A New Perspective on FO Model Checking of Dense Graph Classes

JAKUB GAJARSKÝ, Technical University Berlin, Germany
PETR HLINĚNÝ and JAN OBDRŽÁLEK, Masaryk University, Czech republic
DANIEL LOKSHTANOV, UC Santa Barabara, USA
M. S. RAMANUJAN, University of Warwick, Great Britain

We study the first-order (FO) model checking problem of dense graph classes, namely, those that have FO interpretations in (or are FO transductions of) some sparse graph classes. We give a structural characterization of the graph classes that are FO interpretable in graphs of bounded degree. This characterization allows us to efficiently compute such an FO interpretation for an input graph. As a consequence, we obtain an FPT algorithm for successor-invariant FO model checking on any graph class that is FO interpretable in (or an FO transduction of) a graph class of bounded degree. The approach we use to obtain these results may also be of independent interest.

CCS Concepts: • **Theory of computation → Logic**; *Finite Model Theory*; *Parameterized complexity and exact algorithms*; • **Mathematics of computing →** *Graph algorithms*;

Additional Key Words and Phrases: First-order logic, bounded degree graphs, algorithmic metatheorems, interpretations, fixed parameter tractability

ACM Reference format:
Jakub Gajarský, Petr Hliněný, Jan Obdržálek, Daniel Lokshtanov, and M. S. Ramanujan. 2020. A New Perspective on FO Model Checking of Dense Graph Classes. *ACM Trans. Comput. Logic* 21, 4, Article 28 (July 2020), 23 pages.
https://doi.org/10.1145/3383206

28

Jakub Gajarský has been supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC Consolidator Grant DISTRUCT, Grant No. 648527). P. Hliněný and J. Obdržálek acknowledge support by the Czech Science Foundation, Project No. 17-00837S. Daniel Lokshtanov was supported by Pareto-Optimal Parameterized Algorithms, ERC Starting Grant No. 715744. M. S. Ramanujan acknowledges support from the Bergen Research Foundation, project BeHard - Beating Hardness by preprocessing and from Austrian Science Fund (FWF), Project No. P26696 X-TRACT.
Authors' addresses: J. Gajarský, Logic and Semantics, Technische Universität Berlin, Ernst-Reuter-Platz 7, Berlin, 10587, Germany; email: jakub.gajarsky@tu-berlin.de; P. Hliněný and J. Obdržálek, Faculty of Informatics MU Brno, Botanická 68a, 602 00 Brno, Czech Republic; emails: hlineny@fi.muni.cz, obdrzalek@fi.muni.cz; D. Lokshtanov, Department of Computer Science, University of California Santa Barbara, 2104, USA; email: daniello@ucsb.edu; M. S. Ramanujan, Department of Computer Science, University of Warwick, Coventry CV4 7AL, United Kingdom; email: R.Maadapuzhi-Sridharan@warwick.ac.uk.
Author's current address: J. Gajarsky, Institute of Informatics, University of Warsaw, Stefana Banacha 2, Warsaw, 02-097, Poland; email: gajarsky@mimuw.edu.pl.

1529-3785/2020/07-ART28 $15.00
https://doi.org/10.1145/3383206

1 INTRODUCTION

Algorithmic metatheorems are theorems stating that all problems expressible in a certain logic are efficiently solvable on certain classes of (relational) structures, e.g., on finite graphs. Note that the model-checking problem for *first-order logic*—given a graph G and an FO formula ϕ, we want to decide whether G satisfies ϕ (written as $G \models \phi$)—is trivially solvable in time $|V(G)|^{O(|\phi|)}$. Hence, "efficient solvability," in this context, often means *fixed-parameter tractability* (FPT); that is, solvability in time $f(|\phi|) \cdot |V(G)|^{O(1)}$ for some computable function f.

In the past two decades algorithmic metatheorems for FO logic on sparse graph classes received considerable attention. After the result of Seese [27] establishing fixed-parameter tractability of FO model checking on graphs of bounded degree, there followed a series of results [7, 9, 12] establishing the same result for increasingly rich sparse graph classes. This line of research culminated in the result of Grohe et al. [20], who proved that FO model checking is FPT on nowhere dense graph classes.

The result of Grohe et al. [20] is essentially the best possible of its kind, in the following sense: If a graph class $\mathcal{D}$ is monotone (i.e., closed on taking subgraphs) and not nowhere dense, then the FO model checking problem on $\mathcal{D}$ is as hard as that on all graphs. Possible ways to continue the research into algorithmic metatheorems for FO logic include the following two directions:

First, one can study relational structures other than graphs. This line of research has recently been initiated by Bova et al. [3], who gave an FPT algorithm for existential FO model checking on partially ordered sets of bounded size of a maximum antichain. Their result was first improved upon in Reference [15] and shortly after that followed the result of Gajarský et al. [14], who extended Reference [3] to full FO. Apart from these results, very little is known and it remains to be seen what other types of structures and their parameterizations admit fast FO model checking algorithms.

Second, one may consider metatheorems for FO logic on classes of graphs that are not sparse. Again, little is known along this line of research. Besides the classical result of Courcelle et al. [5] (which actually speaks about more general MSO logic), one can mention the result of Ganian et al. [17] establishing that certain subclasses of interval graphs admit an FPT algorithm for FO model checking, or the result of Grohe [19] about graph classes of low degree. Besides, the aforementioned result of Reference [14] can also be seen as a result about dense (albeit directed) graphs, and Reference [14] actually happens to imply the result of Reference [17].

We would like to initiate a systematic study of dense graph classes for which the FO model checking problem is efficiently solvable. It appears that a natural way to arrive at new graph classes admitting FPT algorithms for FO model checking, is by *means of interpretation, or transduction*. In a simplified setting of interpretations—given a graph G and an FO formula $\psi(x, y)$ with two free variables, we can define a graph $H = I_\psi(G)$ on the same vertex set as G and the edge set determined by $\psi(x, y)$: A pair of distinct vertices u, v is an edge of H iff $G \models \psi(u, v) \vee \psi(v, u)$. We then say that H is interpreted in G using ψ. A graph class $\mathcal{D}$ is FO interpretable in a graph class $\mathcal{C}$ if there exists an FO formula $\psi(x, y)$ such that every member of $\mathcal{D}$ is interpreted in some member of $\mathcal{C}$ using ψ.

For now, let us assume we have an efficient FO model checking algorithm for the previous class $\mathcal{C}$, and consider the FO model checking problem of the class $\mathcal{D}$. If an input graph from $\mathcal{D}$ was given together with the corresponding FO interpretation in a graph from $\mathcal{C}$, then one could easily solve the model checking problem using the existing algorithm for $\mathcal{C}$. This is based on the following natural property of interpretations: If $H \in \mathcal{D}$ is interpreted in $G \in \mathcal{C}$ using formula $\psi(x, y)$, and our question is to decide whether $H \models \phi$, then it is a standard routine to construct from ϕ and ψ a sentence ϕ' such that $H \models \phi$ if and only if $G \models \phi'$. Then, $G \models \phi'$ is decided by the algorithm given for $\mathcal{C}$.

However, if the assumed interpretation (or transduction) is not given, then the situation is markedly harder. In this context, we ask the following question:

Question 1.1. *Let $\mathcal{C}$ be a graph class admitting an FPT algorithm for FO model checking, and $\mathcal{D}$ be a graph class FO interpretable in $\mathcal{C}$. Does there exist an FPT algorithm for FO model checking on $\mathcal{D}$?*

As outlined above, the difficulty of this question lies in the fact that our inputs come from $\mathcal{D}$, without any reference to the respective members of $\mathcal{C}$ in which they are interpreted. Even if the interpretation formula $\psi(x, y)$ is fixed and known beforehand, we have generally no efficient way of obtaining the respective member $G \in \mathcal{C}$ for an input $H \in \mathcal{D}$. Thus, Question 1.1 can be reduced to the following:

Question 1.2. *Let $\mathcal{C}, \mathcal{D}$ be graph classes such that $\mathcal{D}$ is FO interpretable in $\mathcal{C}$. Does there exist an integer s and a polynomial-time algorithm $\mathcal{A}$ such that; given $H \in \mathcal{D}$ as input, $\mathcal{A}$ outputs $G \in \mathcal{C}$ and an FO formula $\psi(x, y)$ of size at most s such that H is interpreted in G using ψ?*

An answer to Question 1.2 is far from being obvious, and it can strongly depend on the choice of ψ. Take, for example, the following particular FO interpretation: A graph H is the *square* of a graph G if the edges of H are those pairs of vertices that are at distance at most 2 in G. Then the problem; given H find G such that H is the square of G, is NP-hard [24]. Another such negative example, specifically tailored to our setting, is discussed in Section 7. These examples show that it is important to choose a suitable interpretation formula ψ (avoiding the hard cases) in an attempt to answer Question 1.2.

Our contribution. We answer both Questions 1.1 and 1.2 in the positive for the case when $\mathcal{C}$ is a class of graphs of bounded degree. Our answers cover also the more general case of FO transductions of bounded-degree classes, and include checking successor-invariant FO properties in addition to ordinary FO ones.

We first define near-uniform graph classes (Definition 4.2), based on a new notion of near-k-twin relation, which generalizes the folklore twin-vertex relation and is related also to the neighbourhood diversity parameter of Reference [22]. The idea behind this approach is to classify pairs of vertices that have almost the same adjacency to the rest of the graph. The approach seems promising and may be of independent use in further investigation of well structured dense graph classes. While the definition of near-uniformity lends itself well to being used in proofs, it is sometimes unnecessarily technical to reason about. We therefore also introduce an equivalent notion of near-covered graph classes (Definition 4.3), which is more intuitive, easier to grasp and offers a slightly different perspective.

We then give an efficient FO model checking algorithm (Theorem 5.1) for the near-uniform graph classes. This algorithm is based upon the above idea of interpretation; briefly, given a graph H, we use the near-k-twin relation for a suitable value of k to partition the vertex set of H and to find a bounded degree graph G, such that H is interpreted in G using a universal formula ψ depending only on the class in question (Theorem 5.5). Then, we employ the aforementioned algorithm of Seese [27]. Furthermore, we extend our algorithm to include also stronger successor-invariant FO properties (see Section 5.3 for more details), for which we can use the recent result of Reference [28].

In the second half of the article, we argue that the concept of near-uniform graph classes is robust and sufficiently rich in content. We prove that the near-covered (and therefore also near-uniform, since the two are equivalent) graph classes are exactly those that are FO interpretable in graphs of bounded degree (Theorem 6.4) and, more generally, that any FO transduction of a graph class of bounded degree is a near-covered graph class (Theorem 6.5). The key tool we use is Gaifman's theorem [13]. Here, we remark that properties of graphs that are FO interpretable in

graphs of bounded degree have already been studied, e.g., by Dong et al. [8], in a different context, but those previous results do not imply our conclusions.

We then complement the previous tractability results with a negative example of a particular FO interpretation that is NP-hard to "reverse" even on the class of graphs of degree at most 3 (Theorem 7.1). We finish by sketching some interesting open directions for future research.

2 DEFINITIONS AND PRELIMINARIES

We begin by clarifying the terminology and recalling some established concepts concerning logic on graphs. We assume that 0 is a natural number, i.e. $0 \in \mathbb{N}$. Let $X \triangle Y$ denote the symmetric difference of two sets.

Graph theory. We work with *finite simple undirected graphs* and use standard graph theoretic notation. We refer to the vertex set of a graph G as to $V(G)$ and to its edge set as to $E(G)$. If $\{u, v\} \in E(G)$, or shortly $uv \in E(G)$, then u and v are adjacent vertices, or neighbours. As it is common in the context of FO logic on graphs, vertices of our graphs can carry arbitrary labels.

A *dominating set* $X \subseteq V(G)$ in a graph G is such that every vertex $v \in V(G) \setminus X$ is a neighbour of some vertex $x \in X$. The *distance* between two vertices u and v of G is the length of a shortest path from u to v in G. The *radius* of a graph G is the smallest integer r such that there exists a vertex $v \in V(G)$ having a distance at most r to every vertex of G. For example, nonempty G has radius at most 1 if and only if G has a dominating set of size 1.

FO logic. The *first-order logic of graphs* (abbreviated as FO) applies the standard language of first-order logic to a graph G viewed as a relational structure with the domain $V(G)$ and the single binary (symmetric) relation $E(G)$. That is, in FO we have got the standard predicate $x = y$, a binary predicate $edge(x, y)$ with the meaning $\{x, y\} \in E(G)$, an arbitrary number of unary predicates $L(x)$ with the meaning that x holds the label L, usual logical connectives $\wedge, \vee, \rightarrow$, and quantifiers $\forall x$, $\exists x$ over the vertex set $V(G)$.

For example, $\phi(x, y) \equiv \exists z(edge(x, z) \wedge edge(y, z) \wedge red(z))$ states that the vertices x, y have a common neighbour in G, which is labeled "red."

Parameterized model checking. The instances of a parameterized problem can be considered as pairs $\langle I, k\rangle$ where I is the main part of the instance and k is the *parameter* of the instance; the latter is usually a non-negative integer. A parameterized problem is *fixed parameter tractable (FPT)* if instances $\langle I, k\rangle$ of size n can be solved in time $O(f(k) \cdot n^c)$ where f is a computable function and c is a constant independent of k. In *parameterized model checking*, instances are considered in the form $\langle (G, \phi), |\phi|\rangle$ where G is a structure, ϕ a formula, the question is whether $G \models \phi$ and the parameter is the size of ϕ.

When speaking about the FO model checking problem in this article, we implicitly consider the formula ϕ (its size) as a parameter.

Interpretations. To simplify our exposition and proofs, we work with a simplified version of FO interpretations (note, however, this does not impact generality of our conclusions, as we will see later).

Let $\psi(x, y)$ be an FO formula with two free variables over the language of (possibly labelled) graphs such that for any graph and any u, v it holds that $G \models \psi(u, v) \Leftrightarrow G \models \psi(v, u)$ and $G \not\models \psi(u, u)$, i.e. the relation on $V(G)$ defined by the formula is symmetric and irreflexive. From now on, we will assume that formulas with two free variables are symmetric and irreflexive (which can easily be enforced). Given a graph G, the formula $\psi(x, y)$ maps G to a graph $H = I_\psi(G)$ defined by $V(H) = V(G)$ and $E(H) = \{\{u, v\} \mid G \models \psi(u, v)\}$. We then say that the graph H is *interpreted* in G.

Notice that even though the graph G can be labelled, our graph H is not. This is to simplify our notation—nevertheless, one may easily inherit labels from G to H if needed.

In the rest of the article, whenever we consider graphs G and H in context of interpretations, graph G will be the graph in which we are interpreting, and graph H will be the "result" of the interpretation.

The notion of interpretation can be extended to graph classes—to a graph class $\mathcal{C}$ the formula $\psi(x, y)$ assigns the graph class $\mathcal{D} = I_\psi(\mathcal{C}) = \{H \mid H = I_\psi(G),\ G \in \mathcal{C}\}$. We say that a graph class $\mathcal{D}$ is *interpretable* in a graph class $\mathcal{C}$ if there exists formula $\psi(x, y)$ such that $\mathcal{D} \subseteq I_\psi(\mathcal{C})$. Note that we do not require $\mathcal{D} = I_\psi(\mathcal{C})$, as we just want every graph from $\mathcal{D}$ to have a preimage in $\mathcal{C}$.

Interpretations are useful for defining new graphs from old using logic (again, we think of H as a result of application of ψ to G), but can also be used to evaluate formulas on H quickly, provided that we have a fast algorithm to evaluate formulas on G. Let $H = I_\psi(G)$, let θ be a sentence and let θ' be a sentence obtained from θ by replacing every occurrence of the atom $edge(x, y)$ by $\psi(x, y)$. Then, obviously, $H \models \theta \iff G \models \theta'$.

FO transductions. While interpretations are restricted in a choice of the target domain (and, in our case, we even require $V(H) = V(G)$), a more general view is provided by so called transductions, see Courcelle and Engelfriet [4]. Informally, in addition to an interpretation this concept allows to add to a graph arbitrary "parameters" (as labels) and to make several disjoint copies of the graph.

Here, we provide a brief definition of a transduction based on Reference [2], simplified to target only the FO graph case. Still, before proceeding to the sought definition, we have to briefly extend the notion of an interpretation toward finite relational structures with finite signatures. A *relational structure* $S = (U, R_1^S, \ldots, R_a^S)$ of the signature $\sigma = \{R_1, \ldots, R_a\}$ consists of a finite domain U and a finite list of relations $R_1^S, \ldots, R_a^S$ over U. For instance, for graphs, $U = V(G)$ is the vertex set and $R_1^G = E(G)$ is the binary symmetric relation of edges of G. The scope of the previous definition of a graph interpretation I_ψ can be naturally extended by allowing ψ to use the predicates from σ when defining $I_\psi(S)$.

A *basic FO graph transduction* τ_0 of a relational structure S is a triple (χ, ν, μ) of FO formulas with 0, 1, and 2 free variables, respectively, such that τ_0 is undefined if $S \not\models \chi$, and otherwise, τ_0 maps the structure S into the graph on the vertex set $\{v \in U \mid S \models \nu(v)\}$ and the edge set $\{\{u, v\} \mid S \models \mu(u, v)\}$ (i.e., to an induced subgraph of $I_\mu(S)$).

The *m-copy operation* maps a graph G to the relational structure G^m on the domain $V(G^m) = V(G) \times \{1, \ldots, m\}$, such that the subset $V(G) \times \{i\}$ for each $i = 1, 2 \ldots, m$ induces a copy of the graph G (there are no edges between distinct copies), and $V(G^m)$ is additionally equipped with a binary relation $\sim$ and unary relations $Q_1, \ldots, Q_m$ such that; $(u, i) \sim (v, j)$ for $u, v \in V(G)$ iff $u = v$, and $Q_i = \{(v, i) : v \in V(G)\}$. The *p-parameter expansion* maps a graph G to the set of all graphs that result by an expansion of $V(G)$ by p unary predicates.

Altogether, a many-valued map τ is an *FO transduction* (of simple undirected graphs) if it is $\tau = \tau_0 \circ \gamma \circ \varepsilon$ where τ_0 is a basic FO graph transduction, γ is a m-copy operation for some m, and ε is a p-parameter expansion for some p.

We remark, once again, that the result of a transduction τ of one graph is generally a set of graphs, due to the involved p-parameter expansion. For a graph class $\mathcal{C}$, the result of a *transduction τ of the class $\mathcal{C}$* is the union of the particular transduction results, precisely, $\tau(\mathcal{C}) := \bigcup_{G \in \mathcal{C}} \tau(G)$.

3 OUTLINE OF OUR APPROACH

Before diving into technical details of our claims and proofs, we give a brief exposition of ideas leading to our results. We start by explaining the core ideas behind our approach to analysing

dense graphs and then we sketch the how interpretations are combined with our approach to dense graphs to obtain the results presented in Sections 4 and 6.

3.1 Locality, Indistinguishability, and the New Approach

The existing FPT algorithms for FO model checking of sparse graph classes we mentioned at the beginning of Section 1 rely heavily on the use of locality of FO logic—i.e., the fact that evaluating FO formulas can be reduced to evaluating *local* FO formulas (cf. Gaifman's theorem [13], also in Section 6). This, together with the fact that in sparse graphs it is possible to evaluate local formulas efficiently, made the locality-based approach suitable for studying FO logic on sparse graphs. The problem with using this approach for dense graphs is obvious—in a dense graph the whole graph can be in the 1-neighbourhood of a single vertex.[1] This makes evaluating local formulas around such a vertex expensive (from the FPT perspective), because this amounts to evaluating them on the whole graph.

An alternative approach to FO model checking, as described in Section 4, is based on the concept of vertex indistinguishability. This approach can be used for dense graphs, but is a bit too limited in its scope. The key notion here is that of twin vertices—two vertices of a graph G are *twins* if they have the same neighbourhood. The fact that two vertices u, v are twins means that they behave in the same way with respect to any other vertex in a graph. Consequently, no FO formula can distinguish between u and v. It is not hard to see that the twin relation is an equivalence on the vertex set of a graph. One may also say that the set of vertex neighbourhoods occurring in G is "covered" by the set of neighbourhoods of representatives of each twin class of G. The number of equivalence classes of the twin relation is called the *neighbourhood diversity* [22] of a graph, and graph classes of bounded neighbourhood diversity admit a very simple FPT algorithm for FO model checking. However, as already mentioned, the problem with this approach is that it is too restrictive—even such simple graph classes as paths have unbounded neighbourhood diversity.

Our approach is based on observing that the locality-based approach, when used on sparse graphs, exploits, in its essence, the indistinguishability of vertices. Take, for example, the graphs of bounded degree. Here any two vertices behave the same way with respect to the rest of the vertex set (they are non-adjacent to it), with only a few exceptions (the vertices in their neighbourhood). In other words, any two vertices have *almost* the same neighbourhood. This leads to a relaxation of the notion of twin vertices. We say that two vertices are *near-k-twins* if their neighbourhoods differ in at most k vertices. To see how this notion works around the issues with locality and indistinguishability explained above, let us consider the near-k-twin relation on the class $\mathcal{D}_d$ of graphs of degree at most d and on the class $\overline{\mathcal{D}_d}$ of its complements. On every graph from these graph classes, the near-$2d$-twin relation is an equivalence with just one class. Yet, graphs from $\overline{\mathcal{D}_d}$ are dense and some of them contain universal vertices.

The above considerations lead us to studying graph classes such that for each graph from these classes there exists a small k such that the near-k-twin relation is an equivalence with a small number of classes—the *near-uniform* graph classes. Though, unlike the ordinary twin relation, the near-k-twin relation is not automatically guaranteed to be an equivalence (this depends heavily on the choice of G and k) and, consequently, dealing with near-uniformity is slightly cumbersome and requires a great care.

However, there is also another (and perhaps simpler to deal with) way to view and formally capture the above informal discussion of diversity of neighbourhoods in a graph—that one can "cover" all distinct neighbourhoods in the graph with only few representative neighbourhoods.

[1] This is also true for some sparse graphs, say stars, but we hope that it is clear that for dense graphs this can cause substantial problems.

This view leads to a new definition—a class of graphs is *near-covered* if there exists a small k such that every graph in this graph class contains a small (of a constant size) set S of vertices such that every vertex is a near-k-twin of at least one vertex from S. It is easily seen that near-uniformity implies near-coveredness—just pick any one representative from each equivalence class. As we shall see, the converse is also true and the two notions are (asymptotically) equivalent. Precisely, we shall prove that for any graph class C the following conditions are equivalent:

(1) C is near-uniform (Definition 4.2);
(2) C is near-covered (Definition 4.3);
(3) C is interpretable in a class of graphs of bounded degree.

Since we can efficiently compute the interpretation claimed in (3), we can then solve FO model checking on near-uniform graph classes in FPT using established tools, such as the algorithm of [27] for FO model checking on graphs of bounded degree. Our proof is structured as follows; we first prove the equivalence between (1) and (2) (Lemma 4.4), and then the implications (1) ⇒ (3) (Theorem 5.5) and (3) ⇒ (2) (Theorem 6.4).

One may, with respect to technical difficulties related to the near-uniformity notion, question whether it is necessary to consider near-uniformity at all and not to go with just near-coveredness alone. However, the equivalence aspect of the near-k-twin relation is crucial in proving that graphs with certain properties are interpretable in graphs of bounded degree. We therefore believe that it deserves a separate definition.

3.2 Interpretability in Graphs of Bounded Degree

Besides dealing with the FO model checking problem via interpretation of certain graph classes into classes of bounded degree, we are also interested in the other direction—to find out which graph classes can be FO interpreted into classes of bounded degree (the direction (3) ⇒ (2) above).

Our characterization of such classes relies on a simple corollary of Gaifman's locality theorem: For a graph G and two vertices $u, v \in V(G)$, which are far apart form each other, the truth value of the formula $\psi(u, v)$ depends only on formulas with one free variable (up to the quantifier rank q, which depends on ψ) valid on u and v (i.e., its logical q-types). This in turn means that when the formula $\psi(u, v)$ is used for interpretation (to obtain the graph H from a graph G of degree at most d) and vertices u and u' satisfy the same formulas with one free variable (again, up to the quantifier rank q), u and u' will be adjacent to the same vertices in the resulting graph, except for a small number of vertices that were in their respective r-neighbourhoods in graph G (here r also depends on $\psi(x, y)$). Any two vertices of the same q-type will therefore be near-k-twins for $k = 2 \cdot d^r$.

While the previous consideration is quite simple, note the following possible pitfall. Since the relation "being of the same q-type" is an equivalence with a bounded number of classes, it is tempting to believe that the near-k-twin relation (for a suitably chosen k) is an equivalence with a bounded number of classes (independent of G) for any graph from a graph class FO interpretable in a class of graphs of bounded degree. This, however, is not true—it can happen that some vertices u and v of different q-types can be near-k-twins and a vertex w of yet different q-type can be near-k-twin of v but not of u, thus failing the transitivity.

Instead, we finish as follows. Since for any q there are finitely many (say m) different q-types, in H there exist at most m vertices such that every vertex is a near-k-twin of (at least) one of them. This in turn means that graph classes FO interpretable in graphs of bounded degree are near-covered, and hence also near-uniform by (1) ⇔ (2) above.

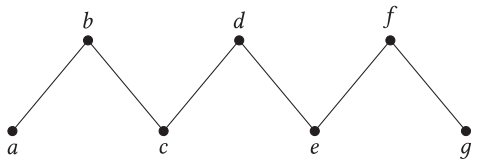

Fig. 1. An example. The near-2-twin relation ρ_2 of this path includes pairs (b, d) and (d, f) but not (b, f), and so ρ_2 is not an equivalence. However, ρ_1 is an equivalence on this path and its near-1-twin classes are $\{a, c\}, \{e, g\}, \{b\}, \{d\}, \{f\}$.

4 NEAR-UNIFORM AND NEAR-COVERED GRAPH CLASSES

In this section, we formally establish the key concepts. For a graph G and a vertex $v \in V(G)$, we define the *neighbourhood* of v as $N^G(v) = \{w \in V(G) \mid \{v, w\} \in E(G)\}$. If the graph G is clear from the context, then we write just $N(v)$. Note that, by definition, $v \notin N(v)$.

A useful concept in graph theory is that of twin vertices. Two vertices $u, v \in V(G)$ are called *false twins* if $N(u) = N(v)$, and they are *true twins* if $N(u) \cup \{u\} = N(v) \cup \{v\}$. We actually follow the concept of false twins, which better suits our purposes, in the next definition.

Definition 4.1 (Near-k-twin Relation). For a graph G and $k \in \mathbb{N}$, the *near-k-twin relation of G* is the relation ρ_k on $V(G)$ defined by $(u, v) \in \rho_k \iff |N(u) \triangle N(v)| \leq k$.

Considering, e.g., k a small parameter and G a large graph then, intuitively, two vertices of G are near-k-twins if they have "almost the same" neighbourhood. This relation, unlike the ordinary twin relations on graph vertices, does not always "behave nicely"; in particular, ρ_k may not be an equivalence relation (see, e.g., the examples below). However, if the near-k-twin relation is an equivalence of bounded index, then we can use it to decompose the vertex set of the graph G into similarly behaving clusters. This leads to the following.

Definition 4.2 (Near-uniform). A graph G is (k_0, p)*-near-uniform* if there exists $k \leq k_0$ for which the near-k-twin relation of G is an equivalence of index at most p.

A graph class $\mathcal{C}$ is (k_0, p)*-near-uniform* if every member of $\mathcal{C}$ is (k_0, p)-near-uniform, and $\mathcal{C}$ is *near-uniform* if there exist integers k_0, p such that $\mathcal{C}$ is (k_0, p)-near-uniform.

To simplify the discussion, we use the following as a shorthand. If ρ_k of Definition 4.1 is an equivalence relation, then we call ρ_k the *near-k-twin equivalence* of G, and the equivalence classes of ρ_k the *near-k-twin classes of G*.

For example, take a class $\mathcal{D}_d$ of the graphs of maximum degree at most d, and let $k = 2d$. Then the near-k-twin relation ρ_k is a trivial equivalence of index one (i.e., with one class) for every graph from $\mathcal{D}_d$. The same holds for the class $\overline{\mathcal{D}}_d$ of the complements of graphs of $\mathcal{D}_d$. Another sort of examples comes, say, with a class $\overline{\mathcal{B}}_d$ of the graphs obtained from complete bipartite graphs by subtracting a subgraph of maximum degree at most d. For $k = 2d$ and every graph of $\overline{\mathcal{B}}_d$, the near-k-twin relation ρ_k is an equivalence of index at most two. However, we can easily see that the near-2-twin relation of, e.g., a path of length 6 is not an equivalence; see Figure 1. Even more, examples such as that of Figure 1 show that, having a near-k-twin equivalence for some k, does not imply that the near-k'-twin relation is an equivalence for $k' > k$. That is why we cannot simply use one universal value of k in Definition 4.2.

The fact that the near-k-twin relation of a graph G is an equivalence on $V(G)$ can used as follows: the neighbourhood of a vertex is represented by the neighbourhood of a selected representative of its class and the (small) difference of these two neighbourhoods. For such purpose of representation it is not always necessary to have a near-k-twin equivalence; just having *at least* one such

representative for every vertex of G may be sufficient (we may not care that there are more than one "close" representatives). This simplified scenario leads to the following definition.

Definition 4.3 (Near-covered). A graph G is (ℓ, q)-*near-covered* if there exist vertices $v_1, \ldots, v_q$ in $V(G)$ such that each vertex $u \in V(G)$ is a near-ℓ-twin of at least one of $v_1, \ldots, v_q$.

A graph class C is (ℓ, q)-*near-covered* if every member of C on at least q vertices is (ℓ, q)-near-covered, and C is *near-covered* if there exist integers ℓ, q such that C is (ℓ, q)-near-covered.

The following lemma establishes that the two notions—being near-uniform and being near-covered—are in fact equivalent. While the definition of being near-covered is less technical and easier to grasp, the definition of near-uniformity is more convenient to work with in the algorithmic context of Section 5, which is the main reason for including both definitions.

Lemma 4.4. *A graph class C is near-uniform if and only if C is near-covered.*

Proof. It is easy to see that if C is (k_0, p)-near-uniform then it is (k_0, p)-near-covered: for every graph $G \in C$ there is $k \leq k_0$ such that the near-k-twin relation is an equivalence relation with p' classes $C_1, \ldots, C_{p'}$ where $p' \leq p$. We pick an arbitrary vertex v_i from from each class C_i to obtain vertices $v_1, \ldots, v_{p'}$. Clearly, each vertex of G is a near-k-twin of one of these vertices.

To prove the opposite direction, consider first the following construction: To any graph G and k, we define the auxiliary graph G_k on the same vertex set by setting $(u, v) \in E(G_k)$ if and only if u and v are near-k-twins in G. Observe the following easy properties of this construction:

(1) Graph G is (ℓ, q)-near-covered if and only if G_ℓ has a dominating set of size at most q.
(2) If for some k the graph G_k is a disjoint union of at most p cliques, then near-k-twin is an equivalence with p classes on G (and so G is (k, p)-near-uniform).
(3) If two vertices are at distance at most p in G_k, then they are pk-near-twins in G.
(4) If G_k contains a component with radius greater than 1, then this component has to be dominated by at least two vertices. Moreover, in any dominating set of such connected component there are two vertices that are at distance at most 3 in G_k.

We now prove that any graph class C that is near-covered with parameters ℓ and q is also near-uniform. The idea behind the proof can be easily summarized as follows. If G is near-covered with parameters ℓ and q, then by Equation (4) above each component of G_ℓ has small radius (bounded in terms of q), and so increasing ℓ to $f(q)\ell$ will turn each component of G_ℓ into a clique in $G_{f(q)\ell}$. This, however, is not enough to conclude the proof, because increasing ℓ to $f(q)\ell$ can also introduce new edges between the components of G_ℓ. In this case, we increase $f(q)\ell$ again to $f(f(q))\ell$, and we repeat this procedure until we reach the desired outcome. To see that this procedure ends after at most $q - 1$ steps, notice that $G_{f(q)\ell}$ has a dominating set of smaller size than G_ℓ.

Formally, we proceed by induction on q. For the case when $q = 1$ the graph G_ℓ has a dominating set of size 1. This means that every two vertices in G_ℓ are at distance at most 2, it follows from Equation (3) that any two vertices of G are near-2ℓ-twins. Graph G is therefore $(2\ell, 1)$-near-uniform, which finishes the induction basis.

For the induction step, we fix $q > 1$ and assume that every $(m, q - 1)$-near-covered graph class, for any m, is (a, b)-near-uniform for some values a, b depending only on m and q. Consider now a graph class C that is (ℓ, q)-near-covered. We will prove that every graph from C is $(2\ell, q)$-near-uniform or $(8\ell, q - 1)$-near-covered. The latter case, from the induction hypothesis, implies that C is (a, b)-near-uniform where a, b depend only on ℓ and q. As a result, every graph in C is $(\max(2\ell, a), \max(q, b))$-near-uniform and so C is near-uniform.

We take a graph $G \in C$ that is (ℓ, q)-near-covered, and consider the derived graph G_ℓ. If G_ℓ has a dominating set of size smaller than q, then it is actually $(\ell, q - 1)$-near-covered, which means it is

also $(8\ell, q-1)$-near-covered as desired. From now on, we therefore assume that G_ℓ has a smallest dominating set $S = \{v_1, \ldots, v_q\}$. We distinguish two cases:

I. G_ℓ contains a connected component C with radius at least 2. By property Equation (4) of the construction, there are two vertices v_i, v_j from S that are at distance at most 3 in C. Consider now the graph $G_{4\ell}$. We claim that $G_{4\ell}$ has a dominating set of size at most $q-1$, which means that G is $(4\ell, q-1)$-near-covered and therefore also $(8\ell, q-1)$-near-covered as desired.

First note that G_ℓ is subgraph of $G_{4\ell}$, so S is a dominating set of $G_{4\ell}$. We claim that $S \setminus v_j$ (of size $q-1$) is also a dominating set of $G_{4\ell}$. To see this, consider any vertex u dominated by v_j in G_ℓ. Since the distance between v_i and v_j in G_ℓ is at most 3, the distance between v_i and u is at most 4 in G_ℓ. This means, by Equation (3), that v_i and u are 4ℓ-near-twins in G. This in turn means that there is an edge between v_i and u in $G_{4\ell}$, and so u is dominated by v_i in $G_{4\ell}$. Since u was an arbitrary neighbour of v_j (in G_ℓ), every vertex dominated by v_j in G_ℓ is dominated by v_i in $G_{4\ell}$. Therefore, $S \setminus v_j$ is a dominating set in $G_{4\ell}$ of size $m-1$.

II. All connected components of G_ℓ have radius at most 1. This means that G_ℓ consists of components $C_1, \ldots, C_q$ such that $v_i \in C_i$ for $i = 1, \ldots, q$. In this case, we consider the graph $G_{2\ell}$. Since every two vertices in the same component C_i of G_ℓ are at distance at most 2, they are 2ℓ-near-twins in G and so there is an edge between them in $G_{2\ell}$, which means that each component C_i forms a clique in $G_{2\ell}$. We distinguish two possibilities:
   (a) There is no pair of distinct indices i, j such that there exists an edge in $G_{2\ell}$ between some vertices $u \in C_i$ and $w \in C_j$. In this case the graph $G_{2\ell}$ is a disjoint union of q cliques, which means that G is $(2\ell, q)$-near-uniform by property Equation (2).
   (b) There exists a pair of distinct indices i, j such $G_{2\ell}$ contains an edge uw between some vertices $u \in C_i$ and $w \in C_j$. Recall that v_i and v_j are the vertices from S that are contained in C_i and C_j, respectively. These vertices are in the same component in $G_{2\ell}$ and at distance at most 4. By the same argument as in the case I, the set $S \setminus v_j$ is a dominating set of size $q-1$ of the graph $G_{8\ell}$, which means that G is $(8\ell, q-1)$-near-covered, as desired. □

5 FO MODEL CHECKING ALGORITHM

This section constitutes the main algorithmic contribution of the article.

Our model checking algorithm for near-uniform graph classes can be shortly summarized as follows. Input is a graph H from a (k_0, p)-near-uniform graph class C and an FO sentence ϕ. Perform the following steps:

(1) For each $k := 0, 1, \ldots, k_0$; compute the near-k-twin relation ρ_k of H, and check whether ρ_k is an equivalence of index at most p. This test has to succeed for some value of k (Definition 4.2).
(2) Compute a universal formula $\psi(x, y)$ depending on k_0 and p, and the graph G_H depending on H and k found in step 1, such that $H = I_\psi(G_H)$ and the vertex degrees in G_H are at most $2k_0p$ (Theorem 5.5).
(3) Run the algorithm of Reference [27] for FO model checking on graphs of bounded degree on G_H and the sentence ϕ', where ϕ' is obtained from ϕ by replacing every occurrence of $edge(z, z')$ with $\psi(z, z')$.

THEOREM 5.1. *Let C be a (k_0, p)-near-uniform graph class for some $k_0, p \in \mathbb{N}$. Then the FO model checking problem of C is fixed-parameter tractable when parameterized by the formula size, i.e., solvable in time $f(|\phi|) \cdot |V(G)|^{O(1)}$ for a computable function f and input G, ϕ.*

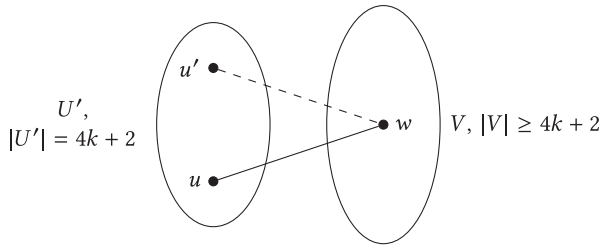

Fig. 2. An illustration; counting the pairs $(w, \{u, u'\})$ such that $w \in V$, $u, u' \in U'$ in the proof of Lemma 5.2, in case $U \neq V$.

The rest of this section is devoted to the proof of this statement.

5.1 Properties of the Near-k-twin Relation

To give details of the algorithm and to prove Theorem 5.1, we study some structural properties of graphs for which the near-k-twin relation is actually an equivalence.

As outlined above in the algorithm, our key step is to show that all near-uniform graph classes are FO interpretable in graph classes of bounded degree. For this, we show that for any two large enough equivalence classes of a near-k-twin equivalence, it holds that every vertex from one class is connected to almost all or to almost none vertices of the other class and vice versa. More precisely:

Lemma 5.2. *Let $k \geq 1$ and G be a graph such that the near-k-twin relation ρ_k of G is an equivalence on $V(G)$. Let U and V be two near-k-twin classes of G with at least $4k + 2$ vertices each (it may be $U = V$). Then, for every $v \in V$, we have*

$$\min\{|U \cap N(v)|,\ |U \setminus N(v)|\} \leq 2k.$$

Note that the claim of Lemma 5.2 universally holds only when both U and V are sufficiently large. A counterexample with small U is a graph consisting of $U = \{u\}$ and V inducing a large clique, such that u is connected to half of the vertices of V. For this graph the near-1-twin classes are exactly U and V, but both $|V \cap N(u)|$ and $|V \setminus N(u)|$ are unbounded.

Proof. For $x \in V(G)$ and $A \subseteq V(G)$, let $\alpha^A(x) = \min\{|N(x) \cap A|,\ |A \setminus N(x)|\}$. Thus, to prove the lemma, we need to show that $\alpha^U(v) \leq 2k$ for $v \in V$.

Toward a contradiction assume $\alpha^U(v) \geq 2k + 1$ for some $v \in V$. Clearly, there is a subset $U' \subseteq U$ such that $|U'| = 4k + 2$ and $\alpha^{U'}(v) = 2k + 1 = \frac{1}{2}|U'|$. Since $|N(w) \triangle N(v)| \leq k$ for any $w \in V$ by the definition of ρ_k, we also get $\alpha^{U'}(w) \geq \alpha^{U'}(v) - k = 2k + 1 - k = k + 1$ for all $w \in V$ and, likewise, $\alpha^{U'}(w) \leq 3k + 1$.

We are going to count the number D of pairs $(w, \{u, u'\})$ such that $w \in V$, $u, u' \in U'$ are distinct vertices and exactly one of wu, wu' is an edge of G. See Figure 2. On the one hand, for any fixed $u, u' \in U'$, every w forming such a desired pair $(w, \{u, u'\})$ belongs to $N(u) \triangle N(u')$, and so we have got an upper bound

$$D \leq \sum_{\{u,u'\} \in \binom{U'}{2}} |N(u) \triangle N(u')| \leq \binom{|U'|}{2} \cdot k = \binom{4k+2}{2} \cdot k < 3k^2(4k + 2)\,, \tag{1}$$

where $|N(u) \triangle N(u')| \leq k$ holds by the definition of ρ_k.

On the other hand, we may fix $w \in V$ and count the number D_w of unordered pairs $u, u' \in U' \setminus \{w\}$ such that exactly one of wu, wu' is an edge of G; this number is equal to $|N(w) \cap U'| \cdot |U' \setminus N(w)| = \alpha^{U'}(w) \cdot (|U'| - \alpha^{U'}(w))$ if $w \notin U'$, and to $\alpha^{U'}(w) \cdot (|U'| - 1 - \alpha^{U'}(w))$ or $(\alpha^{U'}(w) - 1) \cdot (|U'| - \alpha^{U'}(w))$ if $w \in U'$. In any case,

$$\begin{aligned} D_w &\geq \left(\alpha^{U'}(w) - 1\right) \cdot \left(|U'| - 1 - \alpha^{U'}(w)\right) = \left(\alpha^{U'}(w) - 1\right) \cdot \left(4k + 1 - \alpha^{U'}(w)\right) \\ &\geq \min\left((k + 1 - 1)(4k + 1 - k - 1), (3k + 1 - 1)(4k + 1 - 3k - 1)\right) = 3k^2, \end{aligned}$$

since the product of two factors of a constant sum is minimized when their difference is extreme. Therefore,

$$D = \sum_{w \in V} D_w \geq 3k^2 \cdot |V| \geq 3k^2(4k + 2). \tag{2}$$

Now, Equations (1) and (2) are in a contradiction, and hence the sought conclusion follows. □

Corollary 5.3. *Let U and V be the two classes of Lemma 5.2 such that $|U|, |V| \geq 5k + 1$. Then exactly one of the following two possibilities holds:*

(a) every vertex of U is connected to at most $2k$ vertices of V and every vertex of V is connected to at most $2k$ vertices of U, or
(b) every vertex of U is connected to all but *at most $2k$ vertices of V and every vertex of V is connected to* all but *at most $2k$ vertices of U.*

Proof. We first show that either

- every vertex of U is connected to at most $2k$ vertices of V, or
- every vertex of U is connected to all but $2k$ vertices of V.

Indeed, for any vertex $v \in U$ taken separately, only one of these cases can happen, since $|V| > 4k$, and one of these cases has to happen by Lemma 5.2, since $|V| \geq 5k + 1 \geq 4k + 2$. Assume that there exist $v, w \in U$ with v having at most $2k$ neighbours in V while w is connected to all but at most $2k$ vertices of V. Then $|N(v) \triangle N(w)| \geq |V| - 2k - 2k \geq k + 1$, contradicting the definition of ρ_k.

To finish the proof, we have to show the the following case (relevant if $U \neq V$) is impossible: every vertex of U connected to at most $2k$ vertices of V and every vertex of V connected to *all but* at most $2k$ vertices of U. In the argument, we count the total number of edges between U and V; it would be at most $2k \cdot |U|$ and, at the same time, at least $(|U| - 2k) \cdot |V|$. Though, the difference between these lower and upper estimates is

$$\begin{aligned} (|U| - 2k) \cdot |V| - 2k \cdot |U| &= |U| \cdot |V| - 2k \cdot (|U| + |V|) \\ &= \left(|U| - 4k\right)\left(|V| - 4k\right) + 2k\left(|U| + |V|\right) - 16k^2 \\ &> k \cdot k + 2k(5k + 5k) - 16k^2 = 5k^2 > 0, \end{aligned} \tag{3}$$

a contradiction, thus finishing the whole proof. □

Remark 5.4. Note that Corollary 5.3 still applies if $U = V$. I.e., for a single near-k-twin equivalence class U with $|U| > 5k + 1$ either

(a) every vertex of U has at most $2k$ neighbours in U, or
(b) every vertex of U has at least $|U| - 2k$ neighbours in U.

5.2 From Near-k-twins to Bounded Degree

Here, we present the core of our algorithm—a procedure that, given a graph H for which the near-k-twin relation of H is an equivalence of bounded index, produces a (labelled) graph G_H (on the same vertex set) of bounded degree, and a formula $\psi(x, y)$ such that $H = I_\psi(G_H)$.

The idea behind the procedure is the following: We start by dividing the near-k-twin classes of H into "small" and "large" ones (w.r.t. k), dealing with each of these two types of classes separately.

- Each large class (more precisely, the vertices in the class) is assigned a label and each pair of large classes receives another label indicating whether there are "almost all" or "almost none" edges between the two classes. The exceptions to "almost all" or "almost none" rules will be remembered by edges of the graph G_H (by Corollary 5.3 each vertex has a bounded number of such exceptions, hence the bounded degree of G_H). Using these labels and the graph G_H, we properly encode the H-adjacency (i.e., the edges of H) between the vertices in the large classes.
- The H-adjacency of the vertices from small equivalence classes (both within the small classes and also to the large ones) is encoded by assigning a new label to each such vertex and another new label to its neighbourhood. The vertices from small classes have no edges in the graph G_H.

Note that the construction sketched above depends on k and also on the number of near-k-twin equivalence classes of H. Unfortunately, as explained earlier, we cannot fix one universal value of the parameter k beforehand, but at least we can use upper bounds on both k and the number of equivalence classes (as in Definition 4.2). With a slightly more complicated use of labels, we can then give a universal formula $\psi(x, y)$, which depends only on the parameters k_0 and p of a (k_0, p)-near-uniform graph class C, but is independent from particular $H \in C$. This way, we get a result even stronger than what is required for the proof of Theorem 5.1 (see Section 6 for more discussion):

Theorem 5.5. *Let $k_0, p \in \mathbb{N}$, and C be a (k_0, p)-near-uniform graph class. There exists an FO formula $\psi(x, y)$, depending only on k_0 and p, such that $C \subseteq I_\psi(\mathcal{D}_{2k_0p})$ where $\mathcal{D}_d$ denotes the class of (finite) graphs of degree at most d.*

Furthermore, for any $H \in C$ and $k \leq k_0$ such that the near-k-twin relation of H is an equivalence of index at most p, one can in polynomial time compute a graph $G_H \in \mathcal{D}_{2k_0p}$ such that $H = I_\psi(G_H)$.

Proof. We are going to prove the theorem by defining the formula $\psi(x, y)$ and, for each $H \in C$, efficiently constructing a graph $G_H \in \mathcal{D}_{2k_0p}$ such that $H = I_\psi(G_H)$. We give the construction of the graph G_H first, while postponing the definition of ψ to the end of the proof.

Let $0 \leq k \leq k_0$ be such that the near-k-twin relation of H is an equivalence of index at most p. Let $V_1, \ldots, V_m$ where $m \leq p$ be the near-k-twin classes of H with more than $5k$ vertices (possible "small" near-k-twin classes are ignored now). Observe that $W = V_1 \cup \cdots \cup V_m$ contains all but at most $5k(p - m) \leq 5k_0p$ vertices of H. Let $\overline{W} = V(H) \setminus W$ denote the remaining vertices in "small" equivalence classes. See an illustration in Figure 3.

We will construct the graph G_H in three stages. First, we define the graph $G_1 = (W, E_1 \cup E_2)$ on the set W, where the edge sets are given as:

- Let F_1 be the set of those indices i from $\{1, \ldots, m\}$ such that every vertex of V_i has at least $|V_i| - 2k$ neighbours in V_i (case (b) of Remark 5.4). We put $E_1 = \{\{u, v\} \mid u \neq v \wedge \exists i \in F_1 \text{ s.t. } u, v \in V_i\}$.
- Let F_2 be the set of those index pairs $\{i, j\}$ from $\{1, \ldots, m\}$ such that every vertex of V_i is connected to *all but* at most $2k$ vertices of V_j and every vertex of V_j is connected to *all but*

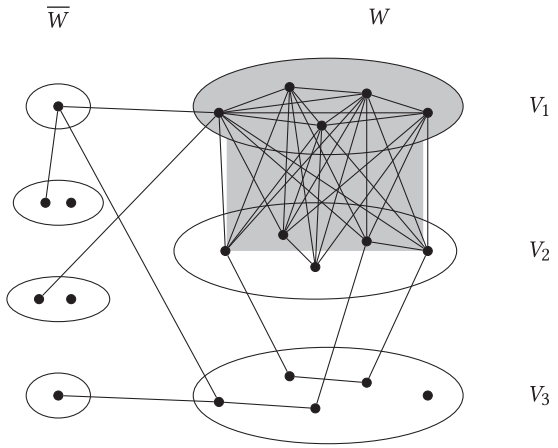

Fig. 3. An illustration; small (on the left, $\overline{W}$) and large (on the right, W) near-k-twin classes of a graph H, and prevailing adjacencies within the large classes remembered by sets $F_1 = \{1\}$ and $F_2 = \{\{1, 2\}\}$, as in the proof of Theorem 5.5.

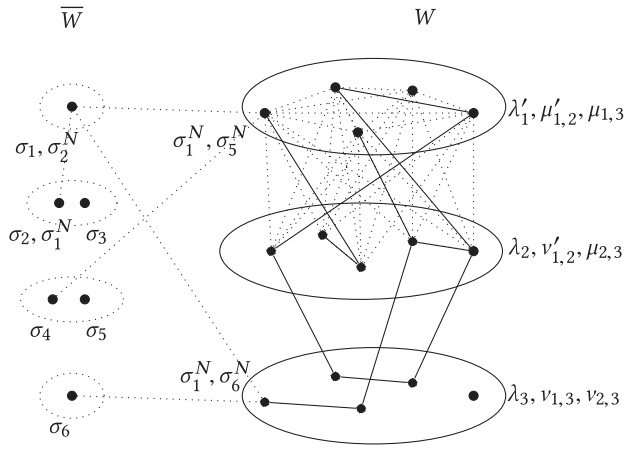

Fig. 4. An illustration; graph G_2 of maximum degree 3 constructed for H (the dotted edges) from Figure 3, and the resulting labelling of $V(G_2) = W \cup \overline{W}$, as in the proof of Theorem 5.5.

at most $2k$ vertices of V_i (case (b) of Corollary 5.3). We put $E_2 = \{\{u, v\} \mid \exists \{i, j\} \in F_2 \text{ s.t. } u \in V_i \wedge v \in V_j\}$.

In the second step, we adjust G_1 by the original edges from H: Let $E_W = \{\{u, v\} \in E(H) \mid u, v \in W\}$. Then, we put $G_2 = (W, E(G_1) \triangle E_W)$. See in Figure 4. Note that every vertex of G_2 has degree at most $2km$ by Corollary 5.3.

In the degenerate case of $k = 0$, we arrive at the same conclusion by the following alternative argument. By the definition, each near-0-twin class is an independent set and each pair of classes is again independent or induces a complete bipartite subgraph—this now defines G_1 and G_2, which is actually edgeless.

In the third step, we add back the vertices from $\overline{W}$ (remember that $V(H) = W \cup \overline{W}$) by putting $G_H = (W \cup \overline{W}, E(G_2))$. Note that $G_H \in \mathcal{D}_{2km} \subseteq \mathcal{D}_{2k_0 p}$.

Finally, we label the vertices of G_H by the following fixed label set, which is independent of particular $H \in C$:

$$\begin{aligned} L := &\{\lambda_i, \lambda'_i : i = 1, \ldots, p\} \\ &\cup \{\mu_{i,j}, \nu_{i,j},\ \mu'_{i,j}, \nu'_{i,j} : 1 \leq i < j \leq p\} \\ &\cup \{\sigma_j, \sigma_j^N : j = 1, \ldots, 5k_0p\}. \end{aligned}$$

The vertices of G_H are labelled as follows (see again Figure 4):

- For $i = 1, \ldots, m \leq p$, each vertex of V_i is assigned label λ'_i if $i \in F_1$, and label λ_i otherwise.
- For $1 \leq i < j \leq m \leq p$, each vertex of V_i is assigned label $\mu'_{i,j}$ and each of V_j label $\nu'_{i,j}$ if $\{i, j\} \in F_2$, and labels $\mu_{i,j}$ and $\nu_{i,j}$, respectively, if $\{i, j\} \notin F_2$.
- Let $\overline{W} = \{w_1, w_2, \ldots, w_r\}$ be indexed in any chosen order. For $j = 1, \ldots, r \leq 5k_0p$, the vertex w_j is assigned label σ_j and each neighbour of w_j in H is assigned label σ_j^N.

With G_H in place, we can now define the formula

$$\psi(x, y) \ \equiv \ (x \neq y) \wedge (\psi'(x, y) \vee \psi'(y, x)),$$

where

$$\begin{aligned} \psi'(x, y) \equiv &\bigvee_{1 \leq i \leq p} \Big(\lambda_i(x) \wedge \lambda_i(y) \wedge edge(x, y)\Big) \\ \vee &\bigvee_{1 \leq i \leq p} \Big(\lambda'_i(x) \wedge \lambda'_i(y) \wedge \neg edge(x, y)\Big) \\ \vee &\bigvee_{1 \leq i < j \leq p} \Big(\mu_{i,j}(x) \wedge \nu_{i,j}(y) \wedge edge(x, y)\Big) \\ \vee &\bigvee_{1 \leq i < j \leq p} \Big(\mu'_{i,j}(x) \wedge \nu'_{i,j}(y) \wedge \neg edge(x, y)\Big) \\ \vee &\bigvee_{1 \leq j \leq 5k_0p} \Big(\sigma_j(x) \wedge \sigma_j^N(y)\Big). \end{aligned}$$

Clearly, $\psi(x, y)$ is independent of particular $H \in C$ and depends only on the parameters k_0 and p. The construction of G_H from H and k is finished in polynomial time and it is also a simple routine to verify that $H = I_\psi(G_H)$. □

This also finishes the proof of Theorem 5.1 via the fixed-parameter tractable algorithm of Seese [27]. However, we also state a refined version of Theorem 5.1, which shows that the precise runtime of our algorithm is only quadratic in the number of vertices of G, assuming fixed class C and formula ϕ.

Proposition 5.6. *Let C be a (k_0, p)-near-uniform graph class for some $k_0, p \in \mathbb{N}$. Then the FO model checking problem of C, for an input $H \in C$ and ϕ, is solvable in time $f'(k_0, p, |\phi|) \cdot |V(H)|^2$ for a computable function f'.*

Proof. We follow the algorithm outline from before Theorem 5.1. The key step is to efficiently verify that, for $k \leq k_0$, the near-k-twin relation ρ_k of H is an equivalence of index at most p. Having done that step, we then easily finish in the claimed runtime along the detailed constructive steps of the proof of Theorem 5.5.

Computing ρ_k of an n-vertex graph H, in general, requires time $O(n^3)$, but assuming fixed parameters k and p, we can do slightly better in time $O(p(p + k)n^2)$ as follows:

(i) We find the at most p candidate equivalence classes of ρ_k, by iteratively going through all vertices of H and comparing the neighborhood of each vertex with the neighborhood of a representative of each of the already (partially) computed classes, or we witness that there must be more than p such classes. This requires time $O(p^2n^2)$.
(ii) Within each of the candidate classes $T \subseteq V(H)$, with a representative $t \in T$, we compute the symmetric differences of the neighbourhoods of $x \in T \setminus \{t\}$ and of t. Since each such difference is of size $O(k)$, we can then pairwise compare them to certify that T indeed is an equivalence class of ρ_k, or to witness a violation of tranzitivity of ρ_k. This part takes time $O(kn^2)$ for each one of the $\leq p$ candidate classes.

Summing these runtimes, we arrive at the aforementioned expression $O(p(p+k)n^2)$ for testing one ρ_k, and this is repeated for all $k \leq k_0$. Hence, in total, the initial step of the algorithm requires time $O(pk_0(p+k_0)n^2)$, which also clearly dominates the runtime of subsequent construction of the graph G_H as in the proof of Theorem 5.5. Finally, FO model checking of G_H is done in time $f''(|\phi|) \cdot n$ by Reference [27]. □

5.3 Successor-invariant FO

Model checking of successor-invariant FO properties is the subject of several recent papers such that Engelmann et al. [11], Ganian et al. [17], Eickmeyer and Kawarabayashi [10], van den Heuvel et al. [28], and others.

In a nutshell, a *successor* relation on a domain X is simply a directed path on the vertex set X (in case of X being the vertex set of a graph, this successor relation is distinct from the graph edges). An FO property over a successor-equipped relational structure is *successor-invariant* if its truth does not change when the same structure is equipped with a different successor relation.

Since successor-invariant FO sentences are generally more expressive than FO sentences [26], it makes good sense to ask whether graph classes with efficient FO model checking algorithms also admit efficient successor-invariant FO model checking. So far, the answers provided by the previously listed works are all positive. Furthermore, since the known examples separating the expressive powers of plain FO and successor-invariant FO are dense (containing large cliques), and the previous works (except Reference [17]) studied sparse graphs, it is especially relevant to ask the question of successor-invariant FO in our dense case.

While it is not clear that successor-invariant FO is indeed more expressive than FO on near-uniform graph classes, the positive algorithmic result for successor-invariant FO follows directly from our Theorem 5.5 and the algorithm of Reference [28]:

Corollary 5.7. *Let C be a (k_0, p)-near-uniform graph class for some $k_0, p \in \mathbb{N}$. Then the successor-invariant FO model checking problem in C is fixed-parameter tractable when parameterized by the formula size.*

Proof. We proceed in the same way as previously. To recapitulate, let ϕ be a successor-invariant FO sentence and $H \in C$ an input graph. By Theorem 5.5, we get formula ψ and compute a graph $G_H \in \mathcal{D}_{2k_0p}$ such that $H = I_\psi(G_H)$. Let ϕ' be obtained from ϕ by replacing every occurrence of $edge(z, z')$ with $\psi(z, z')$. Now, it is important that the domain of G_H is the same as the domain of H, and so any successor relation on G_H is a successor relation on H as well. Consequently, for any successor-equipped H and G_H, we have $H \models \phi$ if and only if $G_H \models \phi'$ (as with previous plain FO logic).

Hence it remains to solve in FPT the successor-invariant FO model checking problem of graphs of bounded degree. This can be done by the algorithm of Reference [28] (which handles more generally graph classes with bounded expansion). □

6 INTERPRETABILITY OF GRAPHS OF BOUNDED DEGREE

Having defined near-uniform graph classes and shown that these classes can be FO interpreted in graph classes of bounded degree, it is a natural question to ask what is the exact relationship between those kinds of classes. As it turns out, we can prove (Theorem 6.4) that each class FO interpretable in a class of graphs of bounded degree is indeed near-covered and therefore also near-uniform. Thus, near-uniform graph classes are exactly those graph classes that are FO interpretable in graph classes of bounded degree. This result can then be easily extended to the more general case of transductions of graph classes of bounded degree, which again result in near-uniform graph classes (Theorem 6.5).

6.1 Adjusted Gaifman's Theorem

In the proof of the main result of this section, we use a corollary of the famous Gaifman's locality theorem [13] (see also Reference [23]) about the local nature of the FO logic.

To keep the article self-contained, in this section, we first recap the notation and statement of Gaifman's theorem and then state a corollary tailored to our needs.

An FO formula $\phi(x_1, \ldots, x_l)$ is *r-local*, sometimes denoted by $\phi^{(r)}(x_1, \ldots, x_l)$, if for every graph G and all $v_1, \ldots, v_l \in V(G)$ it holds $G \models \phi(v_1, \ldots, v_l) \iff \bigcup_{1 \leq i \leq l} N_r^G(v_i) \models \phi(v_1, \ldots, v_l)$, where $N_r^G(v)$ is the subgraph of G induced by v and all vertices of distance at most r from v.

Theorem 6.1 (Gaifman's Theorem). *Every first-order formula with free variables $x_1, \ldots, x_l$ is equivalent to a Boolean combination of the following:*

- *local formulas $\phi^{(r)}(x_1, \ldots, x_l)$ around $x_1, \ldots, x_l$, and*
- *basic local sentences, i.e., sentences of the form*

$$\exists x_1 \ldots \exists x_k \left(\bigwedge_{1 \leq i < j \leq k} dist(x_i, x_j) > 2r \wedge \bigwedge_{1 \leq i \leq k} \phi^{(r)}(x_i) \right).$$

The following statement, known as Gaifman-locality of FO, easily follows from Gaifman's theorem (see Reference [21] and also Chapter 4 of Reference [23]).

Theorem 6.2. *For every FO formula $\psi(x_1, \ldots, x_k)$ there exists r such that for every G and every $v_1, \ldots, v_k \in V(G)$ and $u_1, \ldots, u_k \in V(G)$ it holds that if $N_r(v_1, \ldots, v_k)$ is isomorphic to $N_r(u_1, \ldots, u_k)$, then*

$$G \models \psi(v_1, \ldots, v_k) \iff G \models \psi(u_1, \ldots, u_k).$$

As an immediate corollary, we get the following statement, which will be used in the next section.

Corollary 6.3. *For every FO formula $\psi(y, z)$ of two free variables there exists an integer r such that the following holds true for any graph G: If $u, v_1, v_2 \in V(G)$ such that $dist(u, v_1) > 2r$, $dist(u, v_2) > 2r$ and $N_r(v_1)$ is isomorphic to $N_r(v_2)$, then $G \models \psi(u, v_1)$ if and only if $G \models \psi(u, v_2)$.*

6.2 Characterization of Interpretations

The following theorem provides us with a strong characterization of the classes FO interpreted in graphs of degree at most d, in terms of near-k-twin relation and being near-covered. It amounts to, in an essence, the "opposite direction" to Theorem 5.5.

Theorem 6.4. *Let $\mathcal{D}_d$ be the class of (finite) graphs with maximum degree at most d and let $\psi(x, y)$ be an FO formula with two free variables. Then there exist ℓ and q, depending only on d and ψ, such*

that every graph $H \in I_\psi(\mathcal{D}_d)$ is (ℓ, q)-near-covered. That is, there exists a set $S = \{v_1, \ldots, v_q\} \subseteq V(H)$ such that every $u \in V(H)$ is a near-ℓ-twin of at least one element v_i of S.

Proof. Let $H \in I_\psi(\mathcal{D}_d)$ and $G \in \mathcal{D}_d$ be such that $H = I_\psi(G)$. Recall that $V(H) = V(G)$ and $\{u, v\} \in E(H)$ if and only if $G \models \psi(u, v)$. Fixing G and H, we say that a vertex $x \in V(H)$ is *a-far from $y \in V(H)$* if the graph distance from x to y in G is greater than a.

Let r be the value obtained from application of Corollary 6.3 to ψ. Let T_G be the set of all isomorphism types of r-neighbhorhoods of vertices of G. Since every vertex of G has degree at most d, the set T_G is finite. Set $q := |T_G|$ and $\ell := 2d^r$. For every $t \in T_G$, we pick one vertex in G (and also in H, since $V(H) = V(G)$), which realizes type t to obtain $S := \{v_1, \ldots, v_q\}$. We claim that every vertex u of H is near-ℓ-twin of some vertex in S. For every vertex u of H, there is a vertex v_i in S such that $N_r(u)$ is isomorphic to $N_r(v_i)$ (in G). By Corollary 6.3, for every w that is more than r far from both u and v_i, it holds that $G \models \psi(w, u)$ if and only if $G \models \psi(w, v_i)$. This means that u and v_i will have the same adjacency to all vertices that are more than r far from both of them. Consequently, their neighbourhoods in H can only differ in vertices that are at most r far from either of them, and there are at most $2d^r$ such vertