UNIVERSITY OF LEEDS

This is a repository copy of *On Testability of First-Order Properties in Bounded-Degree Graphs*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/166217/

Version: Accepted Version

Proceedings Paper:
Adler, I orcid.org/0000-0002-9667-9841, Köhler, N and Peng, P (2021) On Testability of First-Order Properties in Bounded-Degree Graphs. In: SODA '21: Proceedings of the Thirty-Second Annual ACM-SIAM Symposium on Discrete Algorithms. SODA '21: ACM-SIAM Symposium on Discrete Algorithms, 10-13 Jan 2021, Online. ACM , pp. 1578-1597. ISBN 9781611976465

https://doi.org/10.1137/1.9781611976465.96

© 2021 by SIAM. This is an author produced version of a conference paper published in SODA '21: Proceedings of the Thirty-Second Annual ACM-SIAM Symposium on Discrete Algorithms. Uploaded in accordance with the publisher's self-archiving policy.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

On Testability of First-Order Properties in Bounded-Degree Graphs*

Isolde Adler† Noleen Köhler‡ 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 a first-order 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 prove testability of some first-order properties that speak about isomorphism types of neighbourhoods, including testability of 1-neighbourhood-freeness, and r-neighbourhood-freeness under a mild assumption on the degrees.

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 P (e. g. triangle-freeness), the goal of a property testing algorithm, called a *property tester*, is to distinguish if a graph satisfies P or is *far* from satisfying P, where the definition of *far* depends on the model. The general notion of property testing was first proposed by Rubinfeld and Sudan [RS96], with the motivation for the study of program checking. Goldreich, Goldwasser and Ron [GGR98] then introduced the property testing for combinatorial objects and graphs. They formalized the *dense graph model* for testing graph properties, in which the algorithm 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 P from the case that one has to modify (delete or insert) more than εn^2 edges to make it satisfy P, for any specified proximity parameter $\varepsilon \in (0, 1]$. A property 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 [GGR98], much effort has been made on the testability of graph properties in this model, culminating in the work by Alon et al. [AFNS09], 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 [GR02] 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* [GR02], 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 P, if one needs to modify more than εdn edges to make it satisfy P. The goal now becomes to distinguish, with probability at least $2/3$, if G satisfies a property P or is ε-far from satisfying P, for any specified proximity parameter $\varepsilon \in (0, 1]$. Again, a property 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 [GR02], minor-freeness [BSS10, HKNO09, KSS19], hyperfinite properties [NS13], k-vertex connectivity [YI12, FNS+20], and subdivision-freeness [KY13].

In this paper, 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

*A full version of the paper is available at: `https://arxiv.org/abs/2008.05800`.

†University of Leeds, UK. Email: `I.M.Adler@leeds.ac.uk`.

‡University of Leeds, UK. Email: `scnk@leeds.ac.uk`.

§University of Sheffield, UK. Email: `p.peng@sheffield.ac.uk`.

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 first-order logic. 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 [AH18], 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 for each edge colour class, querying for both in- and out-neighbours via edges in that class. 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. [AFKS00]. 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 similar to testing subgraph freeness [GR02]. 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 5.1). On the other hand, there is a first-order property in Π_2, that is not testable in the bounded-degree model (Theorem 4.4).

The theorems that we refer to in the above statement are for relational structures, while we also give a lower bound on graphs (Theorem 4.5), 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 [AFKS00], 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 [Han65, Gai82, RFV95]. 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 this in more details in the following. On one hand, Hanf's Theorem [Han65] gives insight into first-order logic on graphs of bounded degree and implies a strong normal form, called *Hanf Normal Form* (HNF) in [BK12], 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 [Han65, RFV95, EF95]. 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 [CPS16] and [GR11]). 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 ev-

[1] Furthermore, previously, typical FO properties are 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. [Gol17]).

ery 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 [AH18]. However, our lower bound gives a negative answer to this question.

Motivated by our above results, we further study testability of graph properties described through Hanf sentences or negated Hanf sentences, which are first-order properties that speak about isomorphism types of neighbourhoods. Given a bounded degree graph, an r-*ball* around a vertex x is the neighbourhood of radius r around x in the graph. We call the isomorphism types of r-balls r-*types*. We consider two basic such properties, called τ-*neighbourhood regularity* and τ-*neighbourhood-freeness*, that correspond to "all vertices have r-type τ" and "no vertex has r-type τ", respectively. (Neighbourhood-regularity can be seen as a generalisation of degree-regularity, which is known to be testable [Gol17].) As we show in Lemma 6.1, there exist 1-types τ, τ' such that neither τ-neighbourhood-freeness nor τ'-neighbourhood regularity can be defined by a formula in Σ_2. Thus, our previous tester for Σ_2 cannot be applied to these properties. We give constant-query testers for them under certain conditions on τ (Theorem 6.1, Theorem 6.2, Theorem 6.3). Both τ-neighbourhood-freeness and τ-neighbourhood regularity can be defined by formulas in Π_2 for any neighbourhood type τ. Thus, our results imply that there are properties defined by formulas in $\Pi_2 \setminus \Sigma_2$ which are testable.

Our techniques To show that every property P defined by a formula φ in Σ_2 (i.e. of the form $\exists^*\forall^*$) is testable, we show that P is equivalent to the union of properties P_i, each of which is 'indistinguishable' from a property Q_i that is defined by a formula of form $\forall^*$. Here the indistinguishability means we can transform any structure satisfying P_i, into a structure satisfying 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 P to testing properties defined by $\forall^*$ formulas. Then the testability of P follows, as any property of the form $\forall^*$ is testable and testable properties are closed under union [Gol17]. 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 4.1). 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 $\mathcal{A}$ satisfying the property φ and a relational structure $\mathcal{B}$ that is ε-far from having the property must be relatively large (see [AH18] which in turn is built upon the so-called "canonical testers" for bounded-degree graphs in [CPS16, GR11]). We then exploit a result of Alon (see Proposition 19.10 in [Lov12]), that the neighbourhood distribution of an arbitrarily large relational structure $\mathcal{A}$ can be approximated by the neighbourhood distribution of a structure $\mathcal{H}$ of small constant size. Thus, for any $\mathcal{A}$ in φ, by taking the union of "many" disjoint copies of the "small" structure $\mathcal{H}$, we obtain another structure $\mathcal{B}$ such that the local neighbourhood distributions of $\mathcal{A}$ and $\mathcal{B}$ have small distance. If the underlying undirected graphs of the structures in φ are expander graphs, it immediately follows that $\mathcal{B}$ 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 [FPS19].

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 [RVW02]. 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 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

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

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. Finally, we extend this construction to simple undirected graphs, by using carefully designed gadgets to encode the different edge-colours and maintain degree regularity.

To give our testers for τ-neighbourhood regularity and τ-neighbourhood-freeness, we show that if a graph G is ε-far from having the property, it contains a linear fraction of constant-size neighbourhoods certifying that G does not satisfy the property. Such a statement may be intuitively true, but it is tricky to prove. Assume we want to test for τ-freeness, for some fixed r-neighbourhood type τ, and assume a graph G has one vertex x with forbidden neighbourhood of type τ. Changing the r-neighbourhood of x by edge modifications, in order to remove τ, might introduce new forbidden neighbourhoods around vertices close to x, triggering a 'chain reaction' of necessary modifications. This means that a graph might be ϵ-far from being τ-free, but we do not see it by sampling constantly many neighbourhoods in the graph. Such a subtle difficulty has already been observed for testing degree-regularity (see Claim 8.5.1 in [Gol17]). We show that under appropriate assumptions, such a 'chain reaction' can be bypassed by carefully fixing the neighbourhood of x without changing the neighbourhood type of the vertices surrounding x. Though fairly simple, it provides non-trivial analysis, handling the subtle difficulty of relating local distance to global distance without triggering a 'chain reaction'.

Other related work Besides the aforementioned works on testing properties with constant query complexity in the bounded-degree graph model, Goldreich and Ron [GR11] 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. [FPS19] showed that every testable property is either finite or contains an infinite hyperfinite subproperty. Ito et al. [IKN19] 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 are also properties (e.g. bipartiteness, expansion, k-clusterability) that require $\Omega(\sqrt{n})$ queries, and properties (e.g. 3-colorability) that require $\Omega(n)$ queries. We refer the reader to Goldreich's recent book [Gol17].

Property testing on relational structures was recently motivated by the application in databases. Besides the aforementioned work [AH18], Chen and Yoshida [CY19] studied the testability of relational database queries for each relational structure in the framework of property testing.

Further discussions and open problems 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.

We believe it is natural to study the problem of testing properties of neighbourhood types. Firstly, our previous results can be seen as an indication that quantifier prefix classes are perhaps less suitable when searching for a dividing line between testable and non-testable first-order properties in the bounded-degree model. Since subgraph-freeness and subgraph containment are testable, Hanf's normal form suggests studying testability of Hanf sentences and their negations, i. e. neighborhood properties, as a next step. Secondly, studying such properties helps us gain more insights on which properties that are defined by distributions of neighbourhood types are testable, which is crucial to solve one of the most important open questions in this area, namely to characterise the combinatorial structure of testable properties in the bounded-degree model.

Furthermore, we remark that our testers for neighbourhood properties have *one-sided error*, i. e. the testers always accept the graphs that satisfy the property. We note that in contrast to subgraph-freeness and induced subgraph-freeness, the properties τ-neighbourhood regularity and τ-neighbourhood-freeness are neither *monotone* nor *hereditary*, which are properties that are closed under edge deletion and closed under vertex deletion, respectively. As we mentioned before, Ito et al. [IKN19] recently characterised one-sided error (constant-query) testable monotone and hereditary graph properties in the bounded-degree (directed and undirected) graph model. In order to give a full characterisation of one-sided error testable properties in the bounded-degree graph model, it is important to take a step beyond monotone and hereditary graph properties.

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 P_φ defined by φ on bounded-degree structures is not constant-query testable. We also provide a Π_2-property of simple, undirected graphs that is non-testable. In Section 5, we show that all Σ_2 properties are testable. In Section 6 we give positive results for some first-order properties speaking about isomorphism types of neighbourhoods.

2 Preliminaries

We refer the reader to the full version of the paper for basics on graphs, relational structures and first-order logic.

2.1 The bounded-degree relational structure model Let $\sigma = \{R_1, \ldots, R_\ell\}$ be a relational signature and let $\mathcal{A}$ be a σ-structure with universe A. The *degree* of an element $a \in A$ denoted by $\deg_{\mathcal{A}}(a)$ is defined to be the number of tuples in $\mathcal{A}$ containing a. We define the *degree* of $\mathcal{A}$ denoted by $\deg(\mathcal{A})$ to be the maximum degree of its elements. For any $d \in \mathbb{N}$ we let C_d be the class of all σ-structures of bounded degree d. Let us remark that $\deg(\mathcal{A})$ and the degree of the Gaifman graph of $\mathcal{A}$ only differ by at most a constant factor (cf. e. g. [DG07]), so the definitions are equivalent in the sense that the same classes have bounded degree. A *property* on any class of structures C is a subset $P \subseteq C$ of structures that is closed under isomorphism. We say that a structure $\mathcal{A} \in C$ has property P if $\mathcal{A} \in P$. On C_d, every FO-sentence φ defines a property $P_\varphi \subseteq C_d$, where $P_\varphi = \{\mathcal{A} \in C_d \mid \mathcal{A} \models \varphi\}$.

We describe the model for bounded-degree relational structures as defined in [AH18]. This extends the bounded-degree model for undirected graphs introduced in [GR02] and conforms with the bidirectional model of [CPS16].

An algorithm that processes a σ-structure $\mathcal{A} \in C_d$ does not obtain an encoding of $\mathcal{A}$ as a bit string in the usual way. Instead, we assume that the algorithm receives the number n of elements of $\mathcal{A}$, and that the elements of $\mathcal{A}$ are numbered $1, 2, \ldots, n$. In addition, the algorithm has direct access to $\mathcal{A}$ using an *oracle* which answers *neighbour queries* in $\mathcal{A}$ in constant time. That is, the oracle accepts queries of the form (i, j, k), for $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, \ell\}$ and $k \in \{1, \ldots, d\}$, to which it responds with the k-th tuple in $R_j^{\mathcal{A}}$ containing i, or with $\perp$ if i is contained in less than k tuples.

The *running time* of the algorithm is defined as usual, i. e. with respect to the size of the structure n. We assume a uniform cost model, i. e. we assume that all basic arithmetic operations including random sampling can be performed in constant time, regardless of the size of the numbers involved.

Distance. For two σ-structures $\mathcal{A}$ and $\mathcal{B}$, both of size n, $\operatorname{dist}(\mathcal{A}, \mathcal{B})$ denotes the minimum number of tuples that have to be modified (i. e. inserted or removed) in $\mathcal{A}$ and $\mathcal{B}$ to make $\mathcal{A}$ and $\mathcal{B}$ isomorphic. For $\epsilon \in (0, 1]$, we say $\mathcal{A}$ and $\mathcal{B}$, both of size n and with degree bound d, are *ϵ-close* if $\operatorname{dist}(\mathcal{A}, \mathcal{B}) \leq \epsilon dn$. If $\mathcal{A}, \mathcal{B}$ are not *ϵ-close*, then they are *ϵ-far*. Note that in particular, $\mathcal{A}$ and $\mathcal{B}$ are ϵ-far if their size differs. A σ-structure $\mathcal{A}$ is *ϵ-close* to a property P if $\mathcal{A}$ is ϵ-close to some $\mathcal{B} \in P$. Otherwise, $\mathcal{A}$ is *ϵ-far* from P.

