minimum index such that $p_k \neq p'_k$. Such an index must exist as the paths are different, and as $p_0 = p'_0 = v$, we also know that $k \ge 1$. Since $A \models \varphi_{\text{tree}}$ for every vertex w of G' there can only be one vertex w' of G' such that $(w', w) \in F(A)$. As $p_{m-1} \notin C$ and $(c_{(i-1) \bmod \ell}, p_m) \in F(A)$ it follows that $(p_m, p_{m-1}) \in F(A)$. Applying the argument inductively we get that $(p_k, p_{k-1}) \in F(A)$. The same argument works for the path $(p'_0, \dots, p'_{m'})$ and therefore $(p'_k, p'_{k-1}) \in F(A)$. By the choice of k we know that $p_{k-1} = p'_{k-1}$ and $p_k \neq p'_k$, which implies that there exists one vertex with two incoming edges. This contradicts the fact that $A \models \varphi_{\text{tree}}$. Thus, for every vertex v, there exists a unique path from v to C. This finishes the proof of the claim. ∎

Let S_0 be the vertex set of the connected component of $\underline{G}(A|_E)$ with $c_0 \in S_0$. Note that S_0 might not be contained in G'.

We now recursively define the infinite sequence of sets $S_i := \{w \in U(A) \mid (v,w) \in F(A), v \in S_{i-1}\}$ for each $i \in \mathbb{N}_{>0}$. Let $m_i := \max_{v \in S_i \cap V} \min_{j \in \{0,\dots,\ell-1\}} \{\mathrm{dist}_{G'}(c_j, v)\}$ and let $v_i \in S_i \cap V$ be a vertex of distance m_i from C in G'. Note here that m_i is well defined as $c_{i \bmod \ell} \in S_i$.

Claim 3.8. *$\underline{G}(A|_E)[S_i] = (\underline{G}(A|_E)[S_{i-1}])^2 \textcircled{z} H$.*

Proof. We show the stronger statement that $\underline{G}(A|_E)[S_i]$ is a connected component of $\underline{G}(A|_E)$, $(\underline{G}(A|_E)[S_i])^2 \textcircled{z} H = \underline{G}(A|_E)[S_{i+1}]$ and $\lambda(\underline{G}(A|_E)[S_i]) < 1$ for $i \in \mathbb{N}$ by induction.

$\underline{G}(A|_E)[S_0]$ is a connected component of $\underline{G}(A|_E)$ by choice of S_0. Let $\tilde{S} := \{w \in U(A) \mid (w,v) \in F(A), v \in S_0\}$.

We now argue that $(\underline{G}(A|_E))^2[\tilde{S}]$ is a connected component of $(\underline{G}(A|_E))^2$. Assuming the contrary, either a connected component of $(\underline{G}(A|_E))^2$ contains vertices from both $\tilde{S}$ and $A \setminus \tilde{S}$ or $(\underline{G}(A|_E))^2[\tilde{S}]$ splits into more than one connected component. Let S' be the vertices of a connected component as in the first case. Then $|S'| > 1$ and hence S' can not contain any root as a root is not in any E-relation with any element different from itself. Hence by Lemma 3.2 we get a connected component of $\underline{G}(A|_E)$ on the children of S' containing vertices both from S_0 and from $U(A) \setminus S_0$, which contradicts S_0 being a connected component of $\underline{G}(A|_E)$. Now let S' be a connected component as in the second case, and pick S' such that it does not contain a root (this is possible as there is at most one root). Then by Lemma 3.2 S_0 must have a non-empty intersection with at least two connected components of $\underline{G}(A|_E)$, which is a contradiction.

Thus, by Lemma 2.14 $\lambda((\underline{G}(A|_E))^2[\tilde{S}]) < 1$. But by Lemma 3.2 $\underline{G}(A|_E)[S_0] = ((\underline{G}(A|_E))^2[\tilde{S}]) \textcircled{z} H$. Then Theorem 2.11 and $\lambda(H) < 1$ ensure that $\lambda(\underline{G}(A|_E)[S_0]) < 1$.

For $i > 1$, by induction it holds that $\lambda(\underline{G}(A|_E)[S_{i-1}]) < 1$, which, together with Lemma 2.9 and Lemma 2.5, implies that $(\underline{G}(A|_E)[S_{i-1}])^2$ is a connected component[4] of $(\underline{G}(A|_E))^2$ and that $(\underline{G}(A|_E))^2[S_{i-1}] = (\underline{G}(A|_E)[S_{i-1}])^2$. Since $c_{i \bmod \ell} \in S_i$, by Lemma 3.2, we have that $\underline{G}(A|_E)[S_i]$ is a connected component of $\underline{G}(A|_E)$ and $\underline{G}(A|_E)[S_i] = (\underline{G}(A|_E)[S_{i-1}])^2 \textcircled{z} H$. Furthermore, using Lemma 2.9 and Theorem 2.11, this proves $\lambda(\underline{G}(A|_E)[S_i]) < 1$. ∎

Claim 3.9. *For every $v \in S_i$ there is $w \in V$ such that $(v,w) \in F(A)$.*

Proof. By Claim 3.8 we have that $\underline{G}(A|_E)[S_{i+1}] = (\underline{G}(A|_E)[S_i])^2 \textcircled{z} H$. This means that by definition of squaring and the zig-zag product we know that $|S_{i+1}| = D^4 \cdot |S_i|$. But as in addition $A \models \varphi_{\mathrm{tree}}$ we know that every element $v \in S_i$ will contribute to no more then D^4 elements to S_{i+1}. This means by construction of S_{i+1} that for every element in S_i there must be $w \in V$ such that $(v,w) \in F(A)$. ∎

Therefore, for every $i \in \mathbb{N}_{>0}$ there is $w_i \in V$ such that $(v_i, w_i) \in F(A)$ where v_i is the vertex of distance m_i from C in G' picked above. Let $(u_0, \dots, u_{m_i})$ be the path in G' from $u_0 = v_i$ to $u_{m_i} \in C$. Note that it is impossible that $w_i = u_1$. This is true as for the path $(u_0, ..., u_{m_i})$, we have that $(u_{j+1}, u_j) \in F(A)$ for all $j \in [m_i]$. Furthermore, since $v_i = u_0 \neq u_1$, assuming that $w_i = u_1$ would imply $(v_i, u_1), (u_2, u_1) \in F(A)$, which contradicts $A \models \varphi_{\mathrm{tree}}$. Then $(w_i, u_0, \dots, u_{m_i})$ is a path in G' from w_i to C.

[4] We remark that the statement that $(\underline{G}(A|_E)[S_{i-1}])^2$ is a connected component does not directly follow from the fact that $\underline{G}(A|_E)[S_{i-1}]$ is a connected component of $\underline{G}(A|_E)$, as the square of a connected bipartite graph is not necessarily connected.

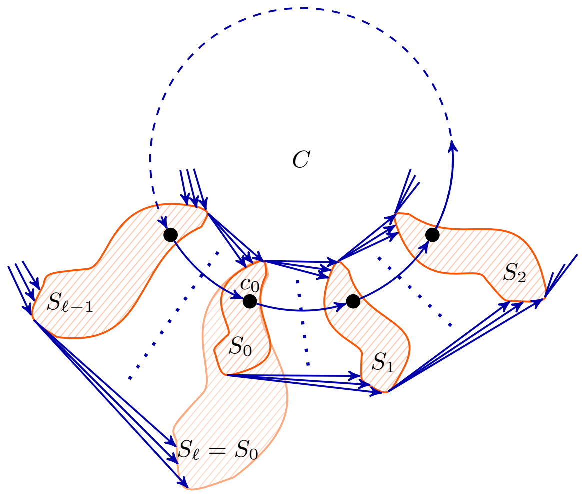


Fig. 3: Illustration of the proof of Lemma 3.5.

Since $w_i \in S_{i+1}$ by construction, Claim 3.7 implies that $m_{i+1} \geq m_i + 1$. Therefore $m_i \geq i + m_0$ inductively. But this yields a contradiction, because $\ell + m_0 \leq m_\ell = m_0$ and $\ell > 0$. See Figure 3 for an illustration. Therefore every connected component of $G(A|_F)$ must contain a root of A. Furthermore, since every connected component of $G(A|_F)$ must contain a root and since $A \models \exists^{\leq 1} x \varphi_{\text{root}}(x)$ there can not be more than one root, $G(A|_F)$ is connected. ◻

We let $\mathcal{P}_{\textcircled{Z}} := \mathcal{P}_{\varphi_{\textcircled{Z}}}$ for the formula $\varphi_{\textcircled{Z}}$ from Section 3.

Lemma 3.10. *Any (finite) model $A \in \mathcal{C}_{\sigma,d}$ of $\varphi_{\textcircled{Z}}$ has the following structure.*

- *Either $U(A) = \emptyset$ or $|U(A)| = \sum_{m=0}^{n} D^{4m}$ for some $n \in \mathbb{N}_{n \geq 1}$.*
- *$G(A|_F)$ is a rooted balanced full D^4-ary tree, where the root is the unique element $r \in U(A)$ for which $A \models \varphi_{\text{root}}(r)$.*
- *$\underline{G}(A|_E)[T_m] \cong G_m$ where G_m is defined as in Definition 2.12 and T_m is the set of vertices of distance m to r in the tree $G(A|_F)$ for any $m \in \{1, \dots, n\}$.*

Furthermore for every $n \in \mathbb{N}_{\geq 1}$ there is a model of $\varphi_{\textcircled{Z}}$ of size $\sum_{m=0}^{n} D^{4m}$.

Proof. First note that the empty structure $A_\emptyset \in \mathcal{P}_{\textcircled{Z}}$ as $A_\emptyset \models \exists^{\leq 1} x \varphi_{\text{root}}(x)$ and therefore $A_\emptyset \models \varphi_{\textcircled{Z}}$ as $\varphi_{\textcircled{Z}}$ is a conjunction of $\exists^{\leq 1} x \varphi_{\text{root}}(x)$ and universally quantified formulas. Hence $U(A) = \emptyset$ is possible. Now assume that A is a model of $\varphi_{\textcircled{Z}}$ and $U(A) \neq \emptyset$. Then Lemma 3.5 implies that $G(A|_F)$ is connected. Combining this with $A \models \varphi_{\text{tree}}$ proves that $G(A|_F)$ is a rooted tree. Let n be the maximum distance of any vertex in $G(A|_F)$ to the root and let T_m be the vertices of distance m to the root for $m \leq n$. Then $\underline{G}(A|_E)[T_1] \cong G_1$ because $A \models \varphi_{\text{base}}$. Now assume towards an inductive proof that $\underline{G}(A|_E)[T_m] \cong G_m$ for some fixed $m \in \mathbb{N}_{>0}$. Since $\lambda(G_m) < 1$ by Lemma 2.9 and Lemma 2.5 we get that $(\underline{G}(A|_E))^2[T_m]$ is a connected component of $(\underline{G}(A|_E))^2$. Hence by Lemma 3.2 we get that $\underline{G}(A|_E)[T_{m+1}] \cong G_{m+1}$. Since G_m has D^{4m} vertices this also proves that A has $\sum_{m=0}^{n} D^{4m}$ vertices. Furthermore, for $n \in \mathbb{N}$ the existence of a model of $\varphi_{\textcircled{Z}}$ of size $\sum_{m=0}^{n} D^{4m}$ is straightforward by the

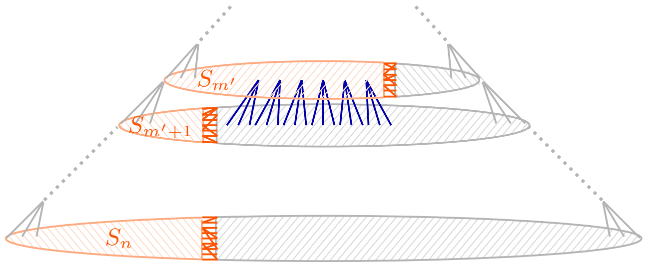


Fig. 4: Schematic representation of S crossing edges (orange and blue) in the underlying undirected graph in the case of $m' < n$.

construction of the formula $\varphi_{\textcircled{z}}$. ∎

Now we are ready to prove Theorem 3.1.

Proof of Theorem 3.1. We will prove that for $\epsilon = D^2/12$ the claimed is true. Let A be the model of $\varphi_{\textcircled{z}}$ of size $\sum_{m=0}^{n} D^{4m}$ and $S \subseteq U(A)$ with $|S| \leq (\sum_{m=0}^{n} D^{4m})/2$. Let T_m be the vertices of distance m to the root of the tree $G(A|_F)$ and let $S_m := T_m \cap S$.

We can assume that $|S| > 1$ as every vertex has degree at least ϵ. Let us first assume that $|S_m| \leq D^{4m}/2$ for all $m \in [n]$. Then because G_m is a $D^2/4$-expander (this follows directly from Theorem 2.7 as $\lambda(G_m) \leq 1/2$ by Proposition 2.13) and $\underline{G}(A|_E)[T_m] \cong G_m$ we know that

$$|\langle S, \overline{S}\rangle_{\underline{G}(A)}| \geq \sum_{m=1}^{n} \frac{D^2}{4}|S_m| \geq \frac{D^2}{12}\sum_{m=0}^{n}|S_m| = \frac{D^2}{12}|S|.$$

Now assume the opposite and choose m' to be the largest index such that

$$|S_{m'}| > \frac{|T_{m'}|}{2} = \frac{D^{4m'}}{2}. \tag{3.1}$$

We will use the following claim.

Claim 3.11. *$\sum_{m=0}^{\tilde{m}-1}|T_m| \leq \frac{1}{2}|T_{\tilde{m}}|$ for all $\tilde{m} \leq n$.*

Proof. Inductively, we argue that

$$\sum_{m=0}^{\tilde{m}-1}|T_m| = \sum_{m=0}^{\tilde{m}-2}|T_m| + |T_{\tilde{m}-1}| \leq \frac{1}{2}(3|T_{\tilde{m}-1}|) \leq \frac{1}{2}|T_{\tilde{m}}|.$$ ∎

Claim 3.11 implies that $\frac{3}{4} \cdot |T_n| \geq \frac{1}{2}|T_n| + \frac{1}{2}\sum_{m=0}^{n-1}|T_m| = \frac{1}{2}|A| \geq |S| \geq |S_n|$. In the case that $m' = n$, using that G_n is a $D^2/4$-expander we get

$$|\langle S, \overline{S}\rangle_{\underline{G}(A)}| \geq \frac{D^2}{4}(|T_n| - |S_n|) \geq \frac{D^2}{16}|T_n| \geq \frac{D^2}{12}|S|.$$

Assume now that $m' < n$. Since S is the disjoint union of all S_m we know that the set $\langle S, \overline{S}\rangle_{\underline{G}(A)}$ contains the set $\langle S_m, T_m \setminus S_m\rangle_{\underline{G}(A)}$, and for all $m \in \{m'+1, \dots, n\}$, the sets $\langle T_{m'} \setminus S_{m'}, T_{m'}\rangle_{\underline{G}(A)}$ and $\langle S_{m'}, T_{m'+1} \setminus S_{m'+1}\rangle_{\underline{G}(A)}$, which are all pairwise disjoint. Since every vertex in $T_{m'}$ has D^4 neighbours in $T_{m'+1}$ and on the other hand every vertex in $T_{m'+1}$ has one neighbour in $T_{m'}$ we know that $|\langle S_{m'}, T_{m'+1} \setminus S_{m'+1}\rangle_{\underline{G}(A)}| = |\langle S_{m'}, T_{m'+1}\rangle_{\underline{G}(A)}| - |\langle S_{m'}, S_{m'+1}\rangle_{\underline{G}(A)}| \geq D^4|S_{m'}| - |S_{m'+1}| \geq D^4(|S_{m'}| - D^{4m'}/2)$. Since additionally G_m is an $D^2/4$-expander for every m we get

$$\begin{aligned}
|\langle S, \overline{S}\rangle_{\underline{G}(A)}| &\geq \sum_{m>m'} \frac{D^2}{4}|S_m| + \frac{D^2}{4}|T_{m'} \setminus S_{m'}| + D^4\Big(|S_{m'}| - \frac{D^{4m'}}{2}\Big) \\
&= \frac{D^2}{4}\sum_{m>m'}|S_m| + \Big(D^4 - \frac{D^2}{2}\Big)|S_{m'}| - \Big(D^4 - \frac{D^2}{2}\Big)\frac{|T_{m'}|}{2} + \frac{D^2}{8}|S_{m'}| + \frac{D^2}{8}|S_{m'}| \\
&\overset{\text{Equation 3.1}}{\geq} \frac{D^2}{4}\sum_{m>m'}|S_m| + \frac{D^2}{8}|S_{m'}| + \frac{D^2}{8}\Big(\frac{|T_{m'}|}{2}\Big) \\
&\overset{\text{Claim 3.11}}{\geq} \frac{D^2}{4}\sum_{m>m'}|S_m| + \frac{D^2}{8}|S_{m'}| + \frac{D^2}{8}\sum_{m<m'}|T_m| \\
&\overset{|T_m|\geq|S_m|}{\geq} \frac{D^2}{12}|S|.
\end{aligned}$$

□

By the choice of ϵ this shows that the models of $\varphi_{\textcircled{z}}$ are a class of ϵ-expanders.

4. On the non-testability of a Π_2-property.

In this section we prove that there exists an FO property on relational structures in Π_2 that is not testable. To do so, we first prove that the property $P_{\varphi_{\textcircled{z}}}$ defined by the formula $\varphi_{\textcircled{z}}$ in Section 3 is not testable. Later we prove that $\varphi_{\textcircled{z}}$ is equivalent to a sentence in Π_2.

4.1. Non-testability.

Recall that r-types are the isomorphism classes of r-balls and that restricted to the class $\mathcal{C}_{\sigma,d}$ there are finitely many r-types. Let $\tau_1, \ldots, \tau_t$ be a list of all r-types of bounded degree d. We let $\rho_{A,r}$ be the r-type distribution of A, i. e.

$$\rho_{A,r}(X) := \frac{\sum_{\tau\in X} |\{a \in U(A) \mid \mathcal{N}_r^A(a) \in \tau\}|}{|U(A)|}$$

for any $X \subseteq \{\tau_1, \ldots, \tau_t\}$. For two σ-structures A and B we define the sampling distance of depth r as $\delta_{\odot}^r(A, B) := \sup_{X\subseteq\{\tau_1,\ldots,\tau_t\}} |\rho_{A,r}(X) - \rho_{B,r}(X)|$. Note that $\delta_{\odot}^r(A, B)$ is just the total variation distance between $\rho_{A,r}, \rho_{B,r}$, and it holds that

$$\delta_{\odot}^r(A, B) = \frac{1}{2}\sum_{i=1}^{t} |\rho_{A,r}(\{\tau_i\}) - \rho_{B,r}(\{\tau_i\})|.$$

Then the sampling distance of A and B is defined as

$$\delta_{\odot}(A, B) := \sum_{r=0}^{\infty} \frac{1}{2^r} \cdot \delta_{\odot}^r(A, B).$$

The following theorem was proven for simple graphs and easily extends to σ-structures.

Theorem 4.1 ([30]). *For every $\lambda > 0$ there is a positive integer n_0 such that for every σ-structure $A \in \mathcal{C}_{\sigma,d}$ there is a σ-structure $H \in \mathcal{C}_{\sigma,d}$ such that $|H| \leq n_0$ and $\delta_{\odot}(A, H) \leq \lambda$.*

We make use of the following definition of repairable properties.

Definition 4.2 ([1]). *Let $\epsilon \in (0, 1]$. A property $\mathcal{P} \subseteq \mathcal{C}_{\sigma,d}$ is* ϵ-repairable[5] *on*

[5] In [1], the notion of repairability is called locality.

$\mathcal{C}_{\sigma,d}$ if there are numbers $r := r(\epsilon) \in \mathbb{N}$, $\lambda := \lambda(\epsilon) > 0$ and $n_0 := n_0(\epsilon) \in \mathbb{N}$ such that for any σ-structure $A \in \mathcal{P}$ and $B \in \mathcal{C}_{\sigma,d}$ both on $n \geq n_0$ vertices, if $\sum_{i=1}^{t} |\rho_{A,r}(\{\tau_i\}) - \rho_{B,r}(\{\tau_i\})| < \lambda$ then B is ϵ-close to P, where $\tau_1, \dots, \tau_t$ is a list of all r-types of bounded degree d.

The property $\mathcal{P}$ is repairable on $\mathcal{C}_{\sigma,d}$ if it is ϵ-repairable on $\mathcal{C}_{\sigma,d}$ for every $\epsilon \in (0, 1]$.

The following theorem relating testable properties and repairable properties was proven in [1].

Theorem 4.3 ([1]). *For every property $\mathcal{P} \in \mathcal{C}_{\sigma,d}$, $\mathcal{P}$ is testable if and only if $\mathcal{P}$ is repairable on $\mathcal{C}_{\sigma,d}$.*

We recall that $\mathcal{P}_{\textcircled{z}} := \mathcal{P}_{\varphi_{\textcircled{z}}}$ where $\varphi_{\textcircled{z}}$ is the formula from Section 3. We also let σ, D and d be as defined in Section 3.

Theorem 4.4. *$\mathcal{P}_{\textcircled{z}}$ is not testable on $\mathcal{C}_{\sigma,d}$.*

Proof. We prove non-repairability for $\mathcal{P}_{\textcircled{z}}$ and get non-testability with Theorem 4.3. Let $\epsilon := 1/(144D^2)$ and let $r \in \mathbb{N}$, $\lambda > 0$ and $n_0 \in \mathbb{N}$ be arbitrary. We set $\lambda' := \lambda/(t2^{r+1})$, where $\tau_1, \dots, \tau_t$ are all r-types of bounded degree d, and let n_0' be the positive integer from Theorem 4.1 corresponding to λ'. We now pick $n \in \mathbb{N}$ such that $n = \sum_{i=0}^{k} D^{4i}$ for some $k \in \mathbb{N}$, $n \geq 4n_0$ and $n \geq 4(n_0'/\lambda)$. Let $A \in \mathcal{C}_{\sigma,d}$ be a model of $\varphi_{\textcircled{z}}$ on n elements. By Theorem 4.1 there is a structure $H \in \mathcal{C}_{\sigma,d}$ on $m \leq n_0'$ elements such that $\delta_{\odot}(A, H) \leq \lambda$. Let B be the structure consisting of $\lfloor n/m \rfloor$ copies of H and $n \bmod m$ isolated elements (elements not being contained in any tuple). Note that we picked B such that $|A| = |B|$.

We will first argue that B is in fact ϵ-far from having the property $\mathcal{P}_{\textcircled{z}}$. First we rename the elements from $U(B)$ in such a way that $U(A) = U(B)$ and the number $\sum_{\tilde{R} \in \sigma} |\tilde{R}(A) \triangle \tilde{R}(B)|$ of edge modifications to turn A and B into the same structure is minimal. Pick a partition $U(A) = U(B) = S \sqcup S'$ in such a way that $(S \times S') \cap \tilde{R}(B) = \emptyset$, $(S' \times S) \cap \tilde{R}(B) = \emptyset$ for any $\tilde{R} \in \sigma$ and $||S| - |S'||$ is minimal among all such partitions. Assume that $|S| \leq |S'|$. Since the connected components of $G(B)$ are of size at most m we know that $||S| - |S'|| \leq m$. This is because otherwise we can get a partition $U(B) = T \sqcup T'$ with $||T| - |T'|| < ||S| - |S'||$ by picking all elements of any connected component of $G(\mathcal{B})$, which is contained in S', and moving these elements from S' to S. Since $|S| \leq |S'|$ and $m \leq n/4$ we know that $n/4 \leq |S| \leq n/2$. Since $(S \times S') \cap \tilde{R}^{\mathcal{B}} = \emptyset$ we know that $\mathcal{A}$ and $\mathcal{B}$ must differ in at least all tuples that correspond to an S and S' crossing edge in $U(\mathcal{A})$ i. e. an edge in $\langle S, S' \rangle_{U(\mathcal{A})}$. Hence

$$\sum_{\tilde{R} \in \sigma} |\tilde{R}(A) \triangle \tilde{R}(B)| \geq |\langle S, S' \rangle_{\underline{G}(A)}| \overset{\text{Def 2.6}}{\geq} |S| \cdot h(A)$$

$$\overset{\text{Thm 3.1}}{\geq} \frac{n}{4} \cdot \frac{D^2}{12} = \frac{1}{48} D^2 n \geq \frac{1}{144D^2} dn.$$

Therefore B is ϵ-far from being in $\mathcal{P}_{\textcircled{z}}$.

However, the neighbourhood distributions of A and B are similar as the following shows, proving that $\mathcal{P}_{\textcircled{z}}$ is not repairable.

$$\sum_{i=1}^{t} |\rho_{A,r}(\{\tau_i\}) - \rho_{B,r}(\{\tau_i\})|$$

$$= \sum_{i=1}^{t} \left| \rho_{A,r}(\{\tau_i\}) - \frac{n \bmod m}{n} \cdot \rho_{K_1,r}(\{\tau_i\}) - \left\lfloor \frac{n}{m} \right\rfloor \cdot \frac{m}{n} \cdot \rho_{H,r}(\{\tau_i\}) \right|$$

$$\leq \sum_{i=1}^{t} \left| \rho_{A,r}(\{\tau_i\}) - \rho_{H,r}(\{\tau_i\}) \right| + \sum_{i=1}^{t} \left| \frac{n \bmod m}{n} \cdot \rho_{K_1,r}(\{\tau_i\}) \right|$$

$$+ \sum_{i=1}^{t} \left| \rho_{H,r}(\{\tau_i\}) - \left\lfloor \frac{n}{m} \right\rfloor \cdot \frac{m}{n} \cdot \rho_{H,r}(\{\tau_i\}) \right|$$

$$\leq \sum_{i=1}^{t} \left| \rho_{A,r}(\{\tau_i\}) - \rho_{H,r}(\{\tau_i\}) \right| + \frac{2m}{n}$$

$$\leq t \cdot \sup_{X \subseteq B_r} |\rho_{A,r}(X) - \rho_{H,r}(X)| + \frac{2m}{n}$$

$$\leq t \cdot 2^r \cdot \delta_{\odot}(A, H) + \frac{2m}{n}$$

$$\leq \frac{\lambda}{2} + \frac{\lambda}{2} = \lambda.$$

Note that in the second inequality we use that $\sum_{i=1}^{t} \rho_{H,r}(\{\tau_i\}) \leq 1$ and the last inequality holds by choice of λ' and Theorem 4.1. ∎

4.2. Every FO property on degree-regular structures is in Π_2. We start with the following observation.

Observation 1. *A Hanf sentence $\exists^{\geq m} x\, \varphi_\tau(x)$ is short for*

$$\exists x_1 \ldots \exists x_m \Big(\bigwedge_{1 \leq i,j \leq m, i \neq j} x_i \neq x_j \wedge \bigwedge_{1 \leq i \leq m} \varphi_\tau(x_i) \Big),$$

and $\varphi_\tau(x_i)$ can be expressed by an $\exists^*\forall$-formula, where the existential quantifiers ensure the existence of the desired r-neighbourhood with all tuples in relations / not in relations as required by τ, and the universal quantifier is used to express that there are no other elements in the r-neighbourhood of x_i.

Note that by the above, any Hanf sentence is in Σ_2. We now show the following lemma.

Lemma 4.5. *Let $d \in \mathbb{N}$ and let φ be an FO sentence. If every model of φ is d-regular, then φ is d-equivalent to a Π_2 sentence.*

The lemma can be equivalently stated by the following syntactic formulation. Let φ^d_{reg} be the FO-sentence expressing that every element has degree d. Then for every FO-sentence φ the sentence $\varphi \wedge \varphi^d_{\text{reg}}$ is d-equivalent to a sentence in Π_2.

Proof. Before we begin, let us define an r-type τ to be *d-regular*, if for all structures A and all elements $a \in U(A)$ of r-type τ, every $b \in U(A)$ with $\text{dist}(a, b) < r$ has $\deg_A(b) = d$.

We first prove the following claim.

Claim 4.6. *Let $d \in \mathbb{N}$, let φ be an FO sentence, and let ψ be in HNF with $\psi \equiv_d \varphi$ such that ψ is in DNF, where the literals are Hanf sentences or negated Hanf sentences. Furthermore, assume that the neighbourhood types in all positive Hanf sentences of ψ are d-regular. Then φ is d-equivalent to a sentence in Π_2.*

Proof. Assume ψ is of the form $\exists^{\geq m} x\, \varphi_\tau(x)$, where τ is d-regular. As in Observation 1, we may assume $\varphi_\tau(x)$ is an $\exists^*\forall$-formula, which arises from a conjunction of an $\exists^*$-formula $\varphi'_\tau(x)$ (expressing that x has an 'induced sub-neighbourhood' of type τ) and a universal formula saying that there are no further elements in the neighbourhood. We now have that $\psi \equiv_d \exists^{\geq m} x\, \varphi'_\tau(x)$. To see this, let $A \models \exists^{\geq m} x \varphi'_\tau(x)$ and $\deg(A) \leq d$. Then $A \models \exists^{\geq m} x \varphi_\tau(x)$ because τ is d-regular. The converse is obvious.

If ψ is of form $\neg\exists^{\geq m} x\, \varphi_\tau(x)$, where $\varphi_\tau(x)$ is an $\exists^*\forall$-formula, then $\neg\exists^{\geq m} x\, \varphi_\tau(x)$ is equivalent to a formula in Π_2. Since Π_2 is closed under disjunction and conjunction, this proves the claim. ∎

Now the proof follows from Claim 4.6, because if φ only has d-regular models, then by Hanf's theorem there is a formula $\psi \equiv_d \varphi$ satisfying the assumptions of the claim.

Existence of a non-testable Π_2*-property.* With Lemma 4.5 and Theorem 4.4, we are ready to prove the following theorem.

THEOREM 4.7. *There is a degree bound $d \in \mathbb{N}$ and a signature σ such that there exists a property on $\mathcal{C}_{\sigma,d}$ definable by a formula in Π_2 that is not testable.*

Proof. Pick $d = 2D^2 + D^4 + 1$ for any large prime power D. Then using the construction from [33] we can find a $(D^4, D, 1/4)$-graph H. By Theorem 4.4, using this base expander H for the construction of the formula $\varphi_{Ⓩ}$ we get a property which is not testable on $\mathcal{C}_{\sigma,d}$. Since all models of $\varphi_{Ⓩ}$ are d-regular by construction, Lemma 4.5 gives us that $\varphi_{Ⓩ}$ is d-equivalent to a formula in Π_2. ∎

5. Reducing to simple undirected graphs. By our previous argument, to show the existence of a non-testable Π_2-property for simple graphs, i.e. undirected graphs without parallel edges and without self-loops, it suffices to construct a non-testable FO graph property of degree regular graphs. To do so, we reduce testing the σ-structure property $\mathcal{P}_{Ⓩ}$ from the previous sections to testing a property $\mathcal{P}_{\text{graph}}$ of simple graphs of bounded degree 3. To construct the reduction we carefully translate the edge-coloured directed graphs (σ-structures) of our previous example in Section 3 to simple graphs. We encode σ-structures by representing each type of directed edge by a constant size graph gadget, maintaining the degree regularity. We then translate the formula $\varphi_{Ⓩ}$ into a formula ψ_{graph} defining the graph property $\mathcal{P}_{\text{graph}}$. This proves the following result.

THEOREM 5.1. *There exists an FO property of simple graphs of bounded degree* 3 *definable by a formula in* Π_2 *that is not testable.*

In the rest of this section, we prove the above theorem via local reductions from a structural property to a graph property, and the non-testable Π_2-property in Theorem 4.7. This technique will also be in the proofs in Section 7.

5.1. Local reductions. We first introduce the following notion of a local reduction between two property testing models. In the following, when the context is clear, we will use $\mathcal{C}$ to denote both a class of structures and the corresponding property testing model, which can be either the bounded-degree model for graphs or bounded-degree model for relational structures.

DEFINITION 5.2 (Local reduction). *Let $\mathcal{C}, \mathcal{C}'$ be two property testing models and let $\mathcal{P} \subseteq \mathcal{C}$, $\mathcal{P}' \subseteq \mathcal{C}'$ be two properties. We say that a function $f : \mathcal{C} \to \mathcal{C}'$ is a local reduction from $\mathcal{P}$ to $\mathcal{P}'$ if there are constants $c_1, c_2 \in \mathbb{N}_{\geq 1}$ such that for every $X \in \mathcal{C}$ the following properties hold.*

1. *If $X \in \mathcal{P}$ then $f(X) \in \mathcal{P}'$.*

2. *If X is ϵ-far from $\mathcal{P}$ then $f(X)$ is (ϵ/c_1)-far from $\mathcal{P}'$.*
3. *For every query to $f(X)$ we can adaptively*[6] *compute c_2 queries to X such that the answer to the query to $f(X)$ can be computed from the answers to the c_2 queries to X.*

The following lemma is known.

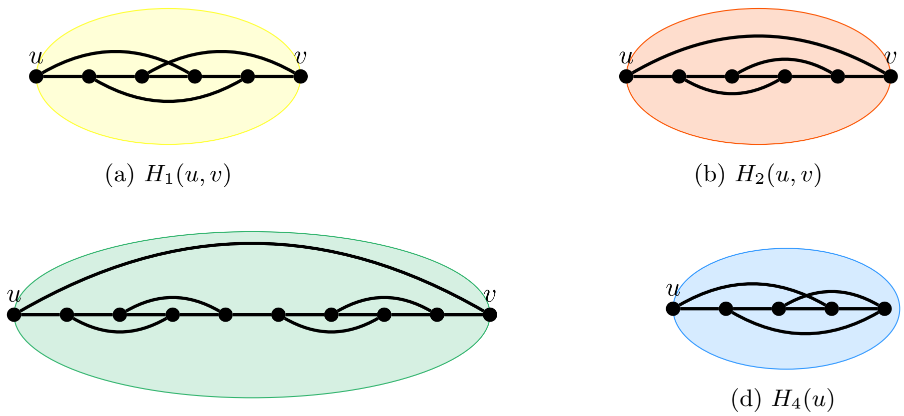
LEMMA 5.3 (Theorem 7.14 in [18]). *Let $\mathcal{C}, \mathcal{C}'$ be two property testing models, $\mathcal{P} \subseteq \mathcal{C}$, $\mathcal{P}' \subseteq \mathcal{C}'$ be two properties and f a local reduction from $\mathcal{P}$ to $\mathcal{P}'$. If $\mathcal{P}'$ is testable then so is $\mathcal{P}$.*

5.2. Constructing the local reduction. Now we construct a property $\mathcal{P}_{\text{graph}}$ of 3-regular graphs from the property $\mathcal{P}_{\textcircled{z}}$. We obtain this graph property as $f(\mathcal{P}_{\textcircled{z}})$ by defining a map $f : \mathcal{C}_{\sigma,d} \to \mathcal{C}_3$. To define f we introduce a distinct arrow-graph gadget for every relation in σ (i. e. for every edge colour). The map f then replaces every tuple in a certain relation (every coloured, directed edge) by the respective arrow-graph gadget. Here all arrow gadgets are designed to allow for 3-regularity of the reduced graph. To obtain 3-regularity we additionally replace every element of a structure in $\mathcal{P}_{\textcircled{z}}$ by a cycle of length d such that each arrow-graph gadget can be incident to a unique vertex of the circle. We further prove that this replacement operation defines a local reduction f from $\mathcal{P}_{\textcircled{z}}$ to $\mathcal{P}_{\text{graph}}$. Recall that a local reduction is a function maintaining distance that can be simulated locally by queries. Since by Lemma 5.3 local reductions preserve testability, we use the local reduction from $\mathcal{P}_{\textcircled{z}}$ to $\mathcal{P}_{\text{graph}}$ to obtain non-testability of the property $\mathcal{P}_{\text{graph}}$ from the non-testability of $\mathcal{P}_{\textcircled{z}}$. We will now define f formally.

We first define building blocks which will be combined to different arrow-graph gadgets. Let $H_1(u, v)$ be the graph with vertex set $\{u = u_0, \dots, v = u_5\}$ and edge set $\{\{u_i, u_{i+3}\} \mid i \in \{0, 1, 2\}\}$. Next we let $H_2(u, v)$ be the graph with vertex set $\{u = u_0, \dots, v = u_5\}$ and edge set $\{\{u_0, u_6\}, \{u_i, u_{i+2}\} \mid i \in \{1, 2\}\}$. Let $H_3(u, v)$ be the graph with vertex set $\{u = u_0, \dots, v = u_9\}$ and edge set $\{\{u_0, u_9\}, \{u_i, u_{i+2}\} \mid i \in \{1, 2, 5, 6\}\}$. Let $H_4(u)$ be the graph with vertex set $\{u = u_0, \dots, u_4\}$ and edge set $\{\{u_0, u_3\}, \{u_1, u_4\}, \{u_2, u_4\}\}$. See Figure 5 for illustration.

Let ℓ be the number of relations (the number of edge colours) in σ. We now introduce the different types of arrow-graph gadgets we need to define the local reduction. For $1 \leq k \leq \ell$, we let $H^k_{\to}(u_0, v_{2\ell})$ be the graph consisting of $2\ell - 1$ vertex disjoint copies $H_1(u_0, v_0), \dots, H_1(u_{k-1}, v_{k-1}), H_1(u_{k+1}, v_{k+1}), \dots, H_1(u_{2\ell-1}, v_{2\ell-1})$, one copy $H_2(u_k, v_k)$, one copy $H_3(u_{2\ell}, v_{2\ell})$ and additional edges $\{v_i, u_{i+1}\}$ for each $i \in [2\ell]$ connecting the respective copies. Note that $H^k_{\to}(u_0, v_{2\ell})$ has $12\ell + 10$ vertices and every vertex apart from $u_0, v_{2\ell}$ has degree 3. We call $H^k_{\to}(u_0, v_{2\ell})$ a *k-arrow*. For any graph G and vertices $u, v \in V(G)$, we say that there is a k-arrow from u to v, denoted $u \xrightarrow{k} v$, if there are $12\ell + 8$ vertices $w_1, \dots, w_{12\ell+8} \in V(G)$ and an isomorphism $g : H^k_{\to}(u_0, v_{2\ell}) \to \mathcal{N}^G_1(w_1, \dots, w_{12\ell+8})$ such that $g(u_0) = u$ and $g(v_{2\ell}) = v$. Note that requiring an isomorphism with these properties guarantees that no vertex contained in a k-arrow has neighbours not contained in the k-arrow with the exception of the end vertices u and v. For any collection $w_1, \dots, w_{12\ell+10}$ of vertices we let $E^k_{\to}(w_1, \dots, w_{12\ell+10})$ be a set of edges such that there is a graph isomorphism $f : H^k_{\to}(u_0, v_{2\ell}) \to \big(\{w_1, \dots, w_{12\ell+10}\}, E^k_{\to}(w_1, \dots, w_{12\ell+10})\big)$ with $f(u_0) = w_1$ and $f(v_{2\ell}) = w_{12\ell+10}$.

[6]By adaptively computing queries we mean that the selection of the next query may depend on the answer to the previous query.



(a) $H_1(u,v)$ (b) $H_2(u,v)$ (c) $H_3(u,v)$ (d) $H_4(u)$

Fig. 5: Illustration of the different building blocks used to define the arrow gadgets.

We now define a second arrow gadget. For $1 \le k \le \ell$, let $H^k_{\circlearrowleft}(u_0)$ be the graph consisting of $\ell - 1$ vertex disjoint copies $H_1(u_0, v_0), \dots, H_1(u_{k-1}, v_{k-1})$, $H_1(u_{k+1}, v_{k+1})$, $\dots, H_1(u_{\ell-1}, v_{\ell-1})$, one copy $H_2(u_k, v_k)$, one copy $H_4(u_\ell)$ and edges $\{v_i, u_{i+1}\}$ for each $i \in [\ell-1]$. Note that $H^k_{\circlearrowleft}(u_0)$ has $6\ell+5$ vertices and every vertex apart from u_0 has degree 3. We call $H^k_{\circlearrowleft}$ a *k-loop*. For any graph G and vertex $u \in V(G)$, we say that there is a k-loop at u, denoted $u \xrightarrow{k} u$, if there are $6\ell+4$ vertices $w_1, \dots, w_{6\ell+4} \in V(G)$ and an isomorphism $g : H^k_{\circlearrowleft}(u_0) \to \mathcal{N}^G_1(w_1, \dots, w_{6\ell+4})$ such that $g(u_0) = u$. For any collection $w_1, \dots, w_{6\ell+5}$ vertices we let $E^k_{\circlearrowleft}(w_1 \dots, w_{6\ell+5})$ be a set of edges for which there is an isomorphism $f : H^k_{\circlearrowleft}(u_0) \to \big(\{w_1, \dots, w_{6\ell+5}\}, E^k_{\circlearrowleft}(w_1, \dots, w_{6\ell+5})\big)$ for which $f(u_0) = w_1$.

Finally, let $H_\perp(u_0)$ be the graph consisting of ℓ vertex disjoint copies $H_1(u_0, v_0)$, $\dots, H_1(u_{\ell-1}, v_{\ell-1})$, one copy $H_4(u_\ell)$ and additional edges $\{v_i, u_{i+1}\}$ for each $i \in [\ell-1]$. Note that $H_\perp(u_0)$ has $6\ell+5$ vertices and every vertex apart from u_0 has degree 3. We call $H_\perp$ a *non-arrow*. For any graph G and vertex $u \in V(G)$, we say that there is a non-arrow at u, denoted $u \not\to$, if there are $6\ell+4$ vertices $w_1, \dots, w_{6\ell+4} \in V(G)$ and an isomorphism $g : H_\perp \to \mathcal{N}^G_1(w_1, \dots, w_{6\ell+4})$ such that $g(u_0) = u$. For any collection $w_1, \dots, w_{6\ell+5}$ vertices we let $E_\perp(w_1 \dots, w_{6\ell+5})$ be a set of edges for which there is an isomorphism $f : H_\perp(u_0) \to \big(\{w_1, \dots, w_{6\ell+5}\}, E^k_{\circlearrowleft}(w_1, \dots, w_{6\ell+5})\big)$ for which $f(u_0) = w_1$.

We now define a function $f : \mathcal{C}_{\sigma,d} \to \mathcal{C}_3$ by $f(A) := G_A$, where G_A is the graph on vertex set $V(G_A) := \{u_{a,i}, v^k_{a,i} \mid 1 \le i \le d, a \in U(A), 1 \le k \le 6\ell+5\}$ and edge set $E(G_A)$ defined by

$$\Big\{\{u_{a,i}, v^1_{a,i}\} \mid a \in U(A), 1 \le i \le d\Big\}$$

$$\cup \Big\{\{u_{a,d}, u_{a,1}\}, \{u_{a,i}, u_{a,i+1}\} \mid a \in U(A), 1 \le i \le d-1\Big\}$$

$$\cup \bigcup_{\substack{\mathrm{ans}(a,i)=\mathrm{ans}(b,j)=(k,a,b)\\ a \neq b}} E^k_{\to}\Big(v^1_{a,i}, \dots, v^{6\ell+5}_{a,i}, v^{6\ell+5}_{b,j}, \dots, v^1_{b,j}\Big)$$

$$\cup \bigcup_{\mathrm{ans}(a,i)=(k,a,a)} E^k_{\circlearrowleft}\left(v^1_{a,i},\ldots,v^{6\ell+5}_{a,i}\right)$$

$$\cup \bigcup_{\mathrm{ans}(a,i)=\perp} E_{\perp}\left(v^1_{a,i},\ldots,v^{6\ell+5}_{a,i}\right),$$

where $\mathrm{ans}(a,i)=(k,a,b)$ denotes that the i-th tuple of a is (a,b) and is in the k-th relation. Hence G_A is defined in such a way that every element $a \in U(A)$ is represented by an induced cycle $(u_{a,1},\ldots,u_{a,d},u_{a,1})$ and if (a,b) is a tuple in the k-th relation of σ in A, then $u_{a,i} \xrightarrow{k} u_{b,j}$ in G_A for some $1 \le i,j \le d$, and $u_{a,i}$ has a non-arrow for every i satisfying that $\mathrm{ans}(a,i)=\perp$ for every k. Note that G_A is 3 regular by construction for every $A \in \mathcal{C}_{\sigma,d}$. For illustration see Figure 6. In the following we refer to vertices of G_A of the form $u_{a,i}$ by *element-vertices* while we call vertices of the form $v^j_{a,i}$ *relation-vertices*. The following is easy to observe from the construction and from the fact that $d = 2D^2 + D^4 + 1 < 3D^4 + 1 = |\sigma| = \ell$ for some large prime power D (see Section 3 for definitions).

FACT 1. *For every $u \in V(G_A)$, u is an element-vertex iff u is contained in a simple cycle of length d. Furthermore, two vertices $u,v \in V(G_A)$ correspond to the same element a of A (i. e. there are $i,j \in \{1,\ldots,d\}$ such that $u = u_{a,i}$ and $v = u_{a,j}$) iff there is a simple cycle of length d containing both u and v.*

Note that we do not need to ask for cycles of length d to be induced because the structure we obtain does not allow for cycles of length d apart from the cycles corresponding to elements.

Now we define property $\mathcal{P}_{\mathrm{graph}} := \{f(A) \mid A \in \mathcal{P}_{\textcircled{z}}\} \subseteq \mathcal{C}_3$.

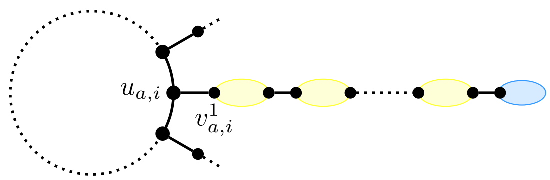
LEMMA 5.4. *The map f is a local reduction from $\mathcal{P}_{\textcircled{z}}$ to $\mathcal{P}_{\mathrm{graph}}$.*

Proof. First note that for any $A \in \mathcal{P}_{\textcircled{z}}$, we have that $f(A) \in \mathcal{P}_{\mathrm{graph}}$ by definition.

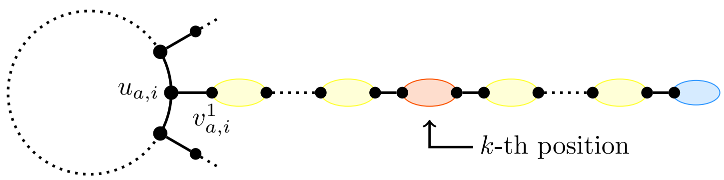
Now let $c_1 = 6d(1+6\ell+5)$. We prove that if $A \in \mathcal{C}_{\sigma,d}$ is ϵ-far from $\mathcal{P}_{\textcircled{z}}$ then $f(A)$ is ϵ/c_1-far from $\mathcal{P}_{\mathrm{graph}}$ by contraposition. Therefore assume that $f(A) =: G_A$ is not ϵ/c_1-far from $\mathcal{P}_{\mathrm{graph}}$ for some $A \in \mathcal{C}_{\sigma,d}$. Then there is a set $E \subseteq \{e \subseteq V(G_A) \mid |e| = 2\}$ of size at most $\epsilon \cdot 3|V(G_A)|/c_1$, and a graph $G \in \mathcal{P}_{\mathrm{graph}}$ such that G is obtained from G_A by modifying the tuples in E. By definition of $\mathcal{P}_{\mathrm{graph}}$, there is a structure $A_G \in \mathcal{P}_{\textcircled{z}}$ such that $f(A_G) = G$. First note that $|U(A_G)| = |U(A)|$, as $d(1+6\ell+5)|U(A)| = |V(G_A)| = |V(G)| = d(1+6\ell+5)|U(A_G)|$. Hence there must be a set R of tuples that need to be modified to make A isomorphic to A_G. First note that R cannot contain a tuple (a,b) where $\{u_{a,i}, v^k_{a,i}, u_{b,i}, v^k_{b,i} \mid 1 \le i \le d, 1 \le k \le \ell\} \cap e = \emptyset$ for every $e \in E$. This is because if (a,b) is a tuple in the k-th relation of A, then $u_{a,i} \xrightarrow{k} u_{b,j}$ in G_A for some $i,j \in \{1,\ldots,d\}$. But since $\{u_{a,i}, v^k_{a,i}, u_{b,i}, v^k_{b,i} \mid 1 \le i \le d, 1 \le k \le \ell\} \cap e = \emptyset$ for every $e \in E$, we have that $u_{a,i} \xrightarrow{k} u_{b,j}$ in G. Further, $(u_{a,1},\ldots,u_{a,d},u_{a,1})$ and $(u_{b,1},\ldots,u_{b,d},u_{b,1})$ are simple cycles of length d in G. Hence by 1 there are elements a,b in A_G corresponding to $(u_{a,1},\ldots,u_{a,d},u_{a,1})$ and $(u_{b,1},\ldots,u_{b,d},u_{b,1})$ such that (a,b) is a tuple in the k-th relation of A_G, and hence (a,b) cannot be in R. The same argument works when assuming that (a,b) is a tuple in A_G. Since for every $e \in E$, there is at most $2d$ tuples (a,b) such that $\{u_{a,i}, v^k_{a,i}, u_{b,i}, v^k_{b,i} \mid 1 \le i \le d, 1 \le k \le \ell\} \cap e \ne \emptyset$, we get that

$$|R| \le 2d\epsilon \cdot 3|V(G_A)|/c_1 = 6d(1+6\ell+5)\epsilon d|U(A)|/c_1 = \epsilon d|U(A)|.$$

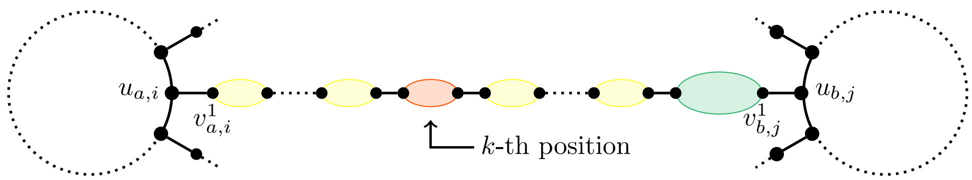
Hence A is not ϵ-far to being in $\mathcal{P}_{\textcircled{z}}$.



(a) Case $\mathrm{ans}(a,i) = \bot$



(b) Case $\mathrm{ans}(a,i) = (k,a,a)$



(c) Case $\mathrm{ans}(a,i) = \mathrm{ans}(b,j) = (k,a,b)$

Fig. 6: Different types of arrows in G_A. Here different coloured ellipses represent a copy of $H_1(u,v), H_2(u,v), H_3(u,v)$ or $H_4(u)$ respectively (see Figure 5 for details).

Let $c_2 := d+1$. Let $A \in \mathcal{C}_{\sigma,d}$ and $G_A := f(A)$. First it is important to observe that we can pick an ordering of the vertices of G_A such that the position of each vertex depends solely on the number of elements of A. Hence we can assume that for any element a of A we can decide for any vertex $v \in V(G_A)$ whether v is of the form $u_{a,i}$ and whether v is of the form $v^k_{a,i}$. Now we argue how we can determine the answer to any neighbour query in G_A. First note that for any $a \in U(A)$ and $i \in \{1,\dots,d\}$ the vertex $u_{a,i}$ is adjacent in G_A to $v^1_{a,i}$ and the two neighbouring vertices on the simple cycle $(u_{a,1},\dots,u_{a,d},u_{a,1})$. Hence any neighbour query in G_A to $u_{a,i}$ can be answered without querying A. Assume $v \in \{v^k_{a,i} \mid 1 \le k \le \ell\}$ for some $a \in U(A)$ and some $1 \le i \le d$. Then we can determine all neighbours of v by querying (a,i) and further if $\mathrm{ans}(a,i) \neq (k,a,a)$, $\mathrm{ans}(a,i) \neq \bot$ and $\mathrm{ans}(a,i) = (k,a,b)$, then we need to query (b,j) for every $1 \le j \le d$ to find out for which j we have $\mathrm{ans}(b,j) = (k,a,b)$. Hence we can determine the answer to any query to G_A by making c_2 queries to A. This proves that f is a local reduction from $\mathcal{P}_{ⓩ}$ to $\mathcal{P}_{\mathrm{graph}}$. ∎

5.3. The property of graphs is definable in FO. In this section we find an FO sentence ψ_{graph} which defines the property $\mathcal{P}_{\mathrm{graph}}$. We do this by defining a formula expressing for two vertices u,v that $u \xrightarrow{k} v$, a formula expressing for vertex u that $u \xrightarrow{k} u$ and a formula expressing for vertex u that $u \not\to$ and replacing formulas of the form $R(u,v), R(v,v)$ and $\neg R(u,v)$ for $R \in \sigma$ by the new formulas appropriately. We additionally restrict the scope of the quantifiers. In the previous subsection we already defined $\ell := |\sigma|$. We further rename the relations in σ in an arbitrary way such that for this section we can assume that $\sigma = \{R_1, \dots, R_\ell\}$.

We now translate the formula $\varphi_{\textcircled{Z}}$ into a formula ψ_{graph} in the language of undirected graphs using the FO formulas defined in the following. We let $\alpha(x)$ be a formula saying 'x is an element-vertex' and $\beta(x,y)$ be a formula saying 'x and y represent the same element of $\mathcal{A}$', which is easy to do by Fact 1. We further let $\gamma(x)$ be a formula saying 'x is an internal vertex of either a k-arrow, a k-loop for any $k \in \{1, \dots, \ell\}$ or a non-arrow'. Here an 'internal vertex' of an arrow refers to any vertex on this arrow except the two endpoints, or the single endpoint in case of a loop or non-arrow. Let $\delta^k_{\rightarrow}(x,y)$ denote '$x \xrightarrow{k} y$' for any $k \in \{1, \dots, \ell\}$, similarly, let $\delta^k_{\circlearrowleft}(x)$ denote '$x \xrightarrow{k} x$' for any $k \in \{1, \dots, \ell\}$. Given $\varphi_{\textcircled{Z}}$, formula ψ_{graph} is obtained as follows. In $\varphi_{\textcircled{Z}}$ we replace each expression $R_k(x,x)$ by $\delta^k_{\circlearrowleft}(x,x)$ and each expression $R_k(x,y)$ by $\delta^k_{\rightarrow}(x,y)$ (for $x \neq y$). In addition, we relativise all quantifiers in the following way. We replace every expression of the form $\exists x\, \chi(x, x_1, \dots, x_m)$ by $\exists x\, (\alpha(x) \wedge \chi(x, x_1, \dots, x_m))$ and every expression of the form $\forall x\, \chi(x, x_1, \dots, x_m)$ by $\forall x\, (\alpha(x) \rightarrow \exists y \beta(x,y) \wedge \chi(y, x_1, \dots, x_m))$. Let us call the resulting formula ψ. Then we set ψ_{graph} to be the conjunction of the formula ψ and the formula $\forall x \Big((\neg\alpha(x) \rightarrow \gamma(x)) \wedge (\alpha(x) \rightarrow \exists y \gamma(y) \wedge E(x,y)) \Big)$.

Lemma 5.5. *For any $A \in \mathcal{C}_{\sigma,d}$ the following proposition is true. $A \models \varphi_{\textcircled{Z}}$ if and only if $f(A) \models \psi_{\text{graph}}$. Additionally we have that if $G \in \mathcal{C}_3$ is a model of ψ_{graph} then $G \cong f(A)$ for some $A \in \mathcal{A}_{\sigma,d}$.*

Proof. First assume that $A \models \varphi_{\textcircled{Z}}$. First observe that by construction of $G_A := f(A)$ and ψ, we get that $A \models \varphi_{\textcircled{Z}}$ if and only if $G_A \models \psi$. Note that for this statement it is important that the set of k-arrows and k-loops for all $k \in \{1, \dots, \ell\}$ is a set of pairwise non-isomorphic graphs. In the construction of G_A, every vertex is either an element-vertex $u_{a,i}$ in which case it is adjacent to the relation-vertex $v^1_{a,i}$, or is an internal vertex of some k-arrow, k-loop or non-arrow. Hence we get that $G_A \models \forall x \Big((\neg\alpha(x) \rightarrow \gamma(x)) \wedge (\alpha(x) \rightarrow \exists y \gamma(y) \wedge E(x,y)) \Big)$, which completes the proof of the first statement.

Towards proving the second statement of Lemma 5.5, let us assume that some graph $G \in \mathcal{C}_3$ is a model of ψ_{graph}. Then $G \models \forall x \Big((\neg\alpha(x) \rightarrow \gamma(x)) \wedge (\alpha(x) \rightarrow \exists y \gamma(y) \wedge E(x,y)) \Big)$. Hence G consists of a set of element-vertices that are connected according to ψ with k-arrow, k-loops or non-arrows. Hence we can reverse the local reduction to obtain A_G which is the corresponding model of $\varphi_{\textcircled{Z}}$ for which $f(A_G) \cong G$ by the following construction. For any maximal set of vertices $X \subseteq V(G)$ such that $\beta(u,v)$ holds for every pair $u, v \in X$, we introduce an element a_X. For $X, Y \subseteq V(G)$, we add a tuple (a_X, a_Y) to the relation $R_k(A_G)$ if there are $u \in X$ and $v \in Y$ such that $u \xrightarrow{k} v$ in G. With a similar argument as above, we get that A_G is a model of $\varphi_{\textcircled{Z}}$ by the construction of ψ. Additionally we get for some ordering of the neighbours of each element of A_G that $f(A_G) \cong G$ (this ordering has to be consistent with the order of k-arrows along the cycle of element-vertices). ◻

Proof of Theorem 5.1. As a consequence from Lemma 5.5, we get that ψ_{graph} defines the property $\mathcal{P}_{\text{graph}}$ on the class $\mathcal{C}_3$. Since we constructed the local reduction f in such a way that $f(\mathcal{A})$ is 3-regular for every $\mathcal{A} \in \mathcal{C}_{\sigma,d}$ by Lemma 4.5, we get that $\mathcal{P}_{\text{graph}}$ can be defined by a sentence in Π_2 on the class $\mathcal{C}_3$. Combining this with Lemma 5.3 and Lemma 5.4, we obtain Theorem 5.1. ∎

We would like to point out here that while we obtain the non-testability of $\mathcal{P}_{\text{graph}}$ using the local reduction f, we can not conclude that $\mathcal{P}_{\text{graph}}$ is a class of expanders. However, we will show that this is true in the following section.

5.4. The property of graphs is a class of expanders. In this subsection we show that $\mathcal{P}_{\text{graph}}$ is a family of expanders and hence prove the following theorem.

THEOREM 5.6. *There exists a universal constant $\xi > 0$ and an (infinite) class of ξ-expanders with maximum degree at most 3 which is definable in FO on undirected graphs.*

Expansion of $\mathcal{P}_{\text{graph}}$ is not needed for the non-testability results in this paper. However, we think that Theorem 5.6 is of independent interest since it gives us new insights into the expressibility of first-order logic. Furthermore, for an expanding property of undirected graphs, its non-testability follows from the main result from [15].

LEMMA 5.7. *The models of ψ_{graph} is a family of ξ-expanders, for some constant $\xi > 0$.*

Proof. Let $A \in \mathcal{P}_{\textcircled{z}}$ and $G_A := f(A)$. For every $a \in A$, we define the vertex set $V_a = \{u_{a,i}, v_{a,i}^k | 1 \leq i \leq d, 1 \leq k \leq 6\ell + 5\}$, encompassing the element vertices of a alongside the vertices of its loops, non-arrows, and half of the vertices of each of its arrows. Considering a set $S \subset V(G_A)$ such that $|S| \leq \frac{|V(G_A)|}{2}$, we partition $U(A)$ into $S^{\text{full}} = \{a \in U(A) : V_a \subseteq S\}$, $S^{\text{disj}} = \{a \in U(A) : V_a \cap S = \emptyset\}$ and $S^{\text{part}} = U(A) \setminus (S^{\text{full}} \cup S^{\text{disj}})$.

First note that $|S^{\text{full}}| \leq \frac{|U(A)|}{2}$. As $\underline{G}(A)$ is an ϵ-expander by Theorem 3.1, there are at least $\epsilon|S^{\text{full}}|$ edges between S^{full} and $S^{\text{part}} \cup S^{\text{disj}}$. Hence, there are at least $\epsilon|S^{\text{full}}| - d|S^{\text{part}}|$ edges between S^{full} and S^{disj}. By the choice of the sets S^{full} and S^{disj}, for every such edge (a, b), there corresponds an edge, i. e. the edge $\{v_{a,i}^{6\ell+5}, v_{b,j}^{6\ell+5}\}$ for some $1 \leq i, j \leq d$, in G_A between S and $V(G_A) \setminus S$. Additionally, for every $a \in S^{\text{part}}$, there is at least one edge in $G_A[V_a]$ between S and $V(G_A) \setminus S$.

We set $c = d(6\ell + 6)$ and observe that $S^{\text{full}} \geq \frac{|S| - |S^{\text{part}}|}{c}$. Let us first consider the case that $|S^{\text{part}}| \leq \frac{\epsilon|S|}{3cd} \leq \frac{|S|}{3}$. In this case, we have at least $\frac{\epsilon|S|}{3c}$ edges between S and $U(A) \setminus S$ on account of edges between S^{full} and S^{disj}. On the other hand, if $|S^{\text{part}}| \geq \frac{\epsilon|S|}{3cd}$ then we have at least $\frac{\epsilon|S|}{3cd}$ edges between S and $U(A) \setminus S$ on account of the edges within each $G_A[V_a]$ for $a \in S^{\text{part}}$. Therefore, $\mathcal{P}_{\text{graph}}$ is a class of ξ-expanders for $\xi = \frac{\epsilon}{3cd}$. ∎

6. On the testability of all Σ_2-properties. Let $\sigma = \{R_1, \ldots, R_m\}$ be any relational signature and $\mathcal{C}_{\sigma,d}$ the set of σ-structures of bounded degree d. We prove the following.

THEOREM 6.1. *Every first-order property defined by a σ-sentence in Σ_2 is testable in the bounded-degree model.*

We adapt the notion of indistinguishability of [4] from the dense model to the bounded-degree model.

DEFINITION 6.2. *Two properties $\mathcal{P}, \mathcal{Q} \subseteq \mathcal{C}_{\sigma,d}$ are called* indistinguishable *if for every $\epsilon \in (0, 1)$ there exists $N = N(\epsilon)$ such that for every structure $A \in \mathcal{P}$ with $|U(A)| > N$ there is a structure $\tilde{A} \in \mathcal{Q}$ with the same universe, that is ϵ-close to A;*

and for every $B \in \mathcal{Q}$ with $|U(B)| > N$ there is a structure $\tilde{B} \in \mathcal{P}$ with the same universe, that is ϵ-close to B.

The following lemma follows from the definitions, and is similar to [4], though we make use of the canonical testers for bounded-degree graphs ([11, 22]).

LEMMA 6.3. *If $\mathcal{P}, \mathcal{Q} \subseteq \mathcal{C}_{\sigma,d}$ are indistinguishable properties, then $\mathcal{P}$ is testable on $\mathcal{C}_{\sigma,d}$ if and only if $\mathcal{Q}$ is testable on $\mathcal{C}_{\sigma,d}$.*

Proof. We show that if $\mathcal{P}$ is testable, then $\mathcal{Q}$ is also testable. The other direction follows by the same argument. Let $\epsilon > 0$. Since $\mathcal{P}$ is testable, there exists an $\frac{\epsilon}{2}$-tester for $\mathcal{P}$ with success probability at least $\frac{2}{3}$. Furthermore, we can assume that the tester (called canonical tester) behaves as follows (see [11, 22]): it first uniformly samples a constant number $c_0 = c_0(\frac{\epsilon}{2}, d)$ of elements, then explores the union of r-balls around all sampled elements for some constant $r = r(\frac{\epsilon}{2}, d) > 0$, and makes a deterministic decision whether to accept, based on an isomorphic copy of the explored substructure. Let $C = C(\frac{\epsilon}{2}, d) = c_0 \cdot (1 + d + \cdots + d^r)$ denote the upper bound on the number of queries the canonical tester made on the input structure. Then there exists some universal constant $c_1 > 0$ such that by repeating the canonical tester c_1 times, and taking the majority vote, we can have a tester T with $c_1 \cdot C$ query complexity and success probability at least $\frac{5}{6}$.

Let N be a number such that if a structure B with $n > N$ elements satisfies $\mathcal{Q}$, then there exists a $\tilde{B} \in \mathcal{P}$ with the same universe such that $\text{dist}(B, \tilde{B}) \leq \min\{\frac{\epsilon}{2}, \frac{1}{c_2 C \cdot d^{C+2}}\}dn$ for some large constant $c_2 > 0$. Now we give an ϵ-tester for $\mathcal{Q}$. If the input structure B has size at most N, we can query the whole input to decide if it satisfies $\mathcal{Q}$ or not. If its size is larger than N, then we use the aforementioned $\frac{\epsilon}{2}$-tester for $\mathcal{P}$ with success probability at least $\frac{5}{6}$. If B satisfies $\mathcal{Q}$, then there exists $\tilde{B} \in \mathcal{P}$ that differs from B in no more than $1/(c_2 C \cdot d^{C+2})dn$ places. Since the algorithm samples $c_0 \cdot c_1$ elements and queries the r-balls around all these sampled elements and makes at most $c_1 \cdot C$ queries in total, we have that with probability at least $1 - \frac{1}{6}$, the algorithm does not query any part where B and $\tilde{B}$ differ, and thus its output is correct with probability at least $\frac{5}{6} - \frac{1}{6} = \frac{2}{3}$. If B is ϵ-far from satisfying $\mathcal{Q}$ then it is $\frac{\epsilon}{2}$-far from satisfying $\mathcal{P}$ and with probability at least $\frac{5}{6} > \frac{2}{3}$, the algorithm will reject B. Thus $\mathcal{Q}$ is also testable. ◻

High-level idea of proof of Theorem 6.1. Let $\varphi \in \Sigma_2$. We prove that the property defined by φ can be written as the union of properties, each of which is defined by another formula φ' in Σ_2 where the structure induced by the existentially quantified variables is a fixed structure M (see Claim 6.6). With some further simplification of φ', we obtain a formula φ'' in Σ_2 which expresses that the structure has to have M as an induced substructure and every set of elements of fixed size ℓ has to induce some structure from a set of structures $\mathfrak{H}$, and – depending on the structure from $\mathfrak{H}$ – there might be some connections to the elements of M (see Claim 6.7). We then define a formula ψ in Π_1 such that the property defined by ψ is indistinguishable from the property defined by φ'' in the sense that we can transform any structure satisfying ψ, into a structure satisfying φ'' by modifying no more than a small fraction of the tuples and vice versa (see Claim 6.10). The intuition behind this is that every structure satisfying φ'' can be made to satisfy ψ by removing the structure M while on the other hand for every structure which satisfies ψ we can plant the structure M to make it satisfy φ''. Since it is a priori unclear how the existentially and universally quantified variables interact, we have to define ψ very carefully. Here it is important to note that the number of occurrences of structures in $\mathfrak{H}$ forcing an interaction with

M is limited because of the degree bound (see Claim 6.8). Thus such structures can not be allowed to occur for models of ψ, as here the number of occurrences can not be limited in any way. Since properties defined by a formula in Π_1 are testable, this implies with the indistinguishability of ψ and φ'' that the property defined by φ'' is testable. Furthermore by the fact that testable properties are closed under union [18], we reach the conclusion that any property defined by a formula in Σ_2 is testable.

We will not directly give a tester for the property $\mathcal{P}_\varphi$ but decompose φ into simpler cases. However, every simplification of φ used is computable, and the proof below yields a construction of an ϵ-tester for $\mathcal{P}_\varphi$ for every $\epsilon \in (0,1)$ and every $\varphi \in \Sigma_2$.

For the full proof of Theorem 6.1, we use the following definition.

Definition 6.4. *Let A be a σ-structure with $U(A) = \{a_1, \dots, a_t\}$. Let $\overline{z} = (z_1, \dots, z_t)$ be a tuple of variables. Then we define $\iota^A(\overline{z})$ as follows.*

$$\iota^A(\overline{z}) := \bigwedge_{R\in\sigma} \Bigg(\bigwedge_{\left(a_{i_1},\dots,a_{i_{\mathrm{ar}(R)}}\right)\in R(A)} R\big(z_{i_1}, \dots, z_{i_{\mathrm{ar}(R)}}\big) \wedge$$

$$\bigwedge_{\left(a_{i_1},\dots,a_{i_{\mathrm{ar}(R)}}\right)\in U(A)^{\mathrm{ar}(R)}\setminus R(A)} \neg R\big(z_{i_1}, \dots, z_{i_{\mathrm{ar}(R)}}\big) \Bigg) \wedge \bigwedge_{\substack{i,j\in[t]\\ i\neq j}} (\neg z_i = z_j).$$

Note that for every σ-structure A' and $\overline{a}' = (a'_1, \dots, a'_t) \in U(A')^t$ we have that $A' \models \iota^A(\overline{a}')$ if and only if $a_i \mapsto a'_i$, $i \in \{1, \dots, t\}$ is an isomorphism from A to $A'[\{a'_1, \dots, a'_t\}]$. In particular, if $A' \models \iota^A(\overline{a}')$, then $\{a'_1, \dots, a'_t\}$ induces a substructure isomorphic to A in A'.

Proof of Theorem 6.1. Let φ be any sentence in Σ_2. Therefore we can assume that φ is of the form $\varphi = \exists \overline{x} \forall \overline{y}\, \chi(\overline{x}, \overline{y})$ where $\overline{x} = (x_1, \dots, x_k)$ is a tuple of $k \in \mathbb{N}$ variables, $\overline{y} = (y_1, \dots, y_\ell)$ is a tuple of $\ell \in \mathbb{N}$ variables and $\chi(\overline{x}, \overline{y})$ is a quantifier-free formula. We can further assume that $\chi(\overline{x}, \overline{y})$ is in disjunctive normal form, and that

$$\varphi = \exists \overline{x} \forall \overline{y} \bigvee_{i\in I} \Big(\alpha^i(\overline{x}) \wedge \beta^i(\overline{y}) \wedge \mathrm{pos}^i(\overline{x}, \overline{y}) \wedge \mathrm{neg}^i(\overline{x}, \overline{y}) \Big), \tag{6.1}$$

where $\alpha^i(\overline{x})$ is a conjunction of literals only containing variables from $\overline{x}$, $\beta^i(\overline{y})$ is a conjunction of literals only containing variables in $\overline{y}$, $\mathrm{neg}^i(\overline{x}, \overline{y})$ is a conjunction of negated atomic formulas containing both variables from $\overline{x}$ and $\overline{y}$ and $\mathrm{pos}^i(\overline{x}, \overline{y})$ is a conjunction of atomic formulas containing both variables from $\overline{x}$ and $\overline{y}$. Now note that if an expression '$x_j = y_{j'}$' appears in a conjunctive clause, then we can replace every occurrence of $y_{j'}$ by x_j in that clause, which will result in an equivalent formula.

We now write the formula φ given in (6.1) as a disjunction over all possible structures in $\mathcal{C}_{\sigma,d}$ the existentially quantified variables could enforce. Since the elements realising the existentially quantified variables will have a certain structure, it is natural to decompose the formula in this way.

Let $\mathfrak{M} \subseteq \mathcal{C}_{\sigma,d}$ be a set of models of φ, such that every model $A \in \mathcal{C}_{\sigma,d}$ of φ contains an isomorphic copy of some $M \in \mathfrak{M}$ as an induced substructure, and $\mathfrak{M}$ is minimal with this property.

Claim 6.5. *Every $M \in \mathfrak{M}$ has at most k elements.*

Proof. Assume there is $M \in \mathfrak{M}$ with $|M| > k$. Since every structure in $\mathfrak{M}$ is a model of φ there must be a tuple $\overline{a} = (a_1, \dots, a_k) \in U(M)^k$ such that $M \models$

$\forall \overline{y} \bigvee_{i\in I} \Big(\alpha^i(\overline{a}) \wedge \beta^i(\overline{y}) \wedge \mathrm{pos}^i(\overline{a},\overline{y}) \wedge \mathrm{neg}^i(\overline{a},\overline{y})\Big)$. This implies that for every tuple $\overline{b} \in U(M)^\ell$ we have $M \models \bigvee_{i\in I} \Big(\alpha^i(\overline{a}) \wedge \beta^i(\overline{b}) \wedge \mathrm{pos}^i(\overline{a},\overline{b}) \wedge \mathrm{neg}^i(\overline{a},\overline{b})\Big)$. Furthermore, since $\{a_1,\dots,a_k\}^\ell \subseteq U(M)^\ell$ we have that $M[\{a_1,\dots,a_k\}] \models \forall \overline{y} \bigvee_{i\in I} \Big(\alpha^i(\overline{a}) \wedge \beta^i(\overline{y}) \wedge \mathrm{pos}^i(\overline{a},\overline{y}) \wedge \mathrm{neg}^i(\overline{a},\overline{y})\Big)$. This means that $M[\{a_1,\dots,a_k\}] \models \varphi$. Hence $\mathfrak{M}$ contains an induced substructure M' of $M[\{a_1,\dots,a_k\}]$. Since every model of φ containing M as an induced substructure must also contain M' as an induced substructure $\mathfrak{M} \setminus \{M\}$ is a strictly smaller set than $\mathfrak{M}$ with all desired properties. This contradicts the minimality $\mathfrak{M}$. ◻

Therefore $\mathfrak{M}$ is finite. For $M \in \mathfrak{M}$ let $J_M := \{j \in I \mid M \models \alpha^j(\overline{m}) \text{ for some } \overline{m} \in U(M)^\ell\} \subseteq I$.

Claim 6.6. *We have*

$$\varphi \equiv_d \bigvee_{M\in\mathfrak{M}} \Big(\exists\overline{x}\forall\overline{y}\Big[\iota^M(\overline{x}) \wedge \bigvee_{j\in J_M} \Big(\beta^j(\overline{y}) \wedge \mathrm{pos}^j(\overline{x},\overline{y}) \wedge \mathrm{neg}^j(\overline{x},\overline{y})\Big)\Big]\Big).$$

Proof. Let $A \in \mathcal{C}_{\sigma,d}$ be a model of φ. Then there is a tuple $\overline{a} = (a_1,\dots,a_k) \in U(A)^k$ such that $A \models \forall y \chi(\overline{a},\overline{y})$. Since $\{a_1,\dots,a_k\}^\ell \subseteq U(A)^\ell$ this implies that $A[\{a_1,\dots,a_k\}] \models \forall y \chi(\overline{a},\overline{y})$ and hence $A[\{a_1,\dots,a_k\}] \models \varphi$. In addition, we may assume that we picked $\overline{a}$ in such a way that for any tuple $\overline{a}' = (a'_1,\dots,a'_k) \in \{a_1,\dots,a_k\}^k$ with $\{a'_1,\dots,a'_k\} \subsetneq \{a_1,\dots,a_k\}$ we have that $A \not\models \forall \overline{y} \chi(\overline{a}',\overline{y})$. (The reason is that if for some tuple $\overline{a}'$ this is not the case then we just replace $\overline{a}$ by $\overline{a}'$ and so on until this property holds). Hence $A[\{a_1,\dots,a_k\}]$ cannot have a proper induced substructure in $\mathfrak{M}$, and it follows that there is $M \in \mathfrak{M}$ such that $M \cong A[\{a_1,\dots,a_k\}]$. By choice of J_M we get $A \models \forall \overline{y}\Big[\iota^M(\overline{a}) \wedge \bigvee_{j\in J_M} \Big(\beta^j(\overline{y}) \wedge \mathrm{pos}^j(\overline{a},\overline{y}) \wedge \mathrm{neg}^j(\overline{a},\overline{y})\Big)\Big]$ and hence

$$A \models \bigvee_{M\in\mathfrak{M}} \Big(\exists\overline{x}\forall\overline{y}\Big[\iota^M(\overline{x}) \wedge \bigvee_{j\in J_M} \Big(\beta^j(\overline{y}) \wedge \mathrm{pos}^j(\overline{x},\overline{y}) \wedge \mathrm{neg}^j(\overline{x},\overline{y})\Big)\Big]\Big).$$

To prove the other direction, we now let the structure $A \in \mathcal{C}_{\sigma,d}$ be a model of the formula $\bigvee_{M\in\mathfrak{M}} \Big(\exists\overline{x}\forall\overline{y}\Big[\iota^M(\overline{x}) \wedge \bigvee_{j\in J_M} \Big(\beta^j(\overline{y}) \wedge \mathrm{pos}^j(\overline{x},\overline{y}) \wedge \mathrm{neg}^j(\overline{x},\overline{y})\Big)\Big]\Big)$. Consequently there is $M \in \mathfrak{M}$ and $\overline{a} \in U(A)^k$ such that $A \models \forall \overline{y}\Big[\iota^M(\overline{a}) \wedge \bigvee_{j\in J_M} \Big(\beta^j(\overline{y}) \wedge \mathrm{pos}^j(\overline{a},\overline{y}) \wedge \mathrm{neg}^j(\overline{a},\overline{y})\Big)\Big]$. By choice of J_M this implies $A \models \forall \overline{y} \bigvee_{j\in J_M} \Big(\alpha^j(\overline{a}) \wedge \beta^j(\overline{y}) \wedge \mathrm{pos}^j(\overline{a},\overline{y}) \wedge \mathrm{neg}^j(\overline{a},\overline{y})\Big)$ and hence $A \models \varphi$. ◻

Since the union of finitely many testable properties is testable (see e.g. [18]), it is sufficient to show that the property $\mathcal{P}_\varphi$ is testable where φ is of the form

$$(6.2) \qquad \varphi = \exists\overline{x}\forall\overline{y}\chi(\overline{x},\overline{y}), \quad \text{where } \chi(\overline{x},\overline{y}) = \Big[\iota^M(\overline{x}) \wedge \bigvee_{j\in J_M} \Big(\beta^j(\overline{y}) \wedge \mathrm{pos}^j(\overline{x},\overline{y}) \wedge \mathrm{neg}^j(\overline{x},\overline{y})\Big)\Big],$$

for some $M \in \mathfrak{M}$. In the following, we will enforce that for every conjunctive clause of the big disjunction of χ, the universally quantified variables induce a specific substructure.

For $j \in J_M$ let $\mathfrak{H}_j \subseteq \mathcal{C}_{\sigma,d}$ be a maximal set of pairwise non-isomorphic structures H such that $H \models \beta^j(\overline{b})$ for some $\overline{b} = (b_1,\dots,b_\ell) \in U(H)^\ell$ with $\{b_1,\dots,b_\ell\} = U(H)$.

CLAIM 6.7. *We have*

$$\varphi \equiv_d \exists \overline{x} \forall \overline{y} \Big[\iota^M(\overline{x}) \wedge \bigvee_{\substack{H \in \mathfrak{H}_j, \\ j \in J_M}} \Big(\iota^H(\overline{y}) \wedge \mathrm{pos}^j(\overline{x}, \overline{y}) \wedge \mathrm{neg}^j(\overline{x}, \overline{y}) \Big) \Big].$$

Proof. Let $A \in \mathcal{C}_{\sigma,d}$ and $\overline{a} = (a_1, \dots, a_k) \in U(A)^k$. First assume that $A \models \forall \overline{y} \chi(\overline{a}, \overline{y})$. Hence for any tuple $\overline{b} \in U(A)^\ell$ there is an index $j \in J_M$ such that $A \models \beta^j(\overline{b}) \wedge \mathrm{pos}^j(\overline{a}, \overline{b}) \wedge \mathrm{neg}^j(\overline{a}, \overline{b})$. Then $A \models \beta^j(\overline{b})$ implies that $A[\{b_1, \dots, b_\ell\}] \cong H$ for some $H \in \mathfrak{H}_j$. Hence $A \models \iota^H(\overline{b})$ and $A \models \Big[\iota^M(\overline{a}) \wedge \bigvee_{\substack{H \in \mathfrak{H}_j, \\ j \in J_M}} \Big(\iota^H(\overline{b}) \wedge \mathrm{pos}^j(\overline{a}, \overline{b}) \wedge \mathrm{neg}^j(\overline{a}, \overline{b}) \Big) \Big]$.

For the other direction, we let $A \models \forall \overline{y} \Big[\iota^M(\overline{a}) \wedge \bigvee_{\substack{H \in \mathfrak{H}_j, \\ j \in J_M}} \Big(\iota^H(\overline{y}) \wedge \mathrm{pos}^j(\overline{a}, \overline{y}) \wedge \mathrm{neg}^j(\overline{a}, \overline{y}) \Big) \Big]$. Then for every tuple $\overline{b} \in U(A)^\ell$ there is an index $j \in J_M$ and $H \in \mathfrak{H}_j$ such that $H \models \iota^H(\overline{b}) \wedge \mathrm{pos}^j(\overline{a}, \overline{b}) \wedge \mathrm{neg}^j(\overline{a}, \overline{b})$. Therefore $A[\{b_1, \dots, b_\ell\}] \cong H$ and we know that $A \models \beta^j(\overline{b})$. Therefore $A \models \beta^j(\overline{b}) \wedge \mathrm{pos}^j(\overline{a}, \overline{b}) \wedge \mathrm{neg}^j(\overline{a}, \overline{b})$ and since this is true for any $\overline{b} \in U(A)^\ell$ we get $A \models \varphi$. ∎

Thus, it suffices to assume that

$$\text{(6.3)} \quad \varphi = \exists \overline{x} \forall \overline{y} \chi(\overline{x}, \overline{y}),$$
$$\text{where } \chi(\overline{x}, \overline{y}) := \Big[\iota^M(\overline{x}) \wedge \bigvee_{\substack{H \in \mathfrak{H}_j, \\ j \in J_M}} \Big(\iota^H(\overline{y}) \wedge \mathrm{pos}^j(\overline{x}, \overline{y}) \wedge \mathrm{neg}^j(\overline{x}, \overline{y}) \Big) \Big]$$

for some $M \in \mathfrak{M}$.

Next we will define a universally quantified formula ψ and show that $\mathcal{P}_\varphi$ is indistinguishable from the property $\mathcal{P}_\psi$. To do so we will need the two claims below. Intuitively, Claim 6.8 says that models of φ of bounded degree do not have many 'interactions' between existential and universal variables – only a constant number of tuples in relations combine both types of variables. Note that for a structure A and tuples $\overline{a} \in U(A)^k$, $\overline{b} = (b_1, \dots, b_\ell) \in U(A)^\ell$ the condition $A \models \iota^H(\overline{b}) \wedge \mathrm{pos}^j(\overline{a}, \overline{b}) \wedge \mathrm{neg}^j(\overline{a}, \overline{b})$ can force an element of $\overline{b}$ to be in a tuple (of a relation of A) with an element of $\overline{a}$, even if $\mathrm{pos}^j(\overline{x}, \overline{y})$ only contains literals of the form $x_i = y_{i'}$. (For example, it may be the case that for some tuple $\overline{b}' \in \{b_1, \dots, b_\ell\}^\ell$, every clause $\iota^{H'}(\overline{y}) \wedge \mathrm{pos}^{j'}(\overline{x}, \overline{y}) \wedge \mathrm{neg}^{j'}(\overline{x}, \overline{y})$ for which $A \models \iota^{H'}(\overline{b}') \wedge \mathrm{pos}^{j'}(\overline{a}, \overline{b}') \wedge \mathrm{neg}^{j'}(\overline{a}, \overline{b}')$ forces a tuple to contain some element of $\overline{b}'$ and some element of $\overline{a}$.) We will now define a set J to pick out the clauses that do not force a tuple to contain both an element from $\overline{a}$ and $\overline{b}$. Note that we still allow elements from $\overline{b}$ to be amongst the elements in $\overline{a}$. In Claim 6.8 we show that for every $A \in \mathcal{C}_{\sigma,d}$, $\overline{a} \in U(A)^k$ for which $A \models \forall \overline{y} \chi(\overline{a}, \overline{y})$ there are a constant number of tuples $\overline{b} \in U(A)^\ell$ that only satisfy clauses which force a tuple to contain both an element from $\overline{a}$ and from $\overline{b}$.

Let $j \in J_M$, $H \in \mathfrak{H}_j$ and $\overline{h} = (h_1, \dots, h_\ell) \in U(H)^\ell$ such that $H \models \iota^H(\overline{h})$. We define the set $P_{j,H} := \{h_i \mid i \in \{1, \dots, \ell\}, \mathrm{pos}^j(\overline{x}, \overline{y}) \text{ does not contain } y_i = x_{i'} \text{ for any } i' \in \{1, \dots, k\}\}$. Now we let $J \subseteq J_M \times \mathcal{C}_{\sigma,d}$ be the set of pairs (j, H), with $H \in \mathfrak{H}_j$ such that the disjoint union $M \sqcup H[P_{j,H}] \models \varphi$. Now J precisely specifies the clauses that can be satisfied by a structure A and tuple $\overline{a} \in U(A)^k$ and $\overline{b} \in U(A)^\ell$ where A does not contain any tuples both containing elements from $\overline{a}$ and $\overline{b}$.

CLAIM 6.8. *Let $A \in \mathcal{C}_{\sigma,d}$ and $\overline{a} = (a_1, \dots, a_k) \in U(A)^k$. If $A \models \forall \overline{y}\, \chi(\overline{a}, \overline{y})$ then there are at most $k \cdot d$ tuples $\overline{b} \in U(A)^\ell$ such that $A \not\models \bigvee_{(j,H) \in J} (\iota^H(\overline{b}) \wedge \mathrm{pos}^j(\overline{a}, \overline{b}) \wedge$*

$\mathrm{neg}^j(\overline{a}, \overline{b}))$.

Proof. Since $A \models \forall \overline{y}\, \chi(\overline{a}, \overline{y})$, it holds that $A \models \forall \overline{y} \bigvee_{\substack{H \in \mathfrak{H}_j,\\ j \in J_M}} \Big(\iota^H(\overline{y}) \wedge \mathrm{pos}^j(\overline{a}, \overline{y}) \wedge \mathrm{neg}^j(\overline{a}, \overline{y})\Big)$ by Equation(6.3). Now let $B := \{\overline{b} \in U(A)^\ell \mid A \not\models \bigvee_{(j,H) \in J} (\iota^H(\overline{b}) \wedge \mathrm{pos}^j(\overline{a}, \overline{b}) \wedge \mathrm{neg}^j(\overline{a}, \overline{b}))\} \subseteq U(A)^\ell$. Then each $\overline{b} \in B$ adds at least one to $\sum_{i=1}^k \deg_A(a_i)$. Since $A \in \mathcal{C}_{\sigma,d}$ implies that $\sum_{i=1}^k \deg_A(a_i) \leq k \cdot d$ we get that $|B| \leq k \cdot d$. ◻

Claim 6.9. *Let ψ be a formula of the form $\psi = \forall \overline{z} \chi(\overline{z})$ where $\overline{z} = (z_1, \dots, z_t)$ is a tuple of variables and $\chi(\overline{z})$ is a quantifier-free formula. Let $A \in \mathcal{C}_{\sigma,d}$ with $|U(A)| > d \cdot \mathrm{ar}(\sigma) \cdot t$ and let $b \in A$ be an arbitrary element. Let $A \models \psi$ and let A' be obtained from A by 'isolating' b, i. e. by deleting all tuples containing b from $R(A)$ for every $R \in \sigma$. Then $A' \models \psi$.*

Proof. First note that $A' \models \chi(\overline{a})$ for any tuple $\overline{a} = (a_1, \dots, a_t) \in (A \setminus \{b\})^t$ as no tuple over the set of elements $\{a_1, \dots, a_t\}$ has been deleted. Let $\overline{a} = (a_1, \dots, a_t) \in U(A)^t$ be a tuple containing b. Pick $b' \in U(A)$ such that $\mathrm{dist}_A(a_j, b') > 1$ for every $j \in \{1, \dots, t\}$. Such an element exists as $|U(A)| > d \cdot \mathrm{ar}(R) \cdot t$. Let $\overline{a}' = (a'_1, \dots, a'_t)$ be the tuple obtained from $\overline{a}$ by replacing any occurrence of b by b'. Hence $a_j \mapsto a'_j$ defines an isomorphism from $A'[\{a_1, \dots, a_t\}]$ to $A[\{a'_1, \dots, a'_t\}]$ since b is an isolated element in $A'[\{a_1, \dots, a_t\}]$ and b' is an isolated element in $A[\{a'_1, \dots, a'_t\}]$. Since $A \models \chi(\overline{a}')$, it follows that $A' \models \chi(\overline{a})$. ◻

Let $J' \subseteq J$ be the set of all pairs (j, H) for which $\mathrm{pos}^j(\overline{x}, \overline{y})$ is the empty conjunction. J' contains (j, H) for which we want to use $\iota^H(\overline{y})$ to define the formula ψ.

Claim 6.10. *The property $\mathcal{P}_\varphi$ with φ as in (6.3) is indistinguishable from the property $\mathcal{P}_\psi$ where $\psi := \forall \overline{y} \bigvee_{(j,H) \in J'} \iota^H(\overline{y})$.*

Proof. Let $\epsilon > 0$ and $N(\epsilon) = N := \frac{k \cdot \ell^2 \cdot d \cdot \mathrm{ar}(R)}{\epsilon}$ and $A \in \mathcal{C}_{\sigma,d}$ be any structure with $|U(A)| > N$.

First assume that $A \models \varphi$. The strategy is to isolate any element b which is contained in a tuple $\overline{b} \in U(A)^\ell$ such that $A \not\models \bigvee_{(j,H) \in J'} \iota^H(\overline{b})$ by deleting all tuples containing b. This will result in a structure which is ϵ-close to A and a model of ψ.

Let $\overline{a} \in U(A)^k$ be a tuple such that $A \models \forall \overline{y} \chi(\overline{a}, \overline{y})$. Let $B \subseteq U(A)^\ell$ be the set of tuples $\overline{b} \in U(A)^\ell$ such that $A \not\models \bigvee_{(j,H) \in J} (\iota^H(\overline{b}) \wedge \mathrm{pos}^j(\overline{a}, \overline{b}) \wedge \mathrm{neg}^j(\overline{a}, \overline{b}))$. Then $|B| \leq k \cdot d$ by Claim 6.8. Hence the structure A' obtained from A by deleting all tuples containing an element of $C := \{a_1, \dots, a_k\} \cup \big\{b \in A \mid \text{ there is } (b_1, \dots, b_\ell) \in B \text{ such that } b \in \{b_1, \dots, b_\ell\}\big\}$ is ϵ-close to A. Since $A \models \forall \overline{y} \chi(\overline{a}, \overline{y})$ implies $A \models \forall \overline{y} \bigvee_{\substack{H \in \mathfrak{H}_j,\\ j \in J_M}} \iota^H(\overline{y})$, by Claim 6.9 we know that $A' \models \forall \overline{y} \bigvee_{\substack{H \in \mathfrak{H}_j,\\ j \in J_M}} \iota^H(\overline{y})$. For any tuple $\overline{b} = (b_1, \dots, b_\ell) \in (U(A) \setminus C)^\ell$ we have by definition of J' that $A \models \iota^H(\overline{b})$ for some $(j, H) \in J'$. Furthermore $A[\{b_1, \dots, b_\ell\}] = A'[\{b_1, \dots, b_\ell\}]$ and hence $A' \models \bigvee_{(j,H) \in J'} \iota^H(\overline{b})$. Let $\overline{b} = (b_1, \dots, b_\ell) \in U(A)^\ell$ be any tuple containing elements from C and let $c_1, \dots, c_t \in C$ be those elements. Pick t elements $c'_1, \dots, c'_t \in U(A) \setminus C$ such that $\mathrm{dist}_A(a_i, c'_{i'}) > 1$, $\mathrm{dist}_A(c'_{i'}, b_i) > 1$ and $\mathrm{dist}_A(c'_i, c'_{i'}) > 1$ for suitable i, i'. This is possible as $|U(A)| > (k + 2\ell) \cdot d \cdot \mathrm{ar}(R)$ which guarantees the existence of $k + 2\ell$ elements of pairwise distance greater than 1. Let $\overline{b}' = (b'_1, \dots, b'_\ell)$ be the vector obtained from $\overline{b}$ by replacing c_i with c'_i. Since $\overline{b}' \in U(A)^\ell$ there must be j', $H' \in \mathfrak{H}_j$ such that $A \models \iota^{H'}(\overline{b}') \wedge \mathrm{pos}^{j'}(\overline{a}, \overline{b}') \wedge \mathrm{neg}^{j'}(\overline{a}, \overline{b}')$. By choice of $c'_1, \dots, c'_t$ we have that $\mathrm{pos}_{j'}(\overline{x}, \overline{y})$ must be the empty conjunction and hence $(j', H') \in J'$. Since additionally $b_i \mapsto b'_i$ defines an isomorphism of $A[\{b'_1, \dots, b'_\ell\}]$ and $A'[\{b_1, \dots, b_\ell\}]$ this implies that

$A' \models \bigvee_{(j,H)\in J'} \iota^H(\overline{b})$ for all $\overline{b} \in U(A)^\ell$ and hence $A' \models \psi$.

Now we prove the other direction. Let $A \models \psi$ with $|U(A)| > N$. The idea here is to plant the structure M somewhere in A. While this takes less then an ϵ-fraction of edge modifications the resulting structure will be a model of φ.

Take any set $B \subseteq A$ of $|U(M)|$ elements. Let A' be the structure obtained from A by deleting all edges incident to any element contained in B. Let A'' be the structure obtained from A' by adding all tuples such that the structure induced by B is isomorphic to M. This takes no more than $2\ell{\cdot}d{\cdot}\mathrm{ar}(R) < \epsilon{\cdot}d{\cdot}|U(A)|$ edge modifications. Let $\overline{a} \in B^k$ be such that $A \models \iota^M(\overline{a})$. By Claim 6.9 we get $A' \models \psi$. Therefore pick any tuple $\overline{b} = (b_1, \ldots, b_\ell) \in (U(A) \setminus B)^\ell$. Since by construction we have that all b_i's are of distance at least two from $\overline{a}$ we have that $A'' \models \bigvee_{(j,H)\in J'}(\iota^H(\overline{b}) \wedge \mathrm{neg}^j(\overline{a}, \overline{b}))$. By choice of M we also know that $A'' \models \bigvee_{H\in\mathfrak{H}_j,\, j\in J_M} \Big(\iota^H(\overline{b}) \wedge \mathrm{pos}^j(\overline{a}, \overline{b}) \wedge \mathrm{neg}^j(\overline{a}, \overline{b})\Big)$ for all $\overline{b} \in B^\ell$. Therefore pick $\overline{b} = (b_1, \ldots, b_\ell)$ containing both elements from B and from $U(A) \setminus B$. Now pick a tuple $\overline{b}' = (b'_1, \ldots, b'_\ell) \in (U(A) \setminus B)^\ell$ that equals $\overline{b}$ in all positions containing an element from $U(A) \setminus B$. As noted before there is $(j, H) \in J'$ such that $A'' \models (\iota^H(\overline{b}') \wedge \mathrm{neg}^j(\overline{a}, \overline{b}'))$. Hence $A''[\{b'_1, \ldots, b'_\ell\}]$ is isomorphic to H and further because $(j, H) \in J'$ the set $P_{j,H}$ (used in the definition of J) is the entire universe of H. Since $J' \subseteq J$ this means that by the definition of J we get $A''[\{a_1, \ldots, a_k, b'_1 \ldots b'_\ell\}] \cong A''[\{a_1, \ldots, a_k\}] \sqcup A''[\{b'_1 \ldots b'_\ell\}] \cong M \sqcup H[P_{j,H}] \models \varphi$. Since $\overline{b} \in \{a_1, \ldots, a_k, b'_1 \ldots b'_\ell\}^\ell$ this implies

$$A''[\{a_1, \ldots, a_k, b'_1 \ldots b'_\ell\}] \models \bigvee_{\substack{H\in\mathfrak{H}_j,\\ j\in J_M}} \Big(\iota^H(\overline{b}) \wedge \mathrm{pos}^j(\overline{a}, \overline{b}) \wedge \mathrm{neg}^j(\overline{a}, \overline{b})\Big).$$

Then $A'' \models \bigvee_{H\in\mathfrak{H}_j,\, j\in J_M} \Big(\iota^H(\overline{b}) \wedge \mathrm{pos}^j(\overline{a}, \overline{b}) \wedge \mathrm{neg}^j(\overline{a}, \overline{b})\Big)$ and hence $A'' \models \varphi$. ◻

Since $\psi \in \Pi_1$ we have that $\mathcal{P}_\psi$ is testable, and hence $\mathcal{P}_\varphi$ is testable by Claim 6.10. ◻

7. GSF-locality is not sufficient for proximity oblivious testing. In this section we show that the property $\mathcal{P}_{\text{graph}}$ can be defined by a generalised notion of forbidden subgraph introduced in [22] (Lemma 7.14). Here a subgraph is only forbidden if it is connected to the rest of the graph in a predefined way, i. e. for a vertex in a forbidden subgraph we can specify that it cannot have neighbours which are not contained in the subgraph itself. Combining our results we show that not every property definable by generalised forbidden subgraphs is testable in the bounded-degree model (Theorem 7.5). This implies a negative answer to a question posed by Goldreich and Ron in [22] (Question 1) which asks whether a small number of appearances of generalised forbidden subgraphs can be fixed with a small number of edge modification or whether any way of fixing the appearances invokes a chain reaction of necessary edge modifications. In the following we introduce the notions and results needed from [22].

7.1. Generalised subgraph freeness. In the following, we present the formal definitions of generalised subgraph freeness, GSF-local properties and the notion of non-propagation, which were introduced in [22].

Definition 7.1 (Generalized subgraph freeness (GSF)). *A* marked *graph is a graph with each vertex marked as either* 'full' *or* 'semifull' *or* 'partial'*. An* embedding *of a marked graph F into a graph G is an injective map $f : V(F) \to V(G)$ such that for every $v \in V(F)$ the following three conditions hold.*

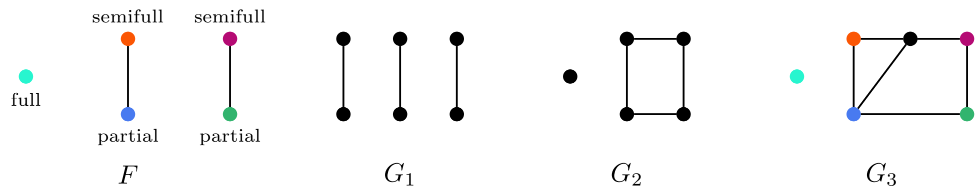


Fig. 7: In the depicted example, G_1 and G_2 are F-free (for G_1 we cannot find an embedding satisfying the condition for the full vertex and for G_2 we cannot find an embedding satisfying the condition for the semiful vertices). On the other hand, F can be embedded into G_3 (the embedding is indicated by colours).

1. *If v is marked 'full', then $N_1^G(f(v)) = f(N_1^F(v))$.*
2. *If v is marked 'semifull', then $N_1^G(f(v)) \cap f(V(F)) = f(N_1^F(v))$.*
3. *If v is marked 'partial', then $N_1^G(f(v)) \supseteq f(N_1^F(v))$.*

The graph G is called F-free if there is no embedding of F into G. For a set of marked graphs $\mathcal{F}$, a graph G is called $\mathcal{F}$-free if it is F-free for every $F \in \mathcal{F}$.

We refer to Figure 7 for an illustration of the definition of GSF. Based on the above definition of generalised subgraph freeness, we can define GSF-local properties.

Definition 7.2 (GSF-local properties). *Let $\mathcal{P} = \bigcup_{n\in\mathbb{N}} \mathcal{P}_n$ be a graph property where $\mathcal{P}_n = \{G \in \mathcal{P} \mid |V(G)| = n\}$ and $\overline{\mathcal{F}} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ a sequence of sets of marked graphs. $\mathcal{P}$ is called* $\overline{\mathcal{F}}$-local *if there exists an integer s such that for every n the following conditions hold.*

1. *$\mathcal{F}_n$ is a set of marked graphs, each of size at most s.*
2. *$\mathcal{P}_n$ equals the set of n-vertex graphs that are $\mathcal{F}_n$-free.*

$\mathcal{P}$ is called GSF-local *if there is a sequence $\overline{\mathcal{F}} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ of sets of marked graphs such that $\mathcal{P}$ is $\overline{\mathcal{F}}$-local.*

The following concept of a non-propagating condition for a sequence of sets of marked graphs was introduced in [22] to investigate constant-query POTs.

Definition 7.3 (Non-propagating). *Let $\overline{\mathcal{F}} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ be a sequence of sets of marked graphs.*

- *For a graph G, a subset $B \subset V(G)$* covers *$\mathcal{F}_n$ in G if for every marked graph $F \in \mathcal{F}_n$ and every embedding of F in G, at least one vertex of F is mapped to a vertex in B.*
- *The sequence $\overline{\mathcal{F}}$ is* non-propagating *if there exists a (monotonically non-decreasing) function $\tau : (0, 1] \to (0, 1]$ such that the following two conditions hold.*
  1. *For every $\epsilon > 0$ there exists $\beta > 0$ such that $\tau(\beta) < \epsilon$.*
  2. *For every graph G and every $B \subset V(G)$ such that B covers $\mathcal{F}_n$ in G, either G is $\tau(|B|/n)$-close to being $\mathcal{F}_n$-free or there are no n-vertex graphs that are $\mathcal{F}_n$-free.*

  A GSF-local property $\mathcal{P}$ is non-propagating *if there exists a non-propagating sequence $\overline{\mathcal{F}}$ such that $\mathcal{P}$ is $\overline{\mathcal{F}}$-local.*

In the above definition, the set B can be viewed as the set involving necessary modifications for repairing a graph G that does not satisfy the property $\mathcal{P}$ that is $\overline{\mathcal{F}}$-local, and the second condition says we do not need to modify G "much beyond" B. In particular, it implies that we can repair G without triggering a global "chain reaction".

Goldreich and Ron gave the following characterization for the proximity-oblivious testable properties in the bounded-degree graph model.

THEOREM 7.4 (Theorem 5.5 in [22]). *A graph property $\mathcal{P}$ has a constant-query proximity-oblivious tester if and only if $\mathcal{P}$ is GSF-local and non-propagating.*

The following open question was raised in [22].

OPEN QUESTION 1 (Are all GSF-local properties non-propagating?). *Is it the case that for every GSF-local property $\mathcal{P} = \bigcup_{n\in\mathbb{N}} \mathcal{P}_n$, there is a sequence $\overline{\mathcal{F}} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ that is non-propagating and $\mathcal{P}$ is $\overline{\mathcal{F}}$-local?*

We are now able to state our theorem answering Question 1. The rest of this section is dedicated to the proof of Theorem 7.5.

THEOREM 7.5. *There exists a GSF-local property of graphs of bounded degree 3 that is not testable in the bounded-degree graph model. Thus, not all GSF-local properties are non-propagating.*

7.2. Relating different notions of locality. In this section we define properties by prescribing upper and lower bounds on the number of occurrences of neighbourhood types. These bounds are given by *neighbourhood profiles* which we will define formally below. We use these properties to give a natural characterization of FO properties of bounded-degree structures in Lemma 7.7, which is a straightforward consequence of Hanf's Theorem (Theorem 2.1). We use this characterization to establish links between FO definability and GSF-locality. This connection is the key ingredient in the proof of our main theorem.

Observe that for fixed $r, d \in \mathbb{N}$ and σ, there are only finitely many r-types in structures in $\mathcal{C}_{\sigma,d}$. For any signature σ and $d, r \in \mathbb{N}$ we let $n_{d,r,\sigma} \in \mathbb{N}$ be the number of different r-types of σ-structures of degree at most d. Assuming that for all $d, r \in \mathbb{N}$ the r-neighbourhood-types of σ-structures of degree at most d are ordered, we let $\tau^i_{d,r,\sigma}$ denote the i-th such neighbourhood type, for $i \in \{1, \ldots, n_{d,r,\sigma}\}$. With each σ-structure $A \in \mathcal{C}_{\sigma,d}$ we associate its *r-histogram vector* $\overline{v}_{d,r,\sigma}(A)$, given by

$$(\overline{v}_{d,r,\sigma}(A))_i := |\{a \in U(A) \mid \mathcal{N}^A_r(a) \in \tau^i_{d,r,\sigma}\}|.$$

We let

$$\mathfrak{I} := \{[k, l] \mid k \leq l \in \mathbb{N}\} \cup \{[k, \infty) \mid k \in \mathbb{N}\}$$

be the set of all closed or half-closed, infinite intervals with natural lower/upper bounds.

DEFINITION 7.6. *Let σ be a signature and $d, r \in \mathbb{N}$.*

1. *An r-*neighbourhood profile *of degree d is a function $\rho : \{1, \ldots, n_{d,r,\sigma}\} \to \mathfrak{I}$.*
2. *For a structure $A \in \mathcal{C}_{\sigma,d}$, we say that A obeys ρ, denoted by $A \sim \rho$, if*

$$(\overline{v}_{d,r,\sigma}(A))_i \in \rho(i) \text{ for all } i \in \{1, \ldots, n_{d,r,\sigma}\}.$$

   Let $\mathcal{P}_\rho$ be the set of structures A that obey ρ, i.e., $\mathcal{P}_\rho = \{A \in \mathcal{C}_{\sigma,d} \mid A \sim \rho\}$.
3. *We say that a property $\mathcal{P}$ is* defined by a finite union of neighbourhood profiles *if there is $k \in \mathbb{N}$ such that $\mathcal{P} = \bigcup_{1\leq i\leq k} \mathcal{P}_{\rho_i}$ where ρ_i is an r_i-neighbourhood profile and $r_i \in \mathbb{N}$ for every $i \in \{1, \ldots, k\}$.*

We let $n_{d,r} := n_{d,r,\sigma_{\text{graph}}}$ denote the total number of r-types of directed graphs of degree at most d. We fix an ordering of the types and let $\tau^i_{d,r} := \tau^i_{d,r,\sigma_{\text{graph}}}$ be the

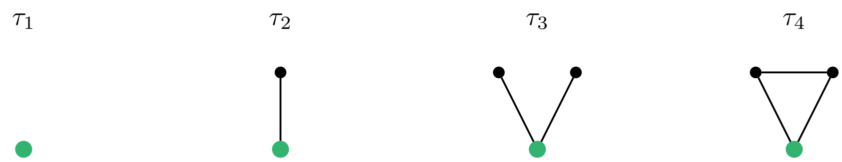


Fig. 8: All 1-types of bounded degree 2, where the centres are the large vertices.

i-th r-type of bounded degree d, for any $i \in \{1, \dots, n_{d,r}\}$. Further, for a graph G let $\overline{v}_{d,r}(G)$ denote the r-histogram vector of G. Note if G is undirected, for any type $\tau^i_{d,r}$ where the edge relation is not symmetric we have that $(\overline{v}_{d,r}(G))_i = 0$ and therefore in any r-neighbourhood profile ρ for graphs we have $\rho(i) = [0, 0]$ for any type $\tau^i_{d,r}$ which is not symmetric. For convenience, for undirected graphs we will ignore the non-symmetric types.

Let us consider the following example in which we find a representation by neighbourhood profiles for an FO-property.

Example 2. *Consider the following FO-sentence.*

$$\varphi := \forall x \forall y \neg E(x, y) \vee \forall x \exists y_1 \exists y_2 \Big(y_1 \neq y_2 \wedge E(x, y_1) \wedge E(x, y_2)$$
$$\wedge \forall z (z \neq y_1 \wedge z \neq y_2) \rightarrow \neg E(x, z) \Big).$$

The property P_φ defined by the sentence φ is the property containing all edgeless graphs and all graphs that are disjoint unions of cycles.

For degree bound 2 all 1-types are listed in Figure 8. Let $\rho_1 : \{1, \dots, 4\} \rightarrow \mathfrak{I}$ be the neighbourhood profile defined by $\rho_1(1) = [0, \infty)$ and $\rho_1(i) = [0, 0]$ for $i \in \{2, 3, 4\}$. Furthermore, let $\rho_2 : \{1, \dots, 4\} \rightarrow \mathfrak{I}$ be the neighbourhood profile defined by $\rho_2(i) = [0, \infty)$ for $i \in \{3, 4\}$ and $\rho_2(j) = [0, 0]$ for $j \in \{1, 2\}$. It is easy to observe that the properties P_φ and $P_{\rho_1} \cup P_{\rho_2}$ are equal.

Indeed representing FO-properties by neighbourhood profiles works in general. We now give a lemma showing that bounded-degree FO properties can be equivalently defined as finite unions of properties defined by neighbourhood profiles. Here the technicalities that arise are due to Hanf normal form not requiring the locality-radius of all Hanf-sentences to be the same.

Lemma 7.7. *For every non-empty property $\mathcal{P} \subseteq \mathcal{C}_{\sigma,d}$, $\mathcal{P}$ is FO definable on $\mathcal{C}_{\sigma,d}$ if and only if $\mathcal{P}$ can be obtained as a finite union of properties defined by neighbourhood profiles.*

Proof. For the first direction assume φ is an FO-sentence. Then by Hanf's Theorem (Theorem 2.1) there is a sentence ψ in Hanf normal form such that $\mathcal{P}_\varphi = \mathcal{P}_\psi$.

We will first convert ψ into a sentence in Hanf normal form where every Hanf sentence appearing has the same locality radius. Let $r \in \mathbb{N}$ be the maximum locality radius appearing in ψ, and let $\varphi^{\geq m}_\tau := \exists^{\geq m} x \varphi_\tau(x)$ be a Hanf sentence, where τ is an r'-type for some $r' \leq r$. Let $\tau_1, \dots, \tau_k$ be a list of all r-types of bounded degree d for which $(\mathcal{N}^B_{r'}(b), b) \in \tau$ for $(B, b) \in \tau_i$, for every $1 \leq i \leq k$. Let Π be the set of all partitions of m into k parts. Let

$$\tilde{\varphi}^{\geq m}_\tau := \bigvee_{(m_1, \dots, m_k) \in \Pi} \bigwedge_{i=1}^{k} \exists^{\geq m_i} x \varphi_{\tau_i}(x).$$

Claim 7.8. *$\varphi_{\tau}^{\geq m}$ is d-equivalent to $\tilde{\varphi}_{\tau}^{\geq m}$.*

Proof. Assume that $A \in \mathcal{C}_d$ satisfies $\varphi_{\tau}^{\geq m}$, and assume that $a_1, \dots, a_m$ are m distinct elements with $(\mathcal{N}_{r'}^{A}(a_j), a_j) \in \tau$, for every $1 \leq j \leq m$. Let $\tilde{\tau}_j$ be the r-type for which $(\mathcal{N}_{r}^{A}(a_j), a_j) \in \tilde{\tau}_j$. By choice of $\tau_1, \dots, \tau_k$, we get that there are indices $i_1, \dots, i_m$ such that $\tilde{\tau}_j = \tau_{i_j}$. For $i \in \{1, \dots, k\}$ let $m_i = |\{j \in \{1, \dots, m\} \mid i_j = i\}|$. Hence $A \models \bigwedge_{i=1}^{k} \exists^{\geq m_i} x \varphi_{\tau_i}(x)$ and since additionally $(m_1, \dots, m_k) \in \Pi$ this implies $A \models \tilde{\varphi}_{\tau}^{\geq m}$.

On the other hand, let $A \in \mathcal{C}_d$ satisfy $\tilde{\varphi}_{\tau}^{\geq m}$, and let $(m_1, \dots, m_k) \in \Pi$ be a partition of m such that $A \models \bigwedge_{i=1}^{k} \exists^{\geq m_i} x \varphi_{\tau_i}(x)$. For every $1 \leq i \leq k$, let $a_1^i, \dots, a_{m_i}^i$ be m_i distinct elements such that $(\mathcal{N}_{r}^{A}(a_j^i), a_j^i) \in \tau_i$, for every $1 \leq j \leq m_i$. By choice of $\tau_1, \dots, \tau_k$, we get that $(\mathcal{N}_{r'}^{A}(a_j^i), a_j^i) \in \tau$, for every pair $1 \leq i \leq k$, $1 \leq j \leq m_i$. But since $m_1 + \dots + m_k = m$ this implies that $A \models \varphi_{\tau}^{\geq m}$. This proves that $\varphi_{\tau}^{\geq m}$ and $\tilde{\varphi}_{\tau}^{\geq m}$ are d-equivalent. ∎

Let ψ' be the formula in which every Hanf-sentence $\varphi_{\tau}^{\geq m}$ for which τ is an r'-type for some $r' < r$ gets replaced by $\tilde{\varphi}_{\tau}^{\geq m}$. By a simple inductive argument using Claim 7.8, we get that ψ is d-equivalent to ψ', and hence $\mathcal{P}_{\varphi} = \mathcal{P}_{\psi} = \mathcal{P}_{\psi'}$. Furthermore since $\tilde{\varphi}_{\tau}^{\geq m}$ is a Boolean combination of Hanf-sentences for every $\varphi_{\tau}^{\geq m}$, and any Boolean combination of Boolean combinations is a Boolean combination itself, ψ' is in Hanf normal form. Furthermore, every Hanf-sentence appearing in ψ' has locality radius r by construction.

Since any Boolean combination can be converted into disjunctive normal form, we can assume that ψ' is a disjunction of sentences ξ of the form

$$\xi = \bigwedge_{j=1}^{k} \exists^{\geq m_j} x \varphi_{\tau_j}(x) \wedge \bigwedge_{j=k+1}^{\ell} \neg \exists^{\geq m_j + 1} x \varphi_{\tau_j}(x),$$

where $\ell \in \mathbb{N}_{\geq 1}$, $1 \leq k \leq \ell$, $m_i \in \mathbb{N}_{\geq 1}$ and τ_i is an r-type for every $1 \leq i \leq \ell$. We can further assume that every sentence in the disjunction ψ' is satisfiable by some $A \in \mathcal{C}_d$, as any sentence with no bounded degree d model can be removed from ψ'.

Let $\tilde{\tau}_1, \dots, \tilde{\tau}_t$ be a list of all r-types of bounded degree d in the order we fixed. Let $k_i := \max(\{m_j \mid 1 \leq j \leq k, \tau_j = \tilde{\tau}_i\} \cup \{0\})$ and $\ell_i := \min(\{m_j \mid k+1 \leq j \leq \ell, \tau_j = \tilde{\tau}_i\} \cup \{\infty\})$ for every $i \in \{1, \dots, t\}$. Since ξ has at least one bounded-degree model, $k_i \leq \ell_i$ for every $i \in \{1, \dots, t\}$. Let $\rho : \{1, \dots, t\} \to \mathfrak{I}$ be the neighbourhood profile defined by $\rho(i) := [k_i, \ell_i]$ if $\ell_i < \infty$ and $\rho(i) := [k_i, \ell_i)$ otherwise. Then by construction, we get that $\mathcal{P}_{\rho} = \mathcal{P}_{\xi}$. Since ψ' is a disjunction of formulas, each of which defines a property which can be defined by some neighbourhood profile, we get that $\mathcal{P}_{\psi'}$ must be a finite union of properties defined by some neighbourhood profile.

On the other hand, for every r-neighbourhood profile ρ of degree d, $\tau_1, \dots, \tau_t$ a list of all r-types of bounded degree d in the order fixed and the formula

$$\varphi_{\rho} := \bigwedge_{\substack{i \in \{1, \dots, t\}, \\ \rho(i) = [k_i, \ell_i]}} \left(\exists^{\geq k_i} x \varphi_{\tau_i}(x) \wedge \neg \exists^{\geq \ell_i + 1} x \varphi_{\tau_i}(x) \right) \wedge \bigwedge_{\substack{i \in \{1, \dots, t\}, \\ \rho(i) = [k_i, \infty)}} \exists^{\geq k_i} x \varphi_{\tau_i}(x)$$

it clearly holds that $\mathcal{P}_{\rho} = \mathcal{P}_{\varphi_{\rho}}$. Hence every finite union of properties defined by neighbourhood profiles can be defined by the disjunction of the formulas φ_{ρ} of all ρ in the finite union. ∎

7.2.1. Relating FO properties to GSF-local properties. We now prove that FO properties which arise as unions of neighbourhood profiles of a particularly simple form are GSF-local. For this let

$$\mathfrak{I}_0 := \{[0,k] \mid k \in \mathbb{N}\} \cup \{[0,\infty)\} \subset \mathfrak{I}.$$

We call any neighbourhood profile ρ with codomain $\mathfrak{I}_0$ a 0-*profile*, as all lower bounds for the occurrence of types are 0.

OBSERVATION 2. *Let ρ be a 0-profile. If two structures $A, A' \in \mathcal{C}_{\sigma,d}$ satisfy $(\overline{v}_{d,r,\sigma}(A))_i \leq (\overline{v}_{d,r,\sigma}(A'))_i$ for every $i \in \{1,\dots,n_{d,r,\sigma}\}$ and $A' \sim \rho$, then $A \sim \rho$.*

In particular, this implies that there cannot be a 0-profile which defines the property of all structures containing at least one occurrence of τ, for any r-type τ.

THEOREM 7.9. *Every finite union of properties of undirected graphs defined by 0-profiles is GSF-local.*

Proof. We prove this in two parts (Claim 7.10 and Claim 7.11). We first argue that every property $\mathcal{P}_\rho$ defined by some 0-profile $\rho : \{1,\dots,n_{d,r}\} \to \mathfrak{I}_0$ is GSF-local. For this it is important to note that we can express a forbidden r-type τ by a forbidden generalised subgraph. For $(B,b) \in \tau$, the set of all graphs with no vertex of neighbourhood type τ is the set of all B-free graphs where every vertex in $V(B)$ of distance less than r to b is marked 'full' and every vertex in $V(B)$ of distance r to b is marked 'semifull'. Since a profile of the form $\rho : \{1,\dots,n_{d,r,\sigma}\} \to \mathfrak{I}_0$ can express that some neighbourhood type τ can appear at most k times for some fixed $k \in \mathbb{N}$, we need to forbid all marked graphs in which type τ appears $k+1$ times. We will formalise this in the following claim.

CLAIM 7.10. *For every r-neighbourhood profile $\rho : \{1,\dots,n_{d,r}\} \to \mathfrak{I}_0$, there is a finite set $\mathcal{F}$ of marked graphs such that $\mathcal{P}_\rho$ is exactly the property of $\mathcal{F}$-free graphs.*

Proof. Assume τ is an r-type and $k \in \mathbb{N}_{>0}$. Then we say that a marked graph F is a k-*realisation* of τ if F has the following properties.

1. There are k distinct vertices $v_1,\dots,v_k$ in F such that $(\mathcal{N}_r^F(v_i), v_i) \in \tau$ for every $i = 1,\dots,k$.
2. Every vertex v in F has distance less or equal to r to at least one vertex v_i.
3. Every vertex v in F of distance less than r to at least one v_i is marked as 'full'.
4. Every vertex v in F of distance greater or equal to r to every v_i is marked as 'semifull'.

We denote by $S^k(\tau)$ the set of all k-realisations of τ.

Now we can define the set $\mathcal{F}$ of forbidden subgraphs to be

$$\mathcal{F} := \bigcup_{k\in\mathbb{N}, 1\leq i\leq n_{d,r,\sigma} : \rho(i)=[0,k]} S^{k+1}(\tau^i_{d,r}).$$

Let $\mathcal{P}$ be the property of all $\mathcal{F}$-free graphs. We first prove that the property $\mathcal{P}$ is contained in $\mathcal{P}_\rho$. Towards a contradiction assume that $G \in \mathcal{C}_d$ is $\mathcal{F}$-free but not contained in $\mathcal{P}_\rho$. As G is not contained in $\mathcal{P}_\rho$ there must be an index $i \in \{1,\dots,n_{d,r}\}$ such that $(\overline{v}_{d,r}(G))_i \notin \rho(i)$. Since $\rho(i) \in \mathfrak{I}_0$ there is $k \in \mathbb{N}$ such that $\rho(i) = [0,k]$ and hence $(\overline{v}_{d,r}(G))_i > k$. Hence there must be $k+1$ vertices $v_1,\dots,v_{k+1}$ in G such that $(\mathcal{N}_r^G(v_i), v_i) \in \tau^i_{d,r}$. We define the marked graph F to be the subgraph of G induced by the r-neighbourhoods of $v_1,\dots,v_{k+1}$, i.e. $G[\bigcup_{1\leq i\leq k+1} N_r^G(v_i)]$, in which every vertex of distance less than k to at least one of the v_i is marked as 'full' and every other vertex is marked 'semifull'. Then F is by definition a $(k+1)$-realisation of $\tau^i_{d,r}$ and hence $F \in \mathcal{F}$. We now argue that F can be embedded into G. Since F is an induced subgraph of G the identity map gives us a natural embedding $f : F \to G$. Let v be any vertex marked 'full' in F. By construction of F, there is $i \in \{1, \dots, k+1\}$ such that $f(v)$ is of distance less than r to v_i in G. But then $N^G_1(f(v))$ is a subset of $N^G_r(v_i)$. As F without the marking is the subgraph of G induced by $\bigcup_{1 \le i \le k+1} N^G_r(v_i)$ this implies that $f(N^F_1(v)) = N^G_1(f(v))$. Furthermore, assume v is a vertex marked 'semifull' in F. Then $f(N^F_1(v)) = N^G_1(f(v)) \cap f(V(F))$ holds as F without the markings is an induced subgraph of G. This proves that G is not F-free by Definition 7.1. This is a contradiction to our assumption that G is $\mathcal{F}$-free and $F \in \mathcal{F}$.

Similarly, we can show that $\mathcal{P}_\rho \subseteq \mathcal{P}$ by assuming $G \in \mathcal{C}_d$ is in $\mathcal{P}_\rho$ but not $\mathcal{F}$-free, and showing that the embedding of any graph of $\mathcal{F}$ into G yields an amount of vertices of a certain type contradicting containment in $\mathcal{P}_\rho$. ∎

Next we prove that classes defined by excluding finitely many marked graphs are closed under finite unions.

Claim 7.11. *Let $\mathcal{F}_1, \mathcal{F}_2$ be two finite sets of marked graphs. For $i \in \{1, 2\}$, let $\mathcal{P}_i$ be the property of $\mathcal{F}_i$-free graphs. Then there is a set $\mathcal{F}$ of generalised subgraphs such that $\mathcal{P}_1 \cup \mathcal{P}_2$ is the property of $\mathcal{F}$-free graphs.*

Proof. We say that a marked graph F is a (not necessarily disjoint) union of marked graphs F_1, F_2 if

1. there is an embedding f_i of F_i into the graph F without its markings as in Definition 7.1 for every $i \in \{1, 2\}$.
2. for every vertex v in F there is $i \in \{1, 2\}$ and a vertex w in F_i such that $f_i(w) = v$.
3. every vertex v in F is marked 'full', if there is $i \in \{1, 2\}$ and a 'full' vertex w in F_i such that $f_i(w) = v$.
4. every vertex v in F is marked 'semifull', if there is $i \in \{1, 2\}$ and a 'semifull' vertex w in F_i such that $f_i(w) = v$ and $f_i(u) \neq v$ for every $i \in \{1, 2\}$ and every 'full' vertex u.
5. every vertex v in F is marked 'partial' if $f_i(u) \neq v$ for every $i \in \{1, 2\}$ and every 'full' or 'semifull' vertex u.

We define $S(F_1, F_2)$ to be the set of all possible (not necessarily disjoint) unions of F_1, F_2. We can now define the set $\mathcal{F}$ to be

$$\mathcal{F} := \bigcup_{F_1 \in \mathcal{F}_1, F_2 \in \mathcal{F}_2} S(F_1, F_2).$$

Let $\mathcal{P}$ be the property of all $\mathcal{F}$-free graphs. Now we prove $\mathcal{P} \subseteq \mathcal{P}_1 \cup \mathcal{P}_2$. Towards a contradiction assume G is $\mathcal{F}$-free but G is in neither $\mathcal{P}_1$ nor in $\mathcal{P}_2$. Then for every $i \in \{1, 2\}$ there is a graph $F_i \in \mathcal{F}_i$ such that G is not F_i-free. It is easy to see that there is a union $F_\cup$ of F_1 and F_2 such that G is not $F_\cup$-free, which contradicts that G is $\mathcal{F}$-free.

Conversely, in order to prove $\mathcal{P}_1 \cup \mathcal{P}_2 \subseteq \mathcal{P}$, if G is $\mathcal{F}_i$ free for some $i \in \{1, 2\}$ then G must be $\mathcal{F}$-free by construction of $\mathcal{F}$. ∎

Combining the two claims above proves Theorem 7.9. ∎

Further discussion of the relation between FO and GSF-locality. First let us remark that it is neither true that every FO definable property is GSF-local, nor that every GSF-local property is FO definable.

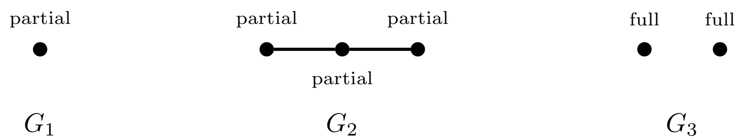


Fig. 9: Marked graphs for Example 5.

EXAMPLE 3. *The property of bounded-degree graphs containing a triangle is FO definable but not GSF-local.*

Indeed, the existence of a fixed number of vertices of certain neighbourhood types can be expressed in FO, while in general, this cannot be expressed by forbidding generalised subgraphs. If a formula has a 0-profile (and hence does not require the existence of any types) then the property defined by that formula is GSF-local, as shown in Theorem 7.9.

EXAMPLE 4. *The class of all bounded-degree graphs with an even number of vertices is GSF-local but not FO definable.*

Let us remark that Theorem 7.9 combined with Lemma 7.7 proves that every finite union of properties definable by 0-profiles is both FO definable and GSF-local. Hence it is natural to ask whether the intersection of FO definable properties and GSF-local properties is precisely the set of finite unions of properties definable by 0-profiles. However, this is not the case. The following example shows that there are properties which are both FO definable and GSF-local but cannot be expressed by 0-profiles.

EXAMPLE 5. *We let $d \geq 2$ and let $B_1 := (\{v\},\{\})$, $B_2 = (\{v,w\},\{\{v,w\}\})$ be two graphs. We further let τ_1, τ_2 be the 1-types of degree d such that $(B_1, v) \in \tau_1$ and $(B_2, v) \in \tau_2$. Consider the property $\mathcal{P}$ defined by the following FO formula*

$$\varphi := \neg\exists x(x = x) \vee \exists^{=1} x\big(\varphi_{\tau_1}(x) \wedge \forall y(x \neq y \to \varphi_{\tau_2}(y))\big).$$

$\mathcal{P}$ contains, besides the empty graph, unions of an arbitrary amount of disjoint edges and one isolated vertex. To define a sequence of forbidden subgraphs we let G_1, G_2, G_3 be the marked graphs in Figure 9. Let $\mathcal{F}_{\text{even}} := \{G_1\}$ and $\mathcal{F}_{\text{odd}} := \{G_2, G_3\}$ and let $\overline{\mathcal{F}} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ where $\mathcal{F}_i = \mathcal{F}_{\text{even}}$ if i is even and $\mathcal{F}_i = \mathcal{F}_{\text{odd}}$ if i is odd. Note that every graph on more than one vertex with an odd number of vertices which is $\mathcal{F}_{\text{odd}}$-free must contain a vertex of neighbourhood type τ_1, and that the set of $\mathcal{F}_{\text{even}}$-free graphs contains only the empty graph. Hence $\mathcal{P}$ is $\overline{\mathcal{F}}$-local. Now assume towards a contradiction that $\mathcal{P} = \bigcup_{1\leq i\leq k} \mathcal{P}_{\rho_i}$ for 0-profiles ρ_i. Let G_m be the graph consisting of m disjoint edges and one isolated vertex and H_m the graph consisting of m disjoint edges. Since $G_m \in \mathcal{P}$ there is $i \in \{1,\dots,k\}$ such that $G_m \sim \rho_i$. By choice of G_m and H_m we have $0 \leq (\overline{v}_{d,r}(H_m))_j \leq (\overline{v}_{d,r}(G_m))_j \in \rho_i(j)$ for every $j \in \{1,\dots,n_{d,r}\}$. Since additionally $\rho_i(j) \in \mathfrak{I}_0$ this implies that $(\overline{v}_{d,r}(H_m))_j \in \rho_i(j)$. But then $H_m \sim \rho_i$ which yields a contradiction as $H_m \notin \mathcal{P}$. Hence $\mathcal{P}$ cannot be defined as a finite union of 0-profiles.

Figure 10 gives a schematic overview of all classes of properties discussed here and their relationship.

7.3. Proving the existence of a GSF-local non-testable property. In this section we prove Theorem 7.5. We show that the property $\mathcal{P}_{ⓩ}$ from Section 4 can be

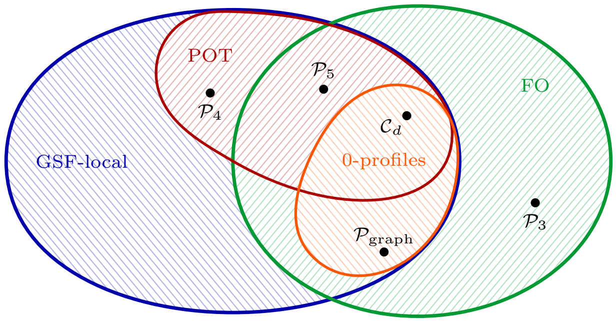


Fig. 10: Overview of the classes of properties, here $\mathcal{P}_i$ refers to the property from Example i, $\mathcal{C}_d$ refers to the property of all graphs of bounded degree d and $\mathcal{P}_{\text{graph}}$ is the property defined in Section 5.2.

expressed by a union of 0-profiles. We then show that the local reduction from $\mathcal{P}_{\textcircled{z}}$ to $\mathcal{P}_{\text{graph}}$ given in Section 5.2 preserves the expressibility by 0-profiles, and hence by Theorem 7.9 $\mathcal{P}_{\text{graph}}$ is GSF-local.

Let σ be the signature, $d \in \mathbb{N}$ and $\mathcal{P}_{\textcircled{z}}$ be the property of d σ-structures of bounded-degree from Section 3.

7.3.1. Characterisation of the relational structure property by neighbourhood profiles. Our aim in this section is to prove that the property $\mathcal{P}_{\textcircled{z}}$ of relational structures can be written as a finite union of properties defined by 0-profiles of radius 2.

For all σ-structures in $\mathcal{P}_{\textcircled{z}}$ (excluding $A_\emptyset$) it is crucial that they are allowed to contain precisely one root element. Hence the neighbourhood profile describing $\mathcal{P}_{\textcircled{z}}$ must restrict the number of occurrences of the 2-type of the root element. But since in $\mathcal{P}_{\textcircled{z}} \setminus \{A_\emptyset\}$, the root elements in different structures may have different 2-types, we partition $\mathcal{P}_{\textcircled{z}} \setminus \{A_\emptyset\}$ into parts $\mathcal{P}_1, \dots, \mathcal{P}_m$ by the 2-type of the root element. Note that the number m of parts is constant as there are at most $n_{d,2,\sigma}$ 2-types in total. For each of these parts we then define a neighbourhood profile ρ_k such that $\mathcal{P}_k \cup \{A_\emptyset\} = \mathcal{P}_{\rho_k}$. We would like to remark here that the roots of all but one structure in $\mathcal{P}_{\textcircled{z}}$ actually have the same 2-types. Hence the partition only contains two parts and one of the two parts only contains one structure. We now define the parts and corresponding profiles formally.

Assume without loss of generality that the 2-types $\tau^1_{d,2,\sigma}, \dots, \tau^{n_{d,2,\sigma}}_{d,2,\sigma}$ of degree d are ordered in such a way that for $(B, b) \in \tau^k_{d,2,\sigma}$, it holds that $B \models \varphi_{\text{root}}(b)$ if and only if $k \in \{1, \dots, m\}$ for some $m \leq n_{d,2,\sigma}$. For $k \in \{1, \dots, m\}$, let

$$\mathcal{P}_k := \{A \in \mathcal{P}_{\textcircled{z}} \setminus \{A_\emptyset\} \mid \text{ there is } a \in U(A) \text{ such that } (\mathcal{N}^A_2(a), a) \in \tau^k_{d,2,\sigma}\}.$$

Since by Lemma 3.5 every $A \in \mathcal{P}_{\textcircled{z}} \setminus \{A_\emptyset\}$ must contain exactly one root we get that

$$\mathcal{P}_{\textcircled{z}} = \bigcup_{1 \leq k \leq m} \mathcal{P}_k \cup \{A_\emptyset\}$$

and this union is disjoint. Furthermore, for $k \in \{1, \dots, m\}$, let $I_k \subseteq \{1, \dots, n_{d,2,\sigma}\}$

be the set of indices j such that there is a structure $A \in \mathcal{P}_k$ and $a \in U(A)$ with $(\mathcal{N}_2^A(a), a) \in \tau^j_{d,2,\sigma}$. For every $k \in \{1, \dots, m\}$ we define the 2-neighbourhood profile $\rho_k : \{1, \dots, n_{d,2,\sigma}\} \to \mathfrak{I}_0$ by

$$\rho_k(i) := \begin{cases} [0,1] & \text{if } i = k, \\ [0,\infty) & \text{if } i \in I_k \setminus \{k\}, \\ [0,0] & \text{otherwise.} \end{cases}$$

To prove that these 0-profiles of radius 2 define the property $\mathcal{P}_{\textcircled{z}}$, the crucial observation is that for every element a of some structure in $\mathcal{C}_{\sigma,d}$, the FO-formula $\varphi_{\textcircled{z}}$ only talks about elements of distance at most 2 to a (i. e. $\varphi_{\textcircled{z}}$ is 2-local). Hence the 2-histogram vector of a structure already captures whether the structure satisfies $\varphi_{\textcircled{z}}$. We will now formally prove this.

Lemma 7.12. *It holds that* $\mathcal{P}_{\textcircled{z}} = \bigcup_{1 \le k \le m} \mathcal{P}_{\rho_k}$.

Proof. We first prove that $\mathcal{P}_{\textcircled{z}} \subseteq \bigcup_{1 \le k \le m} \mathcal{P}_{\rho_k}$. First note that trivially $A_\emptyset \in \bigcup_{1 \le k \le m} \mathcal{P}_{\rho_k}$. Now assume $A \in \mathcal{P}_{\textcircled{z}} \setminus \{A_\emptyset\}$. This implies that there is $k \in \{1, \dots, m\}$ such that $A \in \mathcal{P}_k$. By construction we have that for every $a \in A$, there is $i \in I_k$ such that $(\mathcal{N}_2^A(a), a) \in \tau^i_{d,2,\sigma}$. Furthermore, since $A \models \varphi_{\textcircled{z}}$ and $U(A) \neq \emptyset$, we have by Lemma 3.5 that $A \models \exists^{=1} x \varphi_{\text{root}}(x)$, and that there can be at most one $a \in U(A)$ such that $(\mathcal{N}_2^A(a), a) \in \tau^k_{d,2,\sigma}$. Therefore $A \in \mathcal{P}_{\rho_k}$.

To prove $\bigcup_{1 \le k \le m} \mathcal{P}_{\rho_k} \subseteq \mathcal{P}_{\textcircled{z}}$, we prove that every structure in $\bigcup_{1 \le k \le m} \mathcal{P}_{\rho_k}$ must satisfy $\varphi_{\textcircled{z}}$. We will prove that every $A \in \bigcup_{1 \le k \le m} \mathcal{P}_{\rho_k}$ satisfies $\varphi_{\text{recursion}}$, and refer for the proof that A satisfies $\varphi_{\text{tree}} \wedge \varphi_{\text{rotationMap}} \wedge \varphi_{\text{base}}$ to Claim A.1, Claim A.2 and Claim A.3 in Appendix A. Note that $A_\emptyset \models \varphi_{\textcircled{z}}$ by Lemma 3.10 and hence we exclude $A_\emptyset$ in the following.

Claim 7.13. *Every structure* $A \in \bigcup_{1 \le k \le m} \mathcal{P}_{\rho_k} \setminus \{A_\emptyset\}$ *satisfies* $\varphi_{\text{recursion}}$.

Proof. Let $A \in \bigcup_{1 \le k \le m} \mathcal{P}_{\rho_k} \setminus \{A_\emptyset\}$. Then there is a $k \in \{1, \dots, m\}$ such that $A \in \mathcal{P}_{\rho_k}$.

By definition, $\varphi_{\text{recursion}} := \forall x \forall z \big(\varphi(x,z) \vee \psi(x,z)\big)$ (see Section 3), where

$$\varphi(x,z) := \neg \exists y F(x,y) \wedge \neg \exists y F(z,y) \text{ and}$$

$$\psi(x,z) := \bigwedge_{\substack{k_1',k_2' \in [D]^2 \\ \ell_1',\ell_2' \in [D]^2}} \Bigg(\exists y \big[E_{k_1',\ell_1'}(x,y) \wedge E_{k_2',\ell_2'}(y,z) \big] \to$$

$$\bigwedge_{\substack{i,j,i',j' \in [D], k,\ell \in ([D]^2)^2 \\ \mathrm{ROT}_H(k,i) = ((k_1',k_2'),i') \\ \mathrm{ROT}_H((\ell_2',\ell_1'),j) = (\ell,j')}} \exists x' \exists z' \big[F_k(x,x') \wedge F_\ell(z,z') \wedge E_{(i,j),(j',i')}(x',z') \big] \Bigg).$$

Let $a, c \in U(A)$. Assume first that there is $b \in U(A)$ with $(a,b) \in F(A)$. Hence $A \not\models \varphi(a,c)$. Since $\varphi_{\text{recursion}} := \forall x \forall z \big(\varphi(x,z) \vee \psi(x,z)\big)$ we aim to prove $A \models \psi(a,c)$. By construction of ρ_k, there is an $i \in I_k$ such that $(\mathcal{N}_2^A(a), a) \in \tau^i_{d,2,\sigma}$. Therefore there is a structure $\tilde{A} \models \varphi_{\textcircled{z}}$ and $\tilde{a} \in U(\tilde{A})$ such that $(\mathcal{N}_2^A(a), a) \cong (\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Let f be an isomorphism from $(\mathcal{N}_2^A(a), a)$ to $(\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Since $b \in N_2^A(a)$, we get that $f(b)$

is defined. Since f is an isomorphism mapping a onto $\tilde{a}$, we have that $(a,b) \in F(A)$ implies that $(\tilde{a}, f(b)) \in F(\tilde{A})$. Hence $\tilde{A} \not\models \varphi(\tilde{a},\tilde{c})$, for every $\tilde{c} \in U(\tilde{A})$. But since $\tilde{A} \models \varphi_{\text{recursion}}$, as $\tilde{A} \models \varphi_{\textcircled{Z}}$, this shows that $\tilde{A} \models \psi(\tilde{a},\tilde{c})$ for every $\tilde{c} \in U(\tilde{A})$.

Let $k'_1, k'_2 \in [D]^2$ and $\ell'_1, \ell'_2 \in [D]^2$ be indices such that there is $b' \in U(A)$ with $(a,b') \in E_{k'_1,\ell'_1}(A)$ and $(b',c) \in E_{k'_2,\ell'_2}(A)$. Since $b', c \in N_2^A(a)$, by assumption we get that $f(b')$ and $f(c)$ are defined. Furthermore, $(a,b') \in E_{k'_1,\ell'_1}(A)$ and $(b',c) \in E_{k'_2,\ell'_2}(A)$ imply that $(\tilde{a}, f(b')) \in E_{k'_1,\ell'_1}(\tilde{A})$ and $(f(b'), f(c)) \in E_{k'_2,\ell'_2}(\tilde{A})$, since f is an isomorphism mapping a onto $\tilde{a}$. We proved in the previous paragraph that $\tilde{A} \models \psi(\tilde{a}, f(c))$. Hence we can conclude that for all indices $i,j,i',j' \in [D]$, $k,\ell \in ([D]^2)^2$ for which $\mathrm{ROT}_H(k,i) = ((k'_1,k'_2),i')$ and $\mathrm{ROT}_H((\ell'_2,\ell'_1),j) = (\ell,j')$, there are elements $\tilde{a}',\tilde{c}' \in U(\tilde{A})$ such that $(\tilde{a},\tilde{a}') \in F_k(\tilde{A})$, $(f(c),\tilde{c}') \in F_\ell(\tilde{A})$, and $(\tilde{a}',\tilde{c}') \in E_{(i,j),(j',i')}(\tilde{A})$. Since $\tilde{a}',\tilde{c}' \in N_2^{\tilde{A}}(\tilde{a})$, we get that $a' := f^{-1}(\tilde{a}')$ and $c' := f^{-1}(\tilde{c}')$ are defined. Furthermore, we get that $(a,a') \in F_k(A)$, $(c,c') \in F_\ell(A)$ and $(a',c') \in E_{(i,j),(j',i')}(A)$. This proves that $A \models \psi(a,c)$.

In the case that there is $b \in U(A)$ with $(c,b) \in F(A)$, we can prove similarly that $A \models \psi(a,c)$, by considering that there exist $\tilde{A} \models \varphi_{\textcircled{Z}}$ and $\tilde{c} \in U(\tilde{A})$ such that $(\mathcal{N}_2^A(a), c) \cong (\mathcal{N}_2^{\tilde{A}}(\tilde{c}), \tilde{c})$ by construction of ρ_k. Finally if there is no $b \in U(A)$ such that $(a,b) \in F(A)$ or $(c,b) \in F(A)$ then $A \models \varphi(a,c)$. Since this covers every case we get that $A \models \varphi_{\text{recursion}}$. ◻

Assume $A \in \bigcup_{1\le k\le m} \mathcal{P}_{\rho_k}$. As proved in Claims A.1, A.2, A.3 and 7.13 this implies that $A \models \varphi_{\text{tree}}$, $A \models \varphi_{\text{rotationMap}}$, $A \models \varphi_{\text{base}}$ and $A \models \varphi_{\text{recursion}}$. Since $\varphi_{\textcircled{Z}}$ is a conjunction of these formulas, we get $A \models \varphi_{\textcircled{Z}}$ and hence $A \in \mathcal{P}_{\textcircled{Z}}$. ◻

7.3.2. The graph property is GSF-local. Let $\mathcal{P}_{\text{graph}}$ be the graph property as defined in Section 5.2 and let $f : \mathcal{C}_{\sigma,d} \to \mathcal{C}_3$ be the local reduction from $\mathcal{P}_{\textcircled{Z}}$ to $\mathcal{P}_{\text{graph}}$. We now use this local reduction and the expressibility of $\mathcal{P}_{\textcircled{Z}}$ by 0-profiles to show that $\mathcal{P}_{\text{graph}}$ is GSF-local.

LEMMA 7.14. *The graph property $\mathcal{P}_{\text{graph}}$ is GSF-local.*

Proof. For this we will prove that $\mathcal{P}_{\text{graph}}$ is equal to a finite union of properties defined by 0-profiles, and then use Theorem 7.9 to prove that $\mathcal{P}_{\text{graph}}$ is GSF-local. We define the 0-profiles for $\mathcal{P}_{\text{graph}}$ in a very similar way to the relational structure case, and then use the description of $\mathcal{P}_{\textcircled{Z}}$ by 0-profiles shown in Lemma 7.12. To this end, let $\ell' := 24\ell + 18 + d$ and assume that the ℓ'-types $\tau^1_{d,\ell'}, \dots, \tau^{n_{d,\ell'}}_{d,\ell'}$ are ordered in such a way that $(\mathcal{N}^{f(B)}_{\ell'}(u_{b,1}), u_{b,1}) \in \tau^k_{d,\ell'}$, for every $k \in \{1,\dots,m\}$ and $(B,b) \in \tau^k_{d,2,\sigma}$, where m is the number of parts of the partition of $\mathcal{P}_{\textcircled{Z}}$ defined in Subsection 7.3.1. For $k \in \{1,\dots,m\}$, let $\hat{I}_k$ be the set of indices i such that there is $A \in \mathcal{P}_k$, and $v \in V(f(A))$ for which $(\mathcal{N}^{f(A)}_{\ell'}(v), v) \in \tau^i_{d,\ell'}$. Let $\hat{\rho}_k : \{1,\dots,n_{d,\ell'}\} \to \mathfrak{I}_0$ be defined by

$$\hat{\rho}_k(i) := \begin{cases} [0,1] & \text{if } i = k, \\ [0,\infty) & \text{if } i \in \hat{I}_k \setminus \{k\}, \\ [0,0] & \text{otherwise.} \end{cases}$$

CLAIM 7.15. *It holds that $\mathcal{P}_{\text{graph}} = \bigcup_{1\le k\le m} \mathcal{P}_{\hat{\rho}_k}$.*

Proof. First we prove $\mathcal{P}_{\text{graph}} \subseteq \bigcup_{1\le k\le m} \mathcal{P}_{\hat\rho_k}$. Assume $G \in \mathcal{P}_{\text{graph}}$ and let $A \in \mathcal{P}_{\textcircled{Z}}$ be a structure such that $G = f(A)$. If $A = A_\emptyset$ then clearly $G \in \bigcup_{1\le k\le m} \mathcal{P}_{\hat\rho_k}$. Hence assume $A \neq A_\emptyset$. Then $A \in \mathcal{P}_k$ for some $k \in \{1,\dots,m\}$. By the construction of $\hat I_k$ we know that for every $v \in V(G)$ we have $(\mathcal{N}^G_{\ell'}(v), v) \in \tau^i_{d,\ell'}$ for some $i \in \hat I_k$. Furthermore, since $A \in \mathcal{P}_k$ there is at most one $a \in U(A)$ with $(\mathcal{N}^A_2(a), a) \in \tau^k_{d,2,\sigma}$. We argue that this implies that there can be at most one vertex $v \in V(G)$ with $(\mathcal{N}^G_{\ell'}(v), v) \in \tau^k_{d,\ell'}$. Let us denote the set of vertices of G associated with an element a of A by $V_a = \{u_{a,i}, v^k_{a,i} | 1 \le i \le d, 1 \le k \le 6\ell + 5\}$. By construction of $\hat\rho$ we know that only element-vertices can have type $\tau^k_{d,\ell'}$. Assume that vertex $u_{b,i}$ has type $\tau^k_{d,\ell'}$. By choice of ℓ', we know that the ℓ'-neighbourhood of $u_{b,i}$ must contain the sets V_a for every element a in the 2-neighbourhood of b. Hence, b must have type $\tau^k_{d,2,\sigma}$. We further note that any two element-vertices $u_{b,i}$ and $u_{b,j}$, $i \neq j$ cannot be of the same neighbourhood type. To see this, observe that there are several distinct arrows attached to element b. These include the arrow representing the tuple $(b, b) \in R(A)$ and the arrow representing the unique tuple $(b, b') \in F_1(A)$. These two arrows are attached to distinct vertices amongst $u_{b,1},\dots,u_{b,d}$, say $u_{b,i'}$ and $u_{b,j'}$, $i' \neq j'$. Note that $u_{b,i}$ and $u_{b,j}$ must have different distance (on the cycle $(u_{b,1},\dots,u_{b,d})$) to either $u_{b,i'}$ or $u_{b,j'}$. Since the ℓ-neighbourhoods of both $u_{b,i}$ and $u_{b,j}$ must encompass $u_{b,i'}$ and $u_{b,j'}$ along with the arrow-gadget representing arrows (b, b) and (b, b'), it follows that $u_{b,i}$ and $u_{b,j}$ necessarily possess distinct ℓ-neighbourhoods. Hence $G \in \mathcal{P}_{\hat\rho}$.

Now we prove that $\bigcup_{1\le k\le m} \mathcal{P}_{\hat\rho_k} \subseteq \mathcal{P}_{\text{graph}}$. Let $G \in \bigcup_{1\le k\le m} \mathcal{P}_{\hat\rho_k}$ and let $k \in \{1,\dots,m\}$ be an index such that $G \in \mathcal{P}_{\hat\rho_k}$. Further assume that G is not the empty graph, as $f(A_\emptyset) \in \mathcal{P}_{\text{graph}}$ is the empty graph.

Since for every i for which $\hat\rho(i) \neq [0, 0]$, there is a graph $G' \in \mathcal{P}_{\text{graph}}$ and $v \in V(G')$ such that $(\mathcal{N}^{G'}_{\ell'}(v'), v') \in \tau^i_{d,\ell'}$, we get that the ℓ'-neighbourhood of every vertex in G appears in some graph $G' \in \mathcal{P}_{\text{graph}}$. By choice of ℓ' we get that every vertex $v \in V(G)$ is either contained in a cycle of length d and is the endpoint of some k-arrow, k-loop or non-arrow or v is an internal vertex of a k-arrow, k-loop or non-arrow. Hence, we obtain a σ-structure A with $f(A) \cong G$ by replacing any cycle C of length d by an element a_C and adding a tuple $(a_C, a_{C'})$ to the relation $R_k(A)$ if there are vertices u on C and v on C' such that $u \xrightarrow{k} v$ in G. Let g be an isomorphism from $f(A)$ to G.

Now we argue that $A \in \mathcal{P}_{\rho_k}$. First assume that there are two elements $a, b \in U(A)$ with $(\mathcal{N}^A_2(a), a) \in \tau^k_{d,2,\sigma}$ and $(\mathcal{N}^A_2(b), b) \in \tau^k_{d,2,\sigma}$. By definition, we get that $(\mathcal{N}^{f(A)}_{\ell'}(u_{a,1}), u_{a,1}) \in \tau^k_{d,\ell'}$ and $(\mathcal{N}^{f(A)}_{\ell'}(u_{b,1}), u_{b,1}) \in \tau^k_{d,\ell'}$. Since g is an isomorphism, the restriction of g to $N^{f(A)}_{\ell'}(u_{a,1})$ must be an isomorphism from $\mathcal{N}^{f(A)}_{\ell'}(u_{a,1})$ to $\mathcal{N}^G_{\ell'}(g(u_{a,1}))$, and hence $(\mathcal{N}^G_{\ell'}(g(u_{a,1})), g(u_{a,1})) \cong (\mathcal{N}^{f(A)}_{\ell'}(u_{a,1}), u_{a,1}) \in \tau^k_{d,\ell'}$. But the same holds for the ℓ'-ball of $g(u_{b,1})$, and hence we contradict the assumption that $G \in \mathcal{P}_{\hat\rho_k}$ since $\hat\rho_k(k) = [0, 1]$. Let us further assume that there is an $a \in U(A)$ such that $(\mathcal{N}^A_2(a), a) \in \tau^i_{d,2,\sigma}$ for some $i \notin I_k$. Since $G \in \mathcal{P}_{\hat\rho_k}$ we get $(\mathcal{N}^G_{\ell'}(g(u_{a,1})), g(u_{a,1})) \in \tau^j_{d,\ell'}$ for some $j \in \hat I_k$. But then by construction of $\hat\rho_k$, there must be $G' \in \mathcal{P}_{\text{graph}}$, and a vertex $v \in V(G')$ such that $(\mathcal{N}^{G'}_{\ell'}(v), v) \in \tau^j_{d,\ell'}$. Furthermore, since $\ell' > d$ the vertex v must be contained in cycle of length d. By construction of $\mathcal{P}_{\text{graph}}$, there is a structure $A \in \mathcal{P}_{\textcircled{Z}}$ such that $f(A') = G'$. Since v is contained in a cycle of length d, v must be an element-vertex corresponding to some element $a' \in U(A')$. Since we picked ℓ' in such a way that $f(\mathcal{N}^{A'}_2(a')) \subseteq \mathcal{N}^{G'}_{\ell'}(v)$, we get $(\mathcal{N}^{A'}_2(a'), a') \in \tau^i_{d,2,\sigma}$ by choice of i and j. Hence $A' \notin \mathcal{P}_{\rho_k}$. But this contradicts Lemma 7.12.

Hence we have shown that $A \in P_{\rho_k}$. Then by Lemma 7.12 $A \in \mathcal{P}_{\textcircled{z}}$, and by construction $G \in \mathcal{P}_{\text{graph}}$. ☐

Since by Claim 7.15 we can express $\mathcal{P}_{\text{graph}}$ as a finite union of properties, each defined by a 0-profile, Theorem 7.9 implies that $\mathcal{P}_{\text{graph}}$ is GSF-local. ☐

7.3.3. Putting everything together. Now we prove Theorem 7.5.

Proof of Theorem 7.5. Combining Theorem 4.4, Lemma 5.3 and Lemma 5.4 we obtain that the graph property $\mathcal{P}_{\text{graph}}$ is not testable. Lemma 7.14 shows that $\mathcal{P}_{\text{graph}}$ is also a GSF-local property. Hence there exists a GSF-local property of bounded-degree graphs which is not testable. Furthermore, since having a POT implies being testable, this proves that there is a GSF-local property which has no POT. By Theorem 7.4 this implies that not all GSF-local properties are non-propagating. ☐

7.4. GSF-local properties of graphs of bounded degree 1 and 2 are non-propagating. In this section, we show that the degree 3 from Theorem 7.5 of the example of a GSF-local property which is propagating is optimal, in the sense that all GSF-local properties of graphs of bounded degree 1 and 2 are non-propagating. We note that Ito et al. [27] claimed that every GSF-local sequence of bounded degree at most 2 is non-propagating in the appendix of their paper. However, there is one subtle issue in their proof, as they only considered *connected* forbidden generalized subgraphs (which are called forbidden configurations in [27]). In the following, we resolve this issue. Indeed, the extension from connected forbidden generalised subgraphs to arbitrary forbidden generalized subgraphs is non-trivial and requires an involved proof which we present in this section.

We first observe that even for graphs of bounded degree 1, not every sequence of marked graphs $\overline{\mathcal{F}}$ is non-propagating as the following example shows. A similar example was given in [22].

Example 6. *Let $\mathcal{P} \subseteq \mathcal{C}_1$ be the property of $\overline{\mathcal{F}}$-free graphs, where F is the marked graph depicted in Figure 11, $\mathcal{F}_n = \{F\}$ and $\overline{\mathcal{F}} = (\mathcal{F}_n)_{n\in\mathbb{N}}$. Let G_k be the graph consisting of k edges and one isolated vertex. Then the set B containing the one isolated vertex of G_k covers all embeddings of F (see Figure 11). But the only way to make G_k F-free is to remove all k edges of G_k. Hence G_k is $1/2$-far from being F-free, which implies that $\mathcal{P}$ is propagating for $\overline{\mathcal{F}}$.*

However, the property $\mathcal{P}$ is non-propagating, as we show in the proof of Theorem 7.16. Indeed, consider the alternative sequence of marked graphs $\overline{\mathcal{F}} = (\mathcal{F}_n)_{n\in\mathbb{N}}$, where $\mathcal{F}_n = \{F\}$ for n even and $\mathcal{F}_n = \{F, \tilde{F}\}$ for n odd. Clearly, in G_{2k+1} any set $\tilde{B}$ covering $\mathcal{F}_{2k+1}$ must contain one incident vertex of every edge. Hence the number of necessary modifications is at most $|B|$, suggesting that $\mathcal{P}$ is non-propagating.

Indeed, adding certain redundant marked graphs to the sequence $\overline{\mathcal{F}} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ to control the behaviour of sets covering $\mathcal{F}_n$ as in Example 6 works in general both in the degree 1 and degree 2 case and will be our proof strategy for the following theorem. More precisely, for a property $\mathcal{P}$ of graphs of bounded degree 2, a sequence $\overline{\mathcal{F}}$ of marked graphs such that $\mathcal{P}$ is $\overline{\mathcal{F}}$-local and a bound k on the size of any graph appearing in $\overline{\mathcal{F}}$, we add forbidden generalized subgraphs to $\overline{\mathcal{F}}$ in the following way. In case there is no graph in $\mathcal{P}$ with n vertices containing a set of different small connected components (connected components with at most k vertices) each with frequency at least k, we add a generalised subgraph forbidding precisely this combination of connected components to $\mathcal{F}_n$. Additionally, if no graph in $\mathcal{P}$ with n vertices contains a set of different small connected components each with frequency at least k and

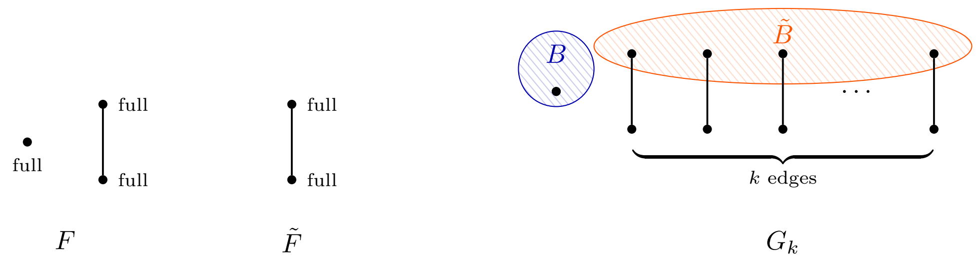


Fig. 11: Marked graphs F and $\tilde{F}$ and graph G_k from Example 6.

one large component (connected component with at least $k+1$ vertices), we add a generalised subgraph forbidding precisely this combination of connected components to $\mathcal{F}_n$. Now for a graph G which is not in $\mathcal{P}$ and a set B covering all forbidden generalised subgraphs in G, we look at what types of connected components appear in the part of G not containing vertices from B. In case G is large enough and B is small enough, we observe that some types of connected components have to appear with high frequency, or there must be a large component in the part of G which is not covered by B. By adding redundant subgraphs as described earlier, this now implies that there must be a graph G' in $\mathcal{P}$, which has the same structure as G on a large subset of the part of G which is not covered by B. Hence we can modify G to obtain a graph satisfying the property $\mathcal{P}$ (by changing G to G') without modifying G much beyond B. The restriction to bounded degree at most 2 is crucial in this argument as it gives us the necessary control over large connected components.

Theorem 7.16. *Any GSF-local property $\mathcal{P} \subseteq \mathcal{C}_d$ for $d \leq 2$ is non-propagating.*

Proof. We only consider the case that $\mathcal{P} \subseteq \mathcal{C}_2$. We can consider any property $\mathcal{P} \subseteq \mathcal{C}_1$ as a property in $\mathcal{C}_2$ by forbidding any vertex to have degree 2, i. e. adding a path of length 2 in which both degree 1 vertices are marked 'partial' and the degree 2 vertex is marked 'full' to every set of forbidden marked graphs in any sequence defining $\mathcal{P}$, and adjusting constants in the following argument to account for the degree being 1 instead of 2.

Let $\mathcal{P} = \bigcup_{n\in\mathbb{N}} \mathcal{P}_n$ and $\overline{\mathcal{F}} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ be a sequence of marked graphs such that $\mathcal{P}$ is $\overline{\mathcal{F}}$-local. By definition there exists $k \in \mathbb{N}$ such that every marked graph appearing in $\overline{\mathcal{F}}$ contains at most k vertices.

For two sets $I \subseteq [k] := \{0, \dots, k-1\}$, $J \subseteq \{3, \dots, k\}$ such that $I \cup J \neq \emptyset$ let $F_{I,J}$ be the marked graph which is the disjoint union of k paths of length i for every $i \in I$ and k cycles of length j for every $j \in J$ in which every vertex is marked as 'full'. Be aware that a path of length i contains $i+1$ vertices and a cycle of length j contains j vertices. Note that graphs that are $F_{I,J}$-free cannot contain at the same time k connected components that are paths of length i for every $i \in I$ and k connected components which are cycles of length j for every $j \in J$. We let $F^{\text{large}}_{\emptyset,\emptyset}$ be a path of length $k+1$ in which both vertices of degree 1 are marked as 'partial' and every other vertex is marked 'full'. We further let $F^{\text{large}}_{I,J}$ be the disjoint union of $F_{I,J}$ and $F^{\text{large}}_{\emptyset,\emptyset}$ for $I \subseteq [k]$, $J \subseteq \{3, \dots, k\}$ with $I \cup J \neq \emptyset$. Note that graphs that are $F^{\text{large}}_{I,J}$-free cannot contain at the same time k connected components that are paths of length i for every $i \in I$ and k connected components which are cycles of length j for every $j \in J$ and one connected component containing at least $k+1$ vertices.

We obtain a sequence $\overline{\mathcal{F}}' = (\overline{\mathcal{F}}'_n)_{n\in\mathbb{N}}$ by setting

$$\mathcal{F}'_n := \mathcal{F}_n \cup \Big\{ F \in \{F^{\text{large}}_{\emptyset,\emptyset}, F_{I,J}, F^{\text{large}}_{I,J} : I \subseteq [k], J \subseteq \{3,\dots,k\},$$
$$I \cup J \neq \emptyset\} : \text{ every } G \in \mathcal{P}_n \text{ is } F\text{-free}\Big\}.$$

First observe that by construction $\mathcal{P}$ must be $\overline{\mathcal{F}}'$-local.

We use the following notation. For a graph $G \in \mathcal{C}_2$, $i \in [k]$ and $j \in \{3,\dots,k\}$ we let

- $p_i(G)$ be the number of connected components of G that are path of length i.
- $c_j(G)$ be the number of connected components of G that are cycles of length j.
- $cc^{\text{large}}(G)$ be the number of connected components of G with more than k vertices.

We choose the following (monotonically non-decreasing) function $\tau(\epsilon) := \min(1, 8k^3\epsilon)$ for $\epsilon \in (0,1]$. First consider the two trivial cases. If $\mathcal{P}_n$ contains all n-vertex graphs in $\mathcal{C}_2$, then every n-vertex graph is close to $\mathcal{P}_n$. On the other hand, if $\mathcal{P}_n$ is empty, then there are no n-vertex $\mathcal{F}_n$-free graphs. Thus, the condition for non-propagation is satisfied. Hence, we may assume that $\mathcal{P}_n$ neither contains all n-vertex graphs in $\mathcal{C}_2$ nor is empty.

Let G be an n-vertex graph which is not $\mathcal{F}'_n$-free. Let $B \subseteq V(G)$ be any set covering $\mathcal{F}'_n$. To show that $\mathcal{F}'$ is non-propagating it is sufficient to show that G is $\tau(|B|/n)$-close to $\mathcal{P}$. By choice of τ this means that we have to argue that we can make G have property $\mathcal{P}_n$ by modifying at most $16k^3|B|$ edges. Hence for the remainder of this proof we argue that G is $\tau(|B|/n)$-close to $\mathcal{P}$.

Assuming $n < 8k^3$, we get that $\tau(|B|/n) = 1$ (since G is not $\mathcal{F}'_n$-free we know that $|B| \geq 1$), which means G is $\tau(|B|/n)$-close to $\mathcal{P}$, as in this case we can modify all edges of G and hence we can make G into any graph in $\mathcal{P}_n$. Hence we now assume that $n \geq 8k^3$.

Now consider the case that $|B| \geq \frac{n}{8k}$. In this case $\tau(|B|/n) = 1$ and G is $\tau(|B|/n)$-close to $\mathcal{P}$ again because we are allowed to modify all edges of G which allows us to make G into any graph in $\mathcal{P}_n$. Hence from now on we only consider the case that $|B| \leq \frac{n}{8k}$.

Let S be the set of vertices for which the k-neighbourhood does not contain any vertex from B. Let $I \subseteq [k]$, $J \subseteq \{3,\dots,k\}$ be the sets of indices such that $i \in I$ if and only if $p_i(G[S]) \geq k$ and $j \in J$ if and only if $c_j(G[S]) \geq k$. Note that $I \cup J$ could be empty.

Case 1: Assume that $F^{\text{large}}_{I,J} \notin \mathcal{F}'_n$.

First note that every component of size at most k which contains a vertex from S cannot contain a vertex from B by definition of S. Hence every connected component of G of size at most k is either fully contained in S or disjoint from S. Since there are at most $2k$ isomorphism types of connected components of size at most k we know that there are at most $2k^2$ connected components X of $G[S]$ such that there are at most $k-1$ other connected components of $G[S]$ isomorphic to X. In other words, there are at most $2k^2$ components X of G containing no element from B such that if X is a path of length i then $i \notin I$ and if X is a cycle of length j then $j \notin J$. We now obtain G' by the following edge modifications from G. For every cycle of length

j where $j \notin J$, we delete one edge (at most $2k^2 + |B|$ edge by our previous argument). Then we add edges connecting all path (including the paths obtained in the last step) of length i for $i \notin I$ to one long cycle C (at most $2k^2 + |B|$ edge additions). If C has length less or equal to k there must be $i \in I$ or $j \in J$ such that $p_i(G) > k$ or $c_j(G) > k$, in which case we include one respective component in C and repeat this until C has length at least $k+1$ (at most $2k$ modifications). Since in total we did at most $4k^2 + 2|B| + 2k \leq 16k^2|B|$ edge modifications, G is $\tau(|B|/n)$-close to G'. The following claim completes the proof of Case 1 by showing that $G' \in \mathcal{P}_n$.

Claim 7.17. *Let $I \subseteq [k]$, $J \subseteq \{3, \ldots, k\}$ and $a_i, b_j \geq k$ where $i \in I$, $j \in J$ be any selection of integers such that*

$$\sum_{i \in I} i \cdot a_i + \sum_{j \in J} j \cdot b_j \leq n - (k+1). \tag{7.1}$$

If $F^{\text{large}}_{I,J} \notin \mathcal{F}'_n$, then any n-vertex graph $H \in \mathcal{C}_2$ with $p_i(H) = a_i$, $c_j(H) = b_j$ for $i \in I$, $j \in J$, and one additional connected component which is a cycle is $\mathcal{F}'_n$-free.

Proof. Assume there is a graph $H \in \mathcal{P}_n$ as given in the statement which is not $\mathcal{F}'_n$-free and let C be the cycle in H of length larger than k. Then there is $F \in \mathcal{F}'_n$ such that there is an embeddings $f : V(F) \to V(H)$. Since $F^{\text{large}}_{I,J} \notin \mathcal{F}'_n$, by construction there is a graph $H' \in \mathcal{P}_n$ with $p_i(H') \geq k$, $c_j(H') \geq k$ for $i \in I$, $j \in J$ and $cc^{\text{large}}(H') \geq 1$. We let C' be a connected component of H' of size larger than k. To find an embedding of F into H', for every connected component X of H of size at most k which contains a vertex from $f(V(F))$, we pick a unique connected component X' of H' which is isomorphic to X. Note that because $|f(V(F))| \leq k$ and $p_i(H') \geq k$, $c_j(H') \geq k$ we can pick the connected component in H' uniquely. For every connected component X of H of size at most k which contains a vertex from $f(V(F))$, we now define f_X to be an isomorphism from X to X'. Furthermore, we pick an injective graph homomorphism $f^{\text{large}} : f(V(F)) \cap C \to C'$. Again, this is possible because $|f(V(F))| \leq k$. We now let $f'(v) := f_X(f(v))$ if $f(v)$ is in the connected component X and $f'(v) := f^{\text{large}}(f(v))$ if $f(v)$ is in C. Note that f' is injective by construction. Furthermore, as a consequence of picking f_X to to be isomorphisms and f^{large} to be a homomorphism we get that f' is an embedding of F into H'. To see this we observe that for any vertex $v \in V(F)$ which is marked as 'full' and for which $f'(v)$ is in a connected component X with at most k vertices we obtain the condition $N_1^{H'}(f'(v)) = f'(N_1^F(v))$ from f_X being an isomorphism. On the other hand, in case $f'(v)$ is in C' and v is marked 'full' we get that v has two neighbours w_1, w_2 in F and $f(w_1)$, $f(w_2)$ are neighbours of $f(v)$ (since $f(v)$ must be on C) which implies that $f'(w_1)$ and $f'(w_2)$ are neighbours of $f'(v)$ (since f^{large} is a homomorphism). Since f' is an embedding of F into H' we obtain a contradiction to $H' \in \mathcal{P}_n$ and hence H is $\mathcal{F}'_n$-free. Therefore H must be $\mathcal{F}'_n$-free as claimed. ∎

Case 2: Assume that $F^{\text{large}}_{I,J} \in \mathcal{F}'_n$. In this case our strategy is to modify the connected components of G containing a vertex from B into paths and cycles of length i for $i \in I$ or $i \in J$, respectively.

Since the k-neighbourhood of every vertex contains no more than $2k+1$ vertices, $|B| \leq \frac{n}{8k}$ implies that $|S| \geq n/2$. Furthermore, since $F^{\text{large}}_{I,J} \in \mathcal{F}'_n$ no vertex in S can be contained in a connected component of size larger than k as otherwise there would be an embedding of $F^{\text{large}}_{I,J}$ into G which is not covered by B. Hence $G[S]$ is the disjoint union of paths of length at most $k-1$ and cycles of length at most k. Since $|S| \geq 4k^3$

and $G[S]$ contains at most $2k$ different isomorphism types of connected components and each of the connected components has at most k vertices we conclude that at least $2k \geq k+1$ of the connected components of $G[S]$ are pairwise isomorphic. Hence $I \cup J \neq \emptyset$. Furthermore, $F_{I,J}$ is defined and not in $\mathcal{F}'_n$ since B covers $\mathcal{F}'_n$.

The next claim is the key to showing that we can modify G into having property $\mathcal{P}$ without modifying more than a constant number of edges in $G[S]$.

CLAIM 7.18. *If for $I \subseteq [k]$, $J \subseteq \{3, \dots, k\}$ with $I \cup J \neq \emptyset$ we have that $F_{I,J}^{\text{large}} \in \mathcal{F}'_n$ and $F_{I,J} \notin \mathcal{F}'_n$ then for any selection of integers $a_i, b_j \geq k$ where $i \in I$, $j \in J$ such that*

$$\sum_{i \in I} i \cdot a_i + \sum_{j \in J} j \cdot b_j \leq n - k^3 \tag{7.2}$$

there is an $\mathcal{F}'_n$-free graph $H \in \mathcal{P}_n$ such that $p_i(H) \geq a_i$ and $c_j(H) \geq b_j$.

Proof. We set $a_i = 0$ for $i \in [k] \setminus I$ and $b_j = 0$ for $j \in \{3, \dots, k\} \setminus J$. Since $F_{I,J}^{\text{large}} \in \mathcal{F}'_n$ and $F_{I,J} \notin \mathcal{F}'_n$ by construction of $\overline{\mathcal{F}}$ there must be a graph in $\mathcal{P}_n$ whose connected components include at least k paths of length i for every $i \in I$, k cycles of length j for every $j \in J$ and no connected component containing more than k vertices. Pick H amongst all graphs in $\mathcal{P}_n$ with these properties such that

$$(*) := \sum_{\substack{i \in [k] \\ p_i(H) < a_i}} a_i - p_i(H) + \sum_{\substack{j \in \{3, \dots, k\} \\ c_j(H) < b_j}} b_j - c_j(H)$$

is minimal. In case $(*) > 0$ there is $i \in [k]$ such that either $p_i(H) < a_i$ or $c_i(H) < b_i$. Combining this with Equation 7.2 we obtain that there must be $j \in [k]$ such that either $p_j(H) - a_j > k$ or $c_j(H) - b_j > k$. We let H' be the graph obtained from H by replacing i connected components which are paths of length j or i connected components which are cycles of length j, respectively, and adding j disjoint paths of length i or j disjoint cycles of length i, respectively. By choice of i, j we get that

$$(*) > \sum_{\substack{i \in [k] \\ p_i(H') < a_i}} a_i - p_i(H') + \sum_{\substack{j \in \{3, \dots, k\} \\ c_j(H') < b_j}} b_j - c_j(H').$$

Furthermore, H' must be $\mathcal{F}'_n$-free which we will argue in the following. Assume this is not the case and there is $F \in \mathcal{F}'_n$ and an embedding $f : V(F) \to V(H')$. We obtain a map $f' : V(F) \to V(H)$ from f as follows. For every connected component X in H' which has been altered we pick a unique connected component X' of H' which is isomorphic to X and contains no vertex in the image $f(V(F))$. This is possible as the assumption that X was altered implies that either $|X| - 1 \in I$ or $|X| \in J$ and hence there are at least k connected components isomorphic to X in H' which were not altered. Since further $|f(V(F))| \leq k$ we can pick the X' uniquely. We now let f_X be an isomorphism from X to X' for every connected component X which has been altered and f_X the identity for every connected component X which has not been altered. We define $f'(v) := f_X(v)$ for $v \in X$. By construction f' is obviously an embedding of F into H. Since $H \in \mathcal{P}_n$ this yields a contradiction. Hence the existence of H' contradicts the assumption that $(*) > 0$ which implies that H has the claimed properties. ◻

First observe that $n \geq 8k^3$ allows us to chose a_i and b_j for every $i \in I$ and $j \in J$ in such a way that $k \leq a_i \leq p_i(G[S])$, $k \leq b_j \leq c_j(G[S])$ and $\sum_{i \in I} i \cdot a_i + \sum_{j \in J} j \cdot b_j \leq n - k^3$.

Amongst all such choices we pick a_i and b_j such that $\sum_{i\in I} i \cdot a_i + \sum_{j\in J} j \cdot b_j$ is maximum. Let M be a set of vertices containing all connected components of G apart from a_i paths and b_j cycles from $G[S]$ for every $i \in I$, $j \in J$. Then $|M| \leq 2k|B| + |B| + 4k^3$ since M consists of $N_k^G(B)$ (at most $2k|B| + |B|$ vertices), all vertices in a connected component which is either a path of length i for $i \notin I$ or a cycle of length j for $j \notin J$ (since there are at most $2k^2$ such paths and cycles (as argued in Case 1) and each contains at most k vertices) or in case $a_i \neq p_i(G)$ or $b_j \neq c_j(G)$ for some $i \in I$, $j \in J$, M consist of at most $k^3 + k$ vertices as we picked a_i, b_j to maximise $\sum_{i\in I} i \cdot a_i + \sum_{j\in J} j \cdot b_j$.

Now we use Claim 7.18 and obtain an $\mathcal{F}'_n$-free graph $H \in \mathcal{P}_n$ such that $p_i(H) \geq a_i$ and $c_j(H) \geq b_j$. Hence we can modify G into a graph G' which is isomorphic to H by only modifying $G[M]$. Since we can modify $G[M]$ into any graph with no more than $4k|B| + 2|B| + 8k^3 \leq 16k^3|B|$ modifications we showed that G is $\tau(|B|/n)$-close to having $\mathcal{P}$. ∎

8. Conclusion. We studied testability of properties definable in first-order logic in the bounded-degree model of property testing for graphs and relational structures, where *testability* of a property means that it is testable with constant query complexity. We showed that all properties in Σ_2 are testable (Theorem 6.1), and we complemented this by exhibiting a property (of relational structures) in Π_2 that is not testable (Theorem 4.7). Using a hardness reduction, we also exhibit a property of undirected, 3-regular graphs in Π_2 that is not testable (Theorem 5.1). The question whether first-order definable properties are testable with a *sublinear* number of queries (e.g. $\sqrt{n}$) in the bounded-degree model is left open.

Similar results (on the separation between Σ_2 and Π_2 properties) were obtained in the dense graph model in [4], albeit with very different methods. Indeed, non-testability of first-order logic in the bounded-degree model is somewhat unexpected: Testing algorithms proceed by sampling vertices and then exploring their local neighbourhoods, and it is well-known that first-order logic can only express 'local' properties. On graphs and structures of bounded degree this is witnessed by Hanf's strong normal form of first-order logic [24], which is built around the absence and presence of different isomorphism types of local neighbourhoods. However, our negative result shows that locality of first-order logic is not sufficient for testability. This also answers an open question from [1].

We obtained our non-testable properties by encoding the zig-zag construction of bounded-degree expanders into first-order logic on relational structures (Theorem 4.4) and then extending this to undirected graphs (Theorem 5.1). We believe that this will be of independent interest. We remark that it might also be possible to use the iterative construction of replacement product graphs of [33] instead of the zig-zag construction to obtain a similar example.

We then used our non-testable graph property to answer a question on *proximity oblivious* testers in the bounded-degree model, asked by Goldreich and Ron more than 10 years ago [22]. Such a tester is particularly simple: it performs a basic test a number of times that may depend on the proximity parameter, whereas the basic test is oblivious of the parameter. In [22], the properties that are testable in this model have been characterised as those that are both *GSF-local*, and *non-propagating*. Roughly speaking, *GSF-local* means that the graph class omits a family of *generalised subgraphs* (i. e. subgraphs with constraints on how the subgraphs interact with the rest of the graph), and *non-propagating* means that graphs in which a forbidden generalised subgraph is unlikely to be detected by sampling vertices are actually close to having the property in terms of edge modifications. In other words, no 'chain reactions' occur, where repairing one edge will produce new unwanted configurations that again need repairing, etc. Goldreich and Ron asked whether 'non-propagating' is necessary. We showed that this is the case. Our proof is based on relating first-order definable properties to GSF-local properties, via a notion that we call neighbourhood profiles, which captures first-order definability.

Appendix A. Deferred Proofs from Section 7.3.1.

Claim A.1. *Every structure* $A \in \bigcup_{1\leq k\leq m} \mathcal{P}_{\rho_k} \setminus \{A_\emptyset\}$ *satisfies* φ_{tree}.

Proof. Let $A \in \bigcup_{1\leq k\leq m} \mathcal{P}_{\rho_k} \setminus \{A_\emptyset\}$. Then there is $k \in \{1,\dots,m\}$ such that $A \in \mathcal{P}_{\rho_k}$.

By definition, $\varphi_{\text{tree}} := \exists^{\leq 1} x\varphi_{\text{root}}(x) \wedge \varphi \wedge \forall x(\psi(x) \vee \chi(x))$, where

$$\varphi := \forall x\Big(\big(\varphi_{\text{root}}(x) \wedge R(x,x)\big) \vee \big(\exists^{=1} yF(y,x) \wedge \neg\exists yR(x,y) \wedge \neg\exists yR(y,x)\big)\Big),$$

$$\psi(x) := \neg\exists yF(x,y) \wedge \bigwedge_{k\in([D]^2)^2} L_k(x,x) \wedge \forall y\Big(y \neq x \rightarrow \bigwedge_{k\in([D]^2)^2} \neg L_k(x,y) \wedge \bigwedge_{k\in([D]^2)^2} \neg L_k(y,x)\Big)$$

and

$$\chi(x) := \neg\exists y \bigvee_{k\in([D]^2)^2} \big(L_k(x,y) \vee L_k(y,x)\big) \wedge \bigwedge_{k\in([D]^2)^2} \exists y_k\Big(x \neq y_k \wedge F_k(x,y_k) \wedge \big(\bigwedge_{k'\in([D]^2)^2, k'\neq k} \neg F_{k'}(x,y_k)\big) \wedge \forall y(y \neq y_k \rightarrow \neg F_k(x,y))\Big).$$

Thus, it is sufficient to prove that $A \models \exists^{\leq 1} x\varphi_{\text{root}}(x)$, $A \models \varphi$ and $A \models \forall x(\psi(x)\vee\chi(x))$.

To prove $A \models \exists^{\leq 1} x\varphi_{\text{root}}(x)$ we note that by construction of ρ_k we have $A \not\models \varphi_{\text{root}}(a)$ for any $a \in U(A)$ for which $(\mathcal{N}_2^A(a), a) \notin \tau_{d,2,\sigma}^k$. Since ρ_k restricts the number of occurrences of elements of neighbourhood type $\tau_{d,2,\sigma}^k$ to at most one, this proves that there is at most one $a \in U(A)$ with $A \models \varphi_{\text{tree}}(a)$ and hence $A \models \exists^{\leq 1} x\varphi_{\text{root}}(x)$.

To prove $A \models \varphi$, let $a \in U(A)$ be an arbitrary element. Since $A \in \mathcal{P}_{\rho_k}$, there is an $i \in I_k$ such that $(\mathcal{N}_2^A(a), a) \in \tau_{d,2,\sigma}^i$. But then by definition, there exist $\tilde{A} \models \varphi_{\textcircled{z}}$ and $\tilde{a} \in U(\tilde{A})$ such that $(\mathcal{N}_2^A(a), a) \cong (\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Assume f is an isomorphism from $(\mathcal{N}_2^A(a), a)$ to $(\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. First consider the case that $A \models \varphi_{\text{root}}(a) := \forall y\neg F(y,a)$. Assume there is $\tilde{b} \in U(\tilde{A})$ such that $(\tilde{b}, \tilde{a}) \in F(\tilde{A})$. Since $\tilde{b} \in N_2^{\tilde{A}}(\tilde{a})$, there must be an element $b \in N_2^A(a)$ such that $f(b) = \tilde{b}$. Since f is an isomorphism mapping a to $\tilde{a}$, this implies $(b,a) \in F(A)$, which contradicts $A \models \varphi_{\text{root}}(a)$. Hence $\tilde{A} \models \varphi_{\text{root}}(\tilde{a})$. Since $\tilde{A} \models \varphi_{\text{tree}}$, it holds that $\tilde{A} \models \varphi$, which means that $(\tilde{a}, \tilde{a}) \in R(\tilde{A})$. But since f is an isomorphism mapping a onto $\tilde{a}$, this implies $(a,a) \in R(A)$. Now consider the case that $A \not\models \varphi_{\text{root}}(a)$. Then there is $b \in U(A)$ with $(b,a) \in F(A)$. Since f is an isomorphism, this implies $(f(b), \tilde{a}) \in F(\tilde{A})$. Hence $\tilde{A} \models \exists^{=1} yF(y,\tilde{a}) \wedge \neg\exists yR(\tilde{a},y) \wedge \neg\exists yR(y,\tilde{a})$, as $\tilde{A} \models \varphi$. Now assume that there is $b' \neq b$ such that $(b',a) \in F(A)$. Then $f(b) \neq f(b')$ and $(f(b), \tilde{a}), (f(b'), \tilde{a}) \in F(\tilde{A})$. Since this contradicts $\tilde{A} \models \exists^{=1} yF(y,\tilde{a})$ we have $A \models \exists^{=1} yF(y,a)$. Furthermore, assume that there is $b' \in U(A)$ such that either $(a,b') \in R(A)$ or $(b',a) \in R(A)$. Then either $(\tilde{a}, f(b')) \in R(\tilde{A'})$ or $(f(b'), \tilde{a}) \in R(\tilde{A})$, which contradicts $\tilde{A} \models \neg\exists R(\tilde{a},y) \wedge \neg\exists yR(y,\tilde{a})$. Therefore $A \models \neg\exists yR(a,y) \wedge \neg\exists yR(y,a)$ which completes the proof of $A \models \varphi$.

We prove $A \models \forall x(\psi(x) \vee \chi(x))$ by considering the two cases $A \models \neg\exists y F(a,y)$ and $A \models \exists y F(a,y)$ for each element $a \in U(A)$. For this, let $a \in U(A)$ be any element. By the construction of ρ_k there is $\tilde{A} \models \varphi_{\textcircled{z}}$ and $\tilde{a} \in U(\tilde{A})$ such that $(\mathcal{N}_2^A(a), a) \cong (\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Let f be an isomorphism from $(\mathcal{N}_2^A(a), a)$ to $(\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. First consider the case that $A \models \neg\exists y F(a,y)$. If there was $\tilde{b} \in U(\tilde{A})$ with $(\tilde{a}, \tilde{b}) \in F(\tilde{A})$ then $(a, f^{-1}(\tilde{b})) \in F(A)$ contradicting our assumption. Hence $\tilde{A} \models \neg\exists y F(\tilde{a}, y)$ which implies that $\tilde{A} \not\models \chi(\tilde{a})$. But since $\tilde{A} \models \varphi_{\textcircled{z}}$, it holds that $\tilde{A} \models \forall x(\psi(x) \vee \chi(x))$, which implies that $\tilde{A} \models \psi(\tilde{a})$. Hence $(\tilde{a}, \tilde{a}) \in L_k(\tilde{A})$ for every $k \in ([D]^2)^2$. Since f is an isomorphism and $f(a) = \tilde{a}$, it holds that $(a,a) \in L_k(A)$ for every $k \in ([D]^2)^2$, and hence $A \models \bigwedge_{k \in ([D]^2)^2} L_k(a,a)$. Furthermore, assume that there is $b \in U(A)$, $b \neq a$ and $k \in ([D]^2)^2$ such that either $(a,b) \in L_k(A)$ or $(b,a) \in L_k(A)$. Since f is an isomorphism this implies that either $(\tilde{a}, f(b)) \in L_k(\tilde{A})$ or $(f(b), \tilde{a}) \in L_k(\tilde{A})$ which contradicts $\tilde{A} \models \chi(\tilde{a})$. Hence $A \models \forall y\Big(y \neq a \to \bigwedge_{k \in ([D]^2)^2} \neg L_k(a,y) \wedge \bigwedge_{k \in ([D]^2)^2} \neg L_k(y,a)\Big)$ proving that $A \models \psi(a)$.

Now consider the case that there is an element $b \in U(A)$ such that $(a,b) \in F(A)$. Since this implies that $(\tilde{a}, f(b)) \in F(\tilde{A})$, we get that $\tilde{A} \not\models \psi(\tilde{a})$, and hence $\tilde{A} \models \chi(\tilde{a})$. Now assume that there is $b \in U(A)$ and $k \in ([D]^2)^2$ such that either $(a,b) \in L_k(A)$ or $(b,a) \in L_k(A)$. But then either $(\tilde{a}, f(b)) \in L_k(\tilde{A})$ or $(f(b), \tilde{a}) \in L_k(\tilde{A})$, which contradicts $\tilde{A} \models \chi(\tilde{a})$. Hence $A \models \neg\exists y \bigvee_{k \in ([D]^2)^2} \big(L_k(a,y) \vee L_k(y,a)\big)$. For each $k \in ([D]^2)^2$, let $\tilde{b}_k \in U(\tilde{A})$ be an element such that $\tilde{A} \models \tilde{a} \neq \tilde{b}_k \wedge F_k(\tilde{a}, \tilde{b}_k) \wedge (\bigwedge_{k' \in ([D]^2)^2, k' \neq k} \neg F_{k'}(\tilde{a}, \tilde{b}_k)) \wedge \forall y(y \neq \tilde{b}_k \to \neg F_k(\tilde{a}, y))$. Since f is an isomorphism, this implies that $a \neq b_k := f^{-1}(\tilde{b}_k)$, $(a, b_k) \in F_k(A)$ and $(a, b_k) \notin F_{k'}(A)$, for each $k' \in ([D]^2)^2, k' \neq k$. Furthermore, assume there is $b \in U(A)$, $b \neq b_k$ such that $(a,b) \in F_k(A)$. Since f is an isomorphism, this implies $f(b) \neq f(b_k) = \tilde{b}_k$ and $(\tilde{a}, \tilde{b}) \in F_k(\tilde{A})$, which contradicts $\tilde{A} \models \forall y(y \neq \tilde{b}_k \to \neg F_k(\tilde{a}, y))$. Hence $A \models \forall y(y \neq b_k \to \neg F_k(a,y))$ and therefore concluding that $A \models \chi(a)$. This proves that in either case $A \models \psi(a) \vee \chi(a)$ and therefore $A \models \forall x(\psi(x) \vee \chi(x))$. ∎

Claim A.2. *Every structure $A \in \bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k} \setminus \{A_\emptyset\}$ satisfies $\varphi_{\text{rotationMap}}$.*

Proof. Let $A \in \bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k} \setminus \{A_\emptyset\}$. Then there is a $k \in \{1, \ldots, m\}$ such that $A \in \mathcal{P}_{\rho_k}$.

By definition, $\varphi_{\text{rotationMap}} = \varphi \wedge \psi$, where

$$\varphi := \forall x \forall y \Big(\bigwedge_{i,j \in [D]^2} (E_{i,j}(x,y) \to E_{j,i}(y,x)) \Big) \text{ and}$$

$$\psi := \forall x \Big(\bigwedge_{i \in [D]^2} \Big(\bigvee_{j \in [D]^2} \big(\exists^{=1} y E_{i,j}(x,y) \wedge \bigwedge_{\substack{j' \in [D]^2 \\ j' \neq j}} \neg\exists y E_{i,j'}(x,y) \big) \Big) \Big).$$

Thus, it is sufficient to prove that $A \models \varphi$ and $A \models \psi$.

To prove $A \models \varphi$, assume towards a contradiction that there are $a, b \in U(A)$ such that for some pair $i,j \in [D]^2$, we have that $(a,b) \in E_{i,j}(A)$, but $(b,a) \notin E_{j,i}(A)$. By construction of $\mathcal{P}_{\rho_k}$, there is a structure $\tilde{A} \models \varphi_{\textcircled{z}}$ and $\tilde{a} \in U(\tilde{A})$ such that $(\mathcal{N}_2^A(a), a) \cong (\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Assume f is an isomorphism from $(\mathcal{N}_2^A(a), a)$ to $(\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Note that $f(b)$ is defined since b is in the 2-neighbourhood of a. Furthermore since f is an isomorphism, $(a,b) \in E_{i,j}(A)$ implies $(\tilde{a}, f(b)) \in E_{i,j}(\tilde{A})$, and $(b,a) \notin E_{j,i}(A)$ implies $(f(b), \tilde{a}) \notin E_{j,i}(\tilde{A})$. Hence $\tilde{A} \not\models \varphi$, which contradicts $\tilde{A} \models \varphi_{\text{rotationMap}}$.

To prove $A \models \psi$, assume towards a contradiction that there is an $a \in U(A)$ and $i \in [D]^2$ such that $A \not\models \exists^{=1} y E_{i,j}(a,y) \wedge \bigwedge_{\substack{j' \in [D]^2 \\ j' \neq j}} \neg \exists y E_{i,j'}(a,y)$ for every $j \in [D]^2$. We know that there is a structure $\tilde{A} \models \varphi_{\textcircled{z}}$ and $\tilde{a} \in U(\tilde{A})$ such that $(\mathcal{N}_2^A(a), a) \cong (\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Let f be an isomorphism from $(\mathcal{N}_2^A(a), a)$ to $(\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Since $\tilde{A} \models \psi$, there must be $j \in [D]^2$ such that $\tilde{A} \models \exists^{=1} y E_{i,j}(\tilde{a}, y) \wedge \bigwedge_{\substack{j' \in [D]^2 \\ j' \neq j}} \neg \exists y E_{i,j'}(\tilde{a}, y)$. Hence there must be $\tilde{b} \in U(\tilde{A})$ such that $(\tilde{a}, \tilde{b}) \in E_{i,j}(\tilde{A})$, which implies that $(a, f^{-1}(\tilde{b})) \in E_{i,j}(A)$. Since we assumed that $A \not\models \exists^{=1} y E_{i,j}(a,y) \wedge \bigwedge_{\substack{j' \in [D]^2 \\ j' \neq j}} \neg \exists y E_{i,j'}(a,y)$, there must be either $b \neq f^{-1}(\tilde{b})$ with $(a,b) \in E_{i,j}(A)$, or there must be $j' \in [D]^2$, $j' \neq j$ and $b' \in U(A)$ such that $(a, b') \in E_{i,j'}(A)$. In the first case $(\tilde{a}, f(b)) \in E_{i,j}(\tilde{A})$, since f is an isomorphism. But then $\tilde{A} \not\models \exists^{=1} y E_{i,j}(\tilde{a}, y)$, which is a contradiction. In the second case, we get that $(\tilde{a}, f(b')) \in E_{i,j'}(\tilde{A})$. But then $\tilde{A} \not\models \bigwedge_{\substack{j' \in [D]^2 \\ j' \neq j}} \neg \exists y E_{i,j'}(\tilde{a}, y)$, which is a contradiction. Hence $A \models \varphi \wedge \psi$. ◻

Claim A.3. *Every structure* $A \in \bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k} \setminus \{A_\emptyset\}$ *satisfies* φ_{base}.

Proof. Let $A \in \bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k} \setminus \{A_\emptyset\}$. Then there is a $k \in \{1, \ldots, m\}$ such that $A \in \mathcal{P}_{\rho_k}$.

By definition, $\varphi_{\text{base}} := \forall x \big(\varphi_{\text{root}}(x) \to (\varphi(x) \wedge \psi(x))\big)$, where

$$\varphi(x) := \bigwedge_{i,j \in [D]^2} \Big(E_{i,j}(x,x) \wedge \forall y \Big(x \neq y \to \big(\neg E_{i,j}(x,y) \wedge \neg E_{i,j}(y,x)\big)\Big)\Big) \text{ and}$$

$$\psi(x) := \bigwedge_{\substack{\mathrm{ROT}_{H^2}(k,i) = (k',i') \\ k,k' \in ([D]^2)^2 \\ i,i' \in [D]^2}} \exists y \exists y' \big(F_k(x,y) \wedge F_{k'}(x,y') \wedge E_{i,i'}(y,y')\big).$$

Thus, it is sufficient to prove that $A \models \varphi(a)$ and $A \models \psi(a)$ for every $a \in U(A)$ for which $A \models \varphi_{\text{root}}(a)$. Therefore assume $a \in U(A)$ is any element such that $A \models \varphi_{\text{root}}(a)$. Because $A \in \mathcal{P}_{\rho_k}$ there is an $i \in I_k$ such that $(\mathcal{N}_2^A(a), a) \in \tau^i_{d,2,\sigma}$. Then by definition there is a structure $\tilde{A} \models \varphi_{\textcircled{z}}$ and $\tilde{a} \in U(\tilde{A})$ such that $(\mathcal{N}_2^A(a), a) \cong (\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Let f be an isomorphism from $(\mathcal{N}_2^A(a), a)$ to $(\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Assume that there is an element $\tilde{b} \in U(\tilde{A})$ such that $(\tilde{b}, \tilde{a}) \in F(\tilde{A})$. Since f is an isomorphism and $\tilde{b} \in N_2^{\tilde{A}}(\tilde{a})$ we get that $(f^{-1}(\tilde{b}), a) \in F(A)$ which contradicts that $A \models \varphi_{\text{root}}(a)$ as $\varphi_{\text{root}}(x) := \forall y \neg F(y,x)$. Hence there is no element $\tilde{b} \in U(\tilde{A})$ such that $(\tilde{b}, \tilde{a}) \in F(\tilde{A})$ which implies that $\tilde{A} \models \varphi_{\text{root}}(\tilde{a})$. But since $\tilde{A} \models \varphi_{\textcircled{z}}$ we have that $\tilde{A} \models \varphi_{\text{base}}$ and hence $\tilde{A} \models \varphi(\tilde{a})$ and $\tilde{A} \models \psi(\tilde{a})$.

To prove $A \models \varphi(a)$ first observe that $(a,a) \in E_{i,j}(A)$ for every $i,j \in [D]^2$ since $\tilde{A} \models \varphi(\tilde{a})$ implies that $(\tilde{a}, \tilde{a}) \in E_{i,j}(\tilde{A})$ for every $i,j \in [D]^2$ and f is an isomorphism mapping a onto $\tilde{a}$. Assume that there is an element $b \in U(A)$, $b \neq a$ and indices $i,j \in [D]^2$ such that either $(a,b) \in E_{i,j}(A)$ or $(b,a) \in E_{i,j}(A)$. Since $b \in N_2^A(a)$ and f is an isomorphism we get that $f(b) \neq f(a) = \tilde{a}$ and either $(\tilde{a}, f(b)) \in E_{i,j}(\tilde{A})$ or $(f(b), \tilde{a}) \in E_{i,j}(\tilde{A})$. But this contradicts $\tilde{A} \models \varphi(\tilde{a})$ and hence $A \models \varphi(a)$.

We now prove that $A \models \psi(a)$. Let $k, k' \in ([D]^2)^2$ and $i, i' \in [D]^2$ such that $\mathrm{ROT}_{H^2}(k,i) = (k',i')$. Since $\tilde{A} \models \psi(\tilde{a})$ there must be elements $\tilde{b}, \tilde{b}' \in U(\tilde{A})$ such that $(\tilde{a}, \tilde{b}) \in F_k(\tilde{A})$, $(\tilde{a}, \tilde{b}') \in F_{k'}(\tilde{A})$ and $(\tilde{b}, \tilde{b}') \in E_{i,i'}(\tilde{A})$. But since $\tilde{b}, \tilde{b}' \in N_2^{\tilde{A}}(\tilde{a})$ we get that $f^{-1}(\tilde{b})$ and $f^{-1}(\tilde{b}')$ are defined and since f is an isomorphism we get that

$(a, f^{-1}(\tilde{b})) \in F_k(A)$, $(a, f^{-1}(\tilde{b}')) \in F_{k'}(A)$ and $(f^{-1}(\tilde{b}), f^{-1}(\tilde{b}')) \in E_{i,i'}(A)$. Hence $A \models \exists y \exists y' \big(F_k(a,y) \wedge F_{k'}(a,y') \wedge E_{i,i'}(y,y')$ for any $k,k' \in ([D]^2)^2$ and $i,i' \in [D]^2$ such that $\mathrm{ROT}_{H^2}(k,i) = (k',i')$ which implies that $A \models \psi(a)$. ∎

Acknowledgments. We thank Sebastian Ordyniak for inspiring discussions and the anonymous reviewers for their detailed comments. The second author thanks Michał Pilipczuk and Pierre Simon for inspiring discussions at the docCourse on Sparsity in Prague 2018.

REFERENCES

[1] I. Adler and F. Harwath, *Property testing for bounded degree databases*, in 35th Symposium on Theoretical Aspects of Computer Science, STACS 2018, February 28 to March 3, 2018, Caen, France, 2018, pp. 6:1–6:14.

[2] I. Adler, N. Köhler, and P. Peng, *GSF-locality is not sufficient for proximity-oblivious testing*, in 36th Computational Complexity Conference, CCC 2021, July 20-23, 2021, Toronto, Ontario, Canada (Virtual Conference), V. Kabanets, ed., vol. 200 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021, pp. 34:1–34:27, https://doi.org/10.4230/LIPIcs.CCC.2021.34, https://doi.org/10.4230/LIPIcs.CCC.2021.34.

[3] I. Adler, N. Köhler, and P. Peng, *On testability of first-order properties in bounded-degree graphs*, in Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021, D. Marx, ed., SIAM, 2021, pp. 1578–1597, https://doi.org/10.1137/1.9781611976465.96, https://doi.org/10.1137/1.9781611976465.96.

[4] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy, *Efficient testing of large graphs*, Combinatorica, 20 (2000), pp. 451–476. Preliminary version in FOCS'99.

[5] N. Alon, E. Fischer, I. Newman, and A. Shapira, *A combinatorial characterization of the testable graph properties: it's all about regularity*, SIAM Journal on Computing, 39 (2009), pp. 143–167.

[6] N. Alon and V. D. Milman, *λ1, isoperimetric inequalities for graphs, and superconcentrators*, Journal of Combinatorial Theory, Series B, 38 (1985), pp. 73–88.

[7] I. Benjamini, O. Schramm, and A. Shapira, *Every minor-closed property of sparse graphs is testable*, Advances in mathematics, 223 (2010), pp. 2200–2218.

[8] M. Blum, M. Luby, and R. Rubinfeld, *Self-testing/correcting with applications to numerical problems*, Journal of computer and system sciences, 47 (1993), pp. 549–595.

[9] B. Bollig and D. Kuske, *An optimal construction of hanf sentences*, Journal of Applied Logic, 10 (2012), pp. 179–186, https://doi.org/10.1016/j.jal.2012.01.002, http://dx.doi.org/10.1016/j.jal.2012.01.002.

[10] H. Chen and Y. Yoshida, *Testability of homomorphism inadmissibility: Property testing meets database theory*, in Proceedings of the 38th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, 2019, pp. 365–382.

[11] A. Czumaj, P. Peng, and C. Sohler, *Relating two property testing models for bounded degree directed graphs*, in Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016, 2016, pp. 1033–1045.

[12] R. Diestel, *Graph Theory, 4th Edition*, vol. 173 of Graduate texts in mathematics, Springer, 2012.

[13] J. Dodziuk, *Difference equations, isoperimetric inequality and transience of certain random walks*, Transactions of the American Mathematical Society, 284 (1984), pp. 787–794.

[14] H. Ebbinghaus and J. Flum, *Finite model theory*, Perspectives in Mathematical Logic, Springer, 1995.

[15] H. Fichtenberger, P. Peng, and C. Sohler, *Every testable (infinite) property of bounded-degree graphs contains an infinite hyperfinite subproperty*, in Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics, 2019, pp. 714–726.

[16] S. Forster, D. Nanongkai, T. Saranurak, L. Yang, and S. Yingchareonthawornchai, *Computing and testing small connectivity in near-linear time and queries via fast local cut algorithms*, SODA, (2020).

[17] H. Gaifman, *On local and non-local properties*, 1982.

[18] O. Goldreich, *Introduction to property testing*, Cambridge University Press, 2017.

[19] O. Goldreich, S. Goldwasser, and D. Ron, *Property testing and its connection to learning*

and approximation, Journal of the ACM (JACM), 45 (1998), pp. 653–750.

[20] O. Goldreich and T. Kaufman, *Proximity oblivious testing and the role of invariances*, in Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation, Springer, 2011, pp. 173–190.

[21] O. Goldreich and D. Ron, *Property testing in bounded degree graphs*, Algorithmica, 32 (2002), pp. 302–343, https://doi.org/10.1007/s00453-001-0078-7, http://dx.doi.org/10.1007/s00453-001-0078-7.

[22] O. Goldreich and D. Ron, *On proximity-oblivious testing*, SIAM Journal on Computing, 40 (2011), pp. 534–566.

[23] O. Goldreich and I. Shinkar, *Two-sided error proximity oblivious testing*, Random Structures & Algorithms, 48 (2016), pp. 341–383.

[24] W. Hanf, *The Theory of Models*, North Holland, 1965, ch. Model-theoretic methods in the study of elementary logic, pp. 132–145.

[25] A. Hassidim, J. A. Kelner, H. N. Nguyen, and K. Onak, *Local graph partitions for approximation and testing*, in 2009 50th Annual IEEE Symposium on Foundations of Computer Science, IEEE, 2009, pp. 22–31.

[26] S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications*, BULL. AMER. MATH. SOC., 43 (2006), pp. 439–561.

[27] H. Ito, A. Khoury, and I. Newman, *On the characterization of* 1*-sided error strongly-testable graph properties for bounded-degree graphs*, (to appear) Journal of Computational Complexity. arXiv:1909.09984, (2019).

[28] K.-i. Kawarabayashi and Y. Yoshida, *Testing subdivision-freeness: property testing meets structural graph theory*, in Proceedings of the forty-fifth annual ACM symposium on Theory of computing, ACM, 2013, pp. 437–446.

[29] A. Kumar, C. Seshadhri, and A. Stolman, *Random walks and forbidden minors ii: a poly (d ε-1)-query tester for minor-closed properties of bounded degree graphs*, in Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, 2019, pp. 559–567.

[30] L. Lovász, *Large Networks and Graph Limits*, vol. 60 of Colloquium Publications, American Mathematical Society, 2012.

[31] I. Newman and C. Sohler, *Every property of hyperfinite graphs is testable*, SIAM Journal on Computing, 42 (2013), pp. 1095–1112.

[32] L. S. R. Fagin and M. Vardi, *On monadic np vs. monadic co-np*, Information and Computation, 120 (1995), pp. 78–92.

[33] O. Reingold, S. Vadhan, and A. Wigderson, *Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors*, in Annals of Mathematics, 2000, pp. 157–187.

[34] R. Rubinfeld and M. Sudan, *Robust characterizations of polynomials with applications to program testing*, SIAM Journal on Computing, 25 (1996), pp. 252–271.

[35] Y. Yoshida and H. Ito, *Property testing on k-vertex-connectivity of graphs*, Algorithmica, 62 (2012), pp. 701–712.

List of Notation

Notation	**Description**
$\mathbb{N}$	the set of natural numbers including 0
$\mathbb{N}_{>0}$	the set of all positive natural numbers
$[n]$	the set $\{0, \ldots, n-1\}$
$\sqcup$	disjoint union
$\triangle$	symmetric difference
$\tau_r(x)$	the neighbourhood of fixed radius r around x, up to isomorphism
$\mathcal{O}(\cdot), o(\cdot)$	asymtotic upper-bounds
$\textcircled{z}$	zig-zag product
$h(G)$	expansion ratio of G
$\langle S, T \rangle_G$	edges crossing S and T in G
σ	signature
σ_{graph}	signature with one binary relation symbol E
$\deg_G(v)$	degree of vertex v in G
$\text{dist}_A(v, w)$	distance between two vertices v and w in A
$\text{dist}(A, \mathcal{P})$	distance between structure A and property $\mathcal{P}$
ROT_G	rotation map of G
$M_{u,v}$	the (u, v) entry of the normalised adjacency matrix M of G
$A[S]$	substracture of A induced by S
$N_r^A(a)$	the r-neighborhood of a in structure A
$\text{ar}(R)$	arity of relation R
$\mathcal{C}_d$	class of graphs of bounded degree d
$\mathcal{C}_{\sigma,d}$	class of σ-structures of bounded degree d
$\mathcal{P}_\varphi$	property defined by formula φ
Σ_i, Π_i, Δ_i	prefix classes with $i-1$ quantifier alterations
$\models$	is a model of
$\equiv$	equivalence of FO-formulas
$\equiv_d$	equivalence of FO-formulas on structures of bounded degree d
$\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$	logical negation, conjunction, disjunction, implication and biimplication
$\exists$, $\forall$	existential and universal quantifier
$\text{ans}_{\mathcal{A}}(q)$	answer to query q to structure $\mathcal{A}$
$E_{i,j}$, F_k, R, L_k	binary relation symbols for $i, j \in [D]^2$ and $k \in ([D]^2)^2$
$\underline{G}(A)$	underlying graph of a model A of $\varphi_{\textcircled{z}}$
$\varphi_{\textcircled{z}}$	the formula definded in Section 3 whose underlying graphs are expanders
$\mathfrak{M}$	a set of models of some sentence in Σ_2
$\mathfrak{H}$	a maximal set of pairwise non-isomorphic structures in $\mathcal{C}_{\sigma,d}$
$\text{pos}^i(\overline{x}, \overline{y}), \text{neg}^i(\overline{x}, \overline{y})$	a conjunction of atomic (resp. negated) formulas containing both variables from tuples $\overline{x}$ and $\overline{y}$

Notation	**Description**
$\rho_{A,r}$	the r-type distribution of A
$\delta_{\odot}^{r}(A, B)$	the sampling distance of depth r between two σ-structures A and B
HNF	Hanf normal form
POT	proximity oblivious tester
GSF	generalised subgraph freeness