



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/212370/>

Version: Accepted Version

Article:

Adler, I., Koehler, N. and Peng, P. (2024) On Testability of First-Order Properties in Bounded-Degree Graphs and Connections to Proximity-Oblivious Testing. *SIAM Journal on Computing (SICOMP)*, 53 (4). 825- 883. ISSN: 0097-5397

<https://doi.org/10.1137/23M1556253>

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.

1 **ON TESTABILITY OF FIRST-ORDER PROPERTIES IN**
2 **BOUNDED-DEGREE GRAPHS AND CONNECTIONS TO**
3 **PROXIMITY-OBVIOUS TESTING ***

4 ISOLDE ADLER[†], NOLEEN KÖHLER[‡], AND PAN PENG[§]

5 **Abstract.** We study property testing of properties that are definable in first-order logic (FO)
6 in the bounded-degree graph and relational structure models. We show that any FO property that
7 is defined by a formula with quantifier prefix $\exists^*\forall^*$ is testable (i.e., testable with constant query
8 complexity), while there exists an FO property that is expressible by a formula with quantifier prefix
9 $\forall^*\exists^*$ that is not testable. In the dense graph model, a similar picture is long known (Alon, Fischer,
10 Krivelevich, Szegedy, *Combinatorica* 2000), despite the very different nature of the two models. In
11 particular, we obtain our lower bound by an FO formula that defines a class of bounded-degree
12 expanders, based on zig-zag products of graphs. We expect this to be of independent interest.

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

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

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

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

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

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

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

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

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

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

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

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

66 1.1. Our contributions.

67 **1.1.1. Non-testability of first-order logic.** We study the testability of prop-
 68 erties definable in first-order logic (FO) in the bounded-degree graph model. Recall
 69 that formulas of first-order logic on graphs are built from predicates for the edge re-
 70 lation and equality, using Boolean connectives \vee, \wedge, \neg and universal and existential
 71 quantifiers \forall, \exists , where the variables represent graph vertices. First-order logic can e. g.
 72 express subgraph-freeness (i. e., no isomorphic copy of some fixed graph H appears
 73 as a subgraph) and subgraph containment (i. e., an isomorphic copy of some fixed H
 74 appears as a subgraph). Note however, that there are constant-query testable prop-
 75 erties, such as connectivity and cycle-freeness, that cannot be expressed in FO. We
 76 study the question of which first-order properties are testable in the bounded-degree
 77 graph model. Our study extends to the bounded-degree *relational structure* model
 78 [1], while we focus on the classes of relational structures with binary relations, i.e.,
 79 edge-coloured directed graphs. In this model for relational structures, one can perform
 80 neighbour queries, querying for both in- and out-neighbours and the edge colour that
 81 connects them. This model is natural in the context of relational databases, where
 82 each (edge-)relation is given by a list of the tuples it contains.

83 We consider the testability of first-order properties in the bounded-degree model
 84 according to quantifier alternation, inspired by a similar study for dense graphs by
 85 Alon et al. [4]. On relational structures of bounded-degree over a fixed finite sig-
 86 nature, we have the following simple observation: Any first-order property definable

87 by a sentence *without* quantifier alternations is testable. This means the sentence
 88 either consists of a quantifier prefix of the form \exists^* (any finite number of existential
 89 quantifications), followed by a quantifier-free formula, or it consists of a quantifier
 90 prefix of the form \forall^* (any finite number of universal quantifications), followed by a
 91 quantifier-free formula. Basically, every property of the form \exists^* is testable because
 92 the structure required by the quantifier-free part of the formula can be planted with
 93 a small number of tuple modifications if the input structure is large enough (depend-
 94 ing on the formula), and we can use an exact algorithm to determine the answer in
 95 constant time otherwise. Every property of the form \forall^* is testable because a formula
 96 of the form $\forall \bar{x} \varphi(\bar{x})$, where φ is quantifier-free, is logically equivalent to a formula
 97 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
 98 testing for the absence of a finite number of induced substructures, which can be done
 99 similarly to testing subgraph freeness [21]. The testability of a property becomes less
 100 clear if it is defined by a sentence *with* quantifier alternations. Formally, we let Π_2
 101 (resp. Σ_2) denote the set of properties that can be expressed by a formula in the
 102 $\forall^* \exists^*$ -prefix (resp. $\exists^* \forall^*$ -prefix) class. We obtain the following.

103 *Every first-order property in Σ_2 is testable in the bounded-degree model (Theorem*
 104 *6.1). On the other hand, there is a first-order property in Π_2 , that is not testable in*
 105 *the bounded-degree model (Theorem 4.7).*

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

112 We remark that our lower bound, i.e., the existence of a property in Π_2 that is
 113 not testable, is somewhat astonishing (on an intuitive level) due to the following two
 114 reasons. Firstly, it is proven by constructing a first-order definable class of structures
 115 that encode a class of expander graphs, which highlights that FO is surprisingly
 116 expressive on bounded-degree graphs, despite its locality [24, 17, 32]. Secondly, it
 117 is known that property testing algorithms in the bounded-degree model proceed by
 118 sampling vertices from the input graph and exploring their local neighbourhoods,
 119 and FO can only express ‘local’ properties, while our lower bound shows that this
 120 is not sufficient for testability. We elaborate on this in the following. On one hand,
 121 Hanf’s Theorem [24] gives insight into first-order logic on graphs of bounded-degree
 122 and implies a strong normal form, called *Hanf Normal Form* (HNF) in [9], which
 123 we briefly sketch. For a graph G of maximum degree d and a vertex x in G , the
 124 neighbourhood of fixed radius r around x in G can be described by a first-order
 125 formula $\tau_r(x)$, up to isomorphism. A *Hanf sentence* is a first-order sentence of the
 126 form ‘there are at least ℓ vertices x of neighbourhood (isomorphism) type $\tau_r(x)$ ’.
 127 A first-order sentence is in HNF, if it is a Boolean combination of Hanf sentences.
 128 By Hanf’s Theorem, every first-order sentence is equivalent to a sentence in HNF
 129 on bounded-degree graphs [24, 32, 14]. Note that Hanf sentences only speak about
 130 local neighbourhoods. Hence this theorem gives evidence that first-order logic can
 131 only express local properties. On the other hand, if a property is constant-query
 132 testable in the bounded-degree graph model, then it can be tested by approximating
 133 the distribution of local neighbourhoods (see [11] and [22]). That is, a constant-
 134 query tester can essentially only test properties that are close to being defined by a

135 distribution of local neighbourhoods. For these reasons¹, a priori, it could be true that
 136 every property that can be expressed in first-order logic is testable in the bounded-
 137 degree model. Indeed, the validity of this statement was raised as an open question
 138 in [1]. However, our lower bound gives a negative answer to this question.

139 **1.1.2. GSF-locality is not sufficient for proximity oblivious testing.**

140 Typical property testers make decisions regarding the global property of the graph
 141 based on local views only. In the extreme case, a tester could make the size of the
 142 local views independent of the distance ε to a predetermined set of graphs. Motivated
 143 by this, Goldreich and Ron [22] initiated the study of (one-sided error) *proximity-*
 144 *oblivious testers (POTs)* for graphs, where a tester simply repeats a basic test for a
 145 number of times that depends on the proximity parameter, while the basic tester is
 146 oblivious of the proximity parameter. They gave characterizations of graph properties
 147 that can be tested with constant query complexity by a POT in both dense graph
 148 model and the bounded-degree model. In each model, it is known that the class of
 149 properties that have constant-query POTs is a strict subset of the class of properties
 150 that are testable (by standard testers).

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

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

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

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

179 We complete the picture by showing the following.

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

180 *Every GSF-local property of graphs of degree at most 2 is non-propagating (The-*
 181 *orem 7.16).*

182 1.2. Our techniques.

183 **1.2.1. On the testability and non-testability of FO properties.** For show-
 184 ing that every property \mathcal{P} defined by a formula φ in Σ_2 (i.e. of the form $\exists^*\forall^*$) is
 185 testable, we show that \mathcal{P} is equivalent to the union of properties \mathcal{P}_i , each of which is
 186 ‘indistinguishable’ from a property \mathcal{Q}_i that is defined by a formula of form \forall^* . Here
 187 the indistinguishability means we can transform any structure satisfying \mathcal{P}_i , into a
 188 structure satisfying \mathcal{Q}_i by modifying a small fraction of the tuples of the structure
 189 and vice versa. This allows us to reduce the problem of testing \mathcal{P} to testing properties
 190 defined by \forall^* formulas. Then the testability of \mathcal{P} follows, as any property of the form
 191 \forall^* is testable and testable properties are closed under union [18]. The main challenge
 192 here is to deal with the interactions between existentially quantified variables and
 193 universally quantified variables. Intuitively, the degree bound limits the structure
 194 that can be imposed by the universally quantified variables. Using this, we are able
 195 to deal with the existential variables together with these interactions by ‘planting’ a
 196 required constant size substructure in such a way, that we are only a constant number
 197 of modifications ‘away’ from a formula of the form \forall^* .

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

220 Now we construct a formula φ , that defines a class of relational structures with
 221 binary relations only whose underlying undirected graphs are expander graphs, arising
 222 from the zig-zag product by Reingold, Vadhan and Wigderson [33]. For expressibil-
 223 ity in FO, we hybridise the zig-zag construction of expanders with a tree structure.
 224 Roughly speaking, we start with a small graph H , which is a good expander, and the
 225 formula φ expresses that each model² looks like a rooted k -ary tree (for a suitable

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

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

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

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

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

1.3. Other related work. Besides the aforementioned works on testing properties with constant query complexity in the bounded-degree graph model, Goldreich and Ron [22] have obtained a characterisation for a class of properties that are testable by a constant-query proximity-oblivious tester in bounded-degree graphs (and dense graphs). Such a class is a rather restricted subset of the class of all constant-query testable properties. Fichtenberger et al. [15] showed that every testable property is

275 either finite or contains an infinite hyperfinite subproperty. Informally, a hyperfinite
 276 subproperty is a subset of graphs that can be partitioned into small connected compo-
 277 nents by removing a small fraction of the edges and is invariant under isomorphism.
 278 Ito et al. [27] gave characterisations of one-sided error (constant-query) testable mono-
 279 tone graph properties, and one-sided error testable hereditary graph properties in the
 280 bounded-degree (directed and undirected) graph model.

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

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

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

293 **1.4. Structure of the paper.** Section 2 contains the preliminaries, including
 294 logic, property testing and the zig-zag construction of expander graphs. In Section 3
 295 we construct the FO formula φ and prove properties of its models. In Section 4, we
 296 prove that there is a Π_2 -property that is not testable, by proving that the property
 297 defined by φ on bounded-degree structures is not constant-query testable. Using a
 298 reduction, in Section 5 we then provide a Π_2 -property of undirected graphs of degree at
 299 most 3 that is non-testable. In Section 6, we show that all Σ_2 properties are testable.
 300 In Section 7 we then turn to POTs, showing that our Π_2 -property of undirected graphs
 301 of degree at most 3 is GSF-local and propagating. We then show that all GSF-local
 302 properties of degree at most 2 are non-propagating. We conclude in Section 8.

303 **2. Preliminaries.** We let \mathbb{N} denote the set of natural numbers including 0, and
 304 $\mathbb{N}_{>0} := \mathbb{N} \setminus \{0\}$. For $n \in \mathbb{N}$ we let $[n] := \{0, 1, \dots, n - 1\}$ denote the set of the first n
 305 natural numbers. For a set S and $k \in \mathbb{N}$ we denote the Cartesian product $S \times \dots \times S$
 306 of k copies of S by S^k . We use $\binom{S}{2}$ to denote the set of all two-element subsets of S ,
 307 we denote the disjoint union of sets by \sqcup and the symmetric difference by Δ .

308 **2.1. Undirected graphs.** Unless otherwise specified we allow graphs to have
 309 self-loops and parallel edges. We represent an undirected graph G as a triple (V, E, f) ,
 310 where V is the set of vertices, E is the set of edges and $f : E \rightarrow V \cup \binom{V}{2}$ is the
 311 incidence map. An isomorphism from $G_1 = (V_1, E_1, f_1)$ to $G_2 = (V_2, E_2, f_2)$ is a
 312 pair of bijective maps (h_V, h_E) , where $h_V : V_1 \rightarrow V_2$ and $h_E : E_1 \rightarrow E_2$, such that
 313 $h_V(f_1(e)) = f_2(h_E(e))$ for any $e \in E_1$, where $h_V(X) := \{h_V(x) \mid x \in X\}$ for any set
 314 $X \subseteq V_1$. Undirected graphs without self-loops and parallel edges are called *simple*.
 315 For a simple graph G , we also represent G as a tuple $G = (V(G), E(G))$, where $V(G)$
 316 is the vertex set and $E(G) \subseteq \binom{V(G)}{2}$. The *degree* $\deg_G(v)$ of a vertex v in a graph G
 317 is the number of edges to which v is incident. In particular, self-loops contribute one to
 318 the degree. We will say that a graph G is *d-regular* for some $d \in \mathbb{N}$ if every vertex
 319 in G has degree d . We specify paths in graphs by tuples of vertices. We further let
 320 all paths and cycles be simple, i.e. no vertex appears twice. The *length* of a path
 321 on n vertices is $n - 1$. We define the distance between two vertices v and w in a
 322 graph G , denoted $\text{dist}_G(v, w)$, as the length of a shortest path from v to w or ∞ if

323 there is no path from v to w in G . Any subset $S \subseteq V$ of vertices *induces* a graph
 324 $G[S] := (S, \{e \in E \mid f(e) \in S \cup \binom{S}{2}\}, f|_S)$. A *connected component* of G is a graph
 325 induced by a maximal set S , such that each pair $v, w \in S$ has finite distance in G .
 326 A graph is connected if it has only one connected component. We refer the reader
 327 to [12] for the basic notions of graph theory.

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

334 **2.2. Relational structures and first-order logic.** We will briefly introduce
 335 structures and first-order logic and point the reader to [14] for a more detailed intro-
 336 duction. A (relational) *signature* is a finite set $\sigma = \{R_1, \dots, R_\ell\}$ of relation symbols
 337 R_i . Every relation symbol R_i , $1 \leq i \leq \ell$ has an arity $\text{ar}(R_i) \in \mathbb{N}_{>0}$. A σ -*structure*
 338 is a tuple $A = (U(A), R_1(A), \dots, R_\ell(A))$, where $U(A)$ is a *finite* set, called the *uni-*
 339 *verse* of A and $R_i(A) \subseteq U(A)^{\text{ar}(R_i)}$ is an $\text{ar}(R_i)$ -ary relation on $U(A)$. Note that
 340 if $\sigma = \{E_1, \dots, E_\ell\}$ is a signature where each E_i is a binary relation symbol, then
 341 σ -structures are directed simple graphs with ℓ edge-colours. Let $\sigma_{\text{graph}} := \{E\}$ be a
 342 signature with one binary relation symbol E . Then we can understand undirected
 343 simple graphs as σ_{graph} -structures for which the relation E is symmetric (every undi-
 344 rected edge is represented by two tuples) and irreflexive. Using this we can transfer all
 345 notions defined below to simple graphs. Typically we name graphs G, H, F , we denote
 346 the set of vertices of a graph G by $V(G)$, the set of edges by $E(G)$ and vertices are typi-
 347 cally named $u, v, w, u', v', w', \dots$. In contrast when we talk about a general relational
 348 structure we use A, B and a, b, a', b', \dots to denote elements from the universe.

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

362 The syntax and semantics of FO logic are defined in the usual way (see *e. g.* [14]).
 363 We use $\exists^{\geq m} x \varphi$ (and $\exists^=m x \varphi$, $\exists^{\leq m} x \varphi$, respectively) as a shortcut for the FO formula
 364 expressing that the number of witnesses x satisfying φ is at least m (exactly m , at
 365 most m , respectively). We say that a variable occurs *freely* in an FO formula if at
 366 least one of its occurrences is not bound by any quantifier. We use $\varphi(x_1, \dots, x_k)$ to
 367 express that the set of variables which occur freely in the FO formula φ is a subset of
 368 $\{x_1, \dots, x_k\}$. For a formula $\varphi(x_1, \dots, x_k)$, a σ -structure A and $a_1, \dots, a_k \in U(A)$ we
 369 write $A \models \varphi(a_1, \dots, a_k)$ if φ evaluates to true after assigning a_i to x_i , for $1 \leq i \leq k$.
 370 A *sentence* of FO is a formula with no free variables. For an FO sentence φ we say
 371 that A is a *model* of φ or A satisfies φ if $A \models \varphi$. Let \mathcal{C} be a class of σ -structures

372 closed under isomorphism. Every FO-sentence φ over σ defines a property $\mathcal{P}_\varphi \subseteq \mathcal{C}$ on
 373 \mathcal{C} , where $\mathcal{P}_\varphi := \{A \in \mathcal{C} \mid A \models \varphi\}$.

374 *Hanf normal form.* The *Gaifman graph* of a σ -structure A is the undirected graph
 375 $G(A) = (U(A), E)$, where $\{v, w\} \in E$, if $v \neq w$ and there is an $R \in \sigma$ and a tuple
 376 $\bar{a} = (a_1, \dots, a_{\text{ar}(R)}) \in R(A)$, such that $v = a_j$ and $w = a_k$ for some $1 \leq k, j \leq \text{ar}(R)$.
 377 We use $G(A)$ to apply graph theoretic notions to relational structures. Note that for
 378 any simple graph the Gaifman graph of the corresponding symmetric σ_{graph} -structure
 379 is the graph itself. We say that a σ -structure A is *connected* if its Gaifman graph $G(A)$
 380 is connected. For two elements $a, b \in U(A)$, we define the *distance* between a and b
 381 in A , denoted by $\text{dist}_A(a, b)$, as the length of a shortest path from a to b in $G(A)$, or
 382 ∞ if there is no such path. For $r \in \mathbb{N}$ and $a \in U(A)$, the *r -neighbourhood* of a is the
 383 set $N_r^A(a) := \{b \in U(A) : \text{dist}_A(a, b) \leq r\}$. We define $\mathcal{N}_r^A(a) := A[N_r^A(a)]$ to be the
 384 substructure of A induced by the r -neighbourhood of a . For $r \in \mathbb{N}$ an *r -ball* is a tuple
 385 (B, b) , where B is a σ -structure, $b \in U(B)$ and $U(B) = N_r^B(b)$, i. e. B has radius r
 386 and b is the centre. Note that by definition $(\mathcal{N}_r^A(a), a)$ is an r -ball for any σ -structure
 387 A and $a \in U(A)$. Two r -balls $(B, b), (B', b')$ are isomorphic if there is an isomorphism
 388 of σ -structure from B to B' that maps b to b' . We call the isomorphism classes of
 389 r -balls *r -types*. For an r -type τ and an element $a \in U(A)$ we say that a *has* (r -)type
 390 τ if $(\mathcal{N}_r^A(a), a) \in \tau$. Moreover, given such an r -type τ , there is a formula $\varphi_\tau(x)$ such
 391 that for every σ -structure A and for every $a \in U(A)$, $A \models \varphi_\tau(a)$ iff $(\mathcal{N}_r^A(a), a) \in \tau$.
 392 A *Hanf-sentence* is a sentence of the form $\exists^{\geq m} x \varphi_\tau(x)$, for some $m \in \mathbb{N}_{>0}$, where τ is
 393 an r -type. An FO sentence is in *Hanf normal form*, if it is a Boolean combination³
 394 of Hanf sentences. Two formulas $\varphi(x_1, \dots, x_k)$ and $\psi(x_1, \dots, x_k)$ of signature σ are
 395 called *d -equivalent*, denoted by $\varphi(x_1, \dots, x_k) \equiv_d \psi(x_1, \dots, x_k)$, if they are equivalent
 396 on $\mathcal{C}_{\sigma, d}$, i. e. for all $A \in \mathcal{C}_{\sigma, d}$ and all $(a_1, \dots, a_k) \in U(A)^k$ we have $A \models \varphi(a_1, \dots, a_k)$
 397 iff $A \models \psi(a_1, \dots, a_k)$. Hanf's locality theorem for first-order logic [24] implies the
 398 following.

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

401 *Quantifier alternations of first-order formulas.* Let σ be any relational signature.
 402 We use the following recursive definition, classifying first-order formulas according to
 403 the number of quantifier alterations in their quantifier prefix. Let $\Sigma_0 = \Pi_0$ be the
 404 class of all quantifier free first-order formulas over σ . Then for every $i \in \mathbb{N}_{>0}$ we let
 405 Σ_i be the set of all FO formulas $\varphi(y_1, \dots, y_\ell)$ for which there is $k \in \mathbb{N}$ and a formula
 406 $\psi(x_1, \dots, x_k, y_1, \dots, y_\ell) \in \Pi_{i-1}$ such that

$$407 \quad \varphi \equiv \exists x_1 \dots \exists x_k \psi(x_1, \dots, x_k, y_1, \dots, y_\ell).$$

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

$$410 \quad \varphi \equiv \forall x_1 \dots \forall x_k \psi(x_1, \dots, x_k, y_1, \dots, y_\ell).$$

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

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

$$415 \quad \mathcal{P} := \{A \in \mathcal{C}_{\sigma, d} \mid A \text{ does not contain } B \text{ as substructure}\}$$

416 *is in Π_1 .*

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

417 **2.3. Property testing.** In the following, we give definitions of two models for
 418 property testing - the bounded-degree model for simple graphs introduced in [21] and
 419 a bounded-degree model for relational structures similar to the model introduced in [1].
 420 The model for relational structures described here is chosen to simplify notation. It
 421 differs from the model in [1] in the way the query access is defined, however, they are
 422 equivalent in the sense that testability in either model implies testability in the other
 423 model. This can be easily seen using a local reduction as defined in Section 5.2. The
 424 bounded-degree model for relational structures extends the bounded-degree model for
 425 undirected graphs introduced in [21] and conforms with the bidirectional model of
 426 [11].

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

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

427 For $\epsilon \in (0, 1)$ we say that a structure A on n vertices/elements is ϵ -close to \mathcal{P} if
 428 $\text{dist}(A, \mathcal{P}) \leq \epsilon$, that is one can modify A into a structure in \mathcal{P} by adding/deleting at
 429 most ϵdn tuples of A . We say that A is ϵ -far from \mathcal{P} if A is not ϵ -close to \mathcal{P} .

430 An algorithm that processes a structure $A \in \mathcal{C}$ does not obtain an encoding of A
 431 as a bit string in the usual way. Instead, we assume that the algorithm receives the
 432 number n of elements/vertices of A , and that the elements/vertices of A are numbered
 433 $1, 2, \dots, n$. In addition, the algorithm has direct access to A using an *oracle* which
 434 answers *neighbour queries* in A in constant time. A *query* to a σ -structure A of
 435 bounded-degree d has the form (a, i) for an element $a \in U(A)$, $i \in \{1, \dots, d\}$ and
 436 is answered by $\text{ans}(a, i) := (R, a_1, \dots, a_{\text{ar}(R)})$ where $(a_1, \dots, a_{\text{ar}(R)})$ is the i -th tuple
 437 (according to some fixed ordering) containing a and $(a_1, \dots, a_{\text{ar}(R)}) \in R(A)$, or a
 438 special symbol “ \perp ” if i is greater than the degree of a . A *query* to a graph G of
 439 bounded-degree d has the form (v, i) for $v \in V(G)$, $i \in \{1, \dots, d\}$ and is answered by
 440 $\text{ans}(v, i) := w$ where w is the i -th neighbour of v .

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

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

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

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

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

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

- 457 • *accepts* A with probability 1, if $A \in \mathcal{P}_n$.
- 458 • *rejects* A with probability at least $\eta(\text{dist}(A, \mathcal{P}_n))$, if $A \notin \mathcal{P}_n$. Equivalently,
- 459 for any A that is not in \mathcal{P}_n , the algorithm accepts A with probability at most
- 460 $1 - \eta(\text{dist}(A, \mathcal{P}_n))$.

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

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

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

466 **2.4. Expansion and the zig-zag product.** In this section we recall a con-
 467 struction of a class of expanders introduced in [33]. This construction uses undirected
 468 graphs with parallel edges and self-loops.

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

$$473 \quad |\{(i, j) \in I \times I \mid \text{ROT}_G(u, i) = (v, j)\}| = 2|\{e \in E \mid f(e) = \{u, v\}\}|$$

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

$$478 \quad M_{u,v} := \frac{1}{D} \cdot |\{e \mid f(e) = \{u, v\}\}|.$$

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

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

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

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

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

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

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

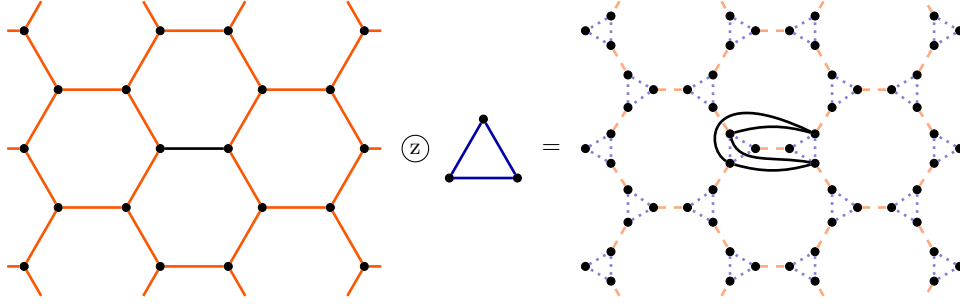


Fig. 1: Zig-zag product of a 3-regular grid with a triangle

500 DEFINITION 2.8. Let G be a D -regular graph on N vertices with rotation map
 501 $\text{ROT}_G : V \times I \rightarrow V \times I$ and I a set of size D . Then the square of G , denoted by
 502 G^2 , is a D^2 -regular graph on N vertices with rotation map $\text{ROT}_{G^2}(u, (k_1, k_2)) :=$
 503 $(w, (\ell_2, \ell_1))$, where

$$504 \quad \text{ROT}_G(u, k_1) = (v, \ell_1) \text{ and } \text{ROT}_G(v, k_2) = (w, \ell_2),$$

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

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

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

510 DEFINITION 2.10. Let $G_1 = (V_1, E_1, f_1)$ be a D_1 -regular graph on N_1 vertices,
 511 I_1 a set of size D_1 and $\text{ROT}_{G_1} : V_1 \times I_1 \rightarrow V_1 \times I_1$ a rotation map of G_1 . Let
 512 $G_2 = (I_1, E_2, f_2)$ be a D_2 -regular graph, let I_2 be a set of size D_2 and $\text{ROT}_{G_2} :$
 513 $I_1 \times I_2 \rightarrow I_1 \times I_2$ be a rotation map of G_2 . Then the zig-zag product of G_1 and G_2 ,
 514 denoted by $G_1 \textcircled{Z} G_2$, is the D_2^2 -regular graph on vertex set $V_1 \times I_1$ with rotation map
 515 given by $\text{ROT}_{G_1 \textcircled{Z} G_2}((v, k), (i, j)) := ((w, \ell), (j', i'))$, where

$$516 \quad \text{ROT}_{G_2}(k, i) = (k', i'), \text{ROT}_{G_1}(v, k') = (w, \ell'), \text{ and } \text{ROT}_{G_2}(\ell', j) = (\ell, j'),$$

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

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

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

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

$$531 \quad g(\lambda_1, \lambda_2) = \frac{1}{2}(1 - \lambda_2^2)\lambda_1 + \frac{1}{2}\sqrt{(1 - \lambda_2^2)^2\lambda_1 + 4\lambda_2^2}.$$

532 This function has the following properties.

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

535 DEFINITION 2.12 ([26]). Let D be a sufficiently large prime power (e.g. $D =$
 536 2^{16}). Let H be a $(D^4, D, 1/4)$ expander (an explicit constructions for H exist, cf. [33]).
 537 We define $\{G_m\}_{m \in \mathbb{N}_{>0}}$ by

$$538 \quad (2.1) \quad G_1 := H^2, \quad G_m := G_{m-1}^2 \otimes H \text{ for } m > 1.$$

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

541 In the next section we will use the following lemma.

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

544 *Proof.* Let $1 = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ be the eigenvalues of $G^2[S]$. Since $G^2[S]$ is
 545 connected, Lemma 2.5 implies that $\lambda_1 > \lambda_2$. Now assume that -1 is an eigenvalue
 546 of $G^2[S]$ with eigenvector \bar{v} . Then the vector \bar{v}' defined by $\bar{v}'_v = \bar{v}_v$ for all $v \in S$ and
 547 $\bar{v}'_v = 0$ otherwise is the eigenvector for eigenvalue -1 of the graph G^2 . But G^2 can
 548 not have a negative eigenvalue as every eigenvalue of G^2 is a square of a real number.
 549 Therefore $\lambda_1 \neq \lambda_N$ and $\lambda(G^2[S]) < 1$ as claimed. \square

550 **3. A class of expanders definable in FO.** In this section we define a formula
 551 such that the underlying graphs of its models are expanders. We start with a high-
 552 level description of the formula. Let $\{G_m\}_{m \in \mathbb{N}_{>0}}$ be as in Definition 2.12. Loosely
 553 speaking, each model of our formula is a structure which consists of the disjoint union
 554 of G_1, \dots, G_n for some $n \in \mathbb{N}_{>0}$ with some underlying tree structure connecting G_{m-1}
 555 to G_m for all $m \in \{2, \dots, n\}$. For illustration see Figure 2. The tree structure enables
 556 us to provide an FO-checkable certificate for the construction of expanders. The tree
 557 structure is a D^4 -ary tree, that is used to connect a vertex v of G_{m-1} to every vertex
 558 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
 559 enforce an ordering on the D^4 children of each vertex. We use additional relations to
 560 encode rotation maps. For $i, j \in [D]^2$ let $E_{i,j}$ be a binary relation. For every pair
 561 $i, j \in [D]^2$ we represent an edge $\{v, w\}$ in G_m by the two tuples $(v, w) \in E_{i,j}(A)$ and
 562 $(w, v) \in E_{j,i}(A)$. This allows us to encode the relationship $\text{ROT}_{G_m}(v, i) = (w, j)$ in
 563 first-order logic using the formula ' $E_{i,j}(v, w)$ '.

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

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

$$569 \quad \sigma := \{ \{E_{i,j}\}_{i,j \in [D]^2}, \{F_k\}_{k \in ([D]^2)^2}, R, \{L_k\}_{k \in ([D]^2)^2} \},$$

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

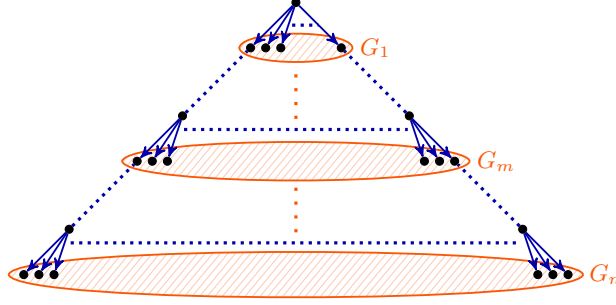


Fig. 2: Schematic representation of a model of $\varphi_{\mathbb{Z}}$, where the parts in red (grey) only contain relations from E and relations in F are blue (black). Relation R and L are omitted.

573 In any formula we can reverse using these auxiliary relations by replacing formulas
 574 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
 575 “ $\bigvee_{k \in ([D]^2)^2} F_k(x, y)$ ” below.

576 We use the following formula $\varphi_{\text{root}}(x) := \forall y \neg F(y, x)$ and we say that an element
 577 $a \in U(A)$ is a root of a structure A if $A \models \varphi_{\text{root}}(a)$.

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

$$\begin{aligned}
 583 \quad \varphi_{\text{tree}} := & \exists^{\leq 1} x \varphi_{\text{root}}(x) \wedge \forall x \left((\varphi_{\text{root}}(x) \wedge R(x, x)) \vee \right. \\
 584 & \left. (\exists^{\leq 1} y F(y, x) \wedge \neg \exists y R(x, y) \wedge \neg \exists y R(y, x)) \right) \wedge \\
 585 & \forall x \left(\left[\neg \exists y F(x, y) \wedge \bigwedge_{k \in ([D]^2)^2} L_k(x, x) \wedge \right. \right. \\
 586 & \left. \left. \forall y (y \neq x \rightarrow \bigwedge_{k \in ([D]^2)^2} \neg L_k(x, y) \wedge \bigwedge_{k \in ([D]^2)^2} \neg L_k(y, x)) \right] \right. \\
 587 & \left. \vee \left[\neg \exists y \bigvee_{k \in ([D]^2)^2} (L_k(x, y) \vee L_k(y, x)) \wedge \bigwedge_{k \in ([D]^2)^2} \exists y_k (x \neq y_k \wedge F_k(x, y_k) \wedge \right. \right. \\
 588 & \left. \left. (\bigwedge_{k' \in ([D]^2)^2, k' \neq k} \neg F_{k'}(x, y_k)) \wedge \forall y (y \neq y_k \rightarrow \neg F_k(x, y)) \right] \right].
 \end{aligned}$$

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

$$594 \quad \varphi_{\text{rotationMap}} := \forall x \forall y \left(\bigwedge_{i,j \in [D]^2} (E_{i,j}(x, y) \rightarrow E_{j,i}(y, x)) \right) \wedge$$

$$595 \quad \forall x \left(\bigwedge_{i \in [D]^2} \left(\bigvee_{j \in [D]^2} (\exists^{-1} y E_{i,j}(x, y) \wedge \bigwedge_{\substack{j' \in [D]^2 \\ j' \neq j}} \neg \exists y E_{i,j'}(x, y)) \right) \right)$$

596 We now define a formula φ_{base} which expresses that every root x of a structure
 597 has a self-loop (x, x) in each relation $E_{i,j}$ and that the D^4 children of a root form
 598 G_1 . Let H be the $(D^4, D, 1/4)$ -graph from Definition 2.12. We assume that H has
 599 vertex set $([D]^2)^2$. We then identify vertex $k \in ([D]^2)^2$ with the element y such that
 600 $(x, y) \in F_k(A)$ for each root x . Let $\text{ROT}_H : ([D]^2)^2 \times [D] \rightarrow ([D]^2)^2 \times [D]$ be any
 601 rotation map of H . Fixing a rotation map for H fixes the rotation map for H^2 . Recall
 602 that $G_1 := H^2$. We can define G_1 by a conjunction over all edges of G_1 .

$$603 \quad \varphi_{\text{base}} := \forall x \left(\varphi_{\text{root}}(x) \rightarrow \left[\bigwedge_{i,j \in [D]^2} \left(E_{i,j}(x, x) \wedge \right. \right. \right. \\ 604 \quad \left. \left. \left. \forall y \left(x \neq y \rightarrow (\neg E_{i,j}(x, y) \wedge \neg E_{i,j}(y, x)) \right) \right) \right] \wedge \right. \\ 605 \quad \left. \bigwedge_{\substack{\text{ROT}_{H^2}(k,i)=(k',i') \\ k,k' \in ([D]^2)^2 \\ i,i' \in [D]^2}} \exists y \exists y' \left(F_k(x, y) \wedge F_{k'}(x, y') \wedge E_{i,i'}(y, y') \right) \right]$$

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

$$613 \quad \varphi_{\text{recursion}} := \forall x \forall z \left[\left(\neg \exists y F(x, y) \wedge \neg \exists y F(z, y) \right) \vee \right. \\ 614 \quad \left. \bigwedge_{\substack{k'_1, k'_2 \in [D]^2 \\ \ell'_1, \ell'_2 \in [D]^2}} \left(\exists y \left[E_{k'_1, \ell'_1}(x, y) \wedge E_{k'_2, \ell'_2}(y, z) \right] \rightarrow \right. \right. \\ 615 \quad \left. \left. \bigwedge_{\substack{i,j,i',j' \in [D], k, \ell \in ([D]^2)^2 \\ \text{ROT}_H(k,i)=(k'_1, k'_2), i' \\ \text{ROT}_H((\ell'_2, \ell'_1), j)=(\ell, j')}} \exists x' \exists z' \left[F_k(x, x') \wedge F_\ell(z, z') \wedge E_{(i,j),(j',i')}(x', z') \right] \right) \right]$$

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

618 **3.1. Proving expansion.** In this section we prove that the formula φ_{\otimes} defines
 619 a property of expanders on bounded-degree relational structures.

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

624 To each model A of φ_{\otimes} we will associate an undirected (with parallel edges
 625 and self loops) graph $\underline{G}(A)$ with vertex set $U(A)$. For every tuple in each of the
 626 relations of A , the graph $\underline{G}(A)$ will have an edge. We will define $\underline{G}(A)$ by a rotation

627 map, which extends the rotation map encoded by the relation E . For this let $I :=$
 628 $\{0\} \sqcup ([D]^2)^2 \sqcup [D]^2$ be an index set. Formally, we define the *underlying graph* $\underline{G}(A)$
 629 of a model A of $\varphi_{\mathbb{Z}}$ to be the undirected graph with vertex set $U(A)$ given by the
 630 rotation map $\text{ROT}_{\underline{G}(A)} : A \times I \rightarrow A \times I$ defined by

$$631 \quad \text{ROT}_{\underline{G}(A)}(v, i) := \begin{cases} (v, 0) & \text{if } i = 0 \text{ and } (v, v) \in R(A) \\ (w, j) & \text{if } i = 0 \text{ and } (w, v) \in F_j(A) \\ (w, 0) & \text{if } i \in ([D]^2)^2 \text{ and } (v, w) \in F_i(A) \\ (v, i) & \text{if } i \in ([D]^2)^2 \text{ and } (v, v) \in L_i(A) \\ (w, j) & \text{if } i \in [D]^2 \text{ and } (v, w) \in E_{i,j}(A). \end{cases}$$

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

639 **THEOREM 3.1.** *There is an $\epsilon > 0$ such that the class $\{\underline{G}(A) \mid A \models \varphi_{\mathbb{Z}}\}$ is a*
 640 *family of ϵ -expanders.*

641 In the rest of this section, we give the proof of Theorem 3.1. Let A be a model of
 642 $\varphi_{\mathbb{Z}}$. Let $A|_F$ be the $\{(F_k)_{k \in ([D]^2)^2}\}$ -structure $(U(A), (F_k(A))_{k \in ([D]^2)^2})$. Recall that
 643 we denote the Gaifman graph of $A|_F$ by $G(A|_F)$. Let $A|_E$ be the $\{(E_{i,j})_{i,j \in [D]^2}\}$ -
 644 structure $(U(A), (E_{i,j}(A))_{i,j \in [D]^2})$. We further define the *underlying graph* $\underline{G}(A|_E)$
 645 of $A|_E$ as the undirected graph specified by the rotation map $\text{ROT}_{\underline{G}(A|_E)}$ which is
 646 defined by $\text{ROT}_{\underline{G}(A|_E)}(v, i) := (w, j)$ if $(v, w) \in E_{i,j}(A)$. This is well defined as
 647 $A \models \varphi_{\text{rotationMap}}$. We use the substructures $G(A|_F)$ and $\underline{G}(A|_E)$ to express the
 648 structural properties of models of $\varphi_{\mathbb{Z}}$. More precisely, we want to prove that $G(A|_F)$
 649 is a rooted balanced full tree and $\underline{G}(A|_E)$ is the disjoint union of the expanders
 650 G_1, \dots, G_n for some $n \in \mathbb{N}$ (Lemma 3.10). To prove this we use two technical lemmas
 651 (Lemma 3.2 and Lemma 3.5). Lemma 3.2 intuitively shows that the children in
 652 $G(A|_F)$ of each connected part of $\underline{G}(A|_E)$ form the zig-zag product with H of the
 653 square of the connected part. Lemma 3.5 shows that $G(A|_F)$ is connected. To prove
 654 Theorem 3.1 we use that a tree with an expander on each level has good expansion.
 655 Loosely speaking, this is true because cutting the tree ‘horizontally’ takes many edge
 656 deletions and for cutting the tree ‘vertically’ we cut many expanders.

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

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

664 *Proof of Lemma 3.2.* Assume that $S' \neq \emptyset$. We first show that $\underline{G}(A|_E)[S'] \cong$
 665 $((\underline{G}(A|_E))^2[S']) \otimes H$. For this we use the following two claims.

CLAIM 3.3. *If*

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

666 *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*
 667 $(u, v) \in E_{k'_1, \ell'_1}(A)$ *and* $(v, w) \in E_{k'_2, \ell'_2}(A)$ *where* $\text{ROT}_H(k, i) = ((k'_1, k'_2), i')$ *and*
 668 $\text{ROT}_H((\ell'_2, \ell'_1), j) = (\ell, j')$.

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

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

676 CLAIM 3.4. *If* $(u, v) \in E_{k'_1, \ell'_1}(A)$ *and* $(v, w) \in E_{k'_2, \ell'_2}(A)$ *for* $u, v, w \in U(A)$,
 677 $k'_1, k'_2, \ell'_1, \ell'_2 \in ([D]^2)^2$ *and there is* $u' \in U(A)$ *with* $(u, u') \in F(A)$ *then there is*
 678 $w' \in U(A)$ *such that* $(w, w') \in F(A)$. *Furthermore for any* $i, i', j, j' \in [D]$ *there are*
 679 $\tilde{u}, \tilde{w} \in U(A)$, $k, \ell \in ([D]^2)^2$ *such that* $(\tilde{u}, \tilde{w}) \in E_{(i, j), (j', i')}(A)$ *for* $(u, \tilde{u}) \in F_k(A)$ *and*
 680 $(w, \tilde{w}) \in F_\ell(A)$ *where* $\text{ROT}_H(k, i) = ((k'_1, k'_2), i')$ *and* $\text{ROT}_H((\ell'_2, \ell'_1), j) = (\ell, j')$.

681 *Proof.* We only use that $A \models \varphi_{\text{recursion}}$. Since $\varphi_{\text{recursion}}$ has the form $\forall x \forall z \psi(x, z)$
 682 for some formula $\psi(x, z)$ we know that $A \models \psi(u, w)$. Since $(u, u') \in F(A)$ we have
 683 $A \not\models \neg \exists y F(u, y) \wedge \neg \exists y F(w, y)$. Since $A \models \exists y [E_{k'_1, \ell'_1}(u, y) \wedge E_{k'_2, \ell'_2}(w, z)]$

$$684 \quad A \models \bigwedge_{\substack{i, j, i', j' \in [D], k, \ell \in ([D]^2)^2 \\ \text{ROT}_H(k, i) = ((k'_1, k'_2), i') \\ \text{ROT}_H((\ell'_2, \ell'_1), j) = (\ell, j')}} \exists x' \exists z' [F_k(u, x') \wedge F_\ell(w, z') \wedge E_{(i, j), (j', i')}(x', z')]$$

685 Since H is D -regular, for every $k'_1, k'_2 \in [D]^2$ and $i, i' \in [D]$, there is $k \in ([D]^2)^2$
 686 such that $\text{ROT}_H(k, i) = ((k'_1, k'_2), i')$ (and the same for ℓ'_1, ℓ'_2, j, j'). Thus, the above
 687 conjunction is not empty. This further implies that for any $i, i', j, j' \in [D]$ there are
 688 $\tilde{u}, \tilde{w} \in U(A)$, $k, \ell \in ([D]^2)^2$ as claimed. In particular there is $w' \in U(A)$ such that
 689 $(w, w') \in F(A)$. \square

690 We will argue that for every element $w \in S$ there is a $w' \in S'$ such that
 691 $(w, w') \in F(A)$. For this pick any $u' \in S'$. Let $u \in S$ be the element such that
 692 $(u, u') \in F(A)$. By combining Lemma 2.14, Theorem 2.11 and Lemma 2.5 it follows
 693 that $((\underline{G}(A|_E))^2[S]) \otimes_H$ is connected. Therefore, there exists a path (u'_0, \dots, u'_m) in
 694 $((\underline{G}(A|_E))^2[S]) \otimes_H$ from $u'_0 = (u, (k_1, k_2))$ to $u'_m = (w, (\ell_1, \ell_2))$ for some k_1, k_2, ℓ_1, ℓ_2
 695 $\in [D]^2$. By Claim 3.3 there is a path $(u_0, v_0, u_1, v_1, \dots, u_{m-1}, v_{m-1}, u_m)$ in $\underline{G}(A|_E)$
 696 from $u_0 = u$ to $u_m = w$. By inductively using Claim 3.4 on the path we find w' such
 697 that $(w, w') \in F(A)$.

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

$$701 \quad \text{ROT}_{(\underline{G}(A|_E))[S']} (f((u, k)), (i, j)) = (f((w, \ell)), (j', i')).$$

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

705 Moreover, $\underline{G}(A|_E)[S'] \cong ((\underline{G}(A|_E))^2[S']) \otimes H$ implies that $\underline{G}(A|_E)[S']$ is connected
 706 and D^2 -regular. Since $A \models \varphi_{\text{rotationMap}}$ enforces that $\underline{G}(A|_E)$ is D^2 -regular, no vertex
 707 $v \in S'$ can have neighbours which are not in S' and therefore $\underline{G}(A|_E)[S']$ is a connected
 708 component of $\underline{G}(A|_E)$. \square

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

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

715 COROLLARY 3.6. *Any model $A \in \mathcal{C}_{\sigma,d}$ of $\varphi_{\mathbb{Z}}$ is connected.*

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

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

727 *Proof.* We first identify a cycle (or a pair as mentioned in the statement) by
 728 traversing along the path of incoming edges until encountering a repeated vertex. Let
 729 v_0 be any vertex in G' and let $S_0 = \{v_0\}$. We will now recursively define v_i to be
 730 the vertex of G' such that $(v_i, v_{i-1}) \in F(A)$. Such a vertex always exists by the
 731 choice of G' (i.e. that no root is in G') and the fact that $A \models \varphi_{\text{tree}}$. Furthermore,
 732 such a vertex is unique as $A \models \varphi_{\text{tree}}$. We let $S_i := S_{i-1} \cup \{v_i\}$. Since $U(A)$ is
 733 finite, the chain $S_0 \subseteq S_1 \subseteq \dots \subseteq S_i \subseteq \dots$ must become stationary at some point.
 734 Let $i \in \mathbb{N}$ be the minimum index such that $S_{i-1} = S_i$ and let $j < i$ be such that
 735 $v_i = v_j$. Then $(v_j, v_{j+1}, \dots, v_{i-1}, v_i)$ is either a cycle in G' or (in case $j = i-1$)
 736 $(v_j, v_i), (v_i, v_j) \in F(A)$. Let $C = \{c_0, \dots, c_{\ell-1}\}$ be the vertices of the cycle or pair of
 737 vertices.

738 We now show that for every vertex v in G' there exists a unique path from v to C .
 739 We first note that since G' is connected, for every v , a path that satisfies the property
 740 as described in the assertion of the claim always exists. Assume that there exists one
 741 vertex v , from which there are two different paths to C , denoted by $(p_0 = v, \dots, p_m)$
 742 and $(p'_0 = v, \dots, p'_{m'})$, respectively. We let $p_m = c_i$ and $p'_{m'} = c_j$. Let $k \leq \min\{m, m'\}$
 743 be the minimum index such that $p_k \neq p'_k$. Such an index must exist as the paths are
 744 different, and as $p_0 = p'_0 = v$, we also know that $k \geq 1$. Since $A \models \varphi_{\text{tree}}$ for every
 745 vertex w of G' there can only be one vertex w' of G' such that $(w', w) \in F(A)$. As
 746 $p_{m-1} \notin C$ and $(c_{(i-1) \bmod \ell}, p_m) \in F(A)$ it follows that $(p_m, p_{m-1}) \in F(A)$. Applying
 747 the argument inductively we get that $(p_k, p_{k-1}) \in F(A)$. The same argument works
 748 for the path $(p'_0, \dots, p'_{m'})$ and therefore $(p'_k, p'_{k-1}) \in F(A)$. By the choice of k we
 749 know that $p_{k-1} = p'_{k-1}$ and $p_k \neq p'_k$, which implies that there exists one vertex with
 750 two incoming edges. This contradicts the fact that $A \models \varphi_{\text{tree}}$. Thus, for every vertex
 751 v , there exists a unique path from v to C . This finishes the proof of the claim. \square

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

754 We now recursively define the infinite sequence of sets $S_i := \{w \in U(A) \mid (v, w) \in$
 755 $F(A), v \in S_{i-1}\}$ for each $i \in \mathbb{N}_{>0}$. Let $m_i := \max_{v \in S_i \cap V} \min_{j \in \{0, \dots, \ell-1\}} \{\text{dist}_{G'}(c_j, v)\}$
 756 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
 757 well defined as $c_{i \bmod \ell} \in S_i$.

758 CLAIM 3.8. $\underline{G}(A|_E)[S_i] = (\underline{G}(A|_E)[S_{i-1}])^2 \otimes H$.

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

762 $\underline{G}(A|_E)[S_0]$ is a connected component of $\underline{G}(A|_E)$ by choice of S_0 . Let $\tilde{S} := \{w \in$
 763 $U(A) \mid (w, v) \in F(A), v \in S_0\}$.

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

776 Thus, by Lemma 2.14 $\lambda((\underline{G}(A|_E))^2[\tilde{S}]) < 1$. But by Lemma 3.2 $\underline{G}(A|_E)[S_0] =$
 777 $((\underline{G}(A|_E))^2[\tilde{S}]) \otimes H$. Then Theorem 2.11 and $\lambda(H) < 1$ ensure that $\lambda(\underline{G}(A|_E)[S_0]) <$
 778 1 .

779 For $i > 1$, by induction it holds that $\lambda(\underline{G}(A|_E)[S_{i-1}]) < 1$, which, together
 780 with Lemma 2.9 and Lemma 2.5, implies that $(\underline{G}(A|_E)[S_{i-1}])^2$ is a connected compo-
 781 nent⁴ of $(\underline{G}(A|_E))^2$ and that $(\underline{G}(A|_E))^2[S_{i-1}] = (\underline{G}(A|_E)[S_{i-1}])^2$. Since $c_{i \bmod \ell} \in S_i$,
 782 by Lemma 3.2, we have that $\underline{G}(A|_E)[S_i]$ is a connected component of $\underline{G}(A|_E)$ and
 783 $\underline{G}(A|_E)[S_i] = (\underline{G}(A|_E)[S_{i-1}])^2 \otimes H$. Furthermore, using Lemma 2.9 and Theorem
 784 2.11, this proves $\lambda(\underline{G}(A|_E)[S_i]) < 1$. \square

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

786 *Proof.* By Claim 3.8 we have that $\underline{G}(A|_E)[S_{i+1}] = (\underline{G}(A|_E)[S_i])^2 \otimes H$. This
 787 means that by definition of squaring and the zig-zag product we know that $|S_{i+1}| =$
 788 $D^4 \cdot |S_i|$. But as in addition $A \models \varphi_{\text{tree}}$ we know that every element $v \in S_i$ will
 789 contribute to no more than D^4 elements to S_{i+1} . This means by construction of S_{i+1}
 790 that for every element in S_i there must be $w \in V$ such that $(v, w) \in F(A)$. \square

791 Therefore, for every $i \in \mathbb{N}_{>0}$ there is $w_i \in V$ such that $(v_i, w_i) \in F(A)$ where v_i is the
 792 vertex of distance m_i from C in G' picked above. Let (u_0, \dots, u_{m_i}) be the path in G'
 793 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
 794 the path (u_0, \dots, u_{m_i}) , we have that $(u_{j+1}, u_j) \in F(A)$ for all $j \in [m_i]$. Furthermore,
 795 since $v_i = u_0 \neq u_1$, assuming that $w_i = u_1$ would imply $(v_i, u_1), (u_2, u_1) \in F(A)$,
 796 which contradicts $A \models \varphi_{\text{tree}}$. Then $(w_i, u_0, \dots, u_{m_i})$ is a path in G' from w_i to C .

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

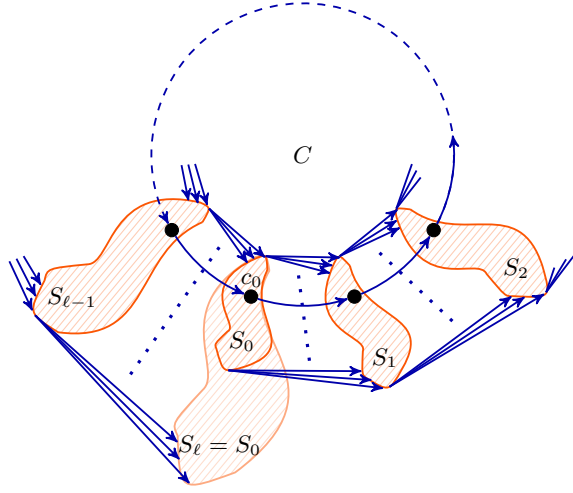


Fig. 3: Illustration of the proof of Lemma 3.5.

797 Since $w_i \in S_{i+1}$ by construction, Claim 3.7 implies that $m_{i+1} \geq m_i + 1$. Therefore
 798 $m_i \geq i + m_0$ inductively. But this yields a contradiction, because $\ell + m_0 \leq m_\ell = m_0$
 799 and $\ell > 0$. See Figure 3 for an illustration. Therefore every connected component of
 800 $G(A|_F)$ must contain a root of A . Furthermore, since every connected component of
 801 $G(A|_F)$ must contain a root and since $A \models \exists^{\leq 1} x \varphi_{\text{root}}(x)$ there can not be more than
 802 one root, $G(A|_F)$ is connected. \square

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

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

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

811 *Proof.* First note that the empty structure $A_\emptyset \in \mathcal{P}_{\mathbb{Z}}$ as $A_\emptyset \models \exists^{\leq 1} x \varphi_{\text{root}}(x)$ and
 812 therefore $A_\emptyset \models \varphi_{\mathbb{Z}}$ as $\varphi_{\mathbb{Z}}$ is a conjunction of $\exists^{\leq 1} x \varphi_{\text{root}}(x)$ and universally quantified
 813 formulas. Hence $U(A) = \emptyset$ is possible. Now assume that A is a model of $\varphi_{\mathbb{Z}}$ and
 814 $U(A) \neq \emptyset$. Then Lemma 3.5 implies that $G(A|_F)$ is connected. Combining this with
 815 $A \models \varphi_{\text{tree}}$ proves that $G(A|_F)$ is a rooted tree. Let n be the maximum distance of
 816 any vertex in $G(A|_F)$ to the root and let T_m be the vertices of distance m to the
 817 root for $m \leq n$. Then $\underline{G}(A|_E)[T_1] \cong G_1$ because $A \models \varphi_{\text{base}}$. Now assume towards an
 818 inductive proof that $\underline{G}(A|_E)[T_m] \cong G_m$ for some fixed $m \in \mathbb{N}_{>0}$. Since $\lambda(G_m) < 1$
 819 by Lemma 2.9 and Lemma 2.5 we get that $(\underline{G}(A|_E))^2[T_m]$ is a connected component
 820 of $(\underline{G}(A|_E))^2$. Hence by Lemma 3.2 we get that $\underline{G}(A|_E)[T_{m+1}] \cong G_{m+1}$. Since G_m
 821 has D^{4m} vertices this also proves that A has $\sum_{m=0}^n D^{4m}$ vertices. Furthermore, for
 822 $n \in \mathbb{N}$ the existence of a model of $\varphi_{\mathbb{Z}}$ of size $\sum_{m=0}^n D^{4m}$ is straightforward by the

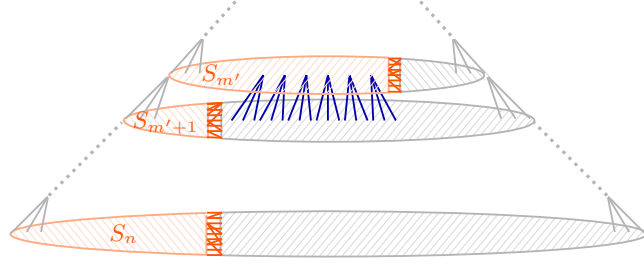


Fig. 4: Schematic representation of S crossing edges (orange and blue) in the underlying undirected graph in the case of $m' < n$.

823 construction of the formula $\varphi_{\mathbb{Z}}$. □

824 Now we are ready to prove Theorem 3.1.

825 *Proof of Theorem 3.1.* We will prove that for $\epsilon = D^2/12$ the claimed is true. Let
 826 A be the model of $\varphi_{\mathbb{Z}}$ of size $\sum_{m=0}^n D^{4m}$ and $S \subseteq U(A)$ with $|S| \leq (\sum_{m=0}^n D^{4m})/2$.
 827 Let T_m be the vertices of distance m to the root of the tree $G(A|_F)$ and let $S_m :=$
 828 $T_m \cap S$.

829 We can assume that $|S| > 1$ as every vertex has degree at least ϵ . Let us first
 830 assume that $|S_m| \leq D^{4m}/2$ for all $m \in [n]$. Then because G_m is a $D^2/4$ -expander
 831 (this follows directly from Theorem 2.7 as $\lambda(G_m) \leq 1/2$ by Proposition 2.13) and
 832 $\underline{G}(A|_E)[T_m] \cong G_m$ we know that

$$833 \quad \langle S, \bar{S} \rangle_{\underline{G}(A)} \geq \sum_{m=1}^n \frac{D^2}{4} |S_m| \geq \frac{D^2}{12} \sum_{m=0}^n |S_m| = \frac{D^2}{12} |S|.$$

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

$$835 \quad (3.1) \quad |S_{m'}| > \frac{|T_{m'}|}{2} = \frac{D^{4m'}}{2}.$$

836 We will use the following claim.

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

838 *Proof.* Inductively, we argue that

$$839 \quad \sum_{m=0}^{\tilde{m}-1} |T_m| = \sum_{m=0}^{\tilde{m}-2} |T_m| + |T_{\tilde{m}-1}| \leq \frac{1}{2} (3|T_{\tilde{m}-1}|) \leq \frac{1}{2} |T_{\tilde{m}}|. \quad \square$$

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

$$842 \quad \langle S, \bar{S} \rangle_{\underline{G}(A)} \geq \frac{D^2}{4} (|T_n| - |S_n|) \geq \frac{D^2}{16} |T_n| \geq \frac{D^2}{12} |S|.$$

843 Assume now that $m' < n$. Since S is the disjoint union of all S_m we know that the
 844 set $\langle S, \bar{S} \rangle_{\underline{G}(A)}$ contains the set $\langle S_m, T_m \setminus S_m \rangle_{\underline{G}(A)}$, and for all $m \in \{m'+1, \dots, n\}$, the
 845 sets $\langle T_{m'} \setminus S_{m'}, T_{m'} \rangle_{\underline{G}(A)}$ and $\langle S_{m'}, T_{m'+1} \setminus S_{m'+1} \rangle_{\underline{G}(A)}$, which are all pairwise disjoint.
 846 Since every vertex in $T_{m'}$ has D^4 neighbours in $T_{m'+1}$ and on the other hand every

847 vertex in $T_{m'+1}$ has one neighbour in $T_{m'}$ we know that $|\langle S_{m'}, T_{m'+1} \setminus S_{m'+1} \rangle_{\underline{G}(A)}| =$
 848 $|\langle S_{m'}, T_{m'+1} \rangle_{\underline{G}(A)}| - |\langle S_{m'}, S_{m'+1} \rangle_{\underline{G}(A)}| \geq D^4 |S_{m'}| - |S_{m'+1}| \geq D^4 (|S_{m'}| - D^{4m'}/2).$
 849 Since additionally G_m is an $D^2/4$ -expander for every m we get

$$\begin{aligned}
 850 \quad & |\langle S, \bar{S} \rangle_{\underline{G}(A)}| \geq \sum_{m > m'} \frac{D^2}{4} |S_m| + \frac{D^2}{4} |T_{m'} \setminus S_{m'}| + D^4 \left(|S_{m'}| - \frac{D^{4m'}}{2} \right) \\
 851 \quad & = \frac{D^2}{4} \sum_{m > m'} |S_m| + \left(D^4 - \frac{D^2}{2} \right) |S_{m'}| - \left(D^4 - \frac{D^2}{2} \right) \frac{|T_{m'}|}{2} + \frac{D^2}{8} |S_{m'}| + \frac{D^2}{8} |S_{m'}| \\
 852 \quad & \stackrel{\text{Equation 3.1}}{\geq} \frac{D^2}{4} \sum_{m > m'} |S_m| + \frac{D^2}{8} |S_{m'}| + \frac{D^2}{8} \left(\frac{|T_{m'}|}{2} \right) \\
 853 \quad & \stackrel{\text{Claim 3.11}}{\geq} \frac{D^2}{4} \sum_{m > m'} |S_m| + \frac{D^2}{8} |S_{m'}| + \frac{D^2}{8} \sum_{m < m'} |T_m| \\
 854 \quad & \stackrel{|T_m| \geq |S_m|}{\geq} \frac{D^2}{12} |S|. \quad \square
 \end{aligned}$$

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

856 **4. On the non-testability of a Π_2 -property.** In this section we prove that
 857 there exists an FO property on relational structures in Π_2 that is not testable. To do
 858 so, we first prove that the property $P_{\varphi_{\mathbb{Z}}}$ defined by the formula $\varphi_{\mathbb{Z}}$ in Section 3 is
 859 not testable. Later we prove that $\varphi_{\mathbb{Z}}$ is equivalent to a sentence in Π_2 .

860 **4.1. Non-testability.** Recall that r -types are the isomorphism classes of r -balls
 861 and that restricted to the class $\mathcal{C}_{\sigma,d}$ there are finitely many r -types. Let τ_1, \dots, τ_t be
 862 a list of all r -types of bounded degree d . We let $\rho_{A,r}$ be the r -type distribution of A ,
 863 i. e.

$$864 \quad \rho_{A,r}(X) := \frac{\sum_{\tau \in X} |\{a \in U(A) \mid \mathcal{N}_r^A(a) \in \tau\}|}{|U(A)|}$$

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

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

Then the sampling distance of A and B is defined as

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

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

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

870 We make use of the following definition of repairable properties.

871 **DEFINITION 4.2** ([1]). *Let $\epsilon \in (0, 1]$. A property $\mathcal{P} \subseteq \mathcal{C}_{\sigma,d}$ is ϵ -repairable⁵ on*

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

872 $\mathcal{C}_{\sigma,d}$ if there are numbers $r := r(\epsilon) \in \mathbb{N}$, $\lambda := \lambda(\epsilon) > 0$ and $n_0 := n_0(\epsilon) \in \mathbb{N}$
 873 such that for any σ -structure $A \in \mathcal{P}$ and $B \in \mathcal{C}_{\sigma,d}$ both on $n \geq n_0$ vertices, if
 874 $\sum_{i=1}^t |\rho_{A,r}(\{\tau_i\}) - \rho_{B,r}(\{\tau_i\})| < \lambda$ then B is ϵ -close to P , where τ_1, \dots, τ_t is a list of
 875 all r -types of bounded degree d .

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

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

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

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

883 **THEOREM 4.4.** *$\mathcal{P}_{\mathbb{Z}}$ is not testable on $\mathcal{C}_{\sigma,d}$.*

884 *Proof.* We prove non-repairability for $\mathcal{P}_{\mathbb{Z}}$ and get non-testability with Theorem
 885 4.3. Let $\epsilon := 1/(144D^2)$ and let $r \in \mathbb{N}$, $\lambda > 0$ and $n_0 \in \mathbb{N}$ be arbitrary. We set
 886 $\lambda' := \lambda/(t2^{r+1})$, where τ_1, \dots, τ_t are all r -types of bounded degree d , and let n'_0 be
 887 the positive integer from Theorem 4.1 corresponding to λ' . We now pick $n \in \mathbb{N}$ such
 888 that $n = \sum_{i=0}^k D^{4i}$ for some $k \in \mathbb{N}$, $n \geq 4n_0$ and $n \geq 4(n'_0/\lambda)$. Let $A \in \mathcal{C}_{\sigma,d}$ be a
 889 model of $\varphi_{\mathbb{Z}}$ on n elements. By Theorem 4.1 there is a structure $H \in \mathcal{C}_{\sigma,d}$ on $m \leq n'_0$
 890 elements such that $\delta_{\circlearrowleft}(A, H) \leq \lambda$. Let B be the structure consisting of $\lfloor n/m \rfloor$ copies
 891 of H and $n \bmod m$ isolated elements (elements not being contained in any tuple).
 892 Note that we picked B such that $|A| = |B|$.

893 We will first argue that B is in fact ϵ -far from having the property $\mathcal{P}_{\mathbb{Z}}$. First we
 894 rename the elements from $U(B)$ in such a way that $U(A) = U(B)$ and the number
 895 $\sum_{\tilde{R} \in \sigma} |\tilde{R}(A) \Delta \tilde{R}(B)|$ of edge modifications to turn A and B into the same structure is
 896 minimal. Pick a partition $U(A) = U(B) = S \sqcup S'$ in such a way that $(S \times S') \cap \tilde{R}(B) =$
 897 \emptyset , $(S' \times S) \cap \tilde{R}(B) = \emptyset$ for any $\tilde{R} \in \sigma$ and $\|S| - |S'|\|$ is minimal among all such
 898 partitions. Assume that $|S| \leq |S'|$. Since the connected components of $G(B)$ are
 899 of size at most m we know that $\|S| - |S'|\| \leq m$. This is because otherwise we can
 900 get a partition $U(B) = T \sqcup T'$ with $\|T| - |T'|\| < \|S| - |S'|\|$ by picking all elements
 901 of any connected component of $G(B)$, which is contained in S' , and moving these
 902 elements from S' to S . Since $|S| \leq |S'|$ and $m \leq n/4$ we know that $n/4 \leq |S| \leq n/2$.
 903 Since $(S \times S') \cap \tilde{R}^B = \emptyset$ we know that \mathcal{A} and \mathcal{B} must differ in at least all tuples that
 904 correspond to an S and S' crossing edge in $U(A)$ i.e. an edge in $\langle S, S' \rangle_{U(A)}$. Hence

$$905 \quad \sum_{\tilde{R} \in \sigma} |\tilde{R}(A) \Delta \tilde{R}(B)| \geq |\langle S, S' \rangle_{\underline{G}(A)}| \stackrel{\text{Def 2.6}}{\geq} |S| \cdot h(A)$$

$$906 \quad \stackrel{\text{Thm 3.1}}{\geq} \frac{n}{4} \cdot \frac{D^2}{12} = \frac{1}{48} D^2 n \geq \frac{1}{144 D^2} dn.$$

907 Therefore B is ϵ -far from being in $\mathcal{P}_{\mathbb{Z}}$.

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

$$910 \quad \sum_{i=1}^t |\rho_{A,r}(\{\tau_i\}) - \rho_{B,r}(\{\tau_i\})|$$

$$\begin{aligned}
911 \quad &= \sum_{i=1}^t \left| \rho_{A,r}(\{\tau_i\}) - \frac{n \bmod m}{n} \cdot \rho_{K_1,r}(\{\tau_i\}) - \left\lfloor \frac{n}{m} \right\rfloor \cdot \frac{m}{n} \cdot \rho_{H,r}(\{\tau_i\}) \right| \\
912 \quad &\leq \sum_{i=1}^t \left| \rho_{A,r}(\{\tau_i\}) - \rho_{H,r}(\{\tau_i\}) \right| + \sum_{i=1}^t \left| \frac{n \bmod m}{n} \cdot \rho_{K_1,r}(\{\tau_i\}) \right| \\
913 \quad &\quad + \sum_{i=1}^t \left| \rho_{H,r}(\{\tau_i\}) - \left\lfloor \frac{n}{m} \right\rfloor \cdot \frac{m}{n} \cdot \rho_{H,r}(\{\tau_i\}) \right| \\
914 \quad &\leq \sum_{i=1}^t \left| \rho_{A,r}(\{\tau_i\}) - \rho_{H,r}(\{\tau_i\}) \right| + \frac{2m}{n} \\
915 \quad &\leq t \cdot \sup_{X \subseteq B_r} |\rho_{A,r}(X) - \rho_{H,r}(X)| + \frac{2m}{n} \\
916 \quad &\leq t \cdot 2^r \cdot \delta_{\odot}(A, H) + \frac{2m}{n} \\
917 \quad &\leq \frac{\lambda}{2} + \frac{\lambda}{2} = \lambda.
\end{aligned}$$

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

920 **4.2. Every FO property on degree-regular structures is in Π_2 .** We start
921 with the following observation.

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

$$923 \quad \exists x_1 \dots \exists x_m \left(\bigwedge_{1 \leq i, j \leq m, i \neq j} x_i \neq x_j \wedge \bigwedge_{1 \leq i \leq m} \varphi_{\tau}(x_i) \right),$$

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

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

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

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

935 *Proof.* Before we begin, let us define an r -type τ to be d -regular, if for all struc-
936 tures A and all elements $a \in U(A)$ of r -type τ , every $b \in U(A)$ with $\text{dist}(a, b) < r$ has
937 $\text{deg}_A(b) = d$.

938 We first prove the following claim.

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

943 *Proof.* Assume ψ is of the form $\exists^{\geq m} x \varphi_\tau(x)$, where τ is d -regular. As in Obser-
 944 vation 1, we may assume $\varphi_\tau(x)$ is an $\exists^* \forall$ -formula, which arises from a conjunction of
 945 an \exists^* -formula $\varphi'_\tau(x)$ (expressing that x has an ‘induced sub-neighbourhood’ of type
 946 τ) and a universal formula saying that there are no further elements in the neighbour-
 947 hood. We now have that $\psi \equiv_d \exists^{\geq m} x \varphi'_\tau(x)$. To see this, let $A \models \exists^{\geq m} x \varphi'_\tau(x)$ and
 948 $\deg(A) \leq d$. Then $A \models \exists^{\geq m} x \varphi_\tau(x)$ because τ is d -regular. The converse is obvious.

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

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

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

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

958 *Proof.* Pick $d = 2D^2 + D^4 + 1$ for any large prime power D . Then using the
 959 construction from [33] we can find a $(D^4, D, 1/4)$ -graph H . By Theorem 4.4, using
 960 this base expander H for the construction of the formula $\varphi_{\mathbb{Z}}$ we get a property which
 961 is not testable on $\mathcal{C}_{\sigma,d}$. Since all models of $\varphi_{\mathbb{Z}}$ are d -regular by construction, Lemma
 962 4.5 gives us that $\varphi_{\mathbb{Z}}$ is d -equivalent to a formula in Π_2 . \square

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

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

976 In the rest of this section, we prove the above theorem via local reductions from a
 977 structural property to a graph property, and the non-testable Π_2 -property in Theorem
 978 4.7. This technique will also be in the proofs in Section 7.

979 **5.1. Local reductions.** We first introduce the following notion of a local re-
 980 duction between two property testing models. In the following, when the context
 981 is clear, we will use \mathcal{C} to denote both a class of structures and the corresponding
 982 property testing model, which can be either the bounded-degree model for graphs or
 983 bounded-degree model for relational structures.

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

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

- 989 2. If X is ϵ -far from \mathcal{P} then $f(X)$ is (ϵ/c_1) -far from \mathcal{P}' .
 990 3. For every query to $f(X)$ we can adaptively⁶ compute c_2 queries to X such
 991 that the answer to the query to $f(X)$ can be computed from the answers to
 992 the c_2 queries to X .

993 The following lemma is known.

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

997 **5.2. Constructing the local reduction.** Now we construct a property $\mathcal{P}_{\text{graph}}$
 998 of 3-regular graphs from the property $\mathcal{P}_{\mathbb{Z}}$. We obtain this graph property as $f(\mathcal{P}_{\mathbb{Z}})$
 999 by defining a map $f : \mathcal{C}_{\sigma,d} \rightarrow \mathcal{C}_3$. To define f we introduce a distinct arrow-graph
 1000 gadget for every relation in σ (i.e. for every edge colour). The map f then replaces
 1001 every tuple in a certain relation (every coloured, directed edge) by the respective
 1002 arrow-graph gadget. Here all arrow gadgets are designed to allow for 3-regularity
 1003 of the reduced graph. To obtain 3-regularity we additionally replace every element
 1004 of a structure in $\mathcal{P}_{\mathbb{Z}}$ by a cycle of length d such that each arrow-graph gadget can
 1005 be incident to a unique vertex of the circle. We further prove that this replacement
 1006 operation defines a local reduction f from $\mathcal{P}_{\mathbb{Z}}$ to $\mathcal{P}_{\text{graph}}$. Recall that a local reduction
 1007 is a function maintaining distance that can be simulated locally by queries. Since by
 1008 Lemma 5.3 local reductions preserve testability, we use the local reduction from $\mathcal{P}_{\mathbb{Z}}$
 1009 to $\mathcal{P}_{\text{graph}}$ to obtain non-testability of the property $\mathcal{P}_{\text{graph}}$ from the non-testability of
 1010 $\mathcal{P}_{\mathbb{Z}}$. We will now define f formally.

1011 We first define building blocks which will be combined to different arrow-graph
 1012 gadgets. Let $H_1(u, v)$ be the graph with vertex set $\{u = u_0, \dots, v = u_5\}$ and edge
 1013 set $\{\{u_i, u_{i+3}\} \mid i \in \{0, 1, 2\}\}$. Next we let $H_2(u, v)$ be the graph with vertex set
 1014 $\{u = u_0, \dots, v = u_5\}$ and edge set $\{\{u_0, u_6\}, \{u_i, u_{i+2}\} \mid i \in \{1, 2\}\}$. Let $H_3(u, v)$ be
 1015 the graph with vertex set $\{u = u_0, \dots, v = u_9\}$ and edge set $\{\{u_0, u_9\}, \{u_i, u_{i+2}\} \mid i \in$
 1016 $\{1, 2, 5, 6\}\}$. Let $H_4(u)$ be the graph with vertex set $\{u = u_0, \dots, u_4\}$ and edge set
 1017 $\{\{u_0, u_3\}, \{u_1, u_4\}, \{u_2, u_4\}\}$. See Figure 5 for illustration.

1018 Let ℓ be the number of relations (the number of edge colours) in σ . We now
 1019 introduce the different types of arrow-graph gadgets we need to define the local re-
 1020 duction. For $1 \leq k \leq \ell$, we let $H_{\rightarrow}^k(u_0, v_{2\ell})$ be the graph consisting of $2\ell - 1$ vertex
 1021 disjoint copies $H_1(u_0, v_0), \dots, H_1(u_{k-1}, v_{k-1}), H_1(u_{k+1}, v_{k+1}), \dots, H_1(u_{2\ell-1}, v_{2\ell-1})$,
 1022 one copy $H_2(u_k, v_k)$, one copy $H_3(u_{2\ell}, v_{2\ell})$ and additional edges $\{v_i, u_{i+1}\}$ for each
 1023 $i \in [2\ell]$ connecting the respective copies. Note that $H_{\rightarrow}^k(u_0, v_{2\ell})$ has $12\ell + 10$ vertices
 1024 and every vertex apart from $u_0, v_{2\ell}$ has degree 3. We call $H_{\rightarrow}^k(u_0, v_{2\ell})$ a k -arrow. For
 1025 any graph G and vertices $u, v \in V(G)$, we say that there is a k -arrow from u to v ,
 1026 denoted $u \xrightarrow{k} v$, if there are $12\ell + 8$ vertices $w_1, \dots, w_{12\ell+8} \in V(G)$ and an isomor-
 1027 phism $g : H_{\rightarrow}^k(u_0, v_{2\ell}) \rightarrow \mathcal{N}_1^G(w_1, \dots, w_{12\ell+8})$ such that $g(u_0) = u$ and $g(v_{2\ell}) = v$.
 1028 Note that requiring an isomorphism with these properties guarantees that no vertex
 1029 contained in a k -arrow has neighbours not contained in the k -arrow with the excep-
 1030 tion of the end vertices u and v . For any collection $w_1, \dots, w_{12\ell+10}$ of vertices we
 1031 let $E_{\rightarrow}^k(w_1, \dots, w_{12\ell+10})$ be a set of edges such that there is a graph isomorphism
 1032 $f : H_{\rightarrow}^k(u_0, v_{2\ell}) \rightarrow (\{w_1, \dots, w_{12\ell+10}\}, E_{\rightarrow}^k(w_1, \dots, w_{12\ell+10}))$ with $f(u_0) = w_1$ and
 1033 $f(v_{2\ell}) = w_{12\ell+10}$.

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

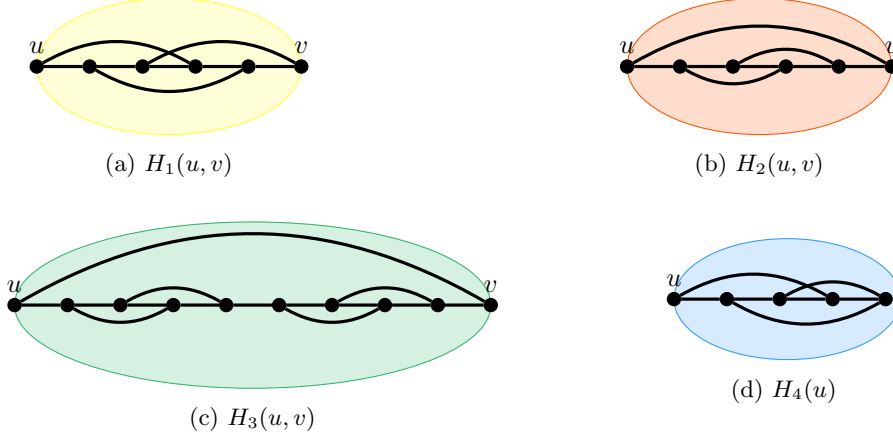


Fig. 5: Illustration of the different building blocks used to define the arrow gadgets.

1034 We now define a second arrow gadget. For $1 \leq k \leq \ell$, let $H_{\odot}^k(u_0)$ be the graph consisting of $\ell - 1$ vertex disjoint copies $H_1(u_0, v_0), \dots, H_1(u_{k-1}, v_{k-1}), H_1(u_{k+1}, v_{k+1}),$
 1035 $\dots, H_1(u_{\ell-1}, v_{\ell-1})$, one copy $H_2(u_k, v_k)$, one copy $H_4(u_{\ell})$ and edges $\{v_i, u_{i+1}\}$ for
 1036 each $i \in [\ell - 1]$. Note that $H_{\odot}^k(u_0)$ has $6\ell + 5$ vertices and every vertex apart from u_0
 1037 has degree 3. We call H_{\odot}^k a k -loop. For any graph G and vertex $u \in V(G)$, we say that
 1038 there is a k -loop at u , denoted $u \xrightarrow{k} u$, if there are $6\ell + 4$ vertices $w_1, \dots, w_{6\ell+4} \in V(G)$

1039 and an isomorphism $g : H_{\odot}^k(u_0) \rightarrow \mathcal{N}_1^G(w_1, \dots, w_{6\ell+4})$ such that $g(u_0) = u$. For
 1040 any collection $w_1, \dots, w_{6\ell+5}$ vertices we let $E_{\odot}^k(w_1, \dots, w_{6\ell+5})$ be a set of edges for
 1041 which there is an isomorphism $f : H_{\odot}^k(u_0) \rightarrow (\{w_1, \dots, w_{6\ell+5}\}, E_{\odot}^k(w_1, \dots, w_{6\ell+5}))$
 1042 for which $f(u_0) = w_1$.
 1043 Finally, let $H_{\perp}(u_0)$ be the graph consisting of ℓ vertex disjoint copies $H_1(u_0, v_0),$
 1044 $\dots, H_1(u_{\ell-1}, v_{\ell-1})$, one copy $H_4(u_{\ell})$ and additional edges $\{v_i, u_{i+1}\}$ for each $i \in [\ell - 1]$.
 1045 Note that $H_{\perp}(u_0)$ has $6\ell + 5$ vertices and every vertex apart from u_0 has degree 3. We
 1046 call H_{\perp} a *non-arrow*. For any graph G and vertex $u \in V(G)$, we say that there is a
 1047 non-arrow at u , denoted $u \not\rightarrow$, if there are $6\ell + 4$ vertices $w_1, \dots, w_{6\ell+4} \in V(G)$ and an
 1048 isomorphism $g : H_{\perp} \rightarrow \mathcal{N}_1^G(w_1, \dots, w_{6\ell+4})$ such that $g(u_0) = u$. For any collection
 1049 $w_1, \dots, w_{6\ell+5}$ vertices we let $E_{\perp}(w_1, \dots, w_{6\ell+5})$ be a set of edges for which there
 1050 is an isomorphism $f : H_{\perp}(u_0) \rightarrow (\{w_1, \dots, w_{6\ell+5}\}, E_{\odot}^k(w_1, \dots, w_{6\ell+5}))$ for which
 1051 $f(u_0) = w_1$.
 1052

1053 We now define a function $f : \mathcal{C}_{\sigma, d} \rightarrow \mathcal{C}_3$ by $f(A) := G_A$, where G_A is the graph
 1054 on vertex set $V(G_A) := \{u_{a,i}, v_{a,i}^k \mid 1 \leq i \leq d, a \in U(A), 1 \leq k \leq 6\ell + 5\}$ and edge set
 1055 $E(G_A)$ defined by

$$\begin{aligned}
 & \{ \{u_{a,i}, v_{a,i}^1\} \mid a \in U(A), 1 \leq i \leq d \} \\
 & \cup \{ \{u_{a,d}, u_{a,1}\}, \{u_{a,i}, u_{a,i+1}\} \mid a \in U(A), 1 \leq i \leq d - 1 \} \\
 & \cup \bigcup_{\substack{\text{ans}(a,i)=\text{ans}(b,j)=(k,a,b) \\ a \neq b}} E_{\rightarrow}^k \left(v_{a,i}^1, \dots, v_{a,i}^{6\ell+5}, v_{b,j}^{6\ell+5}, \dots, v_{b,j}^1 \right)
 \end{aligned}$$

$$\begin{aligned}
1059 \quad & \cup \bigcup_{\text{ans}(a,i)=(k,a,a)} E_{\circlearrowleft}^k \left(v_{a,i}^1, \dots, v_{a,i}^{6\ell+5} \right) \\
1060 \quad & \cup \bigcup_{\text{ans}(a,i)=\perp} E_{\perp} \left(v_{a,i}^1, \dots, v_{a,i}^{6\ell+5} \right),
\end{aligned}$$

1061 where $\text{ans}(a, i) = (k, a, b)$ denotes that the i -th tuple of a is (a, b) and is in the k -th
1062 relation. Hence G_A is defined in such a way that every element $a \in U(A)$ is represented
1063 by an induced cycle $(u_{a,1}, \dots, u_{a,d}, u_{a,1})$ and if (a, b) is a tuple in the k -th relation
1064 of σ in A , then $u_{a,i} \xrightarrow{k} u_{b,j}$ in G_A for some $1 \leq i, j \leq d$, and $u_{a,i}$ has a non-arrow
1065 for every i satisfying that $\text{ans}(a, i) = \perp$ for every k . Note that G_A is 3 regular by
1066 construction for every $A \in \mathcal{C}_{\sigma,d}$. For illustration see Figure 6. In the following we
1067 refer to vertices of G_A of the form $u_{a,i}$ by *element-vertices* while we call vertices of
1068 the form $v_{a,i}^j$ *relation-vertices*. The following is easy to observe from the construction
1069 and from the fact that $d = 2D^2 + D^4 + 1 < 3D^4 + 1 = |\sigma| = \ell$ for some large prime
1070 power D (see Section 3 for definitions).

1071 **FACT 1.** *For every $u \in V(G_A)$, u is an element-vertex iff u is contained in a*
1072 *simple cycle of length d . Furthermore, two vertices $u, v \in V(G_A)$ correspond to the*
1073 *same element a of A (i. e. there are $i, j \in \{1, \dots, d\}$ such that $u = u_{a,i}$ and $v = u_{a,j}$)*
1074 *iff there is a simple cycle of length d containing both u and v .*

1075 Note that we do not need to ask for cycles of length d to be induced because the
1076 structure we obtain does not allow for cycles of length d apart from the cycles corre-
1077 sponding to elements.

1078 Now we define property $\mathcal{P}_{\text{graph}} := \{f(A) \mid A \in \mathcal{P}_{\mathbb{Z}}\} \subseteq \mathcal{C}_3$.

1079 **LEMMA 5.4.** *The map f is a local reduction from $\mathcal{P}_{\mathbb{Z}}$ to $\mathcal{P}_{\text{graph}}$.*

1080 *Proof.* First note that for any $A \in \mathcal{P}_{\mathbb{Z}}$, we have that $f(A) \in \mathcal{P}_{\text{graph}}$ by definition.

1081 Now let $c_1 = 6d(1 + 6\ell + 5)$. We prove that if $A \in \mathcal{C}_{\sigma,d}$ is ϵ -far from $\mathcal{P}_{\mathbb{Z}}$ then
1082 $f(A)$ is ϵ/c_1 -far from $\mathcal{P}_{\text{graph}}$ by contraposition. Therefore assume that $f(A) =: G_A$ is
1083 not ϵ/c_1 -far from $\mathcal{P}_{\text{graph}}$ for some $A \in \mathcal{C}_{\sigma,d}$. Then there is a set $E \subseteq \{e \subseteq V(G_A) \mid$
1084 $|e| = 2\}$ of size at most $\epsilon \cdot 3|V(G_A)|/c_1$, and a graph $G \in \mathcal{P}_{\text{graph}}$ such that G is
1085 obtained from G_A by modifying the tuples in E . By definition of $\mathcal{P}_{\text{graph}}$, there is
1086 a structure $A_G \in \mathcal{P}_{\mathbb{Z}}$ such that $f(A_G) = G$. First note that $|U(A_G)| = |U(A)|$,
1087 as $d(1 + 6\ell + 5)|U(A)| = |V(G_A)| = |V(G)| = d(1 + 6\ell + 5)|U(A_G)|$. Hence there
1088 must be a set R of tuples that need to be modified to make A isomorphic to A_G .
1089 First note that R cannot contain a tuple (a, b) where $\{u_{a,i}, v_{a,i}^k, u_{b,i}, v_{b,i}^k \mid 1 \leq i \leq$
1090 $d, 1 \leq k \leq \ell\} \cap e = \emptyset$ for every $e \in E$. This is because if (a, b) is a tuple in the
1091 k -th relation of A , then $u_{a,i} \xrightarrow{k} u_{b,j}$ in G_A for some $i, j \in \{1, \dots, d\}$. But since
1092 $\{u_{a,i}, v_{a,i}^k, u_{b,i}, v_{b,i}^k \mid 1 \leq i \leq d, 1 \leq k \leq \ell\} \cap e = \emptyset$ for every $e \in E$, we have that
1093 $u_{a,i} \xrightarrow{k} u_{b,j}$ in G . Further, $(u_{a,1}, \dots, u_{a,d}, u_{a,1})$ and $(u_{b,1}, \dots, u_{b,d}, u_{b,1})$ are simple
1094 cycles of length d in G . Hence by 1 there are elements a, b in A_G corresponding
1095 to $(u_{a,1}, \dots, u_{a,d}, u_{a,1})$ and $(u_{b,1}, \dots, u_{b,d}, u_{b,1})$ such that (a, b) is a tuple in the k -th
1096 relation of A_G , and hence (a, b) cannot be in R . The same argument works when
1097 assuming that (a, b) is a tuple in A_G . Since for every $e \in E$, there is at most $2d$ tuples
1098 (a, b) such that $\{u_{a,i}, v_{a,i}^k, u_{b,i}, v_{b,i}^k \mid 1 \leq i \leq d, 1 \leq k \leq \ell\} \cap e \neq \emptyset$, we get that

$$1099 \quad |R| \leq 2d\epsilon \cdot 3|V(G_A)|/c_1 = 6d(1 + 6\ell + 5)\epsilon d|U(A)|/c_1 = \epsilon d|U(A)|.$$

1100 Hence A is not ϵ -far to being in $\mathcal{P}_{\mathbb{Z}}$.

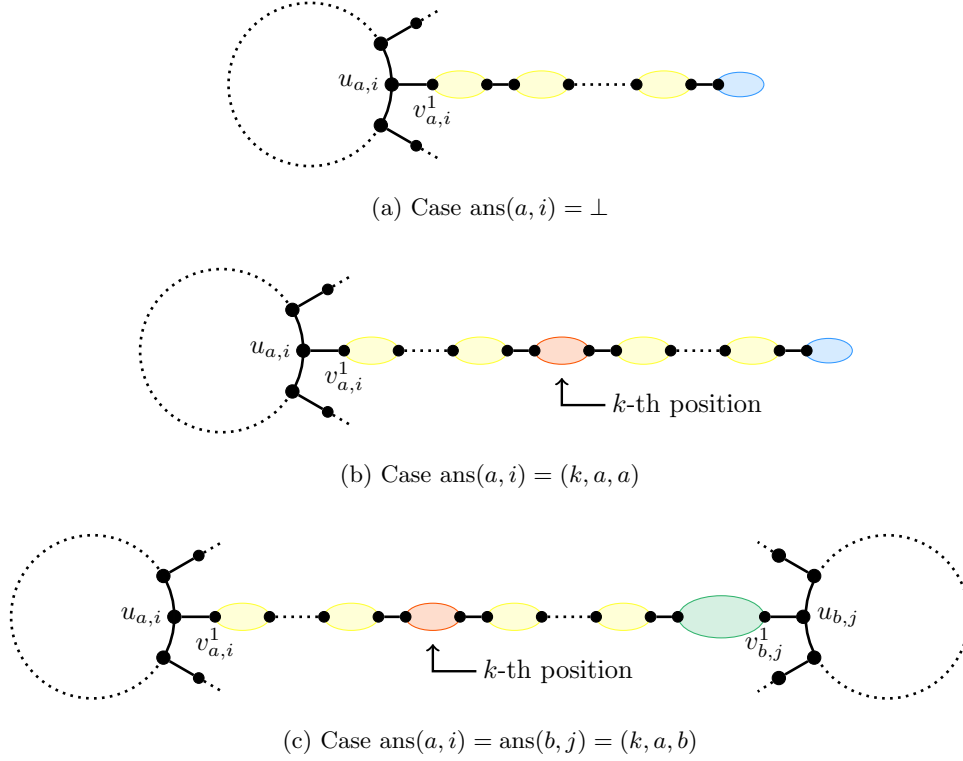


Fig. 6: Different types of arrows in G_A . Here different coloured ellipses represent a copy of $H_1(u, v)$, $H_2(u, v)$, $H_3(u, v)$ or $H_4(u)$ respectively (see Figure 5 for details).

1101 Let $c_2 := d + 1$. Let $A \in \mathcal{C}_{\sigma, d}$ and $G_A := f(A)$. First it is important to observe
 1102 that we can pick an ordering of the vertices of G_A such that the position of each vertex
 1103 depends solely on the number of elements of A . Hence we can assume that for any
 1104 element a of A we can decide for any vertex $v \in V(G_A)$ whether v is of the form $u_{a, i}$
 1105 and whether v is of the form $v_{a, i}^k$. Now we argue how we can determine the answer to
 1106 any neighbour query in G_A . First note that for any $a \in U(A)$ and $i \in \{1, \dots, d\}$ the
 1107 vertex $u_{a, i}$ is adjacent in G_A to $v_{a, i}^1$, and the two neighbouring vertices on the simple
 1108 cycle $(u_{a, 1}, \dots, u_{a, d}, u_{a, 1})$. Hence any neighbour query in G_A to $u_{a, i}$ can be answered
 1109 without querying A . Assume $v \in \{v_{a, i}^k \mid 1 \leq k \leq \ell\}$ for some $a \in U(A)$ and some
 1110 $1 \leq i \leq d$. Then we can determine all neighbours of v by querying (a, i) and further
 1111 if $\text{ans}(a, i) \neq (k, a, a)$, $\text{ans}(a, i) \neq \perp$ and $\text{ans}(a, i) = (k, a, b)$, then we need to query
 1112 (b, j) for every $1 \leq j \leq d$ to find out for which j we have $\text{ans}(b, j) = (k, a, b)$. Hence
 1113 we can determine the answer to any query to G_A by making c_2 queries to A . This
 1114 proves that f is a local reduction from $\mathcal{P}_{\mathbb{Z}}$ to $\mathcal{P}_{\text{graph}}$. \square

1115 **5.3. The property of graphs is definable in FO.** In this section we find
 1116 an FO sentence ψ_{graph} which defines the property $\mathcal{P}_{\text{graph}}$. We do this by defining a
 1117 formula expressing for two vertices u, v that $u \xrightarrow{k} v$, a formula expressing for vertex u
 1118 that $u \xrightarrow{k} u$ and a formula expressing for vertex u that $u \not\xrightarrow{k}$ and replacing formulas of

1119 the form $R(u, v)$, $R(v, v)$ and $\neg R(u, v)$ for $R \in \sigma$ by the new formulas appropriately.
 1120 We additionally restrict the scope of the quantifiers. In the previous subsection we
 1121 already defined $\ell := |\sigma|$. We further rename the relations in σ in an arbitrary way
 1122 such that for this section we can assume that $\sigma = \{R_1, \dots, R_\ell\}$.

1123 We now translate the formula $\varphi_{\mathbb{Z}}$ into a formula ψ_{graph} in the language of undi-
 1124 rected graphs using the FO formulas defined in the following. We let $\alpha(x)$ be a formula
 1125 saying ‘ x is an element-vertex’ and $\beta(x, y)$ be a formula saying ‘ x and y represent the
 1126 same element of \mathcal{A} ’, which is easy to do by Fact 1. We further let $\gamma(x)$ be a formula
 1127 saying ‘ x is an internal vertex of either a k -arrow, a k -loop for any $k \in \{1, \dots, \ell\}$ or
 1128 a non-arrow’. Here an ‘internal vertex’ of an arrow refers to any vertex on this arrow
 1129 except the two endpoints, or the single endpoint in case of a loop or non-arrow. Let
 1130 $\delta_{\rightarrow}^k(x, y)$ denote ‘ $x \xrightarrow{k} y$ ’ for any $k \in \{1, \dots, \ell\}$, similarly, let $\delta_{\circlearrowleft}^k(x)$ denote ‘ $x \xrightarrow{k} x$ ’ for
 1131 any $k \in \{1, \dots, \ell\}$. Given $\varphi_{\mathbb{Z}}$, formula ψ_{graph} is obtained as follows. In $\varphi_{\mathbb{Z}}$ we replace
 1132 each expression $R_k(x, x)$ by $\delta_{\circlearrowleft}^k(x, x)$ and each expression $R_k(x, y)$ by $\delta_{\rightarrow}^k(x, y)$ (for
 1133 $x \neq y$). In addition, we relativise all quantifiers in the following way. We replace every
 1134 expression of the form $\exists x \chi(x, x_1, \dots, x_m)$ by $\exists x (\alpha(x) \wedge \chi(x, x_1, \dots, x_m))$ and every
 1135 expression of the form $\forall x \chi(x, x_1, \dots, x_m)$ by $\forall x (\alpha(x) \rightarrow \exists y \beta(x, y) \wedge \chi(y, x_1, \dots, x_m))$.
 1136 Let us call the resulting formula ψ . Then we set ψ_{graph} to be the conjunction of the
 1137 formula ψ and the formula $\forall x \left((-\alpha(x) \rightarrow \gamma(x)) \wedge (\alpha(x) \rightarrow \exists y \gamma(y) \wedge E(x, y)) \right)$.

1138 **LEMMA 5.5.** *For any $A \in \mathcal{C}_{\sigma, d}$ the following proposition is true. $A \models \varphi_{\mathbb{Z}}$ if and*
 1139 *only if $f(A) \models \psi_{\text{graph}}$. Additionally we have that if $G \in \mathcal{C}_3$ is a model of ψ_{graph} then*
 1140 *$G \cong f(A)$ for some $A \in \mathcal{A}_{\sigma, d}$.*

1141 *Proof.* First assume that $A \models \varphi_{\mathbb{Z}}$. First observe that by construction of $G_A :=$
 1142 $f(A)$ and ψ , we get that $A \models \varphi_{\mathbb{Z}}$ if and only if $G_A \models \psi$. Note that for this
 1143 statement it is important that the set of k -arrows and k -loops for all $k \in \{1, \dots, \ell\}$
 1144 is a set of pairwise non-isomorphic graphs. In the construction of G_A , every vertex is
 1145 either an element-vertex $u_{a, i}$ in which case it is adjacent to the relation-vertex $v_{a, i}^1$,
 1146 or is an internal vertex of some k -arrow, k -loop or non-arrow. Hence we get that
 1147 $G_A \models \forall x \left((-\alpha(x) \rightarrow \gamma(x)) \wedge (\alpha(x) \rightarrow \exists y \gamma(y) \wedge E(x, y)) \right)$, which completes the proof
 1148 of the first statement.

1149 Towards proving the second statement of Lemma 5.5, let us assume that some
 1150 graph $G \in \mathcal{C}_3$ is a model of ψ_{graph} . Then $G \models \forall x \left((-\alpha(x) \rightarrow \gamma(x)) \wedge (\alpha(x) \rightarrow$
 1151 $\exists y \gamma(y) \wedge E(x, y)) \right)$. Hence G consists of a set of element-vertices that are connected
 1152 according to ψ with k -arrow, k -loops or non-arrows. Hence we can reverse the local
 1153 reduction to obtain A_G which is the corresponding model of $\varphi_{\mathbb{Z}}$ for which $f(A_G) \cong G$
 1154 by the following construction. For any maximal set of vertices $X \subseteq V(G)$ such that
 1155 $\beta(u, v)$ holds for every pair $u, v \in X$, we introduce an element a_X . For $X, Y \subseteq V(G)$,
 1156 we add a tuple (a_X, a_Y) to the relation $R_k(A_G)$ if there are $u \in X$ and $v \in Y$ such
 1157 that $u \xrightarrow{k} v$ in G . With a similar argument as above, we get that A_G is a model of
 1158 $\varphi_{\mathbb{Z}}$ by the construction of ψ . Additionally we get for some ordering of the neighbours
 1159 of each element of A_G that $f(A_G) \cong G$ (this ordering has to be consistent with the
 1160 order of k -arrows along the cycle of element-vertices). \square

1161 *Proof of Theorem 5.1.* As a consequence from Lemma 5.5, we get that ψ_{graph}
 1162 defines the property $\mathcal{P}_{\text{graph}}$ on the class \mathcal{C}_3 . Since we constructed the local reduction
 1163 f in such a way that $f(\mathcal{A})$ is 3-regular for every $\mathcal{A} \in \mathcal{C}_{\sigma, d}$ by Lemma 4.5, we get

1164 that $\mathcal{P}_{\text{graph}}$ can be defined by a sentence in Π_2 on the class \mathcal{C}_3 . Combining this with
 1165 Lemma 5.3 and Lemma 5.4, we obtain Theorem 5.1. \square

1166 We would like to point out here that while we obtain the non-testability of $\mathcal{P}_{\text{graph}}$
 1167 using the local reduction f , we can not conclude that $\mathcal{P}_{\text{graph}}$ is a class of expanders.
 1168 However, we will show that this is true in the following section.

1169 **5.4. The property of graphs is a class of expanders.** In this subsection we
 1170 show that $\mathcal{P}_{\text{graph}}$ is a family of expanders and hence prove the following theorem.

1171 **THEOREM 5.6.** *There exists a universal constant $\xi > 0$ and an (infinite) class of*
 1172 *ξ -expanders with maximum degree at most 3 which is definable in FO on undirected*
 1173 *graphs.*

1174 Expansion of $\mathcal{P}_{\text{graph}}$ is not needed for the non-testability results in this paper. How-
 1175 ever, we think that Theorem 5.6 is of independent interest since it gives us new insights
 1176 into the expressibility of first-order logic. Furthermore, for an expanding property of
 1177 undirected graphs, its non-testability follows from the main result from [15].

1178 **LEMMA 5.7.** *The models of ψ_{graph} is a family of ξ -expanders, for some constant*
 1179 *$\xi > 0$.*

1180 *Proof.* Let $A \in \mathcal{P}_{\mathbb{Z}}$ and $G_A := f(A)$. For every $a \in A$, we define the vertex
 1181 set $V_a = \{u_{a,i}, v_{a,i}^k \mid 1 \leq i \leq d, 1 \leq k \leq 6\ell + 5\}$, encompassing the element vertices
 1182 of a alongside the vertices of its loops, non-arrows, and half of the vertices of each
 1183 of its arrows. Considering a set $S \subset V(G_A)$ such that $|S| \leq \frac{|V(G_A)|}{2}$, we partition
 1184 $U(A)$ into $S^{\text{full}} = \{a \in U(A) : V_a \subseteq S\}$, $S^{\text{disj}} = \{a \in U(A) : V_a \cap S = \emptyset\}$ and
 1185 $S^{\text{part}} = U(A) \setminus (S^{\text{full}} \cup S^{\text{disj}})$.

1186 First note that $|S^{\text{full}}| \leq \frac{|U(A)|}{2}$. As $G(A)$ is an ϵ -expander by Theorem 3.1, there
 1187 are at least $\epsilon|S^{\text{full}}|$ edges between S^{full} and $S^{\text{part}} \cup S^{\text{disj}}$. Hence, there are at least
 1188 $\epsilon|S^{\text{full}}| - d|S^{\text{part}}|$ edges between S^{full} and S^{disj} . By the choice of the sets S^{full} and S^{disj} ,
 1189 for every such edge (a, b) , there corresponds an edge, i. e. the edge $\{v_{a,i}^{6\ell+5}, v_{b,j}^{6\ell+5}\}$ for
 1190 some $1 \leq i, j \leq d$, in G_A between S and $V(G_A) \setminus S$. Additionally, for every $a \in S^{\text{part}}$,
 1191 there is at least one edge in $G_A[V_a]$ between S and $V(G_A) \setminus S$.

1192 We set $c = d(6\ell + 6)$ and observe that $|S^{\text{full}}| \geq \frac{|S| - |S^{\text{part}}|}{c}$. Let us first consider
 1193 the case that $|S^{\text{part}}| \leq \frac{\epsilon|S|}{3cd} \leq \frac{|S|}{3}$. In this case, we have at least $\frac{\epsilon|S|}{3c}$ edges between
 1194 S and $U(A) \setminus S$ on account of edges between S^{full} and S^{disj} . On the other hand, if
 1195 $|S^{\text{part}}| \geq \frac{\epsilon|S|}{3cd}$ then we have at least $\frac{\epsilon|S|}{3cd}$ edges between S and $U(A) \setminus S$ on account of
 1196 the edges within each $G_A[V_a]$ for $a \in S^{\text{part}}$. Therefore, $\mathcal{P}_{\text{graph}}$ is a class of ξ -expanders
 1197 for $\xi = \frac{\epsilon}{3cd}$. \square

1198 **6. On the testability of all Σ_2 -properties.** Let $\sigma = \{R_1, \dots, R_m\}$ be any
 1199 relational signature and $\mathcal{C}_{\sigma,d}$ the set of σ -structures of bounded degree d . We prove
 1200 the following.

1201 **THEOREM 6.1.** *Every first-order property defined by a σ -sentence in Σ_2 is testable*
 1202 *in the bounded-degree model.*

1203 We adapt the notion of indistinguishability of [4] from the dense model to the
 1204 bounded-degree model.

1205 **DEFINITION 6.2.** *Two properties $\mathcal{P}, \mathcal{Q} \subseteq \mathcal{C}_{\sigma,d}$ are called indistinguishable if for*
 1206 *every $\epsilon \in (0, 1)$ there exists $N = N(\epsilon)$ such that for every structure $A \in \mathcal{P}$ with*
 1207 *$|U(A)| > N$ there is a structure $\hat{A} \in \mathcal{Q}$ with the same universe, that is ϵ -close to A ;*

1208 and for every $B \in \mathcal{Q}$ with $|U(B)| > N$ there is a structure $\tilde{B} \in \mathcal{P}$ with the same
 1209 universe, that is ϵ -close to B .

1210 The following lemma follows from the definitions, and is similar to [4], though we
 1211 make use of the canonical testers for bounded-degree graphs ([11, 22]).

1212 **LEMMA 6.3.** *If $\mathcal{P}, \mathcal{Q} \subseteq \mathcal{C}_{\sigma,d}$ are indistinguishable properties, then \mathcal{P} is testable on*
 1213 *$\mathcal{C}_{\sigma,d}$ if and only if \mathcal{Q} is testable on $\mathcal{C}_{\sigma,d}$.*

1214 *Proof.* We show that if \mathcal{P} is testable, then \mathcal{Q} is also testable. The other direction
 1215 follows by the same argument. Let $\epsilon > 0$. Since \mathcal{P} is testable, there exists an $\frac{\epsilon}{2}$ -tester
 1216 for \mathcal{P} with success probability at least $\frac{2}{3}$. Furthermore, we can assume that the tester
 1217 (called canonical tester) behaves as follows (see [11, 22]): it first uniformly samples a
 1218 constant number $c_0 = c_0(\frac{\epsilon}{2}, d)$ of elements, then explores the union of r -balls around
 1219 all sampled elements for some constant $r = r(\frac{\epsilon}{2}, d) > 0$, and makes a deterministic
 1220 decision whether to accept, based on an isomorphic copy of the explored substructure.
 1221 Let $C = C(\frac{\epsilon}{2}, d) = c_0 \cdot (1 + d + \dots + d^r)$ denote the upper bound on the number of
 1222 queries the canonical tester made on the input structure. Then there exists some
 1223 universal constant $c_1 > 0$ such that by repeating the canonical tester c_1 times, and
 1224 taking the majority vote, we can have a tester T with $c_1 \cdot C$ query complexity and
 1225 success probability at least $\frac{5}{6}$.

1226 Let N be a number such that if a structure B with $n > N$ elements satis-
 1227 fies \mathcal{Q} , then there exists a $\tilde{B} \in \mathcal{P}$ with the same universe such that $\text{dist}(B, \tilde{B}) \leq$
 1228 $\min\{\frac{\epsilon}{2}, \frac{1}{c_2 C \cdot d^{C+2}}\}dn$ for some large constant $c_2 > 0$. Now we give an ϵ -tester for \mathcal{Q} . If
 1229 the input structure B has size at most N , we can query the whole input to decide if it
 1230 satisfies \mathcal{Q} or not. If its size is larger than N , then we use the aforementioned $\frac{\epsilon}{2}$ -tester
 1231 for \mathcal{P} with success probability at least $\frac{5}{6}$. If B satisfies \mathcal{Q} , then there exists $\tilde{B} \in \mathcal{P}$
 1232 that differs from B in no more than $1/(c_2 C \cdot d^{C+2})dn$ places. Since the algorithm
 1233 samples $c_0 \cdot c_1$ elements and queries the r -balls around all these sampled elements and
 1234 makes at most $c_1 \cdot C$ queries in total, we have that with probability at least $1 - \frac{1}{6}$,
 1235 the algorithm does not query any part where B and \tilde{B} differ, and thus its output is
 1236 correct with probability at least $\frac{5}{6} - \frac{1}{6} = \frac{2}{3}$. If B is ϵ -far from satisfying \mathcal{Q} then it is
 1237 $\frac{\epsilon}{2}$ -far from satisfying \mathcal{P} and with probability at least $\frac{5}{6} > \frac{2}{3}$, the algorithm will reject
 1238 B . Thus \mathcal{Q} is also testable. \square

1239 *High-level idea of proof of Theorem 6.1.* Let $\varphi \in \Sigma_2$. We prove that the property
 1240 defined by φ can be written as the union of properties, each of which is defined by
 1241 another formula φ' in Σ_2 where the structure induced by the existentially quantified
 1242 variables is a fixed structure M (see Claim 6.6). With some further simplification
 1243 of φ' , we obtain a formula φ'' in Σ_2 which expresses that the structure has to have
 1244 M as an induced substructure and every set of elements of fixed size ℓ has to induce
 1245 some structure from a set of structures \mathfrak{H} , and – depending on the structure from \mathfrak{H}
 1246 – there might be some connections to the elements of M (see Claim 6.7). We then
 1247 define a formula ψ in Π_1 such that the property defined by ψ is indistinguishable
 1248 from the property defined by φ'' in the sense that we can transform any structure
 1249 satisfying ψ , into a structure satisfying φ'' by modifying no more than a small fraction
 1250 of the tuples and vice versa (see Claim 6.10). The intuition behind this is that every
 1251 structure satisfying φ'' can be made to satisfy ψ by removing the structure M while
 1252 on the other hand for every structure which satisfies ψ we can plant the structure M
 1253 to make it satisfy φ'' . Since it is a priori unclear how the existentially and universally
 1254 quantified variables interact, we have to define ψ very carefully. Here it is important
 1255 to note that the number of occurrences of structures in \mathfrak{H} forcing an interaction with

1256 M is limited because of the degree bound (see Claim 6.8). Thus such structures can
 1257 not be allowed to occur for models of ψ , as here the number of occurrences can not
 1258 be limited in any way. Since properties defined by a formula in Π_1 are testable, this
 1259 implies with the indistinguishability of ψ and φ'' that the property defined by φ'' is
 1260 testable. Furthermore by the fact that testable properties are closed under union [18],
 1261 we reach the conclusion that any property defined by a formula in Σ_2 is testable.

1262 We will not directly give a tester for the property \mathcal{P}_φ but decompose φ into simpler
 1263 cases. However, every simplification of φ used is computable, and the proof below
 1264 yields a construction of an ϵ -tester for \mathcal{P}_φ for every $\epsilon \in (0, 1)$ and every $\varphi \in \Sigma_2$.
 1265

1266 For the full proof of Theorem 6.1, we use the following definition.

1267 DEFINITION 6.4. Let A be a σ -structure with $U(A) = \{a_1, \dots, a_t\}$. Let $\bar{z} =$
 1268 (z_1, \dots, z_t) be a tuple of variables. Then we define $\iota^A(\bar{z})$ as follows.

$$\begin{aligned}
 1269 \quad \iota^A(\bar{z}) := & \bigwedge_{R \in \sigma} \left(\bigwedge_{(a_{i_1}, \dots, a_{i_{\text{ar}(R)}})} \in R(A) R(z_{i_1}, \dots, z_{i_{\text{ar}(R)}}) \wedge \right. \\
 1270 \quad & \left. \bigwedge_{(a_{i_1}, \dots, a_{i_{\text{ar}(R)}})} \in U(A)^{\text{ar}(R)} \setminus R(A) \neg R(z_{i_1}, \dots, z_{i_{\text{ar}(R)}}) \right) \wedge \bigwedge_{\substack{i, j \in [t] \\ i \neq j}} (\neg z_i = z_j).
 \end{aligned}$$

1271 Note that for every σ -structure A' and $\bar{a}' = (a'_1, \dots, a'_t) \in U(A')^t$ we have that
 1272 $A' \models \iota^A(\bar{a}')$ if and only if $a_i \mapsto a'_i$, $i \in \{1, \dots, t\}$ is an isomorphism from A to
 1273 $A'[\{a'_1, \dots, a'_t\}]$. In particular, if $A' \models \iota^A(\bar{a}')$, then $\{a'_1, \dots, a'_t\}$ induces a substructure
 1274 isomorphic to A in A' .

1275 *Proof of Theorem 6.1.* Let φ be any sentence in Σ_2 . Therefore we can assume
 1276 that φ is of the form $\varphi = \exists \bar{x} \forall \bar{y} \chi(\bar{x}, \bar{y})$ where $\bar{x} = (x_1, \dots, x_k)$ is a tuple of $k \in \mathbb{N}$
 1277 variables, $\bar{y} = (y_1, \dots, y_\ell)$ is a tuple of $\ell \in \mathbb{N}$ variables and $\chi(\bar{x}, \bar{y})$ is a quantifier-free
 1278 formula. We can further assume that $\chi(\bar{x}, \bar{y})$ is in disjunctive normal form, and that

$$1279 \quad (6.1) \quad \varphi = \exists \bar{x} \forall \bar{y} \bigvee_{i \in I} \left(\alpha^i(\bar{x}) \wedge \beta^i(\bar{y}) \wedge \text{pos}^i(\bar{x}, \bar{y}) \wedge \text{neg}^i(\bar{x}, \bar{y}) \right),$$

1280 where $\alpha^i(\bar{x})$ is a conjunction of literals only containing variables from \bar{x} , $\beta^i(\bar{y})$ is a
 1281 conjunction of literals only containing variables in \bar{y} , $\text{neg}^i(\bar{x}, \bar{y})$ is a conjunction of
 1282 negated atomic formulas containing both variables from \bar{x} and \bar{y} and $\text{pos}^i(\bar{x}, \bar{y})$ is a
 1283 conjunction of atomic formulas containing both variables from \bar{x} and \bar{y} . Now note
 1284 that if an expression ' $x_j = y_{j'}$ ' appears in a conjunctive clause, then we can replace
 1285 every occurrence of $y_{j'}$ by x_j in that clause, which will result in an equivalent formula.

1286 We now write the formula φ given in (6.1) as a disjunction over all possible struc-
 1287 tures in $\mathcal{C}_{\sigma, d}$ the existentially quantified variables could enforce. Since the elements
 1288 realising the existentially quantified variables will have a certain structure, it is natural
 1289 to decompose the formula in this way.

1290 Let $\mathfrak{M} \subseteq \mathcal{C}_{\sigma, d}$ be a set of models of φ , such that every model $A \in \mathcal{C}_{\sigma, d}$ of φ
 1291 contains an isomorphic copy of some $M \in \mathfrak{M}$ as an induced substructure, and \mathfrak{M} is
 1292 minimal with this property.

1293 CLAIM 6.5. Every $M \in \mathfrak{M}$ has at most k elements.

1294 *Proof.* Assume there is $M \in \mathfrak{M}$ with $|M| > k$. Since every structure in \mathfrak{M} is
 1295 a model of φ there must be a tuple $\bar{a} = (a_1, \dots, a_k) \in U(M)^k$ such that $M \models$

1296 $\forall \bar{y} \bigvee_{i \in I} \left(\alpha^i(\bar{a}) \wedge \beta^i(\bar{y}) \wedge \text{pos}^i(\bar{a}, \bar{y}) \wedge \text{neg}^i(\bar{a}, \bar{y}) \right)$. This implies that for every tuple
 1297 $\bar{b} \in U(M)^\ell$ we have $M \models \bigvee_{i \in I} \left(\alpha^i(\bar{a}) \wedge \beta^i(\bar{b}) \wedge \text{pos}^i(\bar{a}, \bar{b}) \wedge \text{neg}^i(\bar{a}, \bar{b}) \right)$. Furthermore,
 1298 since $\{a_1, \dots, a_k\}^\ell \subseteq U(M)^\ell$ we have that $M[\{a_1, \dots, a_k\}] \models \forall \bar{y} \bigvee_{i \in I} \left(\alpha^i(\bar{a}) \wedge \beta^i(\bar{y}) \wedge \right.$
 1299 $\left. \text{pos}^i(\bar{a}, \bar{y}) \wedge \text{neg}^i(\bar{a}, \bar{y}) \right)$. This means that $M[\{a_1, \dots, a_k\}] \models \varphi$. Hence \mathfrak{M} contains an
 1300 induced substructure M' of $M[\{a_1, \dots, a_k\}]$. Since every model of φ containing M as
 1301 an induced substructure must also contain M' as an induced substructure $\mathfrak{M} \setminus \{M\}$
 1302 is a strictly smaller set than \mathfrak{M} with all desired properties. This contradicts the
 1303 minimality \mathfrak{M} . \square

1304 Therefore \mathfrak{M} is finite. For $M \in \mathfrak{M}$ let $J_M := \{j \in I \mid M \models \alpha^j(\bar{m}) \text{ for some } \bar{m} \in U(M)^\ell\} \subseteq I$.

1306 CLAIM 6.6. *We have*

$$1307 \quad \varphi \equiv_d \bigvee_{M \in \mathfrak{M}} \left(\exists \bar{x} \forall \bar{y} \left[\iota^M(\bar{x}) \wedge \bigvee_{j \in J_M} \left(\beta^j(\bar{y}) \wedge \text{pos}^j(\bar{x}, \bar{y}) \wedge \text{neg}^j(\bar{x}, \bar{y}) \right) \right] \right).$$

1308 *Proof.* Let $A \in \mathcal{C}_{\sigma, d}$ be a model of φ . Then there is a tuple $\bar{a} = (a_1, \dots, a_k) \in U(A)^k$
 1309 such that $A \models \forall y \chi(\bar{a}, \bar{y})$. Since $\{a_1, \dots, a_k\}^\ell \subseteq U(A)^\ell$ this implies that
 1310 $A[\{a_1, \dots, a_k\}] \models \forall y \chi(\bar{a}, \bar{y})$ and hence $A[\{a_1, \dots, a_k\}] \models \varphi$. In addition, we may
 1311 assume that we picked \bar{a} in such a way that for any tuple $\bar{a}' = (a'_1, \dots, a'_k) \in \{a_1, \dots, a_k\}^k$
 1312 with $\{a'_1, \dots, a'_k\} \subsetneq \{a_1, \dots, a_k\}$ we have that $A \not\models \forall \bar{y} \chi(\bar{a}', \bar{y})$. (The
 1313 reason is that if for some tuple \bar{a}' this is not the case then we just replace \bar{a} by \bar{a}' and
 1314 so on until this property holds). Hence $A[\{a_1, \dots, a_k\}]$ cannot have a proper induced
 1315 substructure in \mathfrak{M} , and it follows that there is $M \in \mathfrak{M}$ such that $M \cong A[\{a_1, \dots, a_k\}]$.
 1316 By choice of J_M we get $A \models \forall \bar{y} \left[\iota^M(\bar{a}) \wedge \bigvee_{j \in J_M} \left(\beta^j(\bar{y}) \wedge \text{pos}^j(\bar{a}, \bar{y}) \wedge \text{neg}^j(\bar{a}, \bar{y}) \right) \right]$ and
 1317 hence

$$1318 \quad A \models \bigvee_{M \in \mathfrak{M}} \left(\exists \bar{x} \forall \bar{y} \left[\iota^M(\bar{x}) \wedge \bigvee_{j \in J_M} \left(\beta^j(\bar{y}) \wedge \text{pos}^j(\bar{x}, \bar{y}) \wedge \text{neg}^j(\bar{x}, \bar{y}) \right) \right] \right).$$

1319 To prove the other direction, we now let the structure $A \in \mathcal{C}_{\sigma, d}$ be a model of the
 1320 formula

1321 $\bigvee_{M \in \mathfrak{M}} \left(\exists \bar{x} \forall \bar{y} \left[\iota^M(\bar{x}) \wedge \bigvee_{j \in J_M} \left(\beta^j(\bar{y}) \wedge \text{pos}^j(\bar{x}, \bar{y}) \wedge \text{neg}^j(\bar{x}, \bar{y}) \right) \right] \right)$. Consequently there
 1322 is $M \in \mathfrak{M}$ and $\bar{a} \in U(A)^k$ such that $A \models \forall \bar{y} \left[\iota^M(\bar{a}) \wedge \bigvee_{j \in J_M} \left(\beta^j(\bar{y}) \wedge \text{pos}^j(\bar{a}, \bar{y}) \wedge \right.$
 1323 $\left. \text{neg}^j(\bar{a}, \bar{y}) \right) \right]$. By choice of J_M this implies $A \models \forall \bar{y} \bigvee_{j \in J_M} \left(\alpha^j(\bar{a}) \wedge \beta^j(\bar{y}) \wedge \text{pos}^j(\bar{a}, \bar{y}) \wedge \right.$
 1324 $\left. \text{neg}^j(\bar{a}, \bar{y}) \right)$ and hence $A \models \varphi$. \square

1325 Since the union of finitely many testable properties is testable (see e.g. [18]), it is
 1326 sufficient to show that the property \mathcal{P}_φ is testable where φ is of the form

$$1327 \quad (6.2) \quad \varphi = \exists \bar{x} \forall \bar{y} \chi(\bar{x}, \bar{y}),$$

$$1328 \quad \text{where } \chi(\bar{x}, \bar{y}) = \left[\iota^M(\bar{x}) \wedge \bigvee_{j \in J_M} \left(\beta^j(\bar{y}) \wedge \text{pos}^j(\bar{x}, \bar{y}) \wedge \text{neg}^j(\bar{x}, \bar{y}) \right) \right],$$

1329 for some $M \in \mathfrak{M}$. In the following, we will enforce that for every conjunctive clause
 1330 of the big disjunction of χ , the universally quantified variables induce a specific sub-
 1331 structure.

1332 For $j \in J_M$ let $\mathfrak{H}_j \subseteq \mathcal{C}_{\sigma, d}$ be a maximal set of pairwise non-isomorphic structures
 1333 H such that $H \models \beta^j(\bar{b})$ for some $\bar{b} = (b_1, \dots, b_\ell) \in U(H)^\ell$ with $\{b_1, \dots, b_\ell\} = U(H)$.

1334 CLAIM 6.7. *We have*

$$1335 \quad \varphi \equiv_d \exists \bar{x} \forall \bar{y} \left[\iota^M(\bar{x}) \wedge \bigvee_{\substack{H \in \mathfrak{H}_j, \\ j \in J_M}} \left(\iota^H(\bar{y}) \wedge \text{pos}^j(\bar{x}, \bar{y}) \wedge \text{neg}^j(\bar{x}, \bar{y}) \right) \right].$$

1336 *Proof.* Let $A \in \mathcal{C}_{\sigma, d}$ and $\bar{a} = (a_1, \dots, a_k) \in U(A)^k$. First assume that $A \models$
 1337 $\forall \bar{y} \chi(\bar{a}, \bar{y})$. Hence for any tuple $\bar{b} \in U(A)^\ell$ there is an index $j \in J_M$ such that $A \models$
 1338 $\beta^j(\bar{b}) \wedge \text{pos}^j(\bar{a}, \bar{b}) \wedge \text{neg}^j(\bar{a}, \bar{b})$. Then $A \models \beta^j(\bar{b})$ implies that $A[\{b_1, \dots, b_\ell\}] \cong H$ for
 1339 some $H \in \mathfrak{H}_j$. Hence $A \models \iota^H(\bar{b})$ and $A \models \left[\iota^M(\bar{a}) \wedge \bigvee_{\substack{H \in \mathfrak{H}_j, \\ j \in J_M}} \left(\iota^H(\bar{b}) \wedge \text{pos}^j(\bar{a}, \bar{b}) \wedge \right.$
 1340 $\left. \text{neg}^j(\bar{a}, \bar{b}) \right) \right]$.

1341 For the other direction, we let $A \models \forall \bar{y} \left[\iota^M(\bar{a}) \wedge \bigvee_{\substack{H \in \mathfrak{H}_j, \\ j \in J_M}} \left(\iota^H(\bar{y}) \wedge \text{pos}^j(\bar{a}, \bar{y}) \wedge \right.$
 1342 $\left. \text{neg}^j(\bar{a}, \bar{y}) \right) \right]$. Then for every tuple $\bar{b} \in U(A)^\ell$ there is an index $j \in J_M$ and $H \in \mathfrak{H}_j$
 1343 such that $H \models \iota^H(\bar{b}) \wedge \text{pos}^j(\bar{a}, \bar{b}) \wedge \text{neg}^j(\bar{a}, \bar{b})$. Therefore $A[\{b_1, \dots, b_\ell\}] \cong H$ and we
 1344 know that $A \models \beta^j(\bar{b})$. Therefore $A \models \beta^j(\bar{b}) \wedge \text{pos}^j(\bar{a}, \bar{b}) \wedge \text{neg}^j(\bar{a}, \bar{b})$ and since this is
 1345 true for any $\bar{b} \in U(A)^\ell$ we get $A \models \varphi$. \square

1346 Thus, it suffices to assume that

$$1347 \quad (6.3) \quad \varphi = \exists \bar{x} \forall \bar{y} \chi(\bar{x}, \bar{y}),$$

$$1348 \quad \text{where } \chi(\bar{x}, \bar{y}) := \left[\iota^M(\bar{x}) \wedge \bigvee_{\substack{H \in \mathfrak{H}_j, \\ j \in J_M}} \left(\iota^H(\bar{y}) \wedge \text{pos}^j(\bar{x}, \bar{y}) \wedge \text{neg}^j(\bar{x}, \bar{y}) \right) \right]$$

1349 for some $M \in \mathfrak{M}$.

1350 Next we will define a universally quantified formula ψ and show that \mathcal{P}_φ is in-
 1351 distinguishable from the property \mathcal{P}_ψ . To do so we will need the two claims below.
 1352 Intuitively, Claim 6.8 says that models of φ of bounded degree do not have many ‘inter-
 1353 actions’ between existential and universal variables – only a constant number of tuples
 1354 in relations combine both types of variables. Note that for a structure A and tuples
 1355 $\bar{a} \in U(A)^k$, $\bar{b} = (b_1, \dots, b_\ell) \in U(A)^\ell$ the condition $A \models \iota^H(\bar{b}) \wedge \text{pos}^j(\bar{a}, \bar{b}) \wedge \text{neg}^j(\bar{a}, \bar{b})$
 1356 can force an element of \bar{b} to be in a tuple (of a relation of A) with an element of \bar{a} , even
 1357 if $\text{pos}^j(\bar{x}, \bar{y})$ only contains literals of the form $x_i = y_{i'}$. (For example, it may be the
 1358 case that for some tuple $\bar{b}' \in \{b_1, \dots, b_\ell\}^\ell$, every clause $\iota^{H'}(\bar{y}) \wedge \text{pos}^{j'}(\bar{x}, \bar{y}) \wedge \text{neg}^{j'}(\bar{x}, \bar{y})$
 1359 for which $A \models \iota^{H'}(\bar{b}') \wedge \text{pos}^{j'}(\bar{a}, \bar{b}') \wedge \text{neg}^{j'}(\bar{a}, \bar{b}')$ forces a tuple to contain some element
 1360 of \bar{b}' and some element of \bar{a} .) We will now define a set J to pick out the clauses that
 1361 do not force a tuple to contain both an element from \bar{a} and \bar{b} . Note that we still allow
 1362 elements from \bar{b} to be amongst the elements in \bar{a} . In Claim 6.8 we show that for every
 1363 $A \in \mathcal{C}_{\sigma, d}$, $\bar{a} \in U(A)^k$ for which $A \models \forall \bar{y} \chi(\bar{a}, \bar{y})$ there are a constant number of tuples
 1364 $\bar{b} \in U(A)^\ell$ that only satisfy clauses which force a tuple to contain both an element
 1365 from \bar{a} and from \bar{b} .

1366 Let $j \in J_M$, $H \in \mathfrak{H}_j$ and $\bar{h} = (h_1, \dots, h_\ell) \in U(H)^\ell$ such that $H \models \iota^H(\bar{h})$.
 1367 We define the set $P_{j, H} := \{h_i \mid i \in \{1, \dots, \ell\}, \text{pos}^j(\bar{x}, \bar{y}) \text{ does not contain } y_i =$
 1368 $x_{i'} \text{ for any } i' \in \{1, \dots, k\}\}$. Now we let $J \subseteq J_M \times \mathcal{C}_{\sigma, d}$ be the set of pairs (j, H) ,
 1369 with $H \in \mathfrak{H}_j$ such that the disjoint union $M \sqcup H[P_{j, H}] \models \varphi$. Now J precisely specifies
 1370 the clauses that can be satisfied by a structure A and tuple $\bar{a} \in U(A)^k$ and $\bar{b} \in U(A)^\ell$
 1371 where A does not contain any tuples both containing elements from \bar{a} and \bar{b} .

1372 CLAIM 6.8. *Let $A \in \mathcal{C}_{\sigma, d}$ and $\bar{a} = (a_1, \dots, a_k) \in U(A)^k$. If $A \models \forall \bar{y} \chi(\bar{a}, \bar{y})$ then*
 1373 *there are at most $k \cdot d$ tuples $\bar{b} \in U(A)^\ell$ such that $A \not\models \bigvee_{(j, H) \in J} (\iota^H(\bar{b}) \wedge \text{pos}^j(\bar{a}, \bar{b}) \wedge$*

1374 $\text{neg}^j(\bar{a}, \bar{b})$.

1375 *Proof.* Since $A \models \forall \bar{y} \chi(\bar{a}, \bar{y})$, it holds that $A \models \forall \bar{y} \bigvee_{\substack{H \in \mathfrak{S}_j \\ j \in J_M}} \left(\iota^H(\bar{y}) \wedge \text{pos}^j(\bar{a}, \bar{y}) \wedge \right.$
 1376 $\left. \text{neg}^j(\bar{a}, \bar{y}) \right)$ by Equation(6.3). Now let $B := \{\bar{b} \in U(A)^\ell \mid A \not\models \bigvee_{(j,H) \in J} (\iota^H(\bar{b}) \wedge$
 1377 $\text{pos}^j(\bar{a}, \bar{b}) \wedge \text{neg}^j(\bar{a}, \bar{b}))\} \subseteq U(A)^\ell$. Then each $\bar{b} \in B$ adds at least one to $\sum_{i=1}^k \text{deg}_A(a_i)$.
 1378 Since $A \in \mathcal{C}_{\sigma,d}$ implies that $\sum_{i=1}^k \text{deg}_A(a_i) \leq k \cdot d$ we get that $|B| \leq k \cdot d$. \square

1379 **CLAIM 6.9.** *Let ψ be a formula of the form $\psi = \forall \bar{z} \chi(\bar{z})$ where $\bar{z} = (z_1, \dots, z_t)$ is*
 1380 *a tuple of variables and $\chi(\bar{z})$ is a quantifier-free formula. Let $A \in \mathcal{C}_{\sigma,d}$ with $|U(A)| >$*
 1381 *$d \cdot \text{ar}(\sigma) \cdot t$ and let $b \in A$ be an arbitrary element. Let $A \models \psi$ and let A' be obtained*
 1382 *from A by ‘isolating’ b , i. e. by deleting all tuples containing b from $R(A)$ for every*
 1383 *$R \in \sigma$. Then $A' \models \psi$.*

1384 *Proof.* First note that $A' \models \chi(\bar{a})$ for any tuple $\bar{a} = (a_1, \dots, a_t) \in (A \setminus \{b\})^t$ as no
 1385 tuple over the set of elements $\{a_1, \dots, a_t\}$ has been deleted. Let $\bar{a} = (a_1, \dots, a_t) \in$
 1386 $U(A)^t$ be a tuple containing b . Pick $b' \in U(A)$ such that $\text{dist}_A(a_j, b') > 1$ for every
 1387 $j \in \{1, \dots, t\}$. Such an element exists as $|U(A)| > d \cdot \text{ar}(R) \cdot t$. Let $\bar{a}' = (a'_1, \dots, a'_t)$
 1388 be the tuple obtained from \bar{a} by replacing any occurrence of b by b' . Hence $a_j \mapsto a'_j$
 1389 defines an isomorphism from $A'[\{a_1, \dots, a_t\}]$ to $A[\{a'_1, \dots, a'_t\}]$ since b is an isolated
 1390 element in $A'[\{a_1, \dots, a_t\}]$ and b' is an isolated element in $A[\{a'_1, \dots, a'_t\}]$. Since
 1391 $A \models \chi(\bar{a}')$, it follows that $A' \models \chi(\bar{a})$. \square

1392 Let $J' \subseteq J$ be the set of all pairs (j, H) for which $\text{pos}^j(\bar{x}, \bar{y})$ is the empty conjunction.
 1393 J' contains (j, H) for which we want to use $\iota^H(\bar{y})$ to define the formula ψ .

1394 **CLAIM 6.10.** *The property \mathcal{P}_φ with φ as in (6.3) is indistinguishable from the*
 1395 *property \mathcal{P}_ψ where $\psi := \forall \bar{y} \bigvee_{(j,H) \in J'} \iota^H(\bar{y})$.*

1396 *Proof.* Let $\epsilon > 0$ and $N(\epsilon) = N := \frac{k \cdot \ell^2 \cdot d \cdot \text{ar}(R)}{\epsilon}$ and $A \in \mathcal{C}_{\sigma,d}$ be any structure with
 1397 $|U(A)| > N$.

1398 First assume that $A \models \varphi$. The strategy is to isolate any element b which is
 1399 contained in a tuple $\bar{b} \in U(A)^\ell$ such that $A \not\models \bigvee_{(j,H) \in J'} \iota^H(\bar{b})$ by deleting all tuples
 1400 containing b . This will result in a structure which is ϵ -close to A and a model of ψ .

1401 Let $\bar{a} \in U(A)^k$ be a tuple such that $A \models \forall \bar{y} \chi(\bar{a}, \bar{y})$. Let $B \subseteq U(A)^\ell$ be the set
 1402 of tuples $\bar{b} \in U(A)^\ell$ such that $A \not\models \bigvee_{(j,H) \in J} (\iota^H(\bar{b}) \wedge \text{pos}^j(\bar{a}, \bar{b}) \wedge \text{neg}^j(\bar{a}, \bar{b}))$. Then
 1403 $|B| \leq k \cdot d$ by Claim 6.8. Hence the structure A' obtained from A by deleting all
 1404 tuples containing an element of $C := \{a_1, \dots, a_k\} \cup \{b \in A \mid \text{there is } (b_1, \dots, b_\ell) \in$
 1405 $B \text{ such that } b \in \{b_1, \dots, b_\ell\}\}$ is ϵ -close to A . Since $A \models \forall \bar{y} \chi(\bar{a}, \bar{y})$ implies $A \models$
 1406 $\forall \bar{y} \bigvee_{\substack{H \in \mathfrak{S}_j \\ j \in J_M}} \iota^H(\bar{y})$, by Claim 6.9 we know that $A' \models \forall \bar{y} \bigvee_{\substack{H \in \mathfrak{S}_j \\ j \in J_M}} \iota^H(\bar{y})$. For any tu-
 1407 ple $\bar{b} = (b_1, \dots, b_\ell) \in (U(A) \setminus C)^\ell$ we have by definition of J' that $A \models \iota^H(\bar{b})$
 1408 for some $(j, H) \in J'$. Furthermore $A[\{b_1, \dots, b_\ell\}] = A'[\{b_1, \dots, b_\ell\}]$ and hence
 1409 $A' \models \bigvee_{(j,H) \in J'} \iota^H(\bar{b})$. Let $\bar{b} = (b_1, \dots, b_\ell) \in U(A)^\ell$ be any tuple containing elements
 1410 from C and let $c_1, \dots, c_t \in C$ be those elements. Pick t elements $c'_1, \dots, c'_t \in U(A) \setminus C$
 1411 such that $\text{dist}_A(a_i, c'_i) > 1$, $\text{dist}_A(c'_i, b_i) > 1$ and $\text{dist}_A(c'_i, c'_i) > 1$ for suitable i, i' .
 1412 This is possible as $|U(A)| > (k + 2\ell) \cdot d \cdot \text{ar}(R)$ which guarantees the existence of
 1413 $k + 2\ell$ elements of pairwise distance greater than 1. Let $\bar{b}' = (b'_1, \dots, b'_\ell)$ be the vector
 1414 obtained from \bar{b} by replacing c_i with c'_i . Since $\bar{b}' \in U(A)^\ell$ there must be $j', H' \in \mathfrak{S}_j$
 1415 such that $A \models \iota^{H'}(\bar{b}') \wedge \text{pos}^{j'}(\bar{a}, \bar{b}') \wedge \text{neg}^{j'}(\bar{a}, \bar{b}')$. By choice of c'_1, \dots, c'_t we have that
 1416 $\text{pos}_{j'}(\bar{x}, \bar{y})$ must be the empty conjunction and hence $(j', H') \in J'$. Since additionally
 1417 $b_i \mapsto b'_i$ defines an isomorphism of $A[\{b_1, \dots, b'_\ell\}]$ and $A'[\{b_1, \dots, b_\ell\}]$ this implies that

1418 $A' \models \bigvee_{(j,H) \in J'} \iota^H(\bar{b})$ for all $\bar{b} \in U(A)^\ell$ and hence $A' \models \psi$.

1419

1420 Now we prove the other direction. Let $A \models \psi$ with $|U(A)| > N$. The idea here is
 1421 to plant the structure M somewhere in A . While this takes less than an ϵ -fraction of
 1422 edge modifications the resulting structure will be a model of φ .

Take any set $B \subseteq A$ of $|U(M)|$ elements. Let A' be the structure obtained
 from A by deleting all edges incident to any element contained in B . Let A'' be the
 structure obtained from A' by adding all tuples such that the structure induced by B is
 isomorphic to M . This takes no more than $2\ell \cdot d \cdot \text{ar}(R) < \epsilon \cdot d \cdot |U(A)|$ edge modifications.
 Let $\bar{a} \in B^k$ be such that $A \models \iota^M(\bar{a})$. By Claim 6.9 we get $A' \models \psi$. Therefore pick
 any tuple $\bar{b} = (b_1, \dots, b_\ell) \in (U(A) \setminus B)^\ell$. Since by construction we have that all b_i 's
 are of distance at least two from \bar{a} we have that $A'' \models \bigvee_{(j,H) \in J'} (\iota^H(\bar{b}) \wedge \text{neg}^j(\bar{a}, \bar{b}))$.

By choice of M we also know that $A'' \models \bigvee_{\substack{H \in \mathfrak{H}_j, \\ j \in J_M}} (\iota^H(\bar{b}) \wedge \text{pos}^j(\bar{a}, \bar{b}) \wedge \text{neg}^j(\bar{a}, \bar{b}))$

for all $\bar{b} \in B^\ell$. Therefore pick $\bar{b} = (b_1, \dots, b_\ell)$ containing both elements from B
 and from $U(A) \setminus B$. Now pick a tuple $\bar{b}' = (b'_1, \dots, b'_\ell) \in (U(A) \setminus B)^\ell$ that equals
 \bar{b} in all positions containing an element from $U(A) \setminus B$. As noted before there is
 $(j, H) \in J'$ such that $A'' \models (\iota^H(\bar{b}') \wedge \text{neg}^j(\bar{a}, \bar{b}'))$. Hence $A''[\{b'_1, \dots, b'_\ell\}]$ is isomorphic
 to H and further because $(j, H) \in J'$ the set $P_{j,H}$ (used in the definition of J) is
 the entire universe of H . Since $J' \subseteq J$ this means that by the definition of J we
 get $A''[\{a_1, \dots, a_k, b'_1 \dots b'_\ell\}] \cong A''[\{a_1, \dots, a_k\}] \sqcup A''[\{b'_1 \dots b'_\ell\}] \cong M \sqcup H[P_{j,H}] \models \varphi$.
 Since $\bar{b} \in \{a_1, \dots, a_k, b'_1 \dots b'_\ell\}^\ell$ this implies

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

1423 Then $A'' \models \bigvee_{\substack{H \in \mathfrak{H}_j, \\ j \in J_M}} (\iota^H(\bar{b}) \wedge \text{pos}^j(\bar{a}, \bar{b}) \wedge \text{neg}^j(\bar{a}, \bar{b}))$ and hence $A'' \models \varphi$. \square

1424 Since $\psi \in \Pi_1$ we have that \mathcal{P}_ψ is testable, and hence \mathcal{P}_φ is testable by Claim 6.10. \square

1425 **7. GSF-locality is not sufficient for proximity oblivious testing.** In this
 1426 section we show that the property $\mathcal{P}_{\text{graph}}$ can be defined by a generalised notion
 1427 of forbidden subgraph introduced in [22] (Lemma 7.14). Here a subgraph is only
 1428 forbidden if it is connected to the rest of the graph in a predefined way, i.e. for a
 1429 vertex in a forbidden subgraph we can specify that it cannot have neighbours which
 1430 are not contained in the subgraph itself. Combining our results we show that not every
 1431 property definable by generalised forbidden subgraphs is testable in the bounded-
 1432 degree model (Theorem 7.5). This implies a negative answer to a question posed
 1433 by Goldreich and Ron in [22] (Question 1) which asks whether a small number of
 1434 appearances of generalised forbidden subgraphs can be fixed with a small number
 1435 of edge modification or whether any way of fixing the appearances invokes a chain
 1436 reaction of necessary edge modifications. In the following we introduce the notions
 1437 and results needed from [22].

1438 **7.1. Generalised subgraph freeness.** In the following, we present the formal
 1439 definitions of generalised subgraph freeness, GSF-local properties and the notion of
 1440 non-propagation, which were introduced in [22].

1441 **DEFINITION 7.1** (Generalized subgraph freeness (GSF)). *A marked graph is a*
 1442 *graph with each vertex marked as either ‘full’ or ‘semifull’ or ‘partial’. An embedding*
 1443 *of a marked graph F into a graph G is an injective map $f : V(F) \rightarrow V(G)$ such that*
 1444 *for every $v \in V(F)$ the following three conditions hold.*

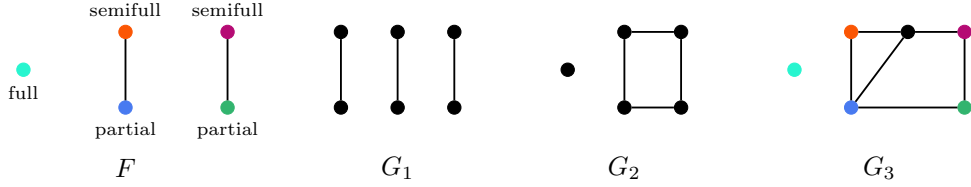


Fig. 7: In the depicted example, G_1 and G_2 are F -free (for G_1 we cannot find an embedding satisfying the condition for the full vertex and for G_2 we cannot find an embedding satisfying the condition for the semiful vertices). On the other hand, F can be embedded into G_3 (the embedding is indicated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

- 1445 1. If v is marked ‘full’, then $N_1^G(f(v)) = f(N_1^F(v))$.
- 1446 2. If v is marked ‘semifull’, then $N_1^G(f(v)) \cap f(V(F)) = f(N_1^F(v))$.
- 1447 3. If v is marked ‘partial’, then $N_1^G(f(v)) \supseteq f(N_1^F(v))$.

1448 The graph G is called F -free if there is no embedding of F into G . For a set of marked
 1449 graphs \mathcal{F} , a graph G is called \mathcal{F} -free if it is F -free for every $F \in \mathcal{F}$.

1450 We refer to Figure 7 for an illustration of the definition of GSF. Based on the above
 1451 definition of generalised subgraph freeness, we can define GSF-local properties.

1452 DEFINITION 7.2 (GSF-local properties). Let $\mathcal{P} = \bigcup_{n \in \mathbb{N}} \mathcal{P}_n$ be a graph property
 1453 where $\mathcal{P}_n = \{G \in \mathcal{P} \mid |V(G)| = n\}$ and $\overline{\mathcal{F}} = (\mathcal{F}_n)_{n \in \mathbb{N}}$ a sequence of sets of marked
 1454 graphs. \mathcal{P} is called $\overline{\mathcal{F}}$ -local if there exists an integer s such that for every n the
 1455 following conditions hold.

- 1456 1. \mathcal{F}_n is a set of marked graphs, each of size at most s .
- 1457 2. \mathcal{P}_n equals the set of n -vertex graphs that are \mathcal{F}_n -free.

1458 \mathcal{P} is called GSF-local if there is a sequence $\overline{\mathcal{F}} = (\mathcal{F}_n)_{n \in \mathbb{N}}$ of sets of marked graphs
 1459 such that \mathcal{P} is $\overline{\mathcal{F}}$ -local.

1460 The following concept of a non-propagating condition for a sequence of sets of marked
 1461 graphs was introduced in [22] to investigate constant-query POTs.

1462 DEFINITION 7.3 (Non-propagating). Let $\overline{\mathcal{F}} = (\mathcal{F}_n)_{n \in \mathbb{N}}$ be a sequence of sets of
 1463 marked graphs.

- 1464 • For a graph G , a subset $B \subset V(G)$ covers \mathcal{F}_n in G if for every marked graph
 1465 $F \in \mathcal{F}_n$ and every embedding of F in G , at least one vertex of F is mapped
 1466 to a vertex in B .
- 1467 • The sequence $\overline{\mathcal{F}}$ is non-propagating if there exists a (monotonically non-
 1468 decreasing) function $\tau : (0, 1] \rightarrow (0, 1]$ such that the following two conditions
 1469 hold.
 - 1470 1. For every $\epsilon > 0$ there exists $\beta > 0$ such that $\tau(\beta) < \epsilon$.
 - 1471 2. For every graph G and every $B \subset V(G)$ such that B covers \mathcal{F}_n in
 1472 G , either G is $\tau(|B|/n)$ -close to being \mathcal{F}_n -free or there are no n -vertex
 1473 graphs that are \mathcal{F}_n -free.

1474 A GSF-local property \mathcal{P} is non-propagating if there exists a non-propagating
 1475 sequence $\overline{\mathcal{F}}$ such that \mathcal{P} is $\overline{\mathcal{F}}$ -local.

1476 In the above definition, the set B can be viewed as the set involving necessary modi-
 1477 fications for repairing a graph G that does not satisfy the property \mathcal{P} that is $\overline{\mathcal{F}}$ -local,
 1478 and the second condition says we do not need to modify G “much beyond” B . In par-
 1479 ticular, it implies that we can repair G without triggering a global “chain reaction”.

1480 Goldreich and Ron gave the following characterization for the proximity-oblivious
 1481 testable properties in the bounded-degree graph model.

1482 **THEOREM 7.4** (Theorem 5.5 in [22]). *A graph property \mathcal{P} has a constant-query
 1483 proximity-oblivious tester if and only if \mathcal{P} is GSF-local and non-propagating.*

1484 The following open question was raised in [22].

1485 **OPEN QUESTION 1** (Are all GSF-local properties non-propagating?). *Is it the
 1486 case that for every GSF-local property $\mathcal{P} = \bigcup_{n \in \mathbb{N}} \mathcal{P}_n$, there is a sequence $\overline{\mathcal{F}} = (\mathcal{F}_n)_{n \in \mathbb{N}}$
 1487 that is non-propagating and \mathcal{P} is $\overline{\mathcal{F}}$ -local?*

1488 We are now able to state our theorem answering Question 1. The rest of this section
 1489 is dedicated to the proof of Theorem 7.5.

1490 **THEOREM 7.5.** *There exists a GSF-local property of graphs of bounded degree 3
 1491 that is not testable in the bounded-degree graph model. Thus, not all GSF-local prop-
 1492 erties are non-propagating.*

1493 **7.2. Relating different notions of locality.** In this section we define proper-
 1494 ties by prescribing upper and lower bounds on the number of occurrences of neigh-
 1495 bourhood types. These bounds are given by *neighbourhood profiles* which we will
 1496 define formally below. We use these properties to give a natural characterization of
 1497 FO properties of bounded-degree structures in Lemma 7.7, which is a straightforward
 1498 consequence of Hanf's Theorem (Theorem 2.1). We use this characterization to es-
 1499 tablish links between FO definability and GSF-locality. This connection is the key
 1500 ingredient in the proof of our main theorem.

1501 Observe that for fixed $r, d \in \mathbb{N}$ and σ , there are only finitely many r -types in
 1502 structures in $\mathcal{C}_{\sigma, d}$. For any signature σ and $d, r \in \mathbb{N}$ we let $n_{d, r, \sigma} \in \mathbb{N}$ be the number
 1503 of different r -types of σ -structures of degree at most d . Assuming that for all $d, r \in \mathbb{N}$
 1504 the r -neighbourhood-types of σ -structures of degree at most d are ordered, we let
 1505 $\tau_{d, r, \sigma}^i$ denote the i -th such neighbourhood type, for $i \in \{1, \dots, n_{d, r, \sigma}\}$. With each
 1506 σ -structure $A \in \mathcal{C}_{\sigma, d}$ we associate its *r -histrogram vector* $\bar{v}_{d, r, \sigma}(A)$, given by

$$1508 \quad (\bar{v}_{d, r, \sigma}(A))_i := |\{a \in U(A) \mid \mathcal{N}_r^A(a) \in \tau_{d, r, \sigma}^i\}|.$$

1509 We let

$$1510 \quad \mathcal{J} := \{[k, l] \mid k \leq l \in \mathbb{N}\} \cup \{[k, \infty) \mid k \in \mathbb{N}\}$$

1511 be the set of all closed or half-closed, infinite intervals with natural lower/upper
 1512 bounds.

1513 **DEFINITION 7.6.** *Let σ be a signature and $d, r \in \mathbb{N}$.*

- 1514 1. *An r -neighbourhood profile of degree d is a function $\rho : \{1, \dots, n_{d, r, \sigma}\} \rightarrow \mathcal{J}$.*
- 1515 2. *For a structure $A \in \mathcal{C}_{\sigma, d}$, we say that A obeys ρ , denoted by $A \sim \rho$, if*

$$1516 \quad (\bar{v}_{d, r, \sigma}(A))_i \in \rho(i) \text{ for all } i \in \{1, \dots, n_{d, r, \sigma}\}.$$

1517 *Let \mathcal{P}_ρ be the set of structures A that obey ρ , i.e., $\mathcal{P}_\rho = \{A \in \mathcal{C}_{\sigma, d} \mid A \sim \rho\}$.*

- 1518 3. *We say that a property \mathcal{P} is defined by a finite union of neighbourhood profiles
 1519 if there is $k \in \mathbb{N}$ such that $\mathcal{P} = \bigcup_{1 \leq i \leq k} \mathcal{P}_{\rho_i}$ where ρ_i is an r_i -neighbourhood
 1520 profile and $r_i \in \mathbb{N}$ for every $i \in \{1, \dots, k\}$.*

1521 We let $n_{d, r} := n_{d, r, \sigma_{\text{graph}}}$ denote the total number of r -types of directed graphs
 1522 of degree at most d . We fix an ordering of the types and let $\tau_{d, r}^i := \tau_{d, r, \sigma_{\text{graph}}}^i$ be the



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

1523 i -th r -type of bounded degree d , for any $i \in \{1, \dots, n_{d,r}\}$. Further, for a graph G let
 1524 $\bar{v}_{d,r}(G)$ denote the r -histogram vector of G . Note if G is undirected, for any type $\tau_{d,r}^i$
 1525 where the edge relation is not symmetric we have that $(\bar{v}_{d,r}(G))_i = 0$ and therefore
 1526 in any r -neighbourhood profile ρ for graphs we have $\rho(i) = [0, 0]$ for any type $\tau_{d,r}^i$
 1527 which is not symmetric. For convenience, for undirected graphs we will ignore the
 1528 non-symmetric types.

1529 Let us consider the following example in which we find a representation by neigh-
 1530 bourhood profiles for an FO-property.

1531 **EXAMPLE 2.** Consider the following FO-sentence.

$$1532 \quad \varphi := \forall x \forall y \neg E(x, y) \vee \forall x \exists y_1 \exists y_2 \left(y_1 \neq y_2 \wedge E(x, y_1) \wedge E(x, y_2) \right. \\ 1533 \quad \left. \wedge \forall z (z \neq y_1 \wedge z \neq y_2) \rightarrow \neg E(x, z) \right).$$

1534 The property P_φ defined by the sentence φ is the property containing all edgeless
 1535 graphs and all graphs that are disjoint unions of cycles.

1536 For degree bound 2 all 1-types are listed in Figure 8. Let $\rho_1 : \{1, \dots, 4\} \rightarrow \mathfrak{J}$ be
 1537 the neighbourhood profile defined by $\rho_1(1) = [0, \infty)$ and $\rho_1(i) = [0, 0]$ for $i \in \{2, 3, 4\}$.
 1538 Furthermore, let $\rho_2 : \{1, \dots, 4\} \rightarrow \mathfrak{J}$ be the neighbourhood profile defined by $\rho_2(i) =$
 1539 $[0, \infty)$ for $i \in \{3, 4\}$ and $\rho_2(j) = [0, 0]$ for $j \in \{1, 2\}$. It is easy to observe that the
 1540 properties P_φ and $P_{\rho_1} \cup P_{\rho_2}$ are equal.

1541 Indeed representing FO-properties by neighbourhood profiles works in general. We
 1542 now give a lemma showing that bounded-degree FO properties can be equivalently
 1543 defined as finite unions of properties defined by neighbourhood profiles. Here the
 1544 technicalities that arise are due to Hanf normal form not requiring the locality-radius
 1545 of all Hanf-sentences to be the same.

1546 **LEMMA 7.7.** For every non-empty property $\mathcal{P} \subseteq \mathcal{C}_{\sigma,d}$, \mathcal{P} is FO definable on $\mathcal{C}_{\sigma,d}$ if
 1547 and only if \mathcal{P} can be obtained as a finite union of properties defined by neighbourhood
 1548 profiles.

1549 *Proof.* For the first direction assume φ is an FO-sentence. Then by Hanf's The-
 1550 orem (Theorem 2.1) there is a sentence ψ in Hanf normal form such that $\mathcal{P}_\varphi = \mathcal{P}_\psi$.

1551 We will first convert ψ into a sentence in Hanf normal form where every Hanf
 1552 sentence appearing has the same locality radius. Let $r \in \mathbb{N}$ be the maximum locality
 1553 radius appearing in ψ , and let $\varphi_\tau^{\geq m} := \exists^{\geq m} x \varphi_\tau(x)$ be a Hanf sentence, where τ is
 1554 an r' -type for some $r' \leq r$. Let τ_1, \dots, τ_k be a list of all r -types of bounded degree d
 1555 for which $(\mathcal{N}_{r'}^B(b), b) \in \tau$ for $(B, b) \in \tau_i$, for every $1 \leq i \leq k$. Let Π be the set of all
 1556 partitions of m into k parts. Let

$$1557 \quad \tilde{\varphi}_\tau^{\geq m} := \bigvee_{(m_1, \dots, m_k) \in \Pi} \bigwedge_{i=1}^k \exists^{\geq m_i} x \varphi_{\tau_i}(x).$$

1558

 1559 CLAIM 7.8. $\varphi_\tau^{\geq m}$ is d -equivalent to $\tilde{\varphi}_\tau^{\geq m}$.

1560 *Proof.* Assume that $A \in \mathcal{C}_d$ satisfies $\varphi_\tau^{\geq m}$, and assume that a_1, \dots, a_m are m
 1561 distinct elements with $(\mathcal{N}_r^A(a_j), a_j) \in \tau$, for every $1 \leq j \leq m$. Let $\tilde{\tau}_j$ be the r -type
 1562 for which $(\mathcal{N}_r^A(a_j), a_j) \in \tilde{\tau}_j$. By choice of τ_1, \dots, τ_k , we get that there are indices
 1563 i_1, \dots, i_m such that $\tilde{\tau}_j = \tau_{i_j}$. For $i \in \{1, \dots, k\}$ let $m_i = |\{j \in \{1, \dots, m\} \mid i_j = i\}|$.
 1564 Hence $A \models \bigwedge_{i=1}^k \exists^{\geq m_i} x \varphi_{\tau_i}(x)$ and since additionally $(m_1, \dots, m_k) \in \Pi$ this implies
 1565 $A \models \tilde{\varphi}_\tau^{\geq m}$.

1566 On the other hand, let $A \in \mathcal{C}_d$ satisfy $\tilde{\varphi}_\tau^{\geq m}$, and let $(m_1, \dots, m_k) \in \Pi$ be a
 1567 partition of m such that $A \models \bigwedge_{i=1}^k \exists^{\geq m_i} x \varphi_{\tau_i}(x)$. For every $1 \leq i \leq k$, let $a_1^i, \dots, a_{m_i}^i$
 1568 be m_i distinct elements such that $(\mathcal{N}_r^A(a_j^i), a_j^i) \in \tau_i$, for every $1 \leq j \leq m_i$. By choice
 1569 of τ_1, \dots, τ_k , we get that $(\mathcal{N}_r^A(a_j^i), a_j^i) \in \tau$, for every pair $1 \leq i \leq k, 1 \leq j \leq m_i$. But
 1570 since $m_1 + \dots + m_k = m$ this implies that $A \models \varphi_\tau^{\geq m}$. This proves that $\varphi_\tau^{\geq m}$ and $\tilde{\varphi}_\tau^{\geq m}$
 1571 are d -equivalent. \square

1572 Let ψ' be the formula in which every Hanf-sentence $\varphi_\tau^{\geq m}$ for which τ is an r' -type for
 1573 some $r' < r$ gets replaced by $\tilde{\varphi}_\tau^{\geq m}$. By a simple inductive argument using Claim 7.8,
 1574 we get that ψ is d -equivalent to ψ' , and hence $\mathcal{P}_\varphi = \mathcal{P}_\psi = \mathcal{P}_{\psi'}$. Furthermore since
 1575 $\tilde{\varphi}_\tau^{\geq m}$ is a Boolean combination of Hanf-sentences for every $\varphi_\tau^{\geq m}$, and any Boolean
 1576 combination of Boolean combinations is a Boolean combination itself, ψ' is in Hanf
 1577 normal form. Furthermore, every Hanf-sentence appearing in ψ' has locality radius r
 1578 by construction.

1579 Since any Boolean combination can be converted into disjunctive normal form,
 1580 we can assume that ψ' is a disjunction of sentences ξ of the form

$$1581 \quad \xi = \bigwedge_{j=1}^k \exists^{\geq m_j} x \varphi_{\tau_j}(x) \wedge \bigwedge_{j=k+1}^{\ell} \neg \exists^{\geq m_j+1} x \varphi_{\tau_j}(x),$$

1582 where $\ell \in \mathbb{N}_{\geq 1}$, $1 \leq k \leq \ell$, $m_i \in \mathbb{N}_{\geq 1}$ and τ_i is an r -type for every $1 \leq i \leq \ell$. We can
 1583 further assume that every sentence in the disjunction ψ' is satisfiable by some $A \in \mathcal{C}_d$,
 1584 as any sentence with no bounded degree d model can be removed from ψ' .

1585 Let $\tilde{\tau}_1, \dots, \tilde{\tau}_t$ be a list of all r -types of bounded degree d in the order we fixed.
 1586 Let $k_i := \max(\{m_j \mid 1 \leq j \leq k, \tau_j = \tilde{\tau}_i\} \cup \{0\})$ and $\ell_i := \min(\{m_j \mid k+1 \leq j \leq$
 1587 $\ell, \tau_j = \tilde{\tau}_i\} \cup \{\infty\})$ for every $i \in \{1, \dots, t\}$. Since ξ has at least one bounded-degree
 1588 model, $k_i \leq \ell_i$ for every $i \in \{1, \dots, t\}$. Let $\rho : \{1, \dots, t\} \rightarrow \mathcal{I}$ be the neighbourhood
 1589 profile defined by $\rho(i) := [k_i, \ell_i]$ if $\ell_i < \infty$ and $\rho(i) := [k_i, \ell_i)$ otherwise. Then by
 1590 construction, we get that $\mathcal{P}_\rho = \mathcal{P}_\xi$. Since ψ' is a disjunction of formulas, each of
 1591 which defines a property which can be defined by some neighbourhood profile, we get
 1592 that $\mathcal{P}_{\psi'}$ must be a finite union of properties defined by some neighbourhood profile.
 1593

1594 On the other hand, for every r -neighbourhood profile ρ of degree d , τ_1, \dots, τ_t a
 1595 list of all r -types of bounded degree d in the order fixed and the formula

$$1596 \quad \varphi_\rho := \bigwedge_{\substack{i \in \{1, \dots, t\}, \\ \rho(i) = [k_i, \ell_i]}} \left(\exists^{\geq k_i} x \varphi_{\tau_i}(x) \wedge \neg \exists^{\geq \ell_i+1} x \varphi_{\tau_i}(x) \right) \wedge \bigwedge_{\substack{i \in \{1, \dots, t\}, \\ \rho(i) = [k_i, \infty)}} \exists^{\geq k_i} x \varphi_{\tau_i}(x)$$

1597 it clearly holds that $\mathcal{P}_\rho = \mathcal{P}_{\varphi_\rho}$. Hence every finite union of properties defined by
 1598 neighbourhood profiles can be defined by the disjunction of the formulas φ_ρ of all ρ
 1599 in the finite union. \square

1600 **7.2.1. Relating FO properties to GSF-local properties.** We now prove
 1601 that FO properties which arise as unions of neighbourhood profiles of a particularly
 1602 simple form are GSF-local. For this let

$$1603 \quad \mathfrak{J}_0 := \{[0, k] \mid k \in \mathbb{N}\} \cup \{[0, \infty)\} \subset \mathfrak{J}.$$

1604 We call any neighbourhood profile ρ with codomain \mathfrak{J}_0 a *0-profile*, as all lower bounds
 1605 for the occurrence of types are 0.

1606 **OBSERVATION 2.** *Let ρ be a 0-profile. If two structures $A, A' \in \mathcal{C}_{\sigma, d}$ satisfy*
 1607 *$(\bar{v}_{d, r, \sigma}(A))_i \leq (\bar{v}_{d, r, \sigma}(A'))_i$ for every $i \in \{1, \dots, n_{d, r, \sigma}\}$ and $A' \sim \rho$, then $A \sim \rho$.*

1608 *In particular, this implies that there cannot be a 0-profile which defines the prop-*
 1609 *erty of all structures containing at least one occurrence of τ , for any r -type τ .*

1610 **THEOREM 7.9.** *Every finite union of properties of undirected graphs defined by*
 1611 *0-profiles is GSF-local.*

1612 *Proof.* We prove this in two parts (Claim 7.10 and Claim 7.11). We first argue
 1613 that every property \mathcal{P}_ρ defined by some 0-profile $\rho : \{1, \dots, n_{d, r}\} \rightarrow \mathfrak{J}_0$ is GSF-
 1614 local. For this it is important to note that we can express a forbidden r -type τ by a
 1615 forbidden generalised subgraph. For $(B, b) \in \tau$, the set of all graphs with no vertex
 1616 of neighbourhood type τ is the set of all B -free graphs where every vertex in $V(B)$ of
 1617 distance less than r to b is marked ‘full’ and every vertex in $V(B)$ of distance r to b
 1618 is marked ‘semifull’. Since a profile of the form $\rho : \{1, \dots, n_{d, r, \sigma}\} \rightarrow \mathfrak{J}_0$ can express
 1619 that some neighbourhood type τ can appear at most k times for some fixed $k \in \mathbb{N}$,
 1620 we need to forbid all marked graphs in which type τ appears $k + 1$ times. We will
 1621 formalise this in the following claim.

1622 **CLAIM 7.10.** *For every r -neighbourhood profile $\rho : \{1, \dots, n_{d, r}\} \rightarrow \mathfrak{J}_0$, there is a*
 1623 *finite set \mathcal{F} of marked graphs such that \mathcal{P}_ρ is exactly the property of \mathcal{F} -free graphs.*

1624 *Proof.* Assume τ is an r -type and $k \in \mathbb{N}_{>0}$. Then we say that a marked graph F
 1625 is a *k -realisation* of τ if F has the following properties.

- 1626 1. There are k distinct vertices v_1, \dots, v_k in F such that $(\mathcal{N}_r^F(v_i), v_i) \in \tau$ for
 1627 every $i = 1, \dots, k$.
- 1628 2. Every vertex v in F has distance less or equal to r to at least one vertex v_i .
- 1629 3. Every vertex v in F of distance less than r to at least one v_i is marked as
 1630 ‘full’.
- 1631 4. Every vertex v in F of distance greater or equal to r to every v_i is marked as
 1632 ‘semifull’.

1633 We denote by $S^k(\tau)$ the set of all k -realisations of τ .

1634 Now we can define the set \mathcal{F} of forbidden subgraphs to be

$$1635 \quad \mathcal{F} := \bigcup_{k \in \mathbb{N}, 1 \leq i \leq n_{d, r, \sigma} : \rho(i) = [0, k]} S^{k+1}(\tau_{d, r}^i).$$

1636 Let \mathcal{P} be the property of all \mathcal{F} -free graphs. We first prove that the property \mathcal{P}
 1637 is contained in \mathcal{P}_ρ . Towards a contradiction assume that $G \in \mathcal{C}_d$ is \mathcal{F} -free but not
 1638 contained in \mathcal{P}_ρ . As G is not contained in \mathcal{P}_ρ there must be an index $i \in \{1, \dots, n_{d, r}\}$
 1639 such that $(\bar{v}_{d, r}(G))_i \notin \rho(i)$. Since $\rho(i) \in \mathfrak{J}_0$ there is $k \in \mathbb{N}$ such that $\rho(i) = [0, k]$ and
 1640 hence $(\bar{v}_{d, r}(G))_i > k$. Hence there must be $k + 1$ vertices v_1, \dots, v_{k+1} in G such that
 1641 $(\mathcal{N}_r^G(v_i), v_i) \in \tau_{d, r}^i$. We define the marked graph F to be the subgraph of G induced by
 1642 the r -neighbourhoods of v_1, \dots, v_{k+1} , i.e. $G[\bigcup_{1 \leq i \leq k+1} \mathcal{N}_r^G(v_i)]$, in which every vertex
 1643 of distance less than k to at least one of the v_i is marked as ‘full’ and every other

1644 vertex is marked as ‘semifull’. Then F is by definition a $(k+1)$ -realisation of $\tau_{d,r}^i$ and
 1645 hence $F \in \mathcal{F}$. We now argue that F can be embedded into G . Since F is an induced
 1646 subgraph of G the identity map gives us a natural embedding $f : F \rightarrow G$. Let v be any
 1647 vertex marked ‘full’ in F . By construction of F , there is $i \in \{1, \dots, k+1\}$ such that
 1648 $f(v)$ is of distance less than r to v_i in G . But then $N_1^G(f(v))$ is a subset of $N_r^G(v_i)$. As
 1649 F without the marking is the subgraph of G induced by $\bigcup_{1 \leq i \leq k+1} N_r^G(v_i)$ this implies
 1650 that $f(N_1^F(v)) = N_1^G(f(v))$. Furthermore, assume v is a vertex marked ‘semifull’ in
 1651 F . Then $f(N_1^F(v)) = N_1^G(f(v)) \cap f(V(F))$ holds as F without the markings is an
 1652 induced subgraph of G . This proves that G is not F -free by Definition 7.1. This is a
 1653 contradiction to our assumption that G is \mathcal{F} -free and $F \in \mathcal{F}$.

1654 Similarly, we can show that $\mathcal{P}_\rho \subseteq \mathcal{P}$ by assuming $G \in \mathcal{C}_d$ is in \mathcal{P}_ρ but not \mathcal{F} -free,
 1655 and showing that the embedding of any graph of \mathcal{F} into G yields an amount of vertices
 1656 of a certain type contradicting containment in \mathcal{P}_ρ . \square

1657 Next we prove that classes defined by excluding finitely many marked graphs are
 1658 closed under finite unions.

1659 CLAIM 7.11. *Let $\mathcal{F}_1, \mathcal{F}_2$ be two finite sets of marked graphs. For $i \in \{1, 2\}$, let \mathcal{P}_i
 1660 be the property of \mathcal{F}_i -free graphs. Then there is a set \mathcal{F} of generalised subgraphs such
 1661 that $\mathcal{P}_1 \cup \mathcal{P}_2$ is the property of \mathcal{F} -free graphs.*

1662 *Proof.* We say that a marked graph F is a (not necessarily disjoint) union of
 1663 marked graphs F_1, F_2 if

- 1664 1. there is an embedding f_i of F_i into the graph F without its markings as in
 1665 Definition 7.1 for every $i \in \{1, 2\}$.
- 1666 2. for every vertex v in F there is $i \in \{1, 2\}$ and a vertex w in F_i such that
 1667 $f_i(w) = v$.
- 1668 3. every vertex v in F is marked ‘full’, if there is $i \in \{1, 2\}$ and a ‘full’ vertex w
 1669 in F_i such that $f_i(w) = v$.
- 1670 4. every vertex v in F is marked ‘semifull’, if there is $i \in \{1, 2\}$ and a ‘semifull’
 1671 vertex w in F_i such that $f_i(w) = v$ and $f_i(u) \neq v$ for every $i \in \{1, 2\}$ and
 1672 every ‘full’ vertex u .
- 1673 5. every vertex v in F is marked ‘partial’ if $f_i(u) \neq v$ for every $i \in \{1, 2\}$ and
 1674 every ‘full’ or ‘semifull’ vertex u .

1675 We define $S(F_1, F_2)$ to be the set of all possible (not necessarily disjoint) unions of
 1676 F_1, F_2 . We can now define the set \mathcal{F} to be

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

1678 Let \mathcal{P} be the property of all \mathcal{F} -free graphs. Now we prove $\mathcal{P} \subseteq \mathcal{P}_1 \cup \mathcal{P}_2$. Towards
 1679 a contradiction assume G is \mathcal{F} -free but G is in neither \mathcal{P}_1 nor in \mathcal{P}_2 . Then for every
 1680 $i \in \{1, 2\}$ there is a graph $F_i \in \mathcal{F}_i$ such that G is not F_i -free. It is easy to see that
 1681 there is a union F_\cup of F_1 and F_2 such that G is not F_\cup -free, which contradicts that
 1682 G is \mathcal{F} -free.

1683 Conversely, in order to prove $\mathcal{P}_1 \cup \mathcal{P}_2 \subseteq \mathcal{P}$, if G is \mathcal{F}_i free for some $i \in \{1, 2\}$ then
 1684 G must be \mathcal{F} -free by construction of \mathcal{F} . \square

1685 Combining the two claims above proves Theorem 7.9. \square

1686 *Further discussion of the relation between FO and GSF-locality.* First let us re-
 1687 mark that it is neither true that every FO definable property is GSF-local, nor that
 1688 every GSF-local property is FO definable.

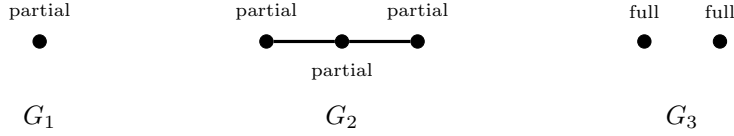


Fig. 9: Marked graphs for Example 5.

1689 **EXAMPLE 3.** *The property of bounded-degree graphs containing a triangle is FO*
 1690 *definable but not GSF-local.*

1691 Indeed, the existence of a fixed number of vertices of certain neighbourhood types
 1692 can be expressed in FO, while in general, this cannot be expressed by forbidding
 1693 generalised subgraphs. If a formula has a 0-profile (and hence does not require the
 1694 existence of any types) then the property defined by that formula is GSF-local, as
 1695 shown in Theorem 7.9.

1696 **EXAMPLE 4.** *The class of all bounded-degree graphs with an even number of ver-*
 1697 *tices is GSF-local but not FO definable.*

1698 Let us remark that Theorem 7.9 combined with Lemma 7.7 proves that every
 1699 finite union of properties definable by 0-profiles is both FO definable and GSF-local.
 1700 Hence it is natural to ask whether the intersection of FO definable properties and
 1701 GSF-local properties is precisely the set of finite unions of properties definable by
 1702 0-profiles. However, this is not the case. The following example shows that there are
 1703 properties which are both FO definable and GSF-local but cannot be expressed by
 1704 0-profiles.

1705 **EXAMPLE 5.** *We let $d \geq 2$ and let $B_1 := (\{v\}, \{\})$, $B_2 = (\{v, w\}, \{\{v, w\}\})$ be*
 1706 *two graphs. We further let τ_1, τ_2 be the 1-types of degree d such that $(B_1, v) \in \tau_1$ and*
 1707 *$(B_2, v) \in \tau_2$. Consider the property \mathcal{P} defined by the following FO formula*

$$1708 \quad \varphi := \neg \exists x(x = x) \vee \exists^{\neq 1} x(\varphi_{\tau_1}(x) \wedge \forall y(x \neq y \rightarrow \varphi_{\tau_2}(y))).$$

1709 \mathcal{P} contains, besides the empty graph, unions of an arbitrary amount of disjoint edges
 1710 and one isolated vertex. To define a sequence of forbidden subgraphs we let G_1, G_2, G_3
 1711 be the marked graphs in Figure 9. Let $\mathcal{F}_{\text{even}} := \{G_1\}$ and $\mathcal{F}_{\text{odd}} := \{G_2, G_3\}$ and let
 1712 $\overline{\mathcal{F}} = (\mathcal{F}_n)_{n \in \mathbb{N}}$ where $\mathcal{F}_i = \mathcal{F}_{\text{even}}$ if i is even and $\mathcal{F}_i = \mathcal{F}_{\text{odd}}$ if i is odd. Note that
 1713 every graph on more than one vertex with an odd number of vertices which is \mathcal{F}_{odd} -
 1714 free must contain a vertex of neighbourhood type τ_1 , and that the set of $\mathcal{F}_{\text{even}}$ -free
 1715 graphs contains only the empty graph. Hence \mathcal{P} is $\overline{\mathcal{F}}$ -local. Now assume towards a
 1716 contradiction that $\mathcal{P} = \bigcup_{1 \leq i \leq k} \mathcal{P}_{\rho_i}$ for 0-profiles ρ_i . Let G_m be the graph consisting
 1717 of m disjoint edges and one isolated vertex and H_m the graph consisting of m disjoint
 1718 edges. Since $G_m \in \mathcal{P}$ there is $i \in \{1, \dots, k\}$ such that $G_m \sim \rho_i$. By choice of G_m
 1719 and H_m we have $0 \leq (\overline{v}_{d,r}(H_m))_j \leq (\overline{v}_{d,r}(G_m))_j \in \rho_i(j)$ for every $j \in \{1, \dots, n_{d,r}\}$.
 1720 Since additionally $\rho_i(j) \in \mathcal{I}_0$ this implies that $(\overline{v}_{d,r}(H_m))_j \in \rho_i(j)$. But then $H_m \sim \rho_i$
 1721 which yields a contradiction as $H_m \notin \mathcal{P}$. Hence \mathcal{P} cannot be defined as a finite union
 1722 of 0-profiles.

1723 Figure 10 gives a schematic overview of all classes of properties discussed here and
 1724 their relationship.

1725 **7.3. Proving the existence of a GSF-local non-testable property.** In this
 1726 section we prove Theorem 7.5. We show that the property $\mathcal{P}_{\mathbb{Z}}$ from Section 4 can be

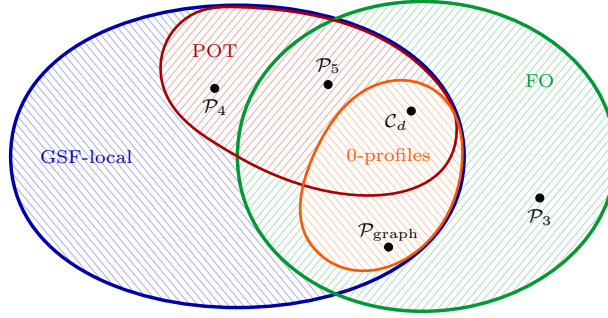


Fig. 10: Overview of the classes of properties, here \mathcal{P}_i refers to the property from Example i , \mathcal{C}_d refers to the property of all graphs of bounded degree d and $\mathcal{P}_{\text{graph}}$ is the property defined in Section 5.2.

1727 expressed by a union of 0-profiles. We then show that the local reduction from $\mathcal{P}_{\mathbb{Z}}$
 1728 to $\mathcal{P}_{\text{graph}}$ given in Section 5.2 preserves the expressibility by 0-profiles, and hence by
 1729 Theorem 7.9 $\mathcal{P}_{\text{graph}}$ is GSF-local.

1730 Let σ be the signature, $d \in \mathbb{N}$ and $\mathcal{P}_{\mathbb{Z}}$ be the property of d σ -structures of
 1731 bounded-degree from Section 3.

1732 **7.3.1. Characterisation of the relational structure property by neigh-**
 1733 **bourhood profiles.** Our aim in this section is to prove that the property $\mathcal{P}_{\mathbb{Z}}$ of
 1734 relational structures can be written as a finite union of properties defined by 0-profiles
 1735 of radius 2.

1736 For all σ -structures in $\mathcal{P}_{\mathbb{Z}}$ (excluding A_{\emptyset}) it is crucial that they are allowed to
 1737 contain precisely one root element. Hence the neighbourhood profile describing $\mathcal{P}_{\mathbb{Z}}$
 1738 must restrict the number of occurrences of the 2-type of the root element. But since
 1739 in $\mathcal{P}_{\mathbb{Z}} \setminus \{A_{\emptyset}\}$, the root elements in different structures may have different 2-types,
 1740 we partition $\mathcal{P}_{\mathbb{Z}} \setminus \{A_{\emptyset}\}$ into parts $\mathcal{P}_1, \dots, \mathcal{P}_m$ by the 2-type of the root element.
 1741 Note that the number m of parts is constant as there are at most $n_{d,2,\sigma}$ 2-types in
 1742 total. For each of these parts we then define a neighbourhood profile ρ_k such that
 1743 $\mathcal{P}_k \cup \{A_{\emptyset}\} = \mathcal{P}_{\rho_k}$. We would like to remark here that the roots of all but one structure
 1744 in $\mathcal{P}_{\mathbb{Z}}$ actually have the same 2-types. Hence the partition only contains two parts
 1745 and one of the two parts only contains one structure. We now define the parts and
 1746 corresponding profiles formally.

1747 Assume without loss of generality that the 2-types $\tau_{d,2,\sigma}^1, \dots, \tau_{d,2,\sigma}^{n_{d,2,\sigma}}$ of degree d
 1748 are ordered in such a way that for $(B, b) \in \tau_{d,2,\sigma}^k$, it holds that $B \models \varphi_{\text{root}}(b)$ if and
 1749 only if $k \in \{1, \dots, m\}$ for some $m \leq n_{d,2,\sigma}$. For $k \in \{1, \dots, m\}$, let

$$1750 \quad \mathcal{P}_k := \{A \in \mathcal{P}_{\mathbb{Z}} \setminus \{A_{\emptyset}\} \mid \text{there is } a \in U(A) \text{ such that } (\mathcal{N}_2^A(a), a) \in \tau_{d,2,\sigma}^k\}.$$

1751 Since by Lemma 3.5 every $A \in \mathcal{P}_{\mathbb{Z}} \setminus \{A_{\emptyset}\}$ must contain exactly one root we get that

$$1752 \quad \mathcal{P}_{\mathbb{Z}} = \bigcup_{1 \leq k \leq m} \mathcal{P}_k \cup \{A_{\emptyset}\}$$

1753 and this union is disjoint. Furthermore, for $k \in \{1, \dots, m\}$, let $I_k \subseteq \{1, \dots, n_{d,2,\sigma}\}$

1754 be the set of indices j such that there is a structure $A \in \mathcal{P}_k$ and $a \in U(A)$ with
 1755 $(\mathcal{N}_2^A(a), a) \in \tau_{d,2,\sigma}^j$. For every $k \in \{1, \dots, m\}$ we define the 2-neighbourhood profile
 1756 $\rho_k : \{1, \dots, n_{d,2,\sigma}\} \rightarrow \mathfrak{I}_0$ by

$$1757 \quad \rho_k(i) := \begin{cases} [0, 1] & \text{if } i = k, \\ [0, \infty) & \text{if } i \in I_k \setminus \{k\}, \\ [0, 0] & \text{otherwise.} \end{cases}$$

1758 To prove that these 0-profiles of radius 2 define the property $\mathcal{P}_{\mathbb{Z}}$, the crucial ob-
 1759 servation is that for every element a of some structure in $\mathcal{C}_{\sigma,d}$, the FO-formula $\varphi_{\mathbb{Z}}$
 1760 only talks about elements of distance at most 2 to a (i. e. $\varphi_{\mathbb{Z}}$ is 2-local). Hence the
 1761 2-histogram vector of a structure already captures whether the structure satisfies $\varphi_{\mathbb{Z}}$.
 1762 We will now formally prove this.

1763 LEMMA 7.12. *It holds that $\mathcal{P}_{\mathbb{Z}} = \bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k}$.*

1764 *Proof.* We first prove that $\mathcal{P}_{\mathbb{Z}} \subseteq \bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k}$. First note that trivially $A_{\emptyset} \in$
 1765 $\bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k}$. Now assume $A \in \mathcal{P}_{\mathbb{Z}} \setminus \{A_{\emptyset}\}$. This implies that there is $k \in \{1, \dots, m\}$
 1766 such that $A \in \mathcal{P}_k$. By construction we have that for every $a \in A$, there is $i \in I_k$ such
 1767 that $(\mathcal{N}_2^A(a), a) \in \tau_{d,2,\sigma}^i$. Furthermore, since $A \models \varphi_{\mathbb{Z}}$ and $U(A) \neq \emptyset$, we have by
 1768 Lemma 3.5 that $A \models \exists^{-1}x\varphi_{\text{root}}(x)$, and that there can be at most one $a \in U(A)$ such
 1769 that $(\mathcal{N}_2^A(a), a) \in \tau_{d,2,\sigma}^k$. Therefore $A \in \mathcal{P}_{\rho_k}$.
 1770

1771 To prove $\bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k} \subseteq \mathcal{P}_{\mathbb{Z}}$, we prove that every structure in $\bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k}$ must
 1772 satisfy $\varphi_{\mathbb{Z}}$. We will prove that every $A \in \bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k}$ satisfies $\varphi_{\text{recursion}}$, and refer
 1773 for the proof that A satisfies $\varphi_{\text{tree}} \wedge \varphi_{\text{rotationMap}} \wedge \varphi_{\text{base}}$ to Claim A.1, Claim A.2 and
 1774 Claim A.3 in Appendix A. Note that $A_{\emptyset} \models \varphi_{\mathbb{Z}}$ by Lemma 3.10 and hence we exclude
 1775 A_{\emptyset} in the following.

1776 CLAIM 7.13. *Every structure $A \in \bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k} \setminus \{A_{\emptyset}\}$ satisfies $\varphi_{\text{recursion}}$.*

1777 *Proof.* Let $A \in \bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k} \setminus \{A_{\emptyset}\}$. Then there is a $k \in \{1, \dots, m\}$ such that
 1778 $A \in \mathcal{P}_{\rho_k}$.

1779 By definition, $\varphi_{\text{recursion}} := \forall x \forall z (\varphi(x, z) \vee \psi(x, z))$ (see Section 3), where

1780 $\varphi(x, z) := \neg \exists y F(x, y) \wedge \neg \exists y F(z, y)$ and

$$1781 \quad \psi(x, z) := \bigwedge_{\substack{k'_1, k'_2 \in [D]^2 \\ \ell'_1, \ell'_2 \in [D]^2}} \left(\exists y [E_{k'_1, \ell'_1}(x, y) \wedge E_{k'_2, \ell'_2}(y, z)] \rightarrow \right.$$

$$1782 \quad \left. \bigwedge_{\substack{i, j, i', j' \in [D], k, \ell \in ([D]^2)^2 \\ \text{ROT}_H(k, i) = ((k'_1, k'_2), i') \\ \text{ROT}_H((\ell'_2, \ell'_1), j) = (\ell, j')}} \exists x' \exists z' [F_k(x, x') \wedge F_\ell(z, z') \wedge E_{(i, j), (j', i')}(x', z')] \right).$$

1783 Let $a, c \in U(A)$. Assume first that there is $b \in U(A)$ with $(a, b) \in F(A)$. Hence
 1784 $A \not\models \varphi(a, c)$. Since $\varphi_{\text{recursion}} := \forall x \forall z (\varphi(x, z) \vee \psi(x, z))$ we aim to prove $A \models \psi(a, c)$.
 1785 By construction of ρ_k , there is an $i \in I_k$ such that $(\mathcal{N}_2^A(a), a) \in \tau_{d,2,\sigma}^i$. Therefore
 1786 there is a structure $\tilde{A} \models \varphi_{\mathbb{Z}}$ and $\tilde{a} \in U(\tilde{A})$ such that $(\mathcal{N}_2^A(a), a) \cong (\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Let f
 1787 be an isomorphism from $(\mathcal{N}_2^A(a), a)$ to $(\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Since $b \in \mathcal{N}_2^A(a)$, we get that $f(b)$

1788 is defined. Since f is an isomorphism mapping a onto \tilde{a} , we have that $(a, b) \in F(A)$
 1789 implies that $(\tilde{a}, f(b)) \in F(\tilde{A})$. Hence $\tilde{A} \not\models \varphi(\tilde{a}, \tilde{c})$, for every $\tilde{c} \in U(\tilde{A})$. But since
 1790 $\tilde{A} \models \varphi_{\text{recursion}}$, as $\tilde{A} \models \varphi_{\mathbb{Z}}$, this shows that $\tilde{A} \models \psi(\tilde{a}, \tilde{c})$ for every $\tilde{c} \in U(\tilde{A})$.

1791 Let $k'_1, k'_2 \in [D]^2$ and $\ell'_1, \ell'_2 \in [D]^2$ be indices such that there is $b' \in U(A)$
 1792 with $(a, b') \in E_{k'_1, \ell'_1}(A)$ and $(b', c) \in E_{k'_2, \ell'_2}(A)$. Since $b', c \in N_2^A(a)$, by assump-
 1793 tion we get that $f(b')$ and $f(c)$ are defined. Furthermore, $(a, b') \in E_{k'_1, \ell'_1}(A)$ and
 1794 $(b', c) \in E_{k'_2, \ell'_2}(A)$ imply that $(\tilde{a}, f(b')) \in E_{k'_1, \ell'_1}(\tilde{A})$ and $(f(b'), f(c)) \in E_{k'_2, \ell'_2}(\tilde{A})$,
 1795 since f is an isomorphism mapping a onto \tilde{a} . We proved in the previous paragraph
 1796 that $\tilde{A} \models \psi(\tilde{a}, f(c))$. Hence we can conclude that for all indices $i, j, i', j' \in [D]$,
 1797 $k, \ell \in ([D]^2)^2$ for which $\text{ROT}_H(k, i) = ((k'_1, k'_2), i')$ and $\text{ROT}_H((\ell'_2, \ell'_1), j) = (\ell, j')$,
 1798 there are elements $\tilde{a}', \tilde{c}' \in U(\tilde{A})$ such that $(\tilde{a}, \tilde{a}') \in F_k(\tilde{A})$, $(f(c), \tilde{c}') \in F_\ell(\tilde{A})$,
 1799 and $(\tilde{a}', \tilde{c}') \in E_{(i,j),(j',i')}(\tilde{A})$. Since $\tilde{a}', \tilde{c}' \in N_2^{\tilde{A}}(\tilde{a})$, we get that $a' := f^{-1}(\tilde{a}')$ and
 1800 $c' := f^{-1}(\tilde{c}')$ are defined. Furthermore, we get that $(a, a') \in F_k(A)$, $(c, c') \in F_\ell(A)$
 1801 and $(a', c') \in E_{(i,j),(j',i')}(A)$. This proves that $A \models \psi(a, c)$.

1802

1803 In the case that there is $b \in U(A)$ with $(c, b) \in F(A)$, we can prove similarly
 1804 that $A \models \psi(a, c)$, by considering that there exist $\tilde{A} \models \varphi_{\mathbb{Z}}$ and $\tilde{c} \in U(\tilde{A})$ such that
 1805 $(\mathcal{N}_2^A(a), c) \cong (\mathcal{N}_2^{\tilde{A}}(\tilde{c}), \tilde{c})$ by construction of ρ_k . Finally if there is no $b \in U(A)$ such
 1806 that $(a, b) \in F(A)$ or $(c, b) \in F(A)$ then $A \models \varphi(a, c)$. Since this covers every case we
 1807 get that $A \models \varphi_{\text{recursion}}$. \square

1808 Assume $A \in \bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k}$. As proved in Claims A.1, A.2, A.3 and 7.13 this implies
 1809 that $A \models \varphi_{\text{tree}}$, $A \models \varphi_{\text{rotationMap}}$, $A \models \varphi_{\text{base}}$ and $A \models \varphi_{\text{recursion}}$. Since $\varphi_{\mathbb{Z}}$ is a
 1810 conjunction of these formulas, we get $A \models \varphi_{\mathbb{Z}}$ and hence $A \in \mathcal{P}_{\mathbb{Z}}$. \square

1811 **7.3.2. The graph property is GSF-local.** Let $\mathcal{P}_{\text{graph}}$ be the graph property
 1812 as defined in Section 5.2 and let $f : \mathcal{C}_{\sigma, d} \rightarrow \mathcal{C}_3$ be the local reduction from $\mathcal{P}_{\mathbb{Z}}$ to
 1813 $\mathcal{P}_{\text{graph}}$. We now use this local reduction and the expressibility of $\mathcal{P}_{\mathbb{Z}}$ by 0-profiles to
 1814 show that $\mathcal{P}_{\text{graph}}$ is GSF-local.

1815 LEMMA 7.14. *The graph property $\mathcal{P}_{\text{graph}}$ is GSF-local.*

1816 *Proof.* For this we will prove that $\mathcal{P}_{\text{graph}}$ is equal to a finite union of properties
 1817 defined by 0-profiles, and then use Theorem 7.9 to prove that $\mathcal{P}_{\text{graph}}$ is GSF-local. We
 1818 define the 0-profiles for $\mathcal{P}_{\text{graph}}$ in a very similar way to the relational structure case,
 1819 and then use the description of $\mathcal{P}_{\mathbb{Z}}$ by 0-profiles shown in Lemma 7.12. To this end,

1820 let $\ell' := 24\ell + 18 + d$ and assume that the ℓ' -types $\tau_{d, \ell'}^1, \dots, \tau_{d, \ell'}^{n_{d, \ell'}}$ are ordered in such
 1821 a way that $(\mathcal{N}_{\ell'}^{f(B)}(u_{b,1}), u_{b,1}) \in \tau_{d, \ell'}^k$, for every $k \in \{1, \dots, m\}$ and $(B, b) \in \tau_{d, 2, \sigma}^k$,
 1822 where m is the number of parts of the partition of $\mathcal{P}_{\mathbb{Z}}$ defined in Subsection 7.3.1.

1823 For $k \in \{1, \dots, m\}$, let \hat{I}_k be the set of indices i such that there is $A \in \mathcal{P}_k$, and
 1824 $v \in V(f(A))$ for which $(\mathcal{N}_{\ell'}^{f(A)}(v), v) \in \tau_{d, \ell'}^i$. Let $\hat{\rho}_k : \{1, \dots, n_{d, \ell'}\} \rightarrow \mathcal{I}_0$ be defined
 1825 by

$$1826 \quad \hat{\rho}_k(i) := \begin{cases} [0, 1] & \text{if } i = k, \\ [0, \infty) & \text{if } i \in \hat{I}_k \setminus \{k\}, \\ [0, 0] & \text{otherwise.} \end{cases}$$

1827

1828 CLAIM 7.15. *It holds that $\mathcal{P}_{\text{graph}} = \bigcup_{1 \leq k \leq m} \mathcal{P}_{\hat{\rho}_k}$.*

1829 *Proof.* First we prove $\mathcal{P}_{\text{graph}} \subseteq \bigcup_{1 \leq k \leq m} \mathcal{P}_{\hat{\rho}_k}$. Assume $G \in \mathcal{P}_{\text{graph}}$ and let $A \in \mathcal{P}_{\mathbb{Z}}$
1830 be a structure such that $G = f(A)$. If $A = A_\emptyset$ then clearly $G \in \bigcup_{1 \leq k \leq m} \mathcal{P}_{\hat{\rho}_k}$. Hence
1831 assume $A \neq A_\emptyset$. Then $A \in \mathcal{P}_k$ for some $k \in \{1, \dots, m\}$. By the construction of
1832 \hat{I}_k we know that for every $v \in V(G)$ we have $(\mathcal{N}_{\ell'}^G(v), v) \in \tau_{d, \ell'}^i$ for some $i \in \hat{I}_k$.
1833 Furthermore, since $A \in \mathcal{P}_k$ there is at most one $a \in U(A)$ with $(\mathcal{N}_2^A(a), a) \in \tau_{d, 2, \sigma}^k$.
1834 We argue that this implies that there can be at most one vertex $v \in V(G)$ with
1835 $(\mathcal{N}_{\ell'}^G(v), v) \in \tau_{d, \ell'}^k$. Let us denote the set of vertices of G associated with an element
1836 a of A by $V_a = \{u_{a, i}, v_{a, i}^k \mid 1 \leq i \leq d, 1 \leq k \leq 6\ell + 5\}$. By construction of $\hat{\rho}$ we know
1837 that only element-vertices can have type $\tau_{d, \ell'}^k$. Assume that vertex $u_{b, i}$ has type $\tau_{d, \ell'}^k$.
1838 By choice of ℓ' , we know that the ℓ' -neighbourhood of $u_{b, i}$ must contain the sets V_a
1839 for every element a in the 2-neighbourhood of b . Hence, b must have type $\tau_{d, 2, \sigma}^k$.
1840 We further note that any two element-vertices $u_{b, i}$ and $u_{b, j}$, $i \neq j$ **cannot** be of the
1841 same neighbourhood type. **To see this, observe that there are several distinct arrows**
1842 **attached to element b . These include the arrow representing the tuple $(b, b) \in R(A)$**
1843 **and the arrow representing the unique tuple $(b, b') \in F_1(A)$. These two arrows are**
1844 **attached to distinct vertices amongst $u_{b, 1}, \dots, u_{b, d}$, say $u_{b, i'}$ and $u_{b, j'}$, $i' \neq j'$. Note**
1845 **that $u_{b, i}$ and $u_{b, j}$ must have different distance (on the cycle $(u_{b, 1}, \dots, u_{b, d})$) to either**
1846 **$u_{b, i'}$ or $u_{b, j'}$. Since the ℓ -neighbourhoods of both $u_{b, i}$ and $u_{b, j}$ must encompass $u_{b, i'}$**
1847 **and $u_{b, j'}$ along with the arrow-gadget representing arrows (b, b) and (b, b') , it follows**
1848 **that $u_{b, i}$ and $u_{b, j}$ necessarily possess distinct ℓ -neighbourhoods. Hence $G \in \mathcal{P}_{\hat{\rho}}$.**
1849

1850 Now we prove that $\bigcup_{1 \leq k \leq m} \mathcal{P}_{\hat{\rho}_k} \subseteq \mathcal{P}_{\text{graph}}$. Let $G \in \bigcup_{1 \leq k \leq m} \mathcal{P}_{\hat{\rho}_k}$ and let $k \in$
1851 $\{1, \dots, m\}$ be an index such that $G \in \mathcal{P}_{\hat{\rho}_k}$. Further assume that G is not the empty
1852 graph, as $f(A_\emptyset) \in \mathcal{P}_{\text{graph}}$ is the empty graph.

1853 Since for every i for which $\hat{\rho}(i) \neq [0, 0]$, there is a graph $G' \in \mathcal{P}_{\text{graph}}$ and $v \in V(G')$
1854 such that $(\mathcal{N}_{\ell'}^{G'}(v'), v') \in \tau_{d, \ell'}^i$, we get that the ℓ' -neighbourhood of every vertex in G
1855 appears in some graph $G' \in \mathcal{P}_{\text{graph}}$. By choice of ℓ' we get that every vertex $v \in V(G)$
1856 is either contained in a cycle of length d and is the endpoint of some k -arrow, k -loop
1857 or non-arrow or v is an internal vertex of a k -arrow, k -loop or non-arrow. Hence, we
1858 obtain a σ -structure A with $f(A) \cong G$ by replacing any cycle C of length d by an
1859 element a_C and adding a tuple $(a_C, a_{C'})$ to the relation $R_k(A)$ if there are vertices u
1860 on C and v on C' such that $u \xrightarrow{k} v$ in G . Let g be an isomorphism from $f(A)$ to G .

1861 Now we argue that $A \in \mathcal{P}_{\rho_k}$. First assume that there are two elements $a, b \in$
1862 $U(A)$ with $(\mathcal{N}_2^A(a), a) \in \tau_{d, 2, \sigma}^k$ and $(\mathcal{N}_2^A(b), b) \in \tau_{d, 2, \sigma}^k$. By definition, we get that
1863 $(\mathcal{N}_{\ell'}^{f(A)}(u_{a, 1}), u_{a, 1}) \in \tau_{d, \ell'}^k$ and $(\mathcal{N}_{\ell'}^{f(A)}(u_{b, 1}), u_{b, 1}) \in \tau_{d, \ell'}^k$. Since g is an isomorphism,
1864 the restriction of g to $\mathcal{N}_{\ell'}^{f(A)}(u_{a, 1})$ must be an isomorphism from $\mathcal{N}_{\ell'}^{f(A)}(u_{a, 1})$ to
1865 $\mathcal{N}_{\ell'}^G(g(u_{a, 1}))$, and hence $(\mathcal{N}_{\ell'}^G(g(u_{a, 1})), g(u_{a, 1})) \cong (\mathcal{N}_{\ell'}^{f(A)}(u_{a, 1}), u_{a, 1}) \in \tau_{d, \ell'}^k$. But
1866 the same holds for the ℓ' -ball of $g(u_{b, 1})$, and hence we contradict the assumption that
1867 $G \in \mathcal{P}_{\hat{\rho}_k}$ since $\hat{\rho}_k(k) = [0, 1]$. Let us further assume that there is an $a \in U(A)$ such that
1868 $(\mathcal{N}_2^A(a), a) \in \tau_{d, 2, \sigma}^i$ for some $i \notin I_k$. Since $G \in \mathcal{P}_{\hat{\rho}_k}$ we get $(\mathcal{N}_{\ell'}^G(g(u_{a, 1})), g(u_{a, 1})) \in$
1869 $\tau_{d, \ell'}^j$ for some $j \in \hat{I}_k$. But then by construction of $\hat{\rho}_k$, there must be $G' \in \mathcal{P}_{\text{graph}}$,
1870 and a vertex $v \in V(G')$ such that $(\mathcal{N}_{\ell'}^{G'}(v), v) \in \tau_{d, \ell'}^j$. Furthermore, since $\ell' > d$ the
1871 vertex v must be contained in cycle of length d . By construction of $\mathcal{P}_{\text{graph}}$, there is
1872 a structure $A' \in \mathcal{P}_{\mathbb{Z}}$ such that $f(A') = G'$. Since v is contained in a cycle of length
1873 d , v must be an element-vertex corresponding to some element $a' \in U(A')$. Since we
1874 picked ℓ' in such a way that $f(\mathcal{N}_2^{A'}(a')) \subseteq \mathcal{N}_{\ell'}^{G'}(v)$, we get $(\mathcal{N}_2^{A'}(a'), a') \in \tau_{d, 2, \sigma}^i$ by
1875 choice of i and j . Hence $A' \notin \mathcal{P}_{\rho_k}$. But this contradicts Lemma 7.12.

1876 Hence we have shown that $A \in P_{\rho_k}$. Then by Lemma 7.12 $A \in \mathcal{P}_{\mathbb{Z}}$, and by
 1877 construction $G \in \mathcal{P}_{\text{graph}}$. \square

1878 Since by Claim 7.15 we can express $\mathcal{P}_{\text{graph}}$ as a finite union of properties, each
 1879 defined by a 0-profile, Theorem 7.9 implies that $\mathcal{P}_{\text{graph}}$ is GSF-local. \square

1880 **7.3.3. Putting everything together.** Now we prove Theorem 7.5.

1881 *Proof of Theorem 7.5.* Combining Theorem 4.4, Lemma 5.3 and Lemma 5.4 we
 1882 obtain that the graph property $\mathcal{P}_{\text{graph}}$ is not testable. Lemma 7.14 shows that $\mathcal{P}_{\text{graph}}$ is
 1883 also a GSF-local property. Hence there exists a GSF-local property of bounded-degree
 1884 graphs which is not testable. Furthermore, since having a POT implies being testable,
 1885 this proves that there is a GSF-local property which has no POT. By Theorem 7.4
 1886 this implies that not all GSF-local properties are non-propagating. \square

1887 **7.4. GSF-local properties of graphs of bounded degree 1 and 2 are non-**
 1888 **propagating.** In this section, we show that the degree 3 from Theorem 7.5 of the
 1889 example of a GSF-local property which is propagating is optimal, in the sense that
 1890 all GSF-local properties of graphs of bounded degree 1 and 2 are non-propagating.
 1891 We note that Ito et al. [27] claimed that every GSF-local sequence of bounded degree
 1892 at most 2 is non-propagating in the appendix of their paper. However, there is one
 1893 subtle issue in their proof, as they only considered *connected* forbidden generalized
 1894 subgraphs (which are called forbidden configurations in [27]). In the following, we
 1895 resolve this issue. Indeed, the extension from connected forbidden generalised sub-
 1896 graphs to arbitrary forbidden generalized subgraphs is non-trivial and requires an
 1897 involved proof which we present in this section.

1898 We first observe that even for graphs of bounded degree 1, not every sequence
 1899 of marked graphs $\overline{\mathcal{F}}$ is non-propagating as the following example shows. A similar
 1900 example was given in [22].

1901 **EXAMPLE 6.** Let $\mathcal{P} \subseteq \mathcal{C}_1$ be the property of $\overline{\mathcal{F}}$ -free graphs, where F is the marked
 1902 graph depicted in Figure 11, $\mathcal{F}_n = \{F\}$ and $\overline{\mathcal{F}} = (\mathcal{F}_n)_{n \in \mathbb{N}}$. Let G_k be the graph
 1903 consisting of k edges and one isolated vertex. Then the set B containing the one
 1904 isolated vertex of G_k covers all embeddings of F (see Figure 11). But the only way
 1905 to make G_k F -free is to remove all k edges of G_k . Hence G_k is $1/2$ -far from being
 1906 F -free, which implies that \mathcal{P} is propagating for $\overline{\mathcal{F}}$.

1907 However, the property \mathcal{P} is non-propagating, as we show in the proof of Theo-
 1908 rem 7.16. Indeed, consider the alternative sequence of marked graphs $\overline{\mathcal{F}} = (\mathcal{F}_n)_{n \in \mathbb{N}}$,
 1909 where $\mathcal{F}_n = \{F\}$ for n even and $\mathcal{F}_n = \{F, \tilde{F}\}$ for n odd. Clearly, in G_{2k+1} any set B
 1910 covering \mathcal{F}_{2k+1} must contain one incident vertex of every edge. Hence the number of
 1911 necessary modifications is at most $|B|$, suggesting that \mathcal{P} is non-propagating.

1912 Indeed, adding certain redundant marked graphs to the sequence $\overline{\mathcal{F}} = (\mathcal{F}_n)_{n \in \mathbb{N}}$ to
 1913 control the behaviour of sets covering \mathcal{F}_n as in Example 6 works in general both
 1914 in the degree 1 and degree 2 case and will be our proof strategy for the following
 1915 theorem. More precisely, for a property \mathcal{P} of graphs of bounded degree 2, a sequence
 1916 $\overline{\mathcal{F}}$ of marked graphs such that \mathcal{P} is $\overline{\mathcal{F}}$ -local and a bound k on the size of any graph
 1917 appearing in $\overline{\mathcal{F}}$, we add forbidden generalized subgraphs to $\overline{\mathcal{F}}$ in the following way. In
 1918 case there is no graph in \mathcal{P} with n vertices containing a set of different small connected
 1919 components (connected components with at most k vertices) each with frequency
 1920 at least k , we add a generalised subgraph forbidding precisely this combination of
 1921 connected components to \mathcal{F}_n . Additionally, if no graph in \mathcal{P} with n vertices contains
 1922 a set of different small connected components each with frequency at least k and

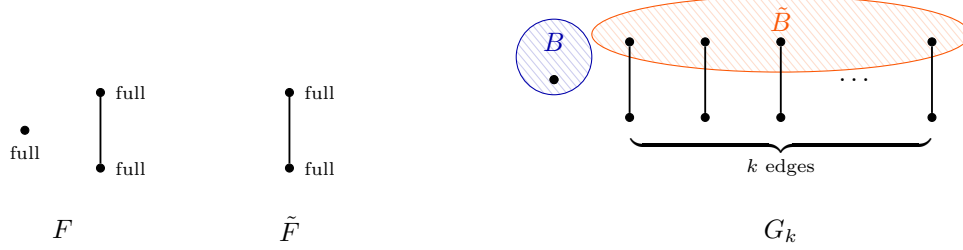


Fig. 11: Marked graphs F and \tilde{F} and graph G_k from Example 6.

1923 one large component (connected component with at least $k + 1$ vertices), we add a
 1924 generalised subgraph forbidding precisely this combination of connected components
 1925 to \mathcal{F}_n . Now for a graph G which is not in \mathcal{P} and a set B covering all forbidden
 1926 generalised subgraphs in G , we look at what types of connected components appear
 1927 in the part of G not containing vertices from B . In case G is large enough and B is
 1928 small enough, we observe that some types of connected components have to appear
 1929 with high frequency, or there must be a large component in the part of G which is not
 1930 covered by B . By adding redundant subgraphs as described earlier, this now implies
 1931 that there must be a graph G' in \mathcal{P} , which has the same structure as G on a large
 1932 subset of the part of G which is not covered by B . Hence we can modify G to obtain
 1933 a graph satisfying the property \mathcal{P} (by changing G to G') without modifying G much
 1934 beyond B . The restriction to bounded degree at most 2 is crucial in this argument as
 1935 it gives us the necessary control over large connected components.

1936 **THEOREM 7.16.** *Any GSF-local property $\mathcal{P} \subseteq \mathcal{C}_d$ for $d \leq 2$ is non-propagating.*

1937 *Proof.* We only consider the case that $\mathcal{P} \subseteq \mathcal{C}_2$. We can consider any property
 1938 $\mathcal{P} \subseteq \mathcal{C}_1$ as a property in \mathcal{C}_2 by forbidding any vertex to have degree 2, i. e. adding a
 1939 path of length 2 in which both degree 1 vertices are marked ‘partial’ and the degree 2
 1940 vertex is marked ‘full’ to every set of forbidden marked graphs in any sequence defining
 1941 \mathcal{P} , and adjusting constants in the following argument to account for the degree being
 1942 1 instead of 2.

1943 Let $\mathcal{P} = \bigcup_{n \in \mathbb{N}} \mathcal{P}_n$ and $\overline{\mathcal{F}} = (\mathcal{F}_n)_{n \in \mathbb{N}}$ be a sequence of marked graphs such that \mathcal{P}
 1944 is $\overline{\mathcal{F}}$ -local. By definition there exists $k \in \mathbb{N}$ such that every marked graph appearing
 1945 in $\overline{\mathcal{F}}$ contains at most k vertices.

1946 For two sets $I \subseteq [k] := \{0, \dots, k-1\}$, $J \subseteq \{3, \dots, k\}$ such that $I \cup J \neq \emptyset$ let $F_{I,J}$
 1947 be the marked graph which is the disjoint union of k paths of length i for every $i \in I$
 1948 and k cycles of length j for every $j \in J$ in which every vertex is marked as ‘full’. Be
 1949 aware that a path of length i contains $i + 1$ vertices and a cycle of length j contains
 1950 j vertices. Note that graphs that are $F_{I,J}$ -free cannot contain at the same time k
 1951 connected components that are paths of length i for every $i \in I$ and k connected
 1952 components which are cycles of length j for every $j \in J$. We let $F_{\emptyset, \emptyset}^{\text{large}}$ be a path
 1953 of length $k + 1$ in which both vertices of degree 1 are marked as ‘partial’ and every
 1954 other vertex is marked ‘full’. We further let $F_{I,J}^{\text{large}}$ be the disjoint union of $F_{I,J}$ and
 1955 $F_{\emptyset, \emptyset}^{\text{large}}$ for $I \subseteq [k]$, $J \subseteq \{3, \dots, k\}$ with $I \cup J \neq \emptyset$. Note that graphs that are $F_{I,J}^{\text{large}}$ -free
 1956 cannot contain at the same time k connected components that are paths of length i
 1957 for every $i \in I$ and k connected components which are cycles of length j for every
 1958 $j \in J$ and one connected component containing at least $k + 1$ vertices.

1959 We obtain a sequence $\overline{\mathcal{F}'} = (\overline{\mathcal{F}'_n})_{n \in \mathbb{N}}$ by setting

$$1960 \quad \mathcal{F}'_n := \mathcal{F}_n \cup \left\{ F \in \{F_{\emptyset, \emptyset}^{\text{large}}, F_{I, J}, F_{I, J}^{\text{large}} : I \subseteq [k], J \subseteq \{3, \dots, k\}, \right. \\ 1961 \quad \left. I \cup J \neq \emptyset\} : \text{every } G \in \mathcal{P}_n \text{ is } F\text{-free} \right\}.$$

1962 First observe that by construction \mathcal{P} must be $\overline{\mathcal{F}'}$ -local.

1963 We use the following notation. For a graph $G \in \mathcal{C}_2$, $i \in [k]$ and $j \in \{3, \dots, k\}$ we
1964 let

- 1965 • $p_i(G)$ be the number of connected components of G that are path of length
1966 i .
- 1967 • $c_j(G)$ be the number of connected components of G that are cycles of length
1968 j .
- 1969 • $cc^{\text{large}}(G)$ be the number of connected components of G with more than k
1970 vertices.

1971 We choose the following (monotonically non-decreasing) function $\tau(\epsilon) := \min(1, 8k^3\epsilon)$
1972 for $\epsilon \in (0, 1]$. First consider the two trivial cases. If \mathcal{P}_n contains all n -vertex graphs
1973 in \mathcal{C}_2 , then every n -vertex graph is close to \mathcal{P}_n . On the other hand, if \mathcal{P}_n is empty,
1974 then there are no n -vertex \mathcal{F}_n -free graphs. **Thus, the condition for non-propagation**
1975 **is satisfied.** Hence, we may assume that \mathcal{P}_n neither contains all n -vertex graphs in \mathcal{C}_2
1976 nor is empty.

1977 Let G be an n -vertex graph which is not \mathcal{F}'_n -free. Let $B \subseteq V(G)$ be any set
1978 covering \mathcal{F}'_n . To show that \mathcal{F}' is non-propagating it is sufficient to show that G is
1979 $\tau(|B|/n)$ -close to \mathcal{P} . By choice of τ this means that we have to argue that we can make
1980 G have property \mathcal{P}_n by modifying at most $16k^3|B|$ edges. Hence for the remainder of
1981 this proof we argue that G is $\tau(|B|/n)$ -close to \mathcal{P} .

1982 Assuming $n < 8k^3$, we get that $\tau(|B|/n) = 1$ (since G is not \mathcal{F}'_n -free we know
1983 that $|B| \geq 1$), which means G is $\tau(|B|/n)$ -close to \mathcal{P} , as in this case we can modify
1984 all edges of G and hence we can make G into any graph in \mathcal{P}_n . Hence we now assume
1985 that $n \geq 8k^3$.

1986 Now consider the case that $|B| \geq \frac{n}{8k}$. In this case $\tau(|B|/n) = 1$ and G is $\tau(|B|/n)$ -
1987 close to \mathcal{P} again because we are allowed to modify all edges of G which allows us to
1988 make G into any graph in \mathcal{P}_n . Hence from now on we only consider the case that
1989 $|B| \leq \frac{n}{8k}$.

1990 Let S be the set of vertices for which the k -neighbourhood does not contain any
1991 vertex from B . Let $I \subseteq [k]$, $J \subseteq \{3, \dots, k\}$ be the sets of indices such that $i \in I$ if
1992 and only if $p_i(G[S]) \geq k$ and $j \in J$ if and only if $c_j(G[S]) \geq k$. Note that $I \cup J$ could
1993 be empty.

1994 **Case 1:** Assume that $F_{I, J}^{\text{large}} \notin \mathcal{F}'_n$.

1995 First note that every component of size at most k which contains a vertex from S
1996 cannot contain a vertex from B by definition of S . Hence every connected component
1997 of G of size at most k is either fully contained in S or disjoint from S . Since there are
1998 at most $2k$ isomorphism types of connected components of size at most k we know
1999 that there are at most $2k^2$ connected components X of $G[S]$ such that there are at
2000 most $k - 1$ other connected components of $G[S]$ isomorphic to X . In other words,
2001 there are at most $2k^2$ components X of G containing no element from B such that
2002 if X is a path of length i then $i \notin I$ and if X is a cycle of length j then $j \notin J$. We
2003 now obtain G' by the following edge modifications from G . For every cycle of length

2004 j where $j \notin J$, we delete one edge (at most $2k^2 + |B|$ edge by our previous argument).
 2005 Then we add edges connecting all path (including the paths obtained in the last step)
 2006 of length i for $i \notin I$ to one long cycle C (at most $2k^2 + |B|$ edge additions). If C
 2007 has length less or equal to k there must be $i \in I$ or $j \in J$ such that $p_i(G) > k$ or
 2008 $c_j(G) > k$, in which case we include one respective component in C and repeat this
 2009 until C has length at least $k + 1$ (at most $2k$ modifications). Since in total we did at
 2010 most $4k^2 + 2|B| + 2k \leq 16k^2|B|$ edge modifications, G is $\tau(|B|/n)$ -close to G' . The
 2011 following claim completes the proof of Case 1 by showing that $G' \in \mathcal{P}_n$.

2012 **CLAIM 7.17.** *Let $I \subseteq [k]$, $J \subseteq \{3, \dots, k\}$ and $a_i, b_j \geq k$ where $i \in I$, $j \in J$ be any*
 2013 *selection of integers such that*

$$2014 \quad (7.1) \quad \sum_{i \in I} i \cdot a_i + \sum_{j \in J} j \cdot b_j \leq n - (k + 1).$$

2015 *If $F_{I,J}^{\text{large}} \notin \mathcal{F}'_n$, then any n -vertex graph $H \in \mathcal{C}_2$ with $p_i(H) = a_i$, $c_j(H) = b_j$ for $i \in I$,*
 2016 *$j \in J$, and one additional connected component which is a cycle is \mathcal{F}'_n -free.*

2017 *Proof.* Assume there is a graph $H \in \mathcal{P}_n$ as given in the statement which is
 2018 not \mathcal{F}'_n -free and let C be the cycle in H of length larger than k . Then there is
 2019 $F \in \mathcal{F}'_n$ such that there is an embeddings $f : V(F) \rightarrow V(H)$. Since $F_{I,J}^{\text{large}} \notin \mathcal{F}'_n$, by
 2020 construction there is a graph $H' \in \mathcal{P}_n$ with $p_i(H') \geq k$, $c_j(H') \geq k$ for $i \in I$, $j \in J$
 2021 and $cc^{\text{large}}(H') \geq 1$. We let C' be a connected component of H' of size larger than
 2022 k . To find an embedding of F into H' , for every connected component X of H of
 2023 size at most k which contains a vertex from $f(V(F))$, we pick a unique connected
 2024 component X' of H' which is isomorphic to X . Note that because $|f(V(F))| \leq k$ and
 2025 $p_i(H') \geq k$, $c_j(H') \geq k$ we can pick the connected component in H' uniquely. For
 2026 every connected component X of H of size at most k which contains a vertex from
 2027 $f(V(F))$, we now define f_X to be an isomorphism from X to X' . Furthermore, we pick
 2028 an injective graph homomorphism $f^{\text{large}} : f(V(F)) \cap C \rightarrow C'$. Again, this is possible
 2029 because $|f(V(F))| \leq k$. We now let $f'(v) := f_X(f(v))$ if $f(v)$ is in the connected
 2030 component X and $f'(v) := f^{\text{large}}(f(v))$ if $f(v)$ is in C . Note that f' is injective by
 2031 construction. Furthermore, as a consequence of picking f_X to to be isomorphisms and
 2032 f^{large} to be a homomorphism we get that f' is an embedding of F into H' . To see
 2033 this we observe that for any vertex $v \in V(F)$ which is marked as ‘full’ and for which
 2034 $f'(v)$ is in a connected component X with at most k vertices we obtain the condition
 2035 $N_1^{H'}(f'(v)) = f'(N_1^F(v))$ from f_X being an isomorphism. On the other hand, in case
 2036 $f'(v)$ is in C' and v is marked ‘full’ we get that v has two neighbours w_1, w_2 in F and
 2037 $f(w_1), f(w_2)$ are neighbours of $f(v)$ (since $f(v)$ must be on C) which implies that
 2038 $f'(w_1)$ and $f'(w_2)$ are neighbours of $f'(v)$ (since f^{large} is a homomorphism). Since f'
 2039 is an embedding of F into H' we obtain a contradiction to $H' \in \mathcal{P}_n$ and hence H is
 2040 \mathcal{F}'_n -free. Therefore H must be \mathcal{F}'_n -free as claimed. \square

2041 **Case 2:** Assume that $F_{I,J}^{\text{large}} \in \mathcal{F}'_n$. In this case our strategy is to modify the
 2042 connected components of G containing a vertex from B into paths and cycles of length
 2043 i for $i \in I$ or $i \in J$, respectively.

2044 Since the k -neighbourhood of every vertex contains no more than $2k + 1$ vertices,
 2045 $|B| \leq \frac{n}{8k}$ implies that $|S| \geq n/2$. Furthermore, since $F_{I,J}^{\text{large}} \in \mathcal{F}'_n$ no vertex in S can be
 2046 contained in a connected component of size larger than k as otherwise there would be
 2047 an embedding of $F_{I,J}^{\text{large}}$ into G which is not covered by B . Hence $G[S]$ is the disjoint
 2048 union of paths of length at most $k - 1$ and cycles of length at most k . Since $|S| \geq 4k^3$

2049 and $G[S]$ contains at most $2k$ different isomorphism types of connected components
 2050 and each of the connected components has at most k vertices we conclude that at
 2051 least $2k \geq k + 1$ of the connected components of $G[S]$ are pairwise isomorphic. Hence
 2052 $I \cup J \neq \emptyset$. Furthermore, $F_{I,J}$ is defined and not in \mathcal{F}'_n since B covers \mathcal{F}'_n .

2053 The next claim is the key to showing that we can modify G into having property
 2054 \mathcal{P} without modifying more than a constant number of edges in $G[S]$.

2055 CLAIM 7.18. *If for $I \subseteq [k]$, $J \subseteq \{3, \dots, k\}$ with $I \cup J \neq \emptyset$ we have that $F_{I,J}^{\text{large}} \in \mathcal{F}'_n$
 2056 and $F_{I,J} \notin \mathcal{F}'_n$ then for any selection of integers $a_i, b_j \geq k$ where $i \in I$, $j \in J$ such
 2057 that*

$$2058 \quad (7.2) \quad \sum_{i \in I} i \cdot a_i + \sum_{j \in J} j \cdot b_j \leq n - k^3$$

2059 *there is an \mathcal{F}'_n -free graph $H \in \mathcal{P}_n$ such that $p_i(H) \geq a_i$ and $c_j(H) \geq b_j$.*

2060 *Proof.* We set $a_i = 0$ for $i \in [k] \setminus I$ and $b_j = 0$ for $j \in \{3, \dots, k\} \setminus J$. Since
 2061 $F_{I,J}^{\text{large}} \in \mathcal{F}'_n$ and $F_{I,J} \notin \mathcal{F}'_n$ by construction of \mathcal{F} there must be a graph in \mathcal{P}_n whose
 2062 connected components include at least k paths of length i for every $i \in I$, k cycles of
 2063 length j for every $j \in J$ and no connected component containing more than k vertices.
 2064 Pick H amongst all graphs in \mathcal{P}_n with these properties such that

$$2065 \quad (*) := \sum_{\substack{i \in [k] \\ p_i(H) < a_i}} a_i - p_i(H) + \sum_{\substack{j \in \{3, \dots, k\} \\ c_j(H) < b_j}} b_j - c_j(H)$$

2066 is minimal. In case $(*) > 0$ there is $i \in [k]$ such that either $p_i(H) < a_i$ or $c_i(H) < b_i$.
 2067 Combining this with Equation 7.2 we obtain that there must be $j \in [k]$ such that
 2068 either $p_j(H) - a_j > k$ or $c_j(H) - b_j > k$. We let H' be the graph obtained from
 2069 H by replacing i connected components which are paths of length j or i connected
 2070 components which are cycles of length j , respectively, and adding j disjoint paths of
 2071 length i or j disjoint cycles of length i , respectively. By choice of i, j we get that

$$2072 \quad (*) > \sum_{\substack{i \in [k] \\ p_i(H') < a_i}} a_i - p_i(H') + \sum_{\substack{j \in \{3, \dots, k\} \\ c_j(H') < b_j}} b_j - c_j(H').$$

2073 Furthermore, H' must be \mathcal{F}'_n -free which we will argue in the following. Assume this
 2074 is not the case and there is $F \in \mathcal{F}'_n$ and an embedding $f : V(F) \rightarrow V(H')$. We obtain
 2075 a map $f' : V(F) \rightarrow V(H)$ from f as follows. For every connected component X in
 2076 H' which has been altered we pick a unique connected component X' of H' which is
 2077 isomorphic to X and contains no vertex in the image $f(V(F))$. This is possible as
 2078 the assumption that X was altered implies that either $|X| - 1 \in I$ or $|X| \in J$ and
 2079 hence there are at least k connected components isomorphic to X in H' which were
 2080 not altered. Since further $|f(V(F))| \leq k$ we can pick the X' uniquely. We now let
 2081 f_X be an isomorphism from X to X' for every connected component X which has
 2082 been altered and f_X the identity for every connected component X which has not
 2083 been altered. We define $f'(v) := f_X(v)$ for $v \in X$. By construction f' is obviously
 2084 an embedding of F into H . Since $H \in \mathcal{P}_n$ this yields a contradiction. Hence the
 2085 existence of H' contradicts the assumption that $(*) > 0$ which implies that H has the
 2086 claimed properties. \square

2087 First observe that $n \geq 8k^3$ allows us to chose a_i and b_j for every $i \in I$ and $j \in J$ in such
 2088 a way that $k \leq a_i \leq p_i(G[S])$, $k \leq b_j \leq c_j(G[S])$ and $\sum_{i \in I} i \cdot a_i + \sum_{j \in J} j \cdot b_j \leq n - k^3$.

2089 Amongst all such choices we pick a_i and b_j such that $\sum_{i \in I} i \cdot a_i + \sum_{j \in J} j \cdot b_j$ is
 2090 maximum. Let M be a set of vertices containing all connected components of G
 2091 apart from a_i paths and b_j cycles from $G[S]$ for every $i \in I, j \in J$. Then $|M| \leq$
 2092 $2k|B| + |B| + 4k^3$ since M consists of $N_k^G(B)$ (at most $2k|B| + |B|$ vertices), all
 2093 vertices in a connected component which is either a path of length i for $i \notin I$ or a
 2094 cycle of length j for $j \notin J$ (since there are at most $2k^2$ such paths and cycles (as
 2095 argued in Case 1) and each contains at most k vertices) or in case $a_i \neq p_i(G)$ or
 2096 $b_j \neq c_j(G)$ for some $i \in I, j \in J$, M consist of at most $k^3 + k$ vertices as we picked
 2097 a_i, b_j to maximise $\sum_{i \in I} i \cdot a_i + \sum_{j \in J} j \cdot b_j$.

2098 Now we use Claim 7.18 and obtain an \mathcal{F}_n -free graph $H \in \mathcal{P}_n$ such that $p_i(H) \geq a_i$
 2099 and $c_j(H) \geq b_j$. Hence we can modify G into a graph G' which is isomorphic to H
 2100 by only modifying $G[M]$. Since we can modify $G[M]$ into any graph with no more
 2101 than $4k|B| + 2|B| + 8k^3 \leq 16k^3|B|$ modifications we showed that G is $\tau(|B|/n)$ -close
 2102 to having \mathcal{P} . \square

2103 **8. Conclusion.** We studied testability of properties definable in first-order logic
 2104 in the bounded-degree model of property testing for graphs and relational structures,
 2105 where *testability* of a property means that it is testable with constant query complex-
 2106 plexity. We showed that all properties in Σ_2 are testable (Theorem 6.1), and we
 2107 complemented this by exhibiting a property (of relational structures) in Π_2 that is
 2108 not testable (Theorem 4.7). Using a hardness reduction, we also exhibit a property
 2109 of undirected, 3-regular graphs in Π_2 that is not testable (Theorem 5.1). The ques-
 2110 tion whether first-order definable properties are testable with a *sublinear* number of
 2111 queries (e.g. \sqrt{n}) in the bounded-degree model is left open.

2112 Similar results (on the separation between Σ_2 and Π_2 properties) were obtained
 2113 in the dense graph model in [4], albeit with very different methods. Indeed, non-
 2114 testability of first-order logic in the bounded-degree model is somewhat unexpected:
 2115 Testing algorithms proceed by sampling vertices and then exploring their local neigh-
 2116 bourhoods, and it is well-known that first-order logic can only express ‘local’ proper-
 2117 ties. On graphs and structures of bounded degree this is witnessed by Hanf’s strong
 2118 normal form of first-order logic [24], which is built around the absence and presence
 2119 of different isomorphism types of local neighbourhoods. However, our negative result
 2120 shows that locality of first-order logic is not sufficient for testability. This also answers
 2121 an open question from [1].

2122 We obtained our non-testable properties by encoding the zig-zag construction of
 2123 bounded-degree expanders into first-order logic on relational structures (Theorem 4.4)
 2124 and then extending this to undirected graphs (Theorem 5.1). We believe that this
 2125 will be of independent interest. We remark that it might also be possible to use the
 2126 iterative construction of replacement product graphs of [33] instead of the zig-zag
 2127 construction to obtain a similar example.

2128 We then used our non-testable graph property to answer a question on *proximity*
 2129 *oblivious* testers in the bounded-degree model, asked by Goldreich and Ron more
 2130 than 10 years ago [22]. Such a tester is particularly simple: it performs a basic
 2131 test a number of times that may depend on the proximity parameter, whereas the
 2132 basic test is oblivious of the parameter. In [22], the properties that are testable in this
 2133 model have been characterised as those that are both *GSF-local*, and *non-propagating*.
 2134 Roughly speaking, *GSF-local* means that the graph class omits a family of *generalised*
 2135 *subgraphs* (i. e. subgraphs with constraints on how the subgraphs interact with the
 2136 rest of the graph), and *non-propagating* means that graphs in which a forbidden
 2137 generalised subgraph is unlikely to be detected by sampling vertices are actually close

2138 to having the property in terms of edge modifications. In other words, no ‘chain
 2139 reactions’ occur, where repairing one edge will produce new unwanted configurations
 2140 that again need repairing, etc. Goldreich and Ron asked whether ‘non-propagating’ is
 2141 necessary. We showed that this is the case. Our proof is based on relating first-order
 2142 definable properties to GSF-local properties, via a notion that we call neighbourhood
 2143 profiles, which captures first-order definability.

2144 **Appendix A. Deferred Proofs from Section 7.3.1.**

2145 CLAIM A.1. *Every structure $A \in \bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k} \setminus \{A_\emptyset\}$ satisfies φ_{tree} .*

2146 *Proof.* Let $A \in \bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k} \setminus \{A_\emptyset\}$. Then there is $k \in \{1, \dots, m\}$ such that
 2147 $A \in \mathcal{P}_{\rho_k}$.

2148 By definition, $\varphi_{\text{tree}} := \exists^{\leq 1} x \varphi_{\text{root}}(x) \wedge \varphi \wedge \forall x (\psi(x) \vee \chi(x))$, where

$$2149 \quad \varphi := \forall x \left((\varphi_{\text{root}}(x) \wedge R(x, x)) \vee (\exists^=1 y F(y, x) \wedge \neg \exists y R(x, y) \wedge \neg \exists y R(y, x)) \right),$$

$$2150 \quad \psi(x) := \neg \exists y F(x, y) \wedge \bigwedge_{k \in ([D]^2)^2} L_k(x, x) \wedge \forall y (y \neq x \rightarrow$$

$$2151 \quad \bigwedge_{k \in ([D]^2)^2} \neg L_k(x, y) \wedge \bigwedge_{k \in ([D]^2)^2} \neg L_k(y, x))$$

2152 and

$$2153 \quad \chi(x) := \neg \exists y \bigvee_{k \in ([D]^2)^2} (L_k(x, y) \vee L_k(y, x)) \wedge \bigwedge_{k \in ([D]^2)^2} \exists y_k (x \neq y_k \wedge F_k(x, y_k)$$

$$2154 \quad \wedge \left(\bigwedge_{k' \in ([D]^2)^2, k' \neq k} \neg F_{k'}(x, y_k) \right) \wedge \forall y (y \neq y_k \rightarrow \neg F_k(x, y)).$$

2155 Thus, it is sufficient to prove that $A \models \exists^{\leq 1} x \varphi_{\text{root}}(x)$, $A \models \varphi$ and $A \models \forall x (\psi(x) \vee \chi(x))$.

2156 To prove $A \models \exists^{\leq 1} x \varphi_{\text{root}}(x)$ we note that by construction of ρ_k we have $A \not\models$
 2157 $\varphi_{\text{root}}(a)$ for any $a \in U(A)$ for which $(\mathcal{N}_2^A(a), a) \notin \tau_{d,2,\sigma}^k$. Since ρ_k restricts the number
 2158 of occurrences of elements of neighbourhood type $\tau_{d,2,\sigma}^k$ to at most one, this proves
 2159 that there is at most one $a \in U(A)$ with $A \models \varphi_{\text{tree}}(a)$ and hence $A \models \exists^{\leq 1} x \varphi_{\text{root}}(x)$.

2160 To prove $A \models \varphi$, let $a \in U(A)$ be an arbitrary element. Since $A \in \mathcal{P}_{\rho_k}$, there is
 2161 an $i \in I_k$ such that $(\mathcal{N}_2^A(a), a) \in \tau_{d,2,\sigma}^i$. But then by definition, there exist $\tilde{A} \models \varphi_{\mathbb{Z}}$

2162 and $\tilde{a} \in U(\tilde{A})$ such that $(\mathcal{N}_2^A(a), a) \cong (\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Assume f is an isomorphism from
 2163 $(\mathcal{N}_2^A(a), a)$ to $(\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. First consider the case that $A \models \varphi_{\text{root}}(a) := \forall y \neg F(y, a)$.

2164 Assume there is $\tilde{b} \in U(\tilde{A})$ such that $(\tilde{b}, \tilde{a}) \in F(\tilde{A})$. Since $\tilde{b} \in \mathcal{N}_2^{\tilde{A}}(\tilde{a})$, there must be an
 2165 element $b \in \mathcal{N}_2^A(a)$ such that $f(b) = \tilde{b}$. Since f is an isomorphism mapping a to \tilde{a} , this
 2166 implies $(b, a) \in F(A)$, which contradicts $A \models \varphi_{\text{root}}(a)$. Hence $\tilde{A} \models \varphi_{\text{root}}(\tilde{a})$. Since

2167 $\tilde{A} \models \varphi_{\text{tree}}$, it holds that $\tilde{A} \models \varphi$, which means that $(\tilde{a}, \tilde{a}) \in R(\tilde{A})$. But since f is an
 2168 isomorphism mapping a onto \tilde{a} , this implies $(a, a) \in R(A)$. Now consider the case that

2169 $A \not\models \varphi_{\text{root}}(a)$. Then there is $b \in U(A)$ with $(b, a) \in F(A)$. Since f is an isomorphism,
 2170 this implies $(f(b), \tilde{a}) \in F(\tilde{A})$. Hence $\tilde{A} \models \exists^=1 y F(y, \tilde{a}) \wedge \neg \exists y R(\tilde{a}, y) \wedge \neg \exists y R(y, \tilde{a})$, as
 2171 $\tilde{A} \models \varphi$. Now assume that there is $b' \neq b$ such that $(b', a) \in F(A)$. Then $f(b) \neq f(b')$

2172 and $(f(b), \tilde{a}), (f(b'), \tilde{a}) \in F(\tilde{A})$. Since this contradicts $\tilde{A} \models \exists^=1 y F(y, \tilde{a})$ we have $A \models$
 2173 $\exists^=1 y F(y, a)$. Furthermore, assume that there is $b' \in U(A)$ such that either $(a, b') \in$

2174 $R(A)$ or $(b', a) \in R(A)$. Then either $(\tilde{a}, f(b')) \in R(\tilde{A})$ or $(f(b'), \tilde{a}) \in R(\tilde{A})$, which
 2175 contradicts $\tilde{A} \models \neg \exists y R(\tilde{a}, y) \wedge \neg \exists y R(y, \tilde{a})$. Therefore $A \models \neg \exists y R(a, y) \wedge \neg \exists y R(y, a)$

2176 which completes the proof of $A \models \varphi$.

2177 We prove $A \models \forall x(\psi(x) \vee \chi(x))$ by considering the two cases $A \models \neg \exists y F(a, y)$
 2178 and $A \models \exists y F(a, y)$ for each element $a \in U(A)$. For this, let $a \in U(A)$ be any
 2179 element. By the construction of ρ_k there is $\tilde{A} \models \varphi_{\mathbb{Z}}$ and $\tilde{a} \in U(\tilde{A})$ such that
 2180 $(\mathcal{N}_2^{\tilde{A}}(a), a) \cong (\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Let f be an isomorphism from $(\mathcal{N}_2^{\tilde{A}}(a), a)$ to $(\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$.
 2181 First consider the case that $A \models \neg \exists y F(a, y)$. If there was $\tilde{b} \in U(\tilde{A})$ with $(\tilde{a}, \tilde{b}) \in$
 2182 $F(\tilde{A})$ then $(a, f^{-1}(\tilde{b})) \in F(A)$ contradicting our assumption. Hence $\tilde{A} \models \neg \exists y F(\tilde{a}, y)$
 2183 which implies that $\tilde{A} \not\models \chi(\tilde{a})$. But since $\tilde{A} \models \varphi_{\mathbb{Z}}$, it holds that $\tilde{A} \models \forall x(\psi(x) \vee$
 2184 $\chi(x))$, which implies that $\tilde{A} \models \psi(\tilde{a})$. Hence $(\tilde{a}, \tilde{a}) \in L_k(\tilde{A})$ for every $k \in ([D]^2)^2$.
 2185 Since f is an isomorphism and $f(a) = \tilde{a}$, it holds that $(a, a) \in L_k(A)$ for every
 2186 $k \in ([D]^2)^2$, and hence $A \models \bigwedge_{k \in ([D]^2)^2} L_k(a, a)$. Furthermore, assume that there is
 2187 $b \in U(A)$, $b \neq a$ and $k \in ([D]^2)^2$ such that either $(a, b) \in L_k(A)$ or $(b, a) \in L_k(A)$.
 2188 Since f is an isomorphism this implies that either $(\tilde{a}, f(b)) \in L_k(\tilde{A})$ or $(f(b), \tilde{a}) \in$
 2189 $L_k(\tilde{A})$ which contradicts $\tilde{A} \models \chi(\tilde{a})$. Hence $A \models \forall y (y \neq a \rightarrow \bigwedge_{k \in ([D]^2)^2} \neg L_k(a, y) \wedge$
 2190 $\bigwedge_{k \in ([D]^2)^2} \neg L_k(y, a))$ proving that $A \models \psi(a)$.

2191 Now consider the case that there is an element $b \in U(A)$ such that $(a, b) \in F(A)$.
 2192 Since this implies that $(\tilde{a}, f(b)) \in F(\tilde{A})$, we get that $\tilde{A} \not\models \psi(\tilde{a})$, and hence $\tilde{A} \models \chi(\tilde{a})$.
 2193 Now assume that there is $b \in U(A)$ and $k \in ([D]^2)^2$ such that either $(a, b) \in L_k(A)$
 2194 or $(b, a) \in L_k(A)$. But then either $(\tilde{a}, f(b)) \in L_k(\tilde{A})$ or $(f(b), \tilde{a}) \in L_k(\tilde{A})$, which
 2195 contradicts $\tilde{A} \models \chi(\tilde{a})$. Hence $A \models \neg \exists y \bigvee_{k \in ([D]^2)^2} (L_k(a, y) \vee L_k(y, a))$. For each
 2196 $k \in ([D]^2)^2$, let $\tilde{b}_k \in U(\tilde{A})$ be an element such that $\tilde{A} \models \tilde{a} \neq \tilde{b}_k \wedge F_k(\tilde{a}, \tilde{b}_k) \wedge$
 2197 $(\bigwedge_{k' \in ([D]^2)^2, k' \neq k} \neg F_{k'}(\tilde{a}, \tilde{b}_k)) \wedge \forall y (y \neq \tilde{b}_k \rightarrow \neg F_k(\tilde{a}, y))$. Since f is an isomorphism,
 2198 this implies that $a \neq b_k := f^{-1}(\tilde{b}_k)$, $(a, b_k) \in F_k(A)$ and $(a, b_k) \notin F_{k'}(A)$, for each
 2199 $k' \in ([D]^2)^2, k' \neq k$. Furthermore, assume there is $b \in U(A)$, $b \neq b_k$ such that
 2200 $(a, b) \in F_k(A)$. Since f is an isomorphism, this implies $f(b) \neq f(b_k) = \tilde{b}_k$ and
 2201 $(\tilde{a}, \tilde{b}) \in F_k(\tilde{A})$, which contradicts $\tilde{A} \models \forall y (y \neq \tilde{b}_k \rightarrow \neg F_k(\tilde{a}, y))$. Hence $A \models \forall y (y \neq$
 2202 $b_k \rightarrow \neg F_k(a, y))$ and therefore concluding that $A \models \chi(a)$. This proves that in either
 2203 case $A \models \psi(a) \vee \chi(a)$ and therefore $A \models \forall x(\psi(x) \vee \chi(x))$. \square

2204 CLAIM A.2. Every structure $A \in \bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k} \setminus \{A_\emptyset\}$ satisfies $\varphi_{\text{rotationMap}}$.

2205 *Proof.* Let $A \in \bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k} \setminus \{A_\emptyset\}$. Then there is a $k \in \{1, \dots, m\}$ such that
 2206 $A \in \mathcal{P}_{\rho_k}$.

2207 By definition, $\varphi_{\text{rotationMap}} = \varphi \wedge \psi$, where

$$2208 \quad \varphi := \forall x \forall y \left(\bigwedge_{i, j \in [D]^2} (E_{i, j}(x, y) \rightarrow E_{j, i}(y, x)) \right) \text{ and}$$

$$2209 \quad \psi := \forall x \left(\bigwedge_{i \in [D]^2} \left(\bigvee_{j \in [D]^2} (\exists^{=1} y E_{i, j}(x, y) \wedge \bigwedge_{\substack{j' \in [D]^2 \\ j' \neq j}} \neg \exists y E_{i, j'}(x, y)) \right) \right).$$

2210 Thus, it is sufficient to prove that $A \models \varphi$ and $A \models \psi$.

2211 To prove $A \models \varphi$, assume towards a contradiction that there are $a, b \in U(A)$ such
 2212 that for some pair $i, j \in [D]^2$, we have that $(a, b) \in E_{i, j}(A)$, but $(b, a) \notin E_{j, i}(A)$.
 2213 By construction of \mathcal{P}_{ρ_k} , there is a structure $\tilde{A} \models \varphi_{\mathbb{Z}}$ and $\tilde{a} \in U(\tilde{A})$ such that
 2214 $(\mathcal{N}_2^{\tilde{A}}(a), a) \cong (\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Assume f is an isomorphism from $(\mathcal{N}_2^{\tilde{A}}(a), a)$ to $(\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$.
 2215 Note that $f(b)$ is defined since b is in the 2-neighbourhood of a . Furthermore since
 2216 f is an isomorphism, $(a, b) \in E_{i, j}(A)$ implies $(\tilde{a}, f(b)) \in E_{i, j}(\tilde{A})$, and $(b, a) \notin E_{j, i}(A)$
 2217 implies $(f(b), \tilde{a}) \notin E_{j, i}(\tilde{A})$. Hence $\tilde{A} \not\models \varphi$, which contradicts $\tilde{A} \models \varphi_{\text{rotationMap}}$.

2218 To prove $A \models \psi$, assume towards a contradiction that there is an $a \in U(A)$ and
 2219 $i \in [D]^2$ such that $A \not\models \exists^{=1} y E_{i,j}(a, y) \wedge \bigwedge_{j' \neq j} \neg \exists y E_{i,j'}(a, y)$ for every $j \in [D]^2$.

2220 We know that there is a structure $\tilde{A} \models \varphi_{\mathbb{Z}}$ and $\tilde{a} \in U(\tilde{A})$ such that $(\mathcal{N}_2^A(a), a) \cong$
 2221 $(\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Let f be an isomorphism from $(\mathcal{N}_2^A(a), a)$ to $(\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Since $\tilde{A} \models \psi$,
 2222 there must be $j \in [D]^2$ such that $\tilde{A} \models \exists^{=1} y E_{i,j}(\tilde{a}, y) \wedge \bigwedge_{j' \neq j} \neg \exists y E_{i,j'}(\tilde{a}, y)$. Hence

2223 there must be $\tilde{b} \in U(\tilde{A})$ such that $(\tilde{a}, \tilde{b}) \in E_{i,j}(\tilde{A})$, which implies that $(a, f^{-1}(\tilde{b})) \in$
 2224 $E_{i,j}(A)$. Since we assumed that $A \not\models \exists^{=1} y E_{i,j}(a, y) \wedge \bigwedge_{j' \neq j} \neg \exists y E_{i,j'}(a, y)$, there

2225 must be either $b \neq f^{-1}(\tilde{b})$ with $(a, b) \in E_{i,j}(A)$, or there must be $j' \in [D]^2$, $j' \neq j$
 2226 and $b' \in U(A)$ such that $(a, b') \in E_{i,j'}(A)$. In the first case $(\tilde{a}, f(b)) \in E_{i,j}(\tilde{A})$, since
 2227 f is an isomorphism. But then $\tilde{A} \not\models \exists^{=1} y E_{i,j}(\tilde{a}, y)$, which is a contradiction. In the
 2228 second case, we get that $(\tilde{a}, f(b')) \in E_{i,j'}(\tilde{A})$. But then $\tilde{A} \not\models \bigwedge_{j' \neq j} \neg \exists y E_{i,j'}(\tilde{a}, y)$,

2229 which is a contradiction. Hence $A \models \varphi \wedge \psi$. \square

2230 CLAIM A.3. *Every structure $A \in \bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k} \setminus \{A_\emptyset\}$ satisfies φ_{base} .*

2231 *Proof.* Let $A \in \bigcup_{1 \leq k \leq m} \mathcal{P}_{\rho_k} \setminus \{A_\emptyset\}$. Then there is a $k \in \{1, \dots, m\}$ such that
 2232 $A \in \mathcal{P}_{\rho_k}$.

2233 By definition, $\varphi_{\text{base}} := \forall x (\varphi_{\text{root}}(x) \rightarrow (\varphi(x) \wedge \psi(x)))$, where

$$2234 \quad \varphi(x) := \bigwedge_{i,j \in [D]^2} \left(E_{i,j}(x, x) \wedge \forall y (x \neq y \rightarrow (\neg E_{i,j}(x, y) \wedge \neg E_{i,j}(y, x))) \right) \text{ and}$$

$$2235 \quad \psi(x) := \bigwedge_{\substack{\text{ROT}_{H^2}(k,i)=(k',i') \\ k,k' \in ([D]^2)^2 \\ i,i' \in [D]^2}} \exists y \exists y' (F_k(x, y) \wedge F_{k'}(x, y') \wedge E_{i,i'}(y, y')).$$

2236 Thus, it is sufficient to prove that $A \models \varphi(a)$ and $A \models \psi(a)$ for every $a \in U(A)$ for
 2237 which $A \models \varphi_{\text{root}}(a)$. Therefore assume $a \in U(A)$ is any element such that $A \models$
 2238 $\varphi_{\text{root}}(a)$. Because $A \in \mathcal{P}_{\rho_k}$ there is an $i \in I_k$ such that $(\mathcal{N}_2^A(a), a) \in \tau_{d,2,\sigma}^i$. Then
 2239 by definition there is a structure $\tilde{A} \models \varphi_{\mathbb{Z}}$ and $\tilde{a} \in U(\tilde{A})$ such that $(\mathcal{N}_2^A(a), a) \cong$

2240 $(\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Let f be an isomorphism from $(\mathcal{N}_2^A(a), a)$ to $(\mathcal{N}_2^{\tilde{A}}(\tilde{a}), \tilde{a})$. Assume that
 2241 there is an element $\tilde{b} \in U(\tilde{A})$ such that $(\tilde{b}, \tilde{a}) \in F(\tilde{A})$. Since f is an isomorphism and
 2242 $\tilde{b} \in N_2^{\tilde{A}}(\tilde{a})$ we get that $(f^{-1}(\tilde{b}), a) \in F(A)$ which contradicts that $A \models \varphi_{\text{root}}(a)$ as
 2243 $\varphi_{\text{root}}(x) := \forall y \neg F(y, x)$. Hence there is no element $\tilde{b} \in U(\tilde{A})$ such that $(\tilde{b}, \tilde{a}) \in F(\tilde{A})$
 2244 which implies that $\tilde{A} \models \varphi_{\text{root}}(\tilde{a})$. But since $\tilde{A} \models \varphi_{\mathbb{Z}}$ we have that $\tilde{A} \models \varphi_{\text{base}}$ and
 2245 hence $\tilde{A} \models \varphi(\tilde{a})$ and $\tilde{A} \models \psi(\tilde{a})$.

2246 To prove $A \models \varphi(a)$ first observe that $(a, a) \in E_{i,j}(A)$ for every $i, j \in [D]^2$ since
 2247 $\tilde{A} \models \varphi(\tilde{a})$ implies that $(\tilde{a}, \tilde{a}) \in E_{i,j}(\tilde{A})$ for every $i, j \in [D]^2$ and f is an isomorphism
 2248 mapping a onto \tilde{a} . Assume that there is an element $b \in U(A)$, $b \neq a$ and indices
 2249 $i, j \in [D]^2$ such that either $(a, b) \in E_{i,j}(A)$ or $(b, a) \in E_{i,j}(A)$. Since $b \in N_2^A(a)$ and
 2250 f is an isomorphism we get that $f(b) \neq f(a) = \tilde{a}$ and either $(\tilde{a}, f(b)) \in E_{i,j}(\tilde{A})$ or
 2251 $(f(b), \tilde{a}) \in E_{i,j}(\tilde{A})$. But this contradicts $\tilde{A} \models \varphi(\tilde{a})$ and hence $A \models \varphi(a)$.

2252 We now prove that $A \models \psi(a)$. Let $k, k' \in ([D]^2)^2$ and $i, i' \in [D]^2$ such that
 2253 $\text{ROT}_{H^2}(k, i) = (k', i')$. Since $\tilde{A} \models \psi(\tilde{a})$ there must be elements $\tilde{b}, \tilde{b}' \in U(\tilde{A})$ such
 2254 that $(\tilde{a}, \tilde{b}) \in F_k(\tilde{A})$, $(\tilde{a}, \tilde{b}') \in F_{k'}(\tilde{A})$ and $(\tilde{b}, \tilde{b}') \in E_{i,i'}(\tilde{A})$. But since $\tilde{b}, \tilde{b}' \in N_2^{\tilde{A}}(\tilde{a})$ we
 2255 get that $f^{-1}(\tilde{b})$ and $f^{-1}(\tilde{b}')$ are defined and since f is an isomorphism we get that

2256 $(a, f^{-1}(\tilde{b})) \in F_k(A)$, $(a, f^{-1}(\tilde{b}')) \in F_{k'}(A)$ and $(f^{-1}(\tilde{b}), f^{-1}(\tilde{b}')) \in E_{i,i'}(A)$. Hence
 2257 $A \models \exists y \exists y' (F_k(a, y) \wedge F_{k'}(a, y') \wedge E_{i,i'}(y, y'))$ for any $k, k' \in ([D]^2)^2$ and $i, i' \in [D]^2$
 2258 such that $\text{ROT}_{H^2}(k, i) = (k', i')$ which implies that $A \models \psi(a)$. \square

2259 **Acknowledgments.** We thank Sebastian Ordyniak for inspiring discussions
 2260 and the anonymous reviewers for their detailed comments. The second author thanks
 2261 Michał Pilipczuk and Pierre Simon for inspiring discussions at the docCourse on
 2262 Sparsity in Prague 2018.

2263

REFERENCES

- 2264 [1] I. ADLER AND F. HARWATH, *Property testing for bounded degree databases*, in 35th Symposium
 2265 on Theoretical Aspects of Computer Science, STACS 2018, February 28 to March 3, 2018,
 2266 Caen, France, 2018, pp. 6:1–6:14.
- 2267 [2] I. ADLER, N. KÖHLER, AND P. PENG, *GSF-locality is not sufficient for proximity-oblivious*
 2268 *testing*, in 36th Computational Complexity Conference, CCC 2021, July 20-23, 2021,
 2269 Toronto, Ontario, Canada (Virtual Conference), V. Kabanets, ed., vol. 200 of LIPIcs,
 2270 Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021, pp. 34:1–34:27, [https://doi.org/](https://doi.org/10.4230/LIPIcs.CCC.2021.34)
 2271 [10.4230/LIPIcs.CCC.2021.34](https://doi.org/10.4230/LIPIcs.CCC.2021.34), <https://doi.org/10.4230/LIPIcs.CCC.2021.34>.
- 2272 [3] I. ADLER, N. KÖHLER, AND P. PENG, *On testability of first-order properties in bounded-*
 2273 *degree graphs*, in Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms,
 2274 SODA 2021, Virtual Conference, January 10 - 13, 2021, D. Marx, ed., SIAM, 2021,
 2275 pp. 1578–1597, <https://doi.org/10.1137/1.9781611976465.96>, <https://doi.org/10.1137/1.9781611976465.96>.
- 2276 [4] N. ALON, E. FISCHER, M. KRIVELEVICH, AND M. SZEGEDY, *Efficient testing of large graphs*,
 2277 *Combinatorica*, 20 (2000), pp. 451–476. Preliminary version in FOCS'99.
- 2278 [5] N. ALON, E. FISCHER, I. NEWMAN, AND A. SHAPIRA, *A combinatorial characterization of the*
 2279 *testable graph properties: it's all about regularity*, *SIAM Journal on Computing*, 39 (2009),
 2280 pp. 143–167.
- 2281 [6] N. ALON AND V. D. MILMAN, λ_1 , *isoperimetric inequalities for graphs, and superconcentrators*,
 2282 *Journal of Combinatorial Theory, Series B*, 38 (1985), pp. 73–88.
- 2283 [7] I. BENJAMINI, O. SCHRAMM, AND A. SHAPIRA, *Every minor-closed property of sparse graphs is*
 2284 *testable*, *Advances in mathematics*, 223 (2010), pp. 2200–2218.
- 2285 [8] M. BLUM, M. LUBY, AND R. RUBINFELD, *Self-testing/correcting with applications to numerical*
 2286 *problems*, *Journal of computer and system sciences*, 47 (1993), pp. 549–595.
- 2287 [9] B. BOLLIG AND D. KUSKE, *An optimal construction of hanf sentences*, *Journal of Applied*
 2288 *Logic*, 10 (2012), pp. 179–186, <https://doi.org/10.1016/j.jal.2012.01.002>, <http://dx.doi.org/10.1016/j.jal.2012.01.002>.
- 2289 [10] H. CHEN AND Y. YOSHIDA, *Testability of homomorphism inadmissibility: Property testing*
 2290 *meets database theory*, in Proceedings of the 38th ACM SIGMOD-SIGACT-SIGAI Sym-
 2291 *posium on Principles of Database Systems*, 2019, pp. 365–382.
- 2292 [11] A. CZUMAJ, P. PENG, AND C. SOHLER, *Relating two property testing models for bounded degree*
 2293 *directed graphs*, in Proceedings of the 48th Annual ACM SIGACT Symposium on Theory
 2294 of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016, 2016, pp. 1033–1045.
- 2295 [12] R. DIESTEL, *Graph Theory, 4th Edition*, vol. 173 of Graduate texts in mathematics, Springer,
 2296 2012.
- 2297 [13] J. DODZIUK, *Difference equations, isoperimetric inequality and transience of certain random*
 2298 *walks*, *Transactions of the American Mathematical Society*, 284 (1984), pp. 787–794.
- 2299 [14] H. EBBINGHAUS AND J. FLUM, *Finite model theory*, *Perspectives in Mathematical Logic*,
 2300 Springer, 1995.
- 2301 [15] H. FICHTENBERGER, P. PENG, AND C. SOHLER, *Every testable (infinite) property of bounded-*
 2302 *degree graphs contains an infinite hyperfinite subproperty*, in Proceedings of the Thirtieth
 2303 Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied
 2304 Mathematics, 2019, pp. 714–726.
- 2305 [16] S. FORSTER, D. NANONGKAI, T. SARANURAK, L. YANG, AND S. YINGCHAREONTHAWORNCHAI,
 2306 *Computing and testing small connectivity in near-linear time and queries via fast local cut*
 2307 *algorithms*, SODA, (2020).
- 2308 [17] H. GAIFMAN, *On local and non-local properties*, 1982.
- 2309 [18] O. GOLDBREICH, *Introduction to property testing*, Cambridge University Press, 2017.
- 2310 [19] O. GOLDBREICH, S. GOLDWASSER, AND D. RON, *Property testing and its connection to learning*

- 2313 *and approximation*, Journal of the ACM (JACM), 45 (1998), pp. 653–750.
- 2314 [20] O. GOLDREICH AND T. KAUFMAN, *Proximity oblivious testing and the role of invariances*, in
2315 Studies in Complexity and Cryptography. Miscellanea on the Interplay between Random-
2316 ness and Computation, Springer, 2011, pp. 173–190.
- 2317 [21] O. GOLDREICH AND D. RON, *Property testing in bounded degree graphs*, Algorithmica, 32
2318 (2002), pp. 302–343, <https://doi.org/10.1007/s00453-001-0078-7>, [http://dx.doi.org/10.](http://dx.doi.org/10.1007/s00453-001-0078-7)
2319 [1007/s00453-001-0078-7](http://dx.doi.org/10.1007/s00453-001-0078-7).
- 2320 [22] O. GOLDREICH AND D. RON, *On proximity-oblivious testing*, SIAM Journal on Computing, 40
2321 (2011), pp. 534–566.
- 2322 [23] O. GOLDREICH AND I. SHINKAR, *Two-sided error proximity oblivious testing*, Random Struc-
2323 tures & Algorithms, 48 (2016), pp. 341–383.
- 2324 [24] W. HANF, *The Theory of Models*, North Holland, 1965, ch. Model-theoretic methods in the
2325 study of elementary logic, pp. 132–145.
- 2326 [25] A. HASSIDIM, J. A. KELNER, H. N. NGUYEN, AND K. ONAK, *Local graph partitions for approx-
2327 imation and testing*, in 2009 50th Annual IEEE Symposium on Foundations of Computer
2328 Science, IEEE, 2009, pp. 22–31.
- 2329 [26] S. HOORY, N. LINIAL, AND A. WIGDERSON, *Expander graphs and their applications*, BULL.
2330 AMER. MATH. SOC., 43 (2006), pp. 439–561.
- 2331 [27] H. ITO, A. KHOURY, AND I. NEWMAN, *On the characterization of 1-sided error strongly-testable
2332 graph properties for bounded-degree graphs*, (to appear) Journal of Computational Com-
2333 plexity. arXiv:1909.09984, (2019).
- 2334 [28] K.-I. KAWARABAYASHI AND Y. YOSHIDA, *Testing subdivision-freeness: property testing meets
2335 structural graph theory*, in Proceedings of the forty-fifth annual ACM symposium on Theory
2336 of computing, ACM, 2013, pp. 437–446.
- 2337 [29] A. KUMAR, C. SESHADHRI, AND A. STOLMAN, *Random walks and forbidden minors ii: a poly (d
2338 ϵ -1)-query tester for minor-closed properties of bounded degree graphs*, in Proceedings of
2339 the 51st Annual ACM SIGACT Symposium on Theory of Computing, 2019, pp. 559–567.
- 2340 [30] L. LOVÁSZ, *Large Networks and Graph Limits*, vol. 60 of Colloquium Publications, American
2341 Mathematical Society, 2012.
- 2342 [31] I. NEWMAN AND C. SOHLER, *Every property of hyperfinite graphs is testable*, SIAM Journal on
2343 Computing, 42 (2013), pp. 1095–1112.
- 2344 [32] L. S. R. FAGIN AND M. VARDI, *On monadic np vs. monadic co - np* , Information and Computa-
2345 tion, 120 (1995), pp. 78–92.
- 2346 [33] O. REINGOLD, S. VADHAN, AND A. WIGDERSON, *Entropy waves, the zig-zag graph product, and
2347 new constant-degree expanders and extractors*, in Annals of Mathematics, 2000, pp. 157–
2348 187.
- 2349 [34] R. RUBINFELD AND M. SUDAN, *Robust characterizations of polynomials with applications to
2350 program testing*, SIAM Journal on Computing, 25 (1996), pp. 252–271.
- 2351 [35] Y. YOSHIDA AND H. ITO, *Property testing on k -vertex-connectivity of graphs*, Algorithmica, 62
2352 (2012), pp. 701–712.

2353 **List of Notation**

Notation	Description
\mathbb{N}	the set of natural numbers including 0
$\mathbb{N}_{>0}$	the set of all positive natural numbers
$[n]$	the set $\{0, \dots, n-1\}$
\sqcup	disjoint union
Δ	symmetric difference
$\tau_r(x)$	the neighbourhood of fixed radius r around x , up to isomorphism
$\mathcal{O}(\cdot), o(\cdot)$	asymptotic upper-bounds
\mathbb{Z}	zig-zag product
$h(G)$	expansion ratio of G
$\langle S, T \rangle_G$	edges crossing S and T in G
σ	signature
σ_{graph}	signature with one binary relation symbol E
$\deg_G(v)$	degree of vertex v in G
$\text{dist}_A(v, w)$	distance between two vertices v and w in A
$\text{dist}(A, \mathcal{P})$	distance between structure A and property \mathcal{P}
ROT_G	rotation map of G
$M_{u,v}$	the (u, v) entry of the normalised adjacency matrix M of G
$A[S]$	substructure of A induced by S
$N_r^A(a)$	the r -neighborhood of a in structure A
2354 $\text{ar}(R)$	arity of relation R
\mathcal{C}_d	class of graphs of bounded degree d
$\mathcal{C}_{\sigma,d}$	class of σ -structures of bounded degree d
\mathcal{P}_φ	property defined by formula φ
$\Sigma_i, \Pi_i, \Delta_i$	prefix classes with $i-1$ quantifier alterations
\models	is a model of
\equiv	equivalence of FO-formulas
\equiv_d	equivalence of FO-formulas on structures of bounded degree d
$\neg, \wedge, \vee, \rightarrow, \leftrightarrow$	logical negation, conjunction, disjunction, implication and biimplication
\exists, \forall	existential and universal quantifier
$\text{ans}_A(q)$	answer to query q to structure A
$E_{i,j}, F_k, R, L_k$	binary relation symbols for $i, j \in [D]^2$ and $k \in ([D]^2)^2$
$\underline{G}(A)$	underlying graph of a model A of $\varphi_{\mathbb{Z}}$
$\varphi_{\mathbb{Z}}$	the formula defined in Section 3 whose underlying graphs are expanders
\mathfrak{M}	a set of models of some sentence in Σ_2
\mathfrak{N}	a maximal set of pairwise non-isomorphic structures in $\mathcal{C}_{\sigma,d}$
$\text{pos}^i(\bar{x}, \bar{y}), \text{neg}^i(\bar{x}, \bar{y})$	a conjunction of atomic (resp. negated) formulas containing both variables from tuples \bar{x} and \bar{y}

Notation **Description** $\rho_{A,r}$ the r -type distribution of A $\delta_{\odot}^r(A, B)$ the sampling distance of depth r between two σ -structures A and B

2355

HNF Hanf normal form

POT proximity oblivious tester

GSF generalised subgraph freeness