Definition 2.1. (ϵ-tester) *Let $P \subseteq C_d$ be a property and $\epsilon \in (0, 1]$. An* ϵ-tester *for P is a probabilistic algorithm with oracle access to an input $\mathcal{A} \in C_d$ and auxiliary input $n := |A|$. The tester does the following.*

1. *If $\mathcal{A} \in P$, then the ϵ-tester accepts with probability at least $2/3$.*
2. *If $\mathcal{A}$ is ϵ-far from P, then the ϵ-tester rejects with probability at least $2/3$.*

The *query complexity* of an ϵ-tester is the maximum number of oracle queries made. A property P is *testable*, if for each $\epsilon \in (0, 1]$ and each n, there is an ϵ-tester for $P \cap \{\mathcal{A} \in C_d \mid |A| = n\}$ on inputs from $\{\mathcal{A} \in C_d \mid |A| = n\}$ with constant query complexity, i. e. the *query complexity* is independent of n. A property P is *uniformly testable*, if for each $\epsilon \in (0, 1]$ there is an ϵ-tester for P, that has *constant query complexity*. Note that 'uniformly' emphasises that this tester must work for all n.

2.2 Quantifier alternations of first-order formulas Let σ be any relational signature and C_d the set of σ-structures of bounded degree d. 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 $\varphi(y_1, \ldots, y_\ell)$ for which there is $k \in \mathbb{N}$ and a formula $\psi(x_1, \ldots, x_k, y_1, \ldots, y_\ell) \in \Pi_{i-1}$ such that $\varphi \equiv \exists x_1 \ldots \exists x_k \psi(x_1, \ldots, x_k, y_1, \ldots, y_\ell)$. Analogously, Π_i consists of all FO formulas $\varphi(y_1, \ldots, y_\ell)$ for which there is $k \in \mathbb{N}$ and a formula $\psi(x_1, \ldots, x_k, y_1, \ldots, y_\ell) \in \Sigma_{i-1}$ such that $\varphi \equiv \exists x_1 \ldots \exists x_k \psi(x_1, \ldots, x_k, y_1, \ldots, y_\ell)$.

Example. [Substructure freeness] Let $\mathcal{B}$ be a σ-structure, and let $d \in \mathbb{N}$. The property

$$P := \{\mathcal{A} \in C_d \mid \mathcal{A} \text{ does not contain } \mathcal{B} \text{ as substructure}\}$$

is in Π_1 and is uniformly testable on C_d with constant running time which can be proven similar to substructure freeness in simple graphs [GR02]. □

As discussed in the introduction, every FO property in Σ_1 or Π_1, i.e., without quantifier alternation, is testable.

2.3 Expansion and the zig-zag product In this section we recall a construction of a class of expanders introduced in [RVW02]. This construction uses undirected graphs with parallel edges and self-loops. We therefore encode a graph G as a triple (G, E, f) where V is a finite sets of vertices, E is a finite set of edges, and f is the incidence map from E to $\{x \subseteq V \mid 1 \leq |x| \leq 2\}$.

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 $u, v \in V$

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

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 for every vertex v fixes an order on all edges incident to v. We let the normalised adjacency matrix M of G be

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

Let $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq -1$ denote the eigenvalues of M. Since M is real, symmetric, contains no negative entries and all columns sum up to 1, all its eigenvalues are in the real interval $[-1, 1]$. 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 2.1. ([HLW06]) *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 $S, T \subseteq V$ 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* S and T.

DEFINITION 2.2. *For any $S \subseteq V$, 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 $m \in \mathbb{N}_{>0}$. There exists the following connection between $h(G)$ and $\lambda(G)$.

THEOREM 2.1. ([DOD84, AM85]) *Let G be a D-regular graph. Then $h(G) \geq D(1 - \lambda(G))/2$.*

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 $\{G_m\}_{m \in \mathbb{N}_{>0}}$ is a family of $D(1 - \varepsilon)/2$-expanders.

DEFINITION 2.3. *Let G be a D-regular graph on N vertices with rotation map ROT_G 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 explained by using Lemma 2.1.

LEMMA 2.2. ([RVW02]) *If G is a (N, D, λ)-graph then G^2 is a (N, D^2, λ^2)-graph.*

DEFINITION 2.4. *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 $G_1 ⓩ G_2$, is the D_2^2-regular graph on $V_1 \times I_1$ with rotation map given by $\mathrm{ROT}_{G_1 ⓩ G_2}((v,k),(i,j)) := ((w,\ell),(j',i'))$, where*

$$\begin{aligned}\mathrm{ROT}_{G_2}(k,i) = (k',i'), \quad &\mathrm{ROT}_{G_1}(v,k') = (w,\ell'), \\ &\mathrm{ROT}_{G_2}(\ell',j) = (\ell,j')\end{aligned}$$

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

The zig-zag product $G_1 ⓩ G_2$ 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 in G_1 of a vertex v will be incident to vertex i of the to-v-corresponding-copy of G_2. 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 $G_1 ⓩ G_2$ 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 2.2. ([RVW02]) *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 a $(N_1 \cdot D_1, D_2^2, g(\lambda_1, \lambda_2))$-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.

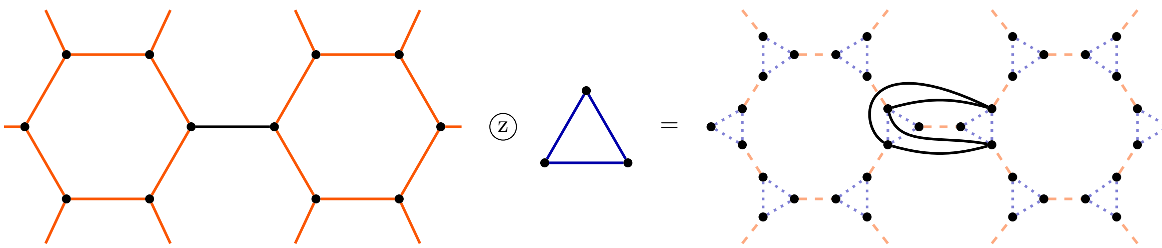

Figure 1: Zig-zag product of a 3-regular grid with a triangle

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 2.5. ([HLW06]) *Let D be a sufficiently large prime power (e.g. $D = 2^{16}$). Let H be a $(D^4, D, 1/4)$ expander (explicit constructions for H exist, cf. [RVW02].) We define $\{G_m\}_{m\in\mathbb{N}_{>0}}$ by*

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

PROPOSITION 2.1. ([HLW06]) *For every $m \in \mathbb{N}_{>0}$, the graph G_m is a $(D^{4m}, D^2, 1/2)$-graph.*

In the next section we will use the following lemma. For a proof of Lemma 2.3 see the full version of the paper.

LEMMA 2.3. *Let G be a D-regular graph, S the vertices of a connected component of G^2. Then $\lambda(G^2[S]) < 1$.*

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 2.5. 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 this construction. 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 $(v, w) \in E_{i,j}^{\mathcal{A}}$ and $(w, v) \in E_{j,i}^{\mathcal{A}}$. This allows us to encode the relationship $\mathrm{ROT}_{G_m}(v, i) = (w, j)$ 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 (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

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

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 let $E := \bigcup_{i,j\in[D]^2} E_{i,j}$ and $F := \bigcup_{k\in([D]^2)^2} F_k$. Note that we can express the relations E and F in our language, 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 to identify the root $\varphi_{\mathrm{root}}(x) := \forall y \neg F(y, x)$.

We now define a formula φ_{tree}, which expresses that the model restricted to the relation F locally looks like a D^4-ary tree. More precisely, the formula defines that the structure has exactly 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.

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

The formula $\varphi_{\mathrm{rotationMap}}$ will define the properties the relations in E need to have in order to encode rotation

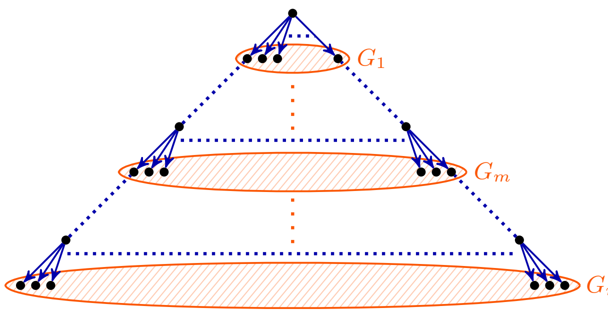

Figure 2: Schematic representation of a model of $\varphi_{ⓩ}$, where the parts in red (grey) only contain relations in E. Relations in F are blue (black). R and L are omitted.

maps of D^2-regular graphs. For this we make sure that the edge colours encode a map, i.e. for any pair x and $i \in [D]^2$ there is only one pair y and $j \in [D]^2$ such that $E_{i,j}(x,y)$ holds, and that the map is self inverse, i.e. if $E_{i,j}(x,y)$ then $E_{j,i}(y,x)$.

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

We now define a formula φ_{base} which expresses that the root r of the tree has a self-loop (r,r) in each relation $E_{i,j}$ and that the D^4 children of the root form G_1. Let H be the $(D^4, D, 1/4)$-graph from Definition 2.5. 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 $(x,y) \in F_k^{\mathcal{A}}$ for the root x. Let $\mathrm{ROT}_H : ([D]^2)^2 \times [D] \to ([D]^2)^2 \times [D]$ 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.

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

We will now define a formula $\varphi_{\text{recursion}}$ which will ensure that level m of the tree contains G_m. Recall that $G_m := G_{m-1}^2 ⓩ H$. 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.

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

We finally let $\varphi_{ⓩ} := \varphi_{\text{tree}} \wedge \varphi_{\text{rotationMap}} \wedge \varphi_{\text{base}} \wedge \varphi_{\text{recursion}}$. This concludes defining the formula. Let $d := 2D^2 + D^4 + 1$, which is chosen in such a way to allow for any element of a σ-structure in C_d to be in $2D^2$ E-relations (G_m is D^2 regular and every edge of G_m is modelled by two directed edges), 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 $\mathcal{A}$ of $\varphi_{ⓩ}$ we will associate an undirected graph $U(\mathcal{A})$ with vertex set A. For every tuple in each of the relations of $\mathcal{A}$, the graph $U(\mathcal{A})$ will have an edge. We will define $U(\mathcal{A})$ 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* $U(\mathcal{A})$ of a model $\mathcal{A}$ of $\varphi_{ⓩ}$ to be the undirected graph with vertex set A given by the rotation map $\mathrm{ROT}_{U(\mathcal{A})} : A \times I \to A \times I$ defined by

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

We can understand this rotation map as labelling the edges incident to a vertex v as follows: $(v,v) \in R^{\mathcal{A}}$ or $(w,v) \in F_k^{\mathcal{A}}$ respectively is labelled by 0, $(v,w) \in F_k^{\mathcal{A}}$ or $(v,v) \in L_k^{\mathcal{A}}$ respectively is labelled by k and $(v,w) \in E_{i,j}^{\mathcal{A}}$ is labelled by i. Note that $U(\mathcal{A})$ 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 [RVW02]. However the Gaifman graph can be obtained from the underlying graph by ignoring self-loops and multiple edges. We use $\mathcal{A} \models \varphi$ to denote that $\mathcal{A}$ is a model of an FO sentence φ and we show the following.

THEOREM 3.1. *There is an $\epsilon > 0$ such that the class $\{U(\mathcal{A}) \mid \mathcal{A} \models \varphi_{\text{Ⓩ}}\}$ is a family of ϵ-expanders.*

In the rest of this section, we give the proof of Theorem 3.1. Let $\mathcal{A}$ be a model of $\varphi_{\text{Ⓩ}}$. Let $\mathcal{A}|_F := (A, (F_k^{\mathcal{A}})_{k\in([D]^2)^2})$ be an $\{(F_k)_{k\in([D]^2)^2}\}$-structure. Recall that we denote the Gaifman graph of $\mathcal{A}|_F$ by $G(\mathcal{A}|_F)$. Let $\mathcal{A}|_E$ be the $\{(E_{i,j})_{i,j\in[D]^2}\}$-structure $(A, (E_{i,j}^{\mathcal{A}})_{i,j\in[D]^2})$. We further define the *underlying graph* $U(\mathcal{A}|_E)$ of $\mathcal{A}|_E$ as the undirected graph specified by the rotation map $\mathrm{ROT}_{U(\mathcal{A}|_E)}$ defined by $\mathrm{ROT}_{U(\mathcal{A}|_E)}(v,i) := (w,j)$ if $(v,w) \in E_{i,j}^{\mathcal{A}}$. This is well defined as $\mathcal{A} \models \varphi_{\text{rotationMap}}$. We use the substructures $G(\mathcal{A}|_F)$ and $U(\mathcal{A}|_E)$ to express the structural properties of models of $\varphi_{\text{Ⓩ}}$. More precisely we want to prove that $G(\mathcal{A}|_F)$ is a rooted complete tree and $U(\mathcal{A}|_E)$ is the disjoint union of the expanders $G_1, \ldots, G_n$ for some $n \in \mathbb{N}$ (Lemma 3.3). To prove this we use two technical lemmas (Lemma 3.1 and Lemma 3.2). Lemma 3.1 intuitively shows that the children in $G(\mathcal{A}|_F)$ of each connected part of $U(\mathcal{A}|_E)$ form the zig-zag product with H of the square of the connected part. Lemma 3.2 shows that $G(\mathcal{A}|_F)$ 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. We define isomorphism for undirected graphs with parallel edges and self-loops in the usual way, and we use $G_1 \cong G_2$ to denote that G_1 is isomorphic to G_2.

LEMMA 3.1. *Let $\mathcal{A}$ be a model of $\varphi_{\text{Ⓩ}}$ and assume S is the set of all vertices belonging to a connected component of $(U(\mathcal{A}|_E))^2$ not containing the root and let $S' := \{w \in A \mid (v,w) \in F^{\mathcal{A}}, v \in S\}$. If $S' \neq \emptyset$ then $U(\mathcal{A}|_E)[S']$ is a connected component of $U(\mathcal{A}|_E)$ and $U(\mathcal{A}|_E)[S'] \cong ((U(\mathcal{A}|_E))^2[S])\,\text{Ⓩ}\,H$.*

We use connected components of $(U(\mathcal{A}|_E))^2$, as the square of a connected component of $U(\mathcal{A}|_E)$ 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 3.1. Assume that $S' \neq \emptyset$. We first show that $U(\mathcal{A}|_E)[S'] \cong ((U(\mathcal{A}|_E))^2[S])\,\text{Ⓩ}\,H$. For this we use the following two claims.

CLAIM 3.1. *If $\mathrm{ROT}_{(U(\mathcal{A}|_E))^2[S]\,\text{Ⓩ}\,H}((u,k),(i,j)) = ((w,\ell),(j',i'))$ 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 $(u,v) \in E_{k_1',\ell_1'}^{\mathcal{A}}$ and $(v,w) \in E_{k_2',\ell_2'}^{\mathcal{A}}$ where $\mathrm{ROT}_H(k,i) = ((k_1',k_2'),i')$ and $\mathrm{ROT}_H((\ell_2',\ell_1'),j) = (\ell,j')$.*

Proof. By definition of the zig-zag product the assumption of the claim imply $\mathrm{ROT}_{(U(\mathcal{A}|_E))^2[S]}(u,(k_1',k_2')) = (w,(\ell_2',\ell_1'))$ for $\mathrm{ROT}_H(k,i) = ((k_1',k_2'),i')$ and $\mathrm{ROT}_H((\ell_2',\ell_1'),j) = (\ell,j')$. Since $\mathrm{ROT}_{(U(\mathcal{A}|_E))^2[S]}$ is equal to $\mathrm{ROT}_{(U(\mathcal{A}|_E))^2}$ restricted to elements of the set S, we have that $\mathrm{ROT}_{(U(\mathcal{A}|_E))^2}(u,(k_1',k_2')) = (w,(\ell_2',\ell_1'))$. By definition of squaring $\mathrm{ROT}_{(U(\mathcal{A}|_E))^2}(u,(k_1',k_2')) = (w,(\ell_2',\ell_1'))$ implies that there is v such that $\mathrm{ROT}_{U(\mathcal{A}|_E)}(u,k_1') = (v,\ell_1')$ and $\mathrm{ROT}_{U(\mathcal{A}|_E)}(v,k_2') = (w,\ell_2')$. ∎

CLAIM 3.2. *If $(u,v) \in E_{k_1',\ell_1'}^{\mathcal{A}}$, $(v,w) \in E_{k_2',\ell_2'}^{\mathcal{A}}$ for $u,v,w \in A$, $k_1',k_2',\ell_1',\ell_2' \in ([D]^2)^2$ and there is $u' \in A$ with $(u,u') \in F^{\mathcal{A}}$ then there is $w' \in A$ with $(w,w') \in F^{\mathcal{A}}$. Moreover for any $i,i',j,j' \in [D]$ there are $\tilde{u},\tilde{w} \in A$, $k,\ell \in ([D]^2)^2$ such that $(\tilde{u},\tilde{w}) \in E_{(i,j),(j'i')}^{\mathcal{A}}$ for $(u,\tilde{u}) \in F_k^{\mathcal{A}}$ and $(w,\tilde{w}) \in F_\ell^{\mathcal{A}}$ 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 $\mathcal{A} \models \varphi_{\text{recursion}}$. Since $\varphi_{\text{recursion}}$ has the form $\forall x \forall z \psi(x,z)$ we know that $\mathcal{A} \models \psi(u,w)$. Then $\mathcal{A} \not\models \big[\neg\exists y F(x,y) \wedge \neg\exists y F(z,y)\big](u,w)$ as $(u,u') \in F^{\mathcal{A}}$. Since $\mathcal{A} \models \exists y\big[E_{k_1',\ell_1'}(x,y) \wedge E_{k_2',\ell_2'}(y,z)\big](u,w)$

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

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 $\tilde{u},\tilde{w} \in A$, $k,\ell \in ([D]^2)^2$ as claimed. In particular there is $w' \in A$ such that $(w,w') \in F^{\mathcal{A}}$. ∎

We will argue that for every element $w \in S$ there is a $w' \in S'$ such that $(w,w') \in F^{\mathcal{A}}$. For this pick any $u' \in S'$. Let $u \in S$ be the element such that $(u,u') \in F^{\mathcal{A}}$. By combining Lemma 2.3 and Theorem 2.2 and Lemma 2.1 it follows that $((U(\mathcal{A}|_E))^2[S])\,\text{Ⓩ}\,H$ is connected. Therefore, there is a path $(u_0',\ldots,u_m')$ in $((U(\mathcal{A}|_E))^2[S])\,\text{Ⓩ}\,H$ 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 3.1 there is a path $(u_0,v_0,u_1,v_1,\ldots u_{m-1},v_{m-1},u_m)$ in $U(\mathcal{A}|_E)$ from $u_0 = u$ to $u_m = w$. By inductively using Claim 3.2 on the path we find w' such that $(w,w') \in F^{\mathcal{A}}$.

Combining this with $\mathcal{A} \models \varphi_{\text{tree}}$ implies that the map $f : S \times ([D]^2)^2 \to S'$, given by $f(v,k) = u$

if $(v,u) \in F_k^{\mathcal{A}}$, is well defined. Furthermore, as a consequence of Claim 3.1 and Claim 3.2 imply that if $\mathrm{ROT}_{(U(\mathcal{A}|_E))^2[S] \textcircled{z} H}((u,k),(i,j)) = ((w,\ell),(j',i'))$ then $\mathrm{ROT}_{(U(\mathcal{A}|_E))[S']}(f((u,k)),(i,j)) = (f((w,\ell)),(j',i'))$. This proves that f maps each edge in $((U(\mathcal{A}|_E))^2[S]) \textcircled{z} H$ injectively to an edge in $U(\mathcal{A}|_E)[S']$. Then the map f together with the corresponding edge map is an isomorphism from $((U(\mathcal{A}|_E))^2[S]) \textcircled{z} H$ to $U(\mathcal{A}|_E)$ as both are D^2-regular.

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

Lemma 3.2. *Let $\mathcal{A} \in C_d$ be a model of $\varphi_{\textcircled{z}}$. Then $G(\mathcal{A}|_F)$ is connected.*

Proof. Assume that this is false and $G(\mathcal{A}|_F)$ has more than one connected component. Since $\mathcal{A} \models \varphi_{\mathrm{tree}}$ there is exactly one element v such that $\mathcal{A} \models \varphi_{\mathrm{root}}(v)$. Therefore we can pick G' to be a connected component of $G(\mathcal{A}|_F)$ which does not contain v. Let V be the set of vertices of G'. For the next claim we should have in mind that $(\mathcal{A}|_F)[V]$ can be understood as a directed graph in which every vertex has in-degree 1 and the corresponding undirected graph G' is connected. Hence $(\mathcal{A}|_F)[V]$ 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 3.3. *G' contains a cycle $(c_0,\ldots,c_{\ell-1})$ and for every vertex v of G' there is exactly one path $(p_0,\ldots,p_m)$ in G' with $p_0 = v$, p_m on the cycle and p_i not on the cycle for all $i \in [m]$.*

Proof. 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 $(v_i, v_{i-1}) \in F^{\mathcal{A}}$. Such a vertex always exists by the choice of G' (i.e. that the root is not in G') and the fact that $\mathcal{A} \models \varphi_{\mathrm{tree}}$. Furthermore, such a vertex is unique as $\mathcal{A} \models \varphi_{\mathrm{tree}}$. We also let $S_i := S_{i-1} \cup \{v_i\}$. Since A is finite the chain $S_0 \subseteq S_1 \subseteq \cdots \subseteq S_i \subseteq \ldots$ must become stationary. 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 $(v_i, v_{i-1}, \ldots, v_{j+1}, v_j)$ is a cycle in G' as by construction $(v_k, v_{k-1}) \in F^{\mathcal{A}}$ which implies that $\{v_k, v_{k-1}\}$ is an edge in $G(\mathcal{A}|_F)$. Let $C = \{c_0, \ldots, c_{\ell-1}\}$ be the vertices of the cycle. Since G' is connected a path that satisfies the property as described in the assertion of the claim always exists. Let us argue that such a path is unique. Assume there are two different such path $(p_0,\ldots,p_m)$ and $(p'_0,\ldots,p'_{m'})$ and assume $p_m = c_i$ 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 $\mathcal{A} \models \varphi_{\mathrm{tree}}$ for every vertex w of G' there can only be one vertex w' of G' such that $(w', w) \in F^{\mathcal{A}}$. As $p_{m-1} \notin C$ and $(c_{(i-1) \bmod \ell}, p_m) \in F^{\mathcal{A}}$ this means that $(p_m, p_{m-1}) \in F^{\mathcal{A}}$. Applying the argument inductively we get that $(p_k, p_{k-1}) \in F^{\mathcal{A}}$. The same argument works for the path $(p'_0,\ldots,p'_{m'})$ and therefore $(p'_k, p'_{k-1}) \in F^{\mathcal{A}}$. By the choice of k we know that $p_{k-1} = p'_{k-1}$ and $p_k \neq p'_k$ which contradicts $\mathcal{A} \models \varphi_{\mathrm{tree}}$. ∎

Let S_0 be the vertex set of the connected component of $U(\mathcal{A}|_E)$ with $c_0 \in S_0$. Note that S_0 might not be contained in G'. We define the infinite sequence $S_i := \{w \in A \mid (v,w) \in F^{\mathcal{A}}, v \in S_{i-1}\}$ for every $i \in \mathbb{N}_{>0}$. Let $m_i := \max_{v \in S_i \cap V} \min_{j \in \{0,\ldots,\ell-1\}} \{\mathrm{dist}_{G'}(c_j, v)\}$ 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 3.4. *$U(\mathcal{A}|_E)[S_i] = (U(\mathcal{A}|_E)[S_{i-1}])^2 \textcircled{z} H$.*

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

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

We now argue that $(U(\mathcal{A}|_E))^2[\tilde{S}]$ is a connected component of $(U(\mathcal{A}|_E))^2$. Assuming the contrary, every connected component of $(U(\mathcal{A}|_E))^2$ either contains vertices from both $\tilde{S}$ and $A \setminus \tilde{S}$, or $(U(\mathcal{A}|_E))^2[\tilde{S}]$ 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 the root as the root is not in any E-relation. Hence by Lemma 3.1 we get a connected component of $U(\mathcal{A}|_E)$ on the children of S' both containing vertices from S_0 and from $A \setminus S_0$ which contradicts S_0 being a connected component of $U(\mathcal{A}|_E)$. Now let S' be a connected component as in the second case, and pick S' such that it does not contain the root. Then by Lemma 3.1 S_0 must have a non-empty intersection with at least two connected components of $U(\mathcal{A}|_E)$ which is a contradiction.

Thus, by Lemma 2.3 $\lambda((U(\mathcal{A}|_E))^2[\tilde{S}]) < 1$. Since $U(\mathcal{A}|_E)[S_0] = ((U(\mathcal{A}|_E))^2[\tilde{S}]) \textcircled{z} H$ by Lemma 3.1, Theorem 2.2 and $\lambda(H) < 1$ ensure that $\lambda(U(\mathcal{A}|_E)[S_0]) < 1$.

By induction we get $\lambda(U(\mathcal{A}|_E)[S_{i-1}]) < 1$ for $i > 1$, which, together with Lemma 2.2 and Lemma 2.1, implies that $(U(\mathcal{A}|_E)[S_{i-1}])^2$ is a connected component[3] of $(U(\mathcal{A}|_E))^2$ and $(U(\mathcal{A}|_E))^2[S_{i-1}] = (U(\mathcal{A}|_E)[S_{i-1}])^2$.

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

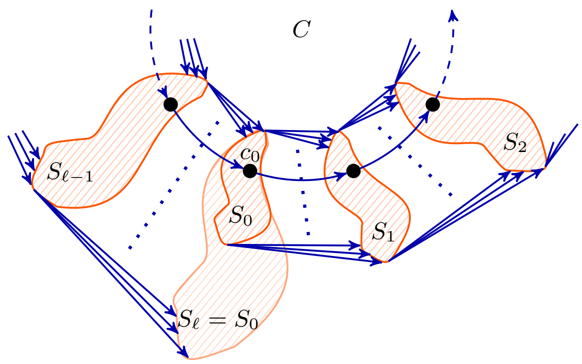

Figure 3: Illustration of the proof of Lemma 3.2.

Since $c_{i \bmod \ell} \in S_i$, by Lemma 3.1, we have that $U(\mathcal{A}|_E)[S_i]$ is a connected component of $U(\mathcal{A}|_E)$ and $U(\mathcal{A}|_E)[S_i] = (U(\mathcal{A}|_E)[S_{i-1}])^2 \textcircled{z} H$. Using Lemma 2.2 and Theorem 2.2 this proves $\lambda(U(\mathcal{A}|_E)[S_i]) < 1$. ∎

Claim 3.5. *For every $v \in S_i$ there is $w \in V$ such that $(v, w) \in F^{\mathcal{A}}$.*

Proof. $U(\mathcal{A}|_E)[S_{i+1}] = (U(\mathcal{A}|_E)[S_i])^2 \textcircled{z} H$ by Claim 3.4 . Therefore by definition of squaring and the zig-zag product we know that $|S_{i+1}| = D^4 \cdot |S_i|$. Since additionally $\mathcal{A} \models \varphi_{\text{tree}}$ we know that every $v \in S_i$ 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 $(v, w) \in F^{\mathcal{A}}$. ∎

Therefore there is $w_i \in V$ such that $(v_i, w_i) \in F^{\mathcal{A}}$. Let $(u_0, \ldots, u_{m_i})$ 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 $(u_{j+1}, u_j) \in F^{\mathcal{A}}$ for all $j \in [m_i]$. Furthermore, since $v_i = u_0 \neq u_1$, assuming that $w_i = u_1$ would imply $(v_i, u_1), (u_2, u_1) \in F^{\mathcal{A}}$, which contradicts $\mathcal{A} \models \varphi_{\text{tree}}$. Then $(w_i, u_0, \ldots, u_{m_i})$ is a path in G' from w_i to C. Since $w_i \in S_{i+1}$ by construction, Claim 3.3 implies that $m_{i+1} \geq m_i + 1$. Therefore $m_i \geq i + m_0$ inductively. But this yields a contradiction since $\ell + m_0 \leq m_\ell = m_0$ and $\ell > 0$. See Figure 3 for illustration. Therefore $G(\mathcal{A}|_F)$ must be connected. □

Lemma 3.3. *Let $\mathcal{A} \in C_d$ be a (finite) model of $\varphi_{\textcircled{z}}$. Then $|A| = \sum_{m=0}^{n} D^{4m}$ for some $n \in \mathbb{N}$, $G(\mathcal{A}|_F)$ is a D^4-ary complete rooted tree, where the root is the unique element $v \in A$ for which $\mathcal{A} \models \varphi_{\text{root}}(v)$, and $U(\mathcal{A}|_E)[T_m] \cong G_m$ where G_m is defined in Definition 2.5 and T_m is the set of vertices of distance m to v in $G(\mathcal{A}|_F)$ for any $m \in \{1, \ldots, n\}$. Moreover for every $n \in \mathbb{N}$ there is a model of $\varphi_{\textcircled{z}}$ of size $\sum_{m=0}^{n} D^{4m}$.*

Proof. Lemma 3.2 combined with $\mathcal{A} \models \varphi_{\text{tree}}$ proves that $G(\mathcal{A}|_F)$ is a rooted tree. Let n be the greatest distance of any vertex in $G(\mathcal{A}|_F)$ to the root and let T_m be the vertices of distance m to the root for $m \leq n$. Then $U(\mathcal{A}|_E)[T_1] \cong G_1$ because $\mathcal{A} \models \varphi_{\text{base}}$. Now assume towards an inductive proof that $U(\mathcal{A}|_E)[T_m] \cong G_m$ for some fixed $m \in \mathbb{N}_{>0}$. As $\lambda(G_m) < 1$ by Lemma 2.2 and Lemma 2.1 we get $(U(\mathcal{A}|_E))^2[T_m]$ is a connected component of $(U(\mathcal{A}|_E))^2$. Hence Lemma 3.1 implies $U(\mathcal{A}|_E)[T_{m+1}] \cong G_{m+1}$. Since G_m has D^{4m} vertices this also shows that $\mathcal{A}$ has $\sum_{m=0}^{n} D^{4m}$ vertices. □

Now we are ready to finish the proof of Theorem 3.1.

Proof of Theorem 3.1.

We prove that for $\epsilon = D^2/12$ the claimed is true. Let $\mathcal{A}$ be the model of $\varphi_{\textcircled{z}}$ of size $\sum_{m=0}^{n} D^{4m}$ and $S \subseteq A$ 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 $G(\mathcal{A}|_F)$ 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 $m \in [n]$. Because G_m is a $D^2/4$-expander (this follows directly from Theorem 2.1 as $\lambda(G_m) \leq 1/2$ by Proposition 2.1) and $U(\mathcal{A}|_E)[T_m] \cong G_m$ we know that

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

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

$$|S_{m'}| > \frac{|T_{m'}|}{2} = \frac{D^{4m'}}{2}. \tag{3.2}$$

Claim 3.6. $\sum_{m=0}^{\tilde{m}-1} |T_m| \leq \frac{1}{2} |T_{\tilde{m}}|$ *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 3.6 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

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

Assume now that $m' < n$. Since S is the disjoint union of all S_m we know that the set $\langle S, \overline{S} \rangle_{U(\mathcal{A})}$ contains the disjoint sets $\langle S_m, T_m \setminus S_m \rangle_{U(\mathcal{A})}$, $\langle T_{m'} \setminus S_{m'}, T_{m'} \rangle_{U(\mathcal{A})}$ and $\langle S_{m'}, T_{m'+1} \setminus S_{m'+1} \rangle_{U(\mathcal{A})}$ for all $m' < m \leq n$. 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 $|\langle S_{m'}, T_{m'+1} \setminus S_{m'+1} \rangle_{U(\mathcal{A})}| =$

$|\langle S_{m'}, T_{m'+1}\rangle_{U(\mathcal{A})}| - |\langle S_{m'}, S_{m'+1}\rangle_{U(\mathcal{A})}| \geq D^4|S_{m'}| - |S_{m'+1}| \geq D^4(|S_{m'}| - D^{4m'}/2)$. Since additionally $|T_{m'}|/2 \geq |T_{m'} \setminus S_{m'}| = D^{4m'} - |S_{m'}|$ and G_m is an $D^2/4$-expander for every m we get

$$
\begin{aligned}
|\langle S, \overline{S}\rangle_{U(\mathcal{A})}| &\geq \sum_{m>m'} \frac{D^2}{4}|S_m| + \frac{D^2}{4}|T_{m'} \setminus S_{m'}| \\
&\qquad + D^4(|S_{m'}| - \frac{D^{4m'}}{2}) \\
&= \frac{D^2}{4}\sum_{m>m'}|S_m| + \Big(D^4 - \frac{D^2}{2}\Big)|S_{m'}| \\
&\qquad - \Big(D^4 - \frac{D^2}{2}\Big)\frac{D^{4m'}}{2} + \frac{D^2}{4}|S_{m'}| \\
&\overset{\text{Equation 3.2}}{\geq} \frac{D^2}{4}\sum_{m>m'}|S_m| + \frac{D^2}{8}|S_{m'}| + \frac{D^2}{8}\Big(\frac{|T_{m'}|}{2}\Big) \\
&\overset{\text{Claim 3.6}}{\geq} \frac{D^2}{4}\sum_{m>m'}|S_m| + \frac{D^2}{8}|S_{m'}| + \frac{D^2}{8}\sum_{m<m'}|T_m| \\
&\overset{|T_m|\geq|S_m|}{\geq} \frac{D^2}{12}|S|. \qquad \square
\end{aligned}
$$

4 On the non-testability of a Π_2-property

In this section we 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 $P_{\varphi_{Ⓩ}}$ defined by the formula $\varphi_{Ⓩ}$ in Section 3 is not testable. Later we prove that $\varphi_{Ⓩ}$ is in Π_2. Finally, we extends our non-testability result to simple graphs.

Non-testability of $P_{\varphi_{Ⓩ}}$. Recall that r-types are the isomorphism class of r-balls and that restricted to the class C_d there are finitely many r-types. Let $\tau_1, \dots, \tau_t$ be a list of all r-types of bounded degree d. We let $\rho_{\mathcal{A},r}$ be the r-type distribution of $\mathcal{A}$, i. e.

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

for any $X \subseteq \{\tau_1, \dots, \tau_t\}$. For two σ-structures $\mathcal{A}$ and $\mathcal{B}$ we define the *sampling distance of depth r* as $\delta_{\odot}^r(\mathcal{A}, \mathcal{B}) := \sup_{X\subseteq\{\tau_1,\dots,\tau_t\}} |\rho_{\mathcal{A},r}(X) - \rho_{\mathcal{B},r}(X)|$. Note that $\delta_{\odot}^r(\mathcal{A}, \mathcal{B})$ is just the total variance distance of $\rho_{\mathcal{A},r}, \rho_{\mathcal{B},r}$, and $\delta_{\odot}^r(\mathcal{A}, \mathcal{B}) = \frac{1}{2}\sum_{i=1}^t |\rho_{\mathcal{A},r}(\{\tau_i\}) - \rho_{\mathcal{B},r}(\{\tau_i\})|$. The *sampling distance* of $\mathcal{A}$ and $\mathcal{B}$ is defined as $\delta_{\odot}(\mathcal{A}, \mathcal{B}) := \sum_{r=0}^{\infty} \frac{1}{2^r} \cdot \delta_{\odot}^r(\mathcal{A}, \mathcal{B})$.

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

Theorem 4.1. ([Lov12]) *For every $\lambda > 0$ there is a positive integer n_0 such that for every σ-structure $\mathcal{A} \in C_d$ there is a σ-structure $\mathcal{H} \in C_d$ such that $|H| \leq n_0$ and $\delta_{\odot}(\mathcal{A}, \mathcal{H}) \leq \lambda$.*

We use the following definition of local properties.

Definition 4.1. ([AH18]) *Let $\epsilon \in (0, 1]$. A property $P \subseteq C_d$ is ϵ-*local *on C_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 $\mathcal{A} \in P$ and $\mathcal{B} \in C_d$ both on $n \geq n_0$ vertices, if $\sum_{i=1}^t |\rho_{\mathcal{A},r}(\{\tau_i\}) - \rho_{\mathcal{B},r}(\{\tau_i\})| < \lambda$ then $\mathcal{B}$ is ϵ-close to P, where $\tau_1, \dots, \tau_t$ is a list of all r-types of bounded degree d. The property P is* local *on C_d if it is ϵ-local on C_d for every $\epsilon \in (0, 1]$.*

The following theorem relating testable properties and local properties was proven in [AH18]

Theorem 4.2. ([AH18]) *For every property $P \in C_d$, P is testable if and only if P is local on C_d.*

We let $P_{Ⓩ} := P_{\varphi_{Ⓩ}}$ for the formula $\varphi_{Ⓩ}$ from Section 3. We also let σ and d be as defined in Section 3.

Theorem 4.3. *$P_{Ⓩ}$ is not testable on C_d.*

Proof. We prove non-locality for $P_{Ⓩ}$ and get non-testability with Theorem 4.2. 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 $\mathcal{A} \in C_d$ be a model of $\varphi_{Ⓩ}$ on n vertices. By Theorem 4.1 there is a structure $\mathcal{H} \in C_d$ on $m \leq n_0'$ vertices such that $\delta_{\odot}(\mathcal{A}, \mathcal{H}) \leq \lambda$. Let $\mathcal{B}$ be the structure consisting of $\lfloor n/m \rfloor$ copies of $\mathcal{H}$ and $n \bmod m$ isolated vertices. Note that we picked $\mathcal{B}$ such that $|A| = |B|$.

We will first argue that $\mathcal{B}$ is in fact ϵ-far from having the property $P_{Ⓩ}$. First we rename the elements from B in such a way that $A = B$ and the number $\sum_{\tilde{R}\in\sigma} |\tilde{R}^{\mathcal{A}} \Delta \tilde{R}^{\mathcal{B}}|$ of edge modifications to turn $\mathcal{A}$ and $\mathcal{B}$ into the same structure is minimal. Pick a partition $A = B = S \sqcup S'$ in such a way that $S \times S' \cap \tilde{R}^{\mathcal{B}} = \emptyset$, $S' \times S \cap \tilde{R}^{\mathcal{B}} = \emptyset$ for any $\tilde{R} \in \sigma$ and $||S| - |S'||$ minimal among all such partitions. Assume that $|S| \leq |S'|$. Since the connected components in $\mathcal{B}$ are of size $\leq m$ we know that $||S| - |S'|| \leq m$ because otherwise we can get a partition $B = T \sqcup T'$ with $||T| - |T'|| < ||S| - |S'||$ by picking a connected component of $\mathcal{B}$ whose elements are contained in S' and moving them from S' to S. Since $|S| \leq |S'|$ and $m \leq n/4$ we know that $n/4 \leq |S| \leq n/2$. This implies that

$$
\begin{aligned}
\sum_{\tilde{R}\in\sigma} |\tilde{R}^{\mathcal{A}} \Delta \tilde{R}^{\mathcal{B}}| \geq |\langle S, S'\rangle_{U(\mathcal{A})}| &\overset{\text{Def 2.2}}{\geq} |S| \cdot h(\mathcal{A}) \\
&\overset{\text{Thm 3.1}}{\geq} \frac{n}{4} \cdot \frac{D^2}{12} = \frac{1}{48}D^2 n \geq \frac{1}{144D^2} dn.
\end{aligned}
$$

Therefore $\mathcal{B}$ is ϵ-far from being in $P_{ⓩ}$. But the neighbourhood distributions of $\mathcal{A}$ and $\mathcal{B}$ are similar as the following shows, proving that $P_{ⓩ}$ is not local.

$$\sum_{i=1}^{t} |\rho_{\mathcal{A},r}(\{\tau_i\}) - \rho_{\mathcal{B},r}(\{\tau_i\})| = \sum_{i=1}^{t} \Big|\rho_{\mathcal{A},r}(\{\tau_i\})$$
$$- \frac{n \bmod m}{n} \cdot \rho_{K_1,r}(\{\tau_i\}) - \Big\lfloor \frac{n}{m} \Big\rfloor \cdot \frac{m}{n} \cdot \rho_{\mathcal{H},r}(\{\tau_i\})\Big|$$
$$\leq \sum_{i=1}^{t} \Big|\rho_{\mathcal{A},r}(\{\tau_i\}) - \rho_{\mathcal{H},r}(\{\tau_i\})\Big| + \sum_{i=1}^{t} \Big|\frac{n \bmod m}{n} \cdot$$
$$\rho_{K_1,r}(\{\tau_i\})\Big| + \sum_{i=1}^{t} \Big|\rho_{\mathcal{H},r}(\{\tau_i\}) - \Big\lfloor \frac{n}{m} \Big\rfloor \cdot \frac{m}{n} \cdot \rho_{\mathcal{H},r}(\{\tau_i\})\Big|$$
$$\leq \sum_{i=1}^{t} \Big|\rho_{\mathcal{A},r}(\{\tau_i\}) - \rho_{\mathcal{H},r}(\{\tau_i\})\Big| + \frac{2m}{n}$$
$$\leq t \cdot \sup_{X \subseteq \mathcal{B}_r} |\rho_{\mathcal{A},r}(X) - \rho_{\mathcal{H},r}(X)| + \frac{2m}{n}$$
$$\leq t \cdot 2^r \cdot \delta_{\odot}(\mathcal{A}, \mathcal{H}) + \frac{2m}{n} \leq \frac{\lambda}{2} + \frac{\lambda}{2} = \lambda.$$

The last inequality holds by choice of λ' and Theorem 4.1. □

Every FO property on degree-regular structures is in Π_2. We first give the following definition.

DEFINITION 4.2. *A* Hanf sentence *is a sentence of the form $\exists^{\geq m} x\, \varphi_\tau(x)$, which is short for*

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

where τ is a neighbourhood type (say, of radius r) and $\varphi_\tau(x_i)$ expresses that x_i has r-neighbourhood type τ.

Note that $\varphi_\tau(x_i)$ 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 or 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. Hence by definition, any Hanf sentence is in Σ_2.

LEMMA 4.1. *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.*

Proof. Before we begin, let us define an r-type τ to be *d-regular*, if for all structures $\mathcal{A}$ and all elements $a \in A$ of r-type τ, every $b \in A$ with $\mathrm{dist}(a, b) < r$ has $\deg_{\mathcal{A}}(b) = d$. We first prove the following claim.

CLAIM 4.1. *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 Definition 4.2, we may assume $\varphi_\tau(x_i)$ is an $\exists^*\forall$-formula, which is a conjunction of an $\exists^*$-formula $\varphi'_\tau(x_i)$ (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 $\mathcal{A} \models \exists^{\geq m} x\, \varphi'_\tau(x)$ and $\deg(\mathcal{A}) \leq d$. Then $\mathcal{A} \models \exists^{\geq m} x\, \varphi_\tau(x)$ because τ is d-regular. The converse is obvious.

If ψ is of the form $\neg\exists^{\geq m} x\, \varphi_\tau(x)$, where $\varphi_\tau(x_i)$ 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 4.1, because if φ only has d-regular models, then by Hanf's Theorem there is $\psi \equiv \varphi$ satisfying the assumptions of the claim. □

Existence of a non-testable Π_2-property. With Lemma 4.1 and Theorem 4.3, we are ready to prove the following theorem.

THEOREM 4.4. *There are degree bounds $d \in \mathbb{N}$ such that there exists a property on C_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. Using the construction from [RVW02] we can find a $(D^4, D, 1/4)$-graph H. By Theorem 4.3, using this base expander H for the construction of $\varphi_{ⓩ}$ we get a property which is not testable on C_d. As all models of $\varphi_{ⓩ}$ are d-regular by construction, Lemma 4.1 implies that $\varphi_{ⓩ}$ is d-equivalent to a formula in Π_2. □

4.1 Extension 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 carefully translate the edge-coloured directed graphs 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 $\varphi_{ⓩ}$

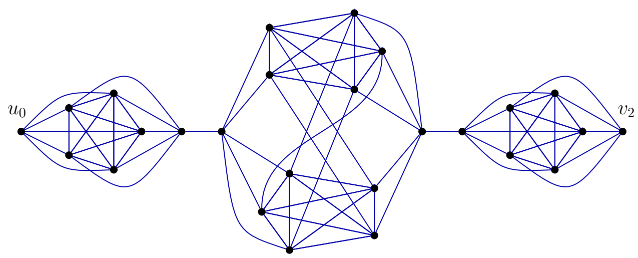

Figure 4: Illustration of $P^6_{2,1}(u_0, v_2)$.

into a formula $\psi_{\text{Ⓩ}}$. We obtain a class of simple expanders, that is defined by an FO sentence, and obtain the analogous Theorem.

THEOREM 4.5. *There are $d \in \mathbb{N}$ and an FO property of simple graphs of bounded degree d that is not testable.*

In the rest of this section, we prove the above theorem.

Let d be as defined in Section 3. Let $G^d(u, v)$ be the graph with vertices $\{u, v, u_0, \dots, u_{d-2}\}$ and edges $\{\{w, u_i\}, \{v, u_i\}, \{u_i, u_j\} \mid i, j \in [d-2], i \neq j\}$. Additionally let $H^d(u, v)$ be the graph with vertices $\{u, v, u_i, u'_j, v_i, v'_j \mid i \in [\lfloor \frac{d-1}{2} \rfloor], j \in [\lceil \frac{d-1}{2} \rceil]\}$ and edges

$$\Big\{\{u, u_i\}, \{v, v_i\}, \{u_i, v_i\} \mid i \in \Big[\Big\lfloor \frac{d-1}{2} \Big\rfloor\Big]\Big\} \cup$$
$$\Big\{\{u, u'_j\}, \{v, v'_j\}, \{u'_j, v'_j\}\} \mid j \in \Big[\Big\lceil \frac{d-1}{2} \Big\rceil\Big]\Big\} \cup$$
$$\Big\{\{u_i, u_k\}, \{v_i, v_k\} \mid i, k \in \Big[\Big\lfloor \frac{d-1}{2} \Big\rfloor\Big], i \neq k\Big\} \cup$$
$$\Big\{\{u'_j, u'_k\}, \{v'_j, v'_k\} \mid j, k \in \Big[\Big\lceil \frac{d-1}{2} \Big\rceil\Big], j \neq k\Big\} \cup$$
$$\Big\{\{u_i, v'_j\}, \{u'_j, v_i\} \mid i \in \Big[\Big\lfloor \frac{d-1}{2} \Big\rfloor\Big], j \in \Big[\Big\lceil \frac{d-1}{2} \Big\rceil\Big]\Big\}.$$

For every $\ell \in \mathbb{N}$, $0 \leq p \leq \ell$, let $P^d_{\ell,p}(u_0, v_\ell)$ be the graph consisting of ℓ copies $G^d(u_0, v_0), \dots, G^d(u_{p-1}, v_{p-1}), G^d(u_{p+1}, v_{p+1}), \dots,$ $G^d(u_\ell, v_\ell)$, one copy $H^d(u_p, v_p)$ and additional edges $\{v_i, u_{i+1}\}$ for each $i \in [\ell]$. Note that $P^d_{\ell,p}(u_0, v_\ell)$ has $\ell \cdot (d+1) + 2d$ vertices, the vertices u_0 and v_ℓ have degree $d-1$ and every other vertex has degree d, see Figure 4 for an example.

Let $\mathcal{A} \in P_{\text{Ⓩ}}$ and let $\ell = 2 \cdot (3D^4 + 1)$. We obtain an undirected graph $G = (V, E)$ from $\mathcal{A}$ as follows.

1. For every $i_0, i_1, i_2, i_3 \in [D]$ we define $p = \sum_{k=0}^{3} i_k \cdot D^k$ and replace every edge $(x, y) \in E^{\mathcal{A}}_{(i_0,i_1),(i_2,i_3)}$ by $P^d_{\ell,p}(u_0, v_\ell)$ and additional edges $\{x, u_0\}$ and $\{v_\ell, y\}$. Here all vertices of $P^d_{\ell,p}(u_0, v_\ell)$ are pairwise distinct and new, and we call them *auxiliary vertices.* Call this gadget graph an $E_{(i_0,i_1),(i_2,i_3)}$*-arrow with end-vertices x and y.*
2. For every $i_0, i_1, i_2, i_3 \in [D]$ we define $p = D^4 + \sum_{k=0}^{3} i_k \cdot D^k$ and replace every edge $(x, y) \in F^{\mathcal{A}}_{((i_0,i_1),(i_2,i_3))}$ by $P^d_{\ell,p}(u_0, v_\ell)$ and additional edges $\{x, u_0\}$ and $\{v_\ell, y\}$. Here all vertices of $P^d_{\ell,p}(u_0, v_\ell)$ are pairwise distinct and new, and we call them *auxiliary vertices.* Call this gadget graph an $F_{((i_0,i_1),(i_2,i_3))}$*-arrow with end-vertices x and y.*
3. For every $i_0, i_1, i_2, i_3 \in [D]$ we define $p = 2D^4 + \sum_{k=0}^{3} i_k \cdot D^k$ and replace every edge $(x, y) \in L^{\mathcal{A}}_{((i_0,i_1),(i_2,i_3))}$ by $P^d_{\ell,p}(u_0, v_\ell)$ and additional edges $\{x, u_0\}$ and $\{v_\ell, y\}$. Here all vertices of $P^d_{\ell,p}(u_0, v_\ell)$ are pairwise distinct and new, and we call them *auxiliary vertices.* Call this gadget graph an $L_{((i_0,i_1),(i_2,i_3))}$*-arrow with end-vertices x and y.*
4. We define $p = 3D^4$ and replace every edge $(x, y) \in R^{\mathcal{A}}$ by $P^d_{\ell,p}(u_0, v_\ell)$ and additional edges $\{x, u_0\}$ and $\{v_\ell, y\}$. Here all vertices of $P^d_{\ell,p}(u_0, v_\ell)$ are pairwise distinct and new, and we call them *auxiliary vertices.* Call this gadget graph an *R-arrow with end-vertices x and y.*

All vertices, that are not *auxiliary*, are called *original vertices.* Note that from the location p of the gadget $H^d(v_0, v_\ell)$ uniquely encodes the colour of the original directed coloured edge. Also note that each arrow defined above has a direction as the gadget $H^d(v_0, v_\ell)$ is always located in the first half of the path $P^d_{\ell,p}(u_0, v_\ell)$. The following is easy to observe from the construction.

FACT 4.1. *For every $x \in V$, x is an original vertex iff x is contained in no triangle.*

We let $\delta(x)$ be a formula in the language of undirected graphs, saying 'x is an original vertex', which is easy to do by Fact 4.1. We further let $\beta(x)$ be a formula saying 'x is an internal vertex of either an $E_{i,j}$-arrow, or an F_k-arrow, or an L_k-arrow, or an R-arrow for any $i, j \in [D]^2$, $k \in ([D]^2)^2$'. Here an 'internal vertex' of an arrow refers to any vertex on this arrow except the two endpoints. We now translate the formula $\varphi_{\text{Ⓩ}}$ into a formula $\psi_{\text{Ⓩ}}$ in the language of undirected graphs using the following first-order formulas $\alpha^E_{i,j}$, α^F_k, α^L_k and α^R. Let $\alpha^E_{i,j}(x, y)$ say 'x and y are the end-vertices of an induced $E_{i,j}$-arrow' for $i, j \in [D]^2$, similarly, let $\alpha^F_k(x, y)$ say 'x and y are the end-vertices of an induced F_k-arrow' for $k \in ([D]^2)^2$. Furthermore let $\alpha^L_k(x, y)$ say 'x and y are the end-vertices of an induced L_k-arrow' for $k \in ([D]^2)^2$ and $\alpha^R(x, y)$ say 'x and y are the end-vertices of an induced R-arrow' . Given $\varphi_{\text{Ⓩ}}$, formula $\psi_{\text{Ⓩ}}$ is obtained as follows. In $\varphi_{\text{Ⓩ}}$ we replace each expression $E_{i,j}(x, y)$ by $\alpha^E_{i,j}(x, y)$, each $F_k(x, y)$ by $\alpha^F_k(x, y)$, each

$L_k(x,y)$ by $\alpha_k^L(x,y)$ and each $R(x,y)$ by $\alpha^R(x,y)$. In addition, we relativise all quantifiers to the original vertices (replacing every expression of the form $\exists x\,\chi$ by $\exists x\,(\delta(x) \wedge \chi)$ and every expression of the form $\forall x\,\chi$ by $\forall x\,(\delta(x) \rightarrow \chi)$). Let us call the resulting formula $\psi'_{Ⓩ}$. Then we set $\psi_{Ⓩ}$ to be the conjunction of the formula $\psi'_{Ⓩ}$ and the formula $\forall x(\neg\delta(x) \rightarrow \beta(x))$. Let $\mathcal{P}_\psi := \{G \in C_d \mid G \models \psi_{Ⓩ}\}$. In the following, we show that $\mathcal{P}_\psi$ is a family of expanders to prove non-testability of $\mathcal{P}_\psi$. We remark that one could also prove the non-testability of $\mathcal{P}_\psi$ by showing that the aforementioned transformation (from σ-structures to simple graphs) is a (local) reduction that preserves testability.

Lemma 4.2. *The models of $\psi_{Ⓩ}$ is a family of ξ-expanders, for some constant $\xi > 0$.*

Proof. Let $G = (V,E)$ be a model of $\psi_{Ⓩ}$ and let $\mathcal{A}$ be the corresponding model of $\varphi_{Ⓩ}$. Let $S \subset V$ such that $|S| \le \frac{|V|}{2}$. Let $V_{\text{original}} \sqcup V_{\text{auxiliary}} = V$ be the partition of V into original and auxiliary vertices. Let $S_{\text{original}} := V_{\text{original}} \cap S$ and $S_{\text{auxiliary}} := V_{\text{auxiliary}} \cap S$.

First note that by the above definitions every directed coloured edge in $\mathcal{A}$ corresponds to a constant number $c_D := 2 \cdot (3D^4+1) \cdot ((d+1)+2d)$ of auxiliary vertices in $V_{\text{auxiliary}}$, where $d = 2D^2 + D^4 + 1$.

Assume $|S_{\text{original}}| > \frac{2}{dc_D} \cdot |S|$. Then there are $|S| - |S_{\text{original}}| < \frac{dc_D - 2}{2} \cdot |S_{\text{original}}|$ vertices in $S_{\text{auxiliary}}$. Hence at least $\frac{d}{2} \cdot |S_{\text{original}}| - \frac{dc_D-2}{2c_D} \cdot |S_{\text{original}}|$ of the arrows have at least one vertex that is not in S. Hence

$$\begin{aligned}\langle S, V \setminus S\rangle_G &\ge \frac{d}{2} \cdot |S_{\text{original}}| - \frac{dc_D - 2}{2c_D} \cdot |S_{\text{original}}| \\ &= \frac{1}{c_D} \cdot |S_{\text{original}}| \ge \frac{2}{dc_D^2} \cdot |S|.\end{aligned}$$

Assume $\frac{1}{2dc_D}|S| < |S_{\text{original}}| \le \frac{2}{dc_D} \cdot |S|$. Let $\epsilon = \frac{D^2}{12}$ as defined in the proof of Theorem 3.1. Since each edge in $U(\mathcal{A})$ corresponds to exactly one arrow in G we get that $\langle S, V \setminus S\rangle_G \ge \langle S_{\text{original}}, V_{\text{original}} \setminus S_{\text{original}}\rangle_{U(\mathcal{A})}$. Since $\mathcal{A}$ is d-regular and every edge gets replaced by c_D auxiliary vertices we get $|V| = (1 + \frac{dc_D}{2})|A|$. Then

$$|S_{\text{original}}| \le \frac{2}{dc_D} \cdot |S| \le \frac{1}{dc_D} \cdot |V| = \frac{2 + dc_D}{2dc_D}|A|$$

and $|A \setminus S_{\text{original}}| \ge (\frac{2dc_D}{2+dc_D} - 1)|S_{\text{original}}|$. Then from Theorem 3.1 we directly get

$$\begin{aligned}\langle S, V \setminus S\rangle_G &\ge \langle S_{\text{original}}, V_{\text{original}} \setminus S_{\text{original}}\rangle_{U(\mathcal{A})} \\ &= \epsilon \min\{|S_{\text{original}}|, |A \setminus S_{\text{original}}|\} \\ &\ge \epsilon \cdot \frac{1}{2dc_D} \cdot \frac{dc_D - 2}{2 + dc_D} \cdot |S|.\end{aligned}$$

Now assume $|S_{\text{original}}| \le \frac{1}{2dc_D} \cdot |S|$. Therefore there are $|S| - |S_{\text{original}}| \ge |S| - \frac{1}{2dc_D} \cdot |S|$ vertices in $S_{\text{auxiliary}}$. Of these, at least $\frac{2dc_D - 1}{2dc_D} \cdot |S| - |S_{\text{original}}|c_D \ge \frac{2dc_D - 1 - c_D}{2dc_D}|S|$ vertices are not in a connected component with any element from S_{original} in the graph $G[S]$. Since any connected component of $G[S]$ with no vertices in S_{original} contains at most c_d vertices, we get that

$$\langle S, V \setminus S\rangle_G \ge \frac{2dc_D - c_D - 1}{2dc_D^2}|S|.$$

By setting $\xi = \min\{\frac{2dc_D - c_D - 1}{2dc_D^2}, \epsilon \cdot \frac{1}{2dc_D} \cdot \frac{dc_D - 2}{2 + dc_D}, \frac{2}{dc_D^2}\} > 0$ we proved the claimed. □

One can then prove that the property $P_{\psi_{Ⓩ}}$ is not testable by using analogous arguments as in the proof of Theorem 4.3. In the full version of the paper we give an alternative proof using a result from [FPS19].

5 On the testability of all Σ_2-properties

In this section we let $\sigma = \{R_1, \ldots, R_m\}$ be any relational signature and C_d the set of σ-structures of bounded degree d. We prove the following.

Theorem 5.1. *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 [AFKS00] from the dense model to the bounded degree model.

Definition 5.1. *Two properties $P, Q \subseteq C_d$ are called* indistinguishable *if for every $\epsilon \in (0,1)$ there exists $N = N(\epsilon)$ such that for every structure $\mathcal{A} \in P$ with $|A| > N$ there is a structure $\tilde{\mathcal{A}} \in Q$ with the same universe, that is ϵ-close to $\mathcal{A}$; and for every $\mathcal{B} \in Q$ with $|B| > N$ there is a structure $\tilde{\mathcal{B}} \in P$ with the same universe, that is ϵ-close to $\mathcal{B}$.*

The following lemma follows from the definitions, and is similar to [AFKS00], though we use the canonical testers for bounded degree graphs ([CPS16, GR11]).

Lemma 5.1. *If $P, Q \subseteq C_d$ are indistinguishable properties, then P is testable on C_d if and only if Q is testable on C_d.*

Proof. We show that if P is testable, then Q is also testable. The other direction follows by the same argument. Let $\epsilon > 0$. Since P is testable, there exists an $\frac{\epsilon}{2}$-tester for P with success probability $\ge \frac{2}{3}$. Furthermore, we can assume that the tester (called canonical tester) behaves as follows (see [CPS16, GR11]): it first uniformly samples a constant number of elements, then explores the union of r-balls around all sampled elements for some constant $r > 0$, and makes a deterministic decision whether to accept, based on an isomorphic copy of the explored substructure. Let $C = C(\frac{\epsilon}{2}, d)$ denote the number of queries the tester made on the input structure. By repeating this tester and taking the majority, we get a tester T with $c_1 \cdot C$ queries and success probability at least $\frac{5}{6}$ for some integer $c_1 > 0$.

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

High-level idea of proof of Theorem 5.1. 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 $\mathcal{M}$ (see Claim 5.2). With some further simplification of φ', we obtain a formula φ'' in Σ_2 which expresses that the structure has to have $\mathcal{M}$ as an induced substructure (see Claim 5.3) and every set of elements of fixed size ℓ has to induce some structure from a set of structures $\mathfrak{B}$, and – depending on the structure from $\mathfrak{B}$ – there might be some connections to the elements of $\mathcal{M}$. 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 then a small fraction of the tuples and vice versa (see Claim 5.6). The intuition behind this is that every structure satisfying φ'' can be made to satisfy ψ by removing the structure $\mathcal{M}$ while on the other hand for every structure which satisfies ψ we can plant the structure $\mathcal{M}$ to make it satisfy φ''. Since it is a priori unclear how the existential and universal quantified variables interact, we have to define ψ very carefully. Here it is important to note that the existence of occurrences of structures in $\mathfrak{B}$ forcing an interaction with $\mathcal{M}$ is limited because of the degree bound (see Claim 5.4). 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 [Gol17], we reach the conclusion that any property defined by a formula in Σ_2 is testable.

Especially we will not directly give a tester for the property P_φ 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 P_φ for every $\epsilon \in (0,1)$ and every $\varphi \in \Sigma_2$.

For the proof of Theorem 5.1, we use the following.

Definition 5.2. *Let $\mathcal{A}$ be a σ-structure with $A = \{a_1, \dots, a_t\}$. Let $\overline{z} = (z_1, \dots, z_t)$ be a tuple of variables. Then we define $\iota^{\mathcal{A}}(\overline{z})$ as follows.*

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

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

Proof of Theorem 5.1. Let φ be any σ-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 quantifier free. We can further assume that $\chi(\overline{x}, \overline{y})$ is in disjunctive normal form and

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

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 write the formula φ as a disjunction over all possible structures in C_d the existentially quantified variables could enforce. Since the elements realising the existentially quantified variables have a certain structure, it is a natural way to decompose the formula.

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

Claim 5.1. *Every $\mathcal{M} \in \mathfrak{M}$ has at most k elements.*

Proof. Assume there is $\mathcal{M} \in \mathfrak{M}$ with $|M| > k$. Since every structure in $\mathfrak{M}$ is a model of φ there must be a tuple $\overline{a} = (a_1, \dots, a_k) \in M^k$ such that $\mathcal{M} \models \forall \overline{y} \bigvee_{i \in I} \Big(\alpha^i(\overline{a}) \wedge \beta^i(\overline{y}) \wedge \mathrm{pos}^i(\overline{a}, \overline{y}) \wedge \mathrm{neg}^i(\overline{a}, \overline{y})\Big)$. This implies that for every tuple $\bar{b} \in M^\ell$ we have $\mathcal{M} \models \bigvee_{i \in I} \Big(\alpha^i(\overline{a}) \wedge \beta^i(\bar{b}) \wedge \mathrm{pos}^i(\overline{a}, \bar{b}) \wedge \mathrm{neg}^i(\overline{a}, \bar{b})\Big)$. Since $\{a_1, \dots, a_k\}^\ell \subseteq M^\ell$ we get that $\mathcal{M}[\{a_1, \dots, a_k\}]$ is a model of $\forall \overline{y} \bigvee_{i \in I} \Big(\alpha^i(\overline{a}) \wedge \beta^i(\overline{y}) \wedge \mathrm{pos}^i(\overline{a}, \overline{y}) \wedge \mathrm{neg}^i(\overline{a}, \overline{y})\Big)$. This means that $\mathcal{M}[\{a_1, \dots, a_k\}] \models \varphi$. Hence by definition, $\mathfrak{M}$ contains an induced substructure $\mathcal{M}'$ of $\mathcal{M}[\{a_1, \dots, a_k\}]$. But then $\mathcal{M}'$ is an induced substructure of $\mathcal{M}$ with strictly fewer elements than $\mathcal{M}$, a contradiction to the definition of $\mathfrak{M}$. ∎

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

Claim 5.2. *We have* $\varphi \equiv_d \bigvee_{\mathcal{M} \in \mathfrak{M}} \Big(\exists \overline{x} \forall \overline{y} \Big[\iota^{\mathcal{M}}(\overline{x}) \wedge \bigvee_{j \in J_{\mathcal{M}}} \Big(\beta^j(\overline{y}) \wedge \mathrm{pos}^j(\overline{x}, \overline{y}) \wedge \mathrm{neg}^j(\overline{x}, \overline{y})\Big)\Big]\Big)$.

Proof. Let $\mathcal{A} \in C_d$ be a model of φ. Then there is $\overline{a} = (a_1, \dots, a_k) \in A^k$ such that $\mathcal{A} \models \forall y \chi(\overline{a}, \overline{y})$. Since $\{a_1, \dots, a_k\}^\ell \subseteq A^\ell$ this implies that $\mathcal{A}[\{a_1, \dots, a_k\}] \models \forall y \chi(\overline{a}, \overline{y})$ and hence $\mathcal{A}[\{a_1, \dots, a_k\}] \models \varphi$. Furthermore we may assume that we picked $\overline{a}$ such 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 $\mathcal{A} \not\models \forall \overline{y} \, \chi(\overline{a}', \overline{y})$. (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 $\mathcal{A}[\{a_1, \dots, a_k\}]$ cannot have a proper induced substructure in $\mathfrak{M}$, and it follows that there is $\mathcal{M} \in \mathfrak{M}$ such that $\mathcal{M} \cong \mathcal{A}[\{a_1, \dots, a_k\}]$. By choice of $J_{\mathcal{M}}$ we get $\mathcal{A} \models \forall \overline{y} \Big[\iota^{\mathcal{M}}(\overline{a}) \wedge \bigvee_{j \in J_{\mathcal{M}}} \Big(\beta^j(\overline{y}) \wedge \mathrm{pos}^j(\overline{a}, \overline{y}) \wedge \mathrm{neg}^j(\overline{a}, \overline{y})\Big)\Big]$ and hence $\mathcal{A} \models \bigvee_{\mathcal{M} \in \mathfrak{M}} \Big(\exists \overline{x} \forall \overline{y} \Big[\iota^{\mathcal{M}}(\overline{x}) \wedge \bigvee_{j \in J_{\mathcal{M}}} \Big(\beta^j(\overline{y}) \wedge \mathrm{pos}^j(\overline{x}, \overline{y}) \wedge \mathrm{neg}^j(\overline{x}, \overline{y})\Big)\Big]\Big)$.

Now let $\mathcal{A} \in C_d$ be a model of $\bigvee_{\mathcal{M} \in \mathfrak{M}} \Big(\exists \overline{x} \forall \overline{y} \Big[\iota^{\mathcal{M}}(\overline{x}) \wedge \bigvee_{j \in J_{\mathcal{M}}} \Big(\beta^j(\overline{y}) \wedge \mathrm{pos}^j(\overline{x}, \overline{y}) \wedge \mathrm{neg}^j(\overline{x}, \overline{y})\Big)\Big]\Big)$. Consequently there is $\mathcal{M} \in \mathfrak{M}$ and $\overline{a} \in A^k$ such that $\mathcal{A} \models \forall \overline{y} \Big[\iota^{\mathcal{M}}(\overline{a}) \wedge \bigvee_{j \in J_{\mathcal{M}}} \Big(\beta^j(\overline{y}) \wedge \mathrm{pos}^j(\overline{a}, \overline{y}) \wedge \mathrm{neg}^j(\overline{a}, \overline{y})\Big)\Big]$. By choice of $J_{\mathcal{M}}$ this implies $\mathcal{A} \models \forall \overline{y} \bigvee_{j \in J_{\mathcal{M}}} \Big(\alpha^j(\overline{a}) \wedge \beta^j(\overline{y}) \wedge \mathrm{pos}^j(\overline{a}, \overline{y}) \wedge \mathrm{neg}^j(\overline{a}, \overline{y})\Big)$ and hence $\mathcal{A} \models \varphi$. ∎

Since the union of finitely many testable properties is testable (see e.g. [Gol17]), it suffices to show that P_φ where φ is of the form $\varphi = \exists \overline{x} \forall \overline{y} \, \chi(\overline{x}, \overline{y})$, where

$$\chi(\overline{x}, \overline{y}) = \iota^{\mathcal{M}}(\overline{x}) \wedge \bigvee_{j \in J_{\mathcal{M}}} \Big(\beta^j(\overline{y}) \wedge \mathrm{pos}^j(\overline{x}, \overline{y}) \wedge \mathrm{neg}^j(\overline{x}, \overline{y})\Big)$$

for some $\mathcal{M} \in \mathfrak{M}$ is testable. 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 $j \in J_{\mathcal{M}}$ let $\mathfrak{H}_j \subseteq C_d$ be a maximal set of pairwise non-isomorphic structures $\mathcal{H}$ such that $\mathcal{H} \models \beta^j(\bar{b})$ for some $\bar{b} = (b_1, \dots, b_\ell) \in H^\ell$ with $\{b_1, \dots, b_\ell\} = H$.

Claim 5.3. *The following holds.* $\varphi \equiv_d \exists \overline{x} \forall \overline{y} \Big[\iota^{\mathcal{M}}(\overline{x}) \wedge \bigvee_{\substack{\mathcal{H} \in \mathfrak{H}_j,\\ j \in J_{\mathcal{M}}}} \Big(\iota^{\mathcal{H}}(\overline{y}) \wedge \mathrm{pos}^j(\overline{x}, \overline{y}) \wedge \mathrm{neg}^j(\overline{x}, \overline{y})\Big)\Big]$.

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

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

Thus, it suffices to assume $\varphi = \exists \overline{x} \forall \overline{y} \, \chi(\overline{x}, \overline{y})$, where

$$\chi(\overline{x}, \overline{y}) := \Big[\iota^{\mathcal{M}}(\overline{x}) \wedge \bigvee_{\substack{\mathcal{H} \in \mathfrak{H}_j,\\ j \in J_{\mathcal{M}}}} \Big(\iota^{\mathcal{H}}(\overline{y}) \wedge \mathrm{pos}^j(\overline{x}, \overline{y}) \wedge \mathrm{neg}^j(\overline{x}, \overline{y})\Big)\Big] \tag{5.3}$$

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

Next we will define a universally quantified formula ψ and show that P_φ is indistinguishable from the property P_ψ. To do so we will need the two claims below. Intuitively, Claim 5.4 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 $\mathcal{A}$ and $\overline{a} \in A^k$, $\overline{b} = (b_1, \dots, b_\ell) \in A^\ell$ the condition $\mathcal{A} \models \iota^{\mathcal{H}}(\overline{b}) \wedge \mathrm{pos}^j(\overline{a}, \overline{b}) \wedge \mathrm{neg}^j(\overline{a}, \overline{b})$ can force an element of $\overline{b}$ to be in a tuple (of a relation of $\mathcal{A}$) with an element of $\overline{a}$, even if $\mathrm{pos}^j(\overline{x}, \overline{y})$ only contains literals of the form $x_i = y_{i'}$ (*e. g.* it may be that for some tuple $\overline{b}' \in \{b_1, \dots, b_\ell\}^\ell$, every clause $\iota^{\mathcal{H}'}(\overline{y}) \wedge \mathrm{pos}^{j'}(\overline{x}, \overline{y}) \wedge \mathrm{neg}^{j'}(\overline{x}, \overline{y})$ for which $\mathcal{A} \models \iota^{\mathcal{H}'}(\overline{b}') \wedge \mathrm{pos}^{j'}(\overline{a}, \overline{b}') \wedge \mathrm{neg}^{j'}(\overline{a}, \overline{b}')$ enforces 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 enforce 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 5.4 we show that for every $\mathcal{A} \in C_d$, $\overline{a} \in A^k$ with $\mathcal{A} \models \forall \overline{y}\, \chi(\overline{a}, \overline{y})$ there are only a constant number of tuples $\overline{b} \in A^\ell$ that only satisfy clauses which enforce a tuple to contain both an element from $\overline{a}$ and from $\overline{b}$.

Let $j \in \mathcal{M}$, $\mathcal{H} \in \mathfrak{H}_j$ and $\overline{h} = (h_1, \dots, h_\ell) \in H^\ell$ with $\mathcal{H} \models \iota^{\mathcal{H}}(\overline{h})$. Let $P_{j,\mathcal{H}} := \{h_i \mid i \in \{1, \dots, \ell\}, \mathrm{pos}^j(\overline{x}, \overline{y})$ does not contain $y_i = x_{i'}$ for any $i' \in \{1, \dots, k\}\}$. Now we let $J \subseteq J_{\mathcal{M}} \times C_d$ be the set of pairs $(j, \mathcal{H})$, with $\mathcal{H} \in \mathfrak{H}_j$ with the following two properties. Firstly $\mathrm{pos}^j(\overline{x}, \overline{y})$ only contains literals of the form $x_{i'} = y_i$ for some $i \in \{1, \dots, \ell\}$, $i' \in \{1, \dots, k\}$. Secondly the disjoint union $\mathcal{M} \sqcup \mathcal{H}[P_{j,\mathcal{H}}] \models \varphi$. J now precisely specifies the clauses that can be satisfied by a structure $\mathcal{A}$, $\overline{a} \in A^k$ and $\overline{b} \in A^\ell$ where $\mathcal{A}$ does not contain any tuples with both elements from $\overline{a}$ and $\overline{b}$.

Claim 5.4. *Let $\mathcal{A} \in C_d$ and $\overline{a} = (a_1, \dots, a_k) \in A^k$. If $\mathcal{A} \models \forall \overline{y}\, \chi(\overline{a}, \overline{y})$ then there are at most $k \cdot d$ tuples $\overline{b} \in A^\ell$ such that $\mathcal{A} \not\models \bigvee_{(j,\mathcal{H}) \in J} (\iota^{\mathcal{H}}(\overline{b}) \wedge \mathrm{pos}^j(\overline{a}, \overline{b}) \wedge \mathrm{neg}^j(\overline{a}, \overline{b}))$.*

Proof. Since $\mathcal{A} \models \forall \overline{y}\, \chi(\overline{a}, \overline{y})$, by Equation(5.3) it holds that $\mathcal{A} \models \forall \overline{y} \bigvee_{\substack{\mathcal{H} \in \mathfrak{H}_j, \\ j \in J_{\mathcal{M}}}} \Big(\iota^{\mathcal{H}}(\overline{y}) \wedge \mathrm{pos}^j(\overline{a}, \overline{y}) \wedge \mathrm{neg}^j(\overline{a}, \overline{y})\Big)$. Now let $B := \{\overline{b} \in A^\ell \mid \mathcal{A} \not\models \bigvee_{(j,\mathcal{H}) \in J} (\iota^{\mathcal{H}}(\overline{b}) \wedge \mathrm{pos}^j(\overline{a}, \overline{b}) \wedge \mathrm{neg}^j(\overline{a}, \overline{b}))\} \subseteq A^\ell$. Then every $\overline{b} \in B$ adds at least one to $\sum_{i=1}^k \deg_{\mathcal{A}}(a_i)$. Since $\mathcal{A} \in C_d$ implies that $\sum_{i=1}^k \deg_{\mathcal{A}}(a_i) \leq k \cdot d$ we get that $|B| \leq k \cdot d$. ∎

Claim 5.5. *Let ψ be a formula of the form $\psi = \forall \overline{z} \bigvee_{i \in I} c^i(\overline{z})$ where $\overline{z} = (z_1, \dots, z_t)$ is a tuple of variables and c^i is a conjunction of literals. Let $\mathcal{A} \in C_d$ with $|A| > d \cdot \mathrm{ar}(\sigma) \cdot t$ and let $b \in A$ be an arbitrary element. Let $\mathcal{A} \models \psi$ and let $\mathcal{A}'$ be obtained from $\mathcal{A}$ by 'isolating' b, i. e. by deleting all tuples containing b from $R^{\mathcal{A}}$ for every $R \in \sigma$. Then $\mathcal{A}' \models \psi$.*

Proof. First note that $\mathcal{A}' \models \bigvee_{i \in I} c^i(\overline{a})$ 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 $\overline{a} = (a_1, \dots, a_t) \in A^t$ be a tuple containing b. Pick $b' \in A$ such that $\mathrm{dist}_{\mathcal{A}}(a_j, b') > 1$ for every $j \in \{1, \dots, t\}$. Such an element exists as $|A| > d \cdot \mathrm{ar}(R) \cdot t$. 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 $\mathcal{A}'[\{a_1, \dots, a_t\}]$ to $\mathcal{A}[\{a'_1, \dots, a'_t\}]$ since b is an isolated element in $\mathcal{A}'[\{a_1, \dots, a_t\}]$ and b' is an isolated element in $\mathcal{A}[\{a'_1, \dots, a'_t\}]$. Since $\mathcal{A} \models \bigvee_{i \in I} c^i(\overline{a}')$, it follows that $\mathcal{A}' \models \bigvee_{i \in I} c^i(\overline{a})$. ∎

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

Claim 5.6. *The property P_φ with φ as in (5.3) is indistinguishable from the property P_ψ where $\psi := \forall \overline{y} \bigvee_{(j,\mathcal{H}) \in J'} \iota^{\mathcal{H}}(\overline{y})$.*

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

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

Let $\overline{a} \in A^k$ be a tuple such that $\mathcal{A} \models \forall \overline{y} \chi(\overline{a}, \overline{y})$. Let $B \subseteq A^\ell$ be the set of tuples $\overline{b} \in A^\ell$ such that $\mathcal{A} \not\models \bigvee_{(j,\mathcal{H}) \in J} (\iota^{\mathcal{H}}(\overline{b}) \wedge \mathrm{pos}^j(\overline{a}, \overline{b}) \wedge \mathrm{neg}^j(\overline{a}, \overline{b}))$. Then $|B| \leq \ell \cdot d \cdot \mathrm{ar}(R)$ by Claim 5.4. Hence the structure $\mathcal{A}'$ obtained from $\mathcal{A}$ by deleting all tuples containing an element of $C := \{a_1, \dots, a_k, b_1, \dots, b_\ell \mid (b_1, \dots, b_\ell) \in B\}$ is ϵ-close to $\mathcal{A}$. Since $\mathcal{A} \models \forall \overline{y} \chi(\overline{a}, \overline{y})$ implies $\mathcal{A} \models \forall \overline{y} \bigvee_{\substack{\mathcal{H} \in \mathfrak{H}_j, \\ j \in J_{\mathcal{M}}}} \iota^{\mathcal{H}}(\overline{y})$ by Claim 5.5 we know that $\mathcal{A}' \models \forall \overline{y} \bigvee_{\substack{\mathcal{H} \in \mathfrak{H}_j, \\ j \in J_{\mathcal{M}}}} \iota^{\mathcal{H}}(\overline{y})$. For any tuple $\overline{b} = (b_1, \dots, b_\ell) \in (A \setminus C)^\ell$ we have by definition of J' that $\mathcal{A} \models \iota^{\mathcal{H}}(\overline{b})$ for some $(j, \mathcal{H}) \in J'$. Furthermore $\mathcal{A}[\{b_1, \dots, b_\ell\}] = \mathcal{A}'[\{b_1, \dots, b_\ell\}]$ and hence $\mathcal{A}' \models \bigvee_{(j,\mathcal{H}) \in J'} \iota^{\mathcal{H}}(\overline{b})$. Let $\overline{b} = (b_1, \dots, b_\ell) \in A^\ell$ be any tuple containing element from C and let $c_1, \dots, c_t \in C$ be those elements. Pick t elements $c'_1, \dots, c'_t \in A \setminus C$ such that $\mathrm{dist}_{\mathcal{A}}(a_i, c'_{i'}) > 1$ and $\mathrm{dist}_{\mathcal{A}}(c'_{i'}, b_i) > 1$ for suitable i, i'. This is possible as $|A| > (k + 2\ell) \cdot d \cdot \mathrm{ar}(R)$ which guarantees the existence of $k + 2\ell$ elements of pairwise distance 1. Let $\overline{b}' = (b'_1, \dots, b'_\ell)$ be the vector obtained from $\overline{b}$ by replacing c_i with c'_i. Since $\overline{b}' \in A^\ell$ there must be j', $\mathcal{H}' \in \mathfrak{H}_j$ such that $\mathcal{A} \models \iota^{\mathcal{H}'}(\overline{b}') \wedge \mathrm{pos}^{j'}(\overline{a}, \overline{b}') \wedge \mathrm{neg}^{j'}(\overline{a}, \overline{b}')$. 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 $(j', \mathcal{H}') \in J'$. Since additionally $b_i \mapsto b'_i$ defines an isomorphism of $\mathcal{A}[\{b'_1, \dots, b'_\ell\}]$ and $\mathcal{A}'[\{b_1, \dots, b_\ell\}]$ this implies that $\mathcal{A}' \models \bigvee_{(j,\mathcal{H}) \in J'} \iota^{\mathcal{H}}(\overline{b})$

for all $\bar{b} \in A^\ell$ and hence $\mathcal{A}' \models \psi$.

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

Take any set $B \subseteq A$ of $|M|$ elements. Let $\mathcal{A}'$ be the structure obtained from $\mathcal{A}$ by deleting all edges incident to any element contained in B. Let $\mathcal{A}''$ be the structure obtained from $\mathcal{A}'$ by adding all tuples such that the structure induced by B is isomorphic to $\mathcal{M}$. This takes no more then $2\ell \cdot d \cdot \mathrm{ar}(R) < \epsilon \cdot d \cdot |A|$ edge modifications Let $\bar{a} \in B^k$ be such that $\mathcal{A} \models \iota^{\mathcal{M}}(\bar{a})$. By Claim 5.5 we get $\mathcal{A}' \models \psi$. Therefore pick any $\bar{b} = (b_1, \dots, b_\ell) \in (A \setminus B)^\ell$. Since by construction we have that all b_i's are of distance ≥ 1 from $\bar{a}$ we have that $\mathcal{A}'' \models \bigvee_{(j,\mathcal{H}) \in J'} (\iota^{\mathcal{H}}(\bar{b}) \wedge \mathrm{neg}^j(\bar{a}, \bar{b}))$. By choice of $\mathcal{M}$ we get $\mathcal{A}'' \models \bigvee_{\substack{\mathcal{H} \in \mathfrak{H}_j, \\ j \in J_{\mathcal{M}}}} \Big(\iota^{\mathcal{H}}(\bar{b}) \wedge \mathrm{pos}^j(\bar{a}, \bar{b}) \wedge \mathrm{neg}^j(\bar{a}, \bar{b})\Big)$ for all $\bar{b} \in B^\ell$. Therefore pick $\bar{b} = (b_1, \dots, b_\ell)$ containing both elements from B and from $A \setminus B$. Now pick $\bar{b}' = (b'_1, \dots, b'_\ell) \in (A \setminus B)^\ell$ that equals $\bar{b}$ in all positions containing an element from $A \setminus B$. As noted before there is $(j, \mathcal{H}) \in J'$ with $\mathcal{A}'' \models (\iota^{\mathcal{H}}(\bar{b}') \wedge \mathrm{neg}^j(\bar{a}, \bar{b}'))$. By definition of J, J' this implies $\mathcal{A}''[\{a_1, \dots, a_k, b'_1 \dots b'_\ell\}] \models \varphi$. As $\bar{b} \in \{a_1, \dots, a_k, b'_1 \dots b'_\ell\}^\ell$ we get $\mathcal{A}''[\{a_1, \dots, a_k, b'_1 \dots b'_\ell\}] \models \bigvee_{\substack{\mathcal{H} \in \mathfrak{H}_j, \\ j \in J_{\mathcal{M}}}} \Big(\iota^{\mathcal{H}}(\bar{b}) \wedge \mathrm{pos}^j(\bar{a}, \bar{b}) \wedge \mathrm{neg}^j(\bar{a}, \bar{b})\Big)$. Then $\mathcal{A}'' \models \bigvee_{\substack{\mathcal{H} \in \mathfrak{H}_j, \\ j \in J_{\mathcal{M}}}} \Big(\iota^{\mathcal{H}}(\bar{b}) \wedge \mathrm{pos}^j(\bar{a}, \bar{b}) \wedge \mathrm{neg}^j(\bar{a}, \bar{b})\Big)$ and hence $\mathcal{A}'' \models \varphi$. ∎

Since $\psi \in \Pi_1$ we have that P_ψ is testable, and hence P_φ is testable by Claim 5.6. □

6 Testing properties of neighbourhoods

In this section we only consider simple graphs, i. e. undirected graphs without self-loops and without parallel edges, and for any $d \in \mathbb{N}$ let C_d be the class of simple graphs of bounded degree d. We view simple graphs as structures over the signature $\sigma_{\mathrm{graph}} := \{E\}$, where E encodes a binary, symmetric and irreflexive relation. This allows transferring the notions from Section 2 to graphs.

Let $r \geq 1$ and let τ be an r-type and let $\varphi_\tau(x)$ be a FO formula saying that x has r-type τ. We say that a graph G is *τ-neighbourhood regular*, if $G \models \forall x \varphi_\tau(x)$. We say that a graph G is *τ-neighbourhood free*, if $G \models \neg\exists x \varphi_\tau(x)$. Let $\tau_1, \dots, \tau_t$ be a list of all r-types in C_d. If $F \subseteq \{\tau_1, \dots, \tau_t\}$ we say that G is F-free, if G is τ-neighbourhood free for all $\tau \in F$.

Observe that both τ-neighbourhood-freeness and τ-neighbourhood regularity can be defined by formulas in Π_2 for any neighbourhood type τ. Hence the next Lemma shows that there exist neighbourhood properties that are in Π_2, but not in Σ_2. See the full version of the paper for a proof.

Lemma 6.1. *There exist* 1*-types τ, τ' such that neither τ-neighbourhood freeness nor τ'-neighbourhood regularity can be defined by a formula in Σ_2.*

Note that the above lemma implies that we cannot simply invoke the testers for testing Σ_2 properties from Theorem 5.1 to test these two properties.

Now we state our main algorithmic results in this section. The first result shows that if τ is an r-type with degree smaller than the degree bound of the class of graphs, then the τ-neighbourhood-freeness is testable.

Theorem 6.1. *Let τ be an r-type, where $r \geq 1$. If $\tau \subseteq C_{d'}$ and $d' < d$, then τ-neighbourhood freeness is uniformly testable on the class C_d with constant running time.*

The second result shows if τ is a 1-type, then τ-neighbourhood-freeness is testable.

Theorem 6.2. *For every* 1*-type τ, τ-neighbourhood freeness is uniformly testable on the class C_d with constant time.*

The third result says that τ-neighbourhood regularity is testable for every 1-type τ consisting of cliques, which only overlap in the centre vertex.

Theorem 6.3. *Let τ be a* 1*-type such that vertex a having* 1*-type τ in B implies that $B \setminus \{a\}$ is a union of disjoint cliques for every* 1*-ball B with centre a. Then τ-neighbourhood regularity is uniformly testable on C_d in constant time.*

By previous discussions, the above theorems imply that there are formulas in $\Pi_2 \setminus \Sigma_2$ which are testable. For the proofs of Theorems 6.2, 6.2 and 6.3, we refer the reader to the full version of the paper.

Acknowledgements. We thank Sebastian Ordyniak for inspiring discussions. The second author thanks Michał Pilipczuk and Pierre Simon for inspiring discussions at the docCourse on Sparsity in Prague 2018.

References

[AFKS00] Noga Alon, Eldar Fischer, Michael Krivelevich, and Mario Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000. Preliminary version in FOCS'99.

[AFNS09] Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira. A combinatorial characterization of the testable graph properties: it's all about regularity. *SIAM Journal on Computing*, 39(1):143–167, 2009.

[AH18] Isolde Adler and Frederik 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*, pages 6:1–6:14, 2018.

[AM85] Noga Alon and Vitali D Milman. $\lambda 1$, isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38(1):73–88, 1985.

[BK12] Benedikt Bollig and Dietrich Kuske. An optimal construction of hanf sentences. *Journal of Applied Logic*, 10(2):179–186, 2012.

[BSS10] Itai Benjamini, Oded Schramm, and Asaf Shapira. Every minor-closed property of sparse graphs is testable. *Advances in mathematics*, 223(6):2200–2218, 2010.

[CPS16] Artur Czumaj, Pan Peng, and Christian 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*, pages 1033–1045, 2016.

[CY19] Hubie Chen and Yuichi 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*, pages 365–382, 2019.

[DG07] Arnaud Durand and Etienne Grandjean. First-order queries on structures of bounded degree are computable with constant delay. *ACM Trans. Comput. Log.*, 8(4):21, 2007.

[Dod84] Jozef Dodziuk. Difference equations, isoperimetric inequality and transience of certain random walks. *Transactions of the American Mathematical Society*, 284(2):787–794, 1984.

[EF95] Heinz-Dieter Ebbinghaus and Jörg Flum. *Finite model theory*. Perspectives in Mathematical Logic. Springer, 1995.

[FNS+20] Sebastian Forster, Danupon Nanongkai, Thatchaphol Saranurak, Liu Yang, and Sorrachai Yingchareonthawornchai. Computing and testing small connectivity in near-linear time and queries via fast local cut algorithms. *SODA*, 2020.

[FPS19] Hendrik Fichtenberger, Pan Peng, and Christian 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*, pages 714–726, 2019.

[Gai82] Haim Gaifman. On local and non-local properties, 1982.

[GGR98] Oded Goldreich, Shari Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM (JACM)*, 45(4):653–750, 1998.

[Gol17] Oded Goldreich. *Introduction to property testing*. Cambridge University Press, 2017.

[GR02] Oded Goldreich and Dana Ron. Property testing in bounded degree graphs. *Algorithmica*, 32(2):302–343, 2002.

[GR11] Oded Goldreich and Dana Ron. On proximity-oblivious testing. *SIAM Journal on Computing*, 40(2):534–566, 2011.

[Han65] William Hanf. *The Theory of Models*, chapter Model-theoretic methods in the study of elementary logic, pages 132–145. North Holland, 1965.

[HKNO09] Avinatan Hassidim, Jonathan A Kelner, Huy N Nguyen, and Krzysztof Onak. Local graph partitions for approximation and testing. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 22–31. IEEE, 2009.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *BULL. AMER. MATH. SOC.*, 43(4):439–561, 2006.

[IKN19] Hiro Ito, Areej Khoury, and Ilan 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.

[KSS19] Akash Kumar, C Seshadhri, and Andrew 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*, pages 559–567, 2019.

[KY13] Ken-ichi Kawarabayashi and Yuichi Yoshida. Testing subdivision-freeness: property testing meets structural graph theory. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 437–446. ACM, 2013.

[Lov12] László Lovász. *Large Networks and Graph Limits*, volume 60 of *Colloquium Publications*. American Mathematical Society, 2012.

[NS13] Ilan Newman and Christian Sohler. Every property of hyperfinite graphs is testable. *SIAM Journal on Computing*, 42(3):1095–1112, 2013.

[RFV95] L. Stockmeyer R. Fagin and M. Vardi. On monadic np vs. monadic co-np. *Information and Computation*, 120(1):78–92, 1995.

[RS96] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.

[RVW02] Omer Reingold, Salil Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Annals of mathematics*, pages 157–187, 2002.

[YI12] Yuichi Yoshida and Hiro Ito. Property testing on k-vertex-connectivity of graphs. *Algorithmica*, 62(3-4):701–712, 2012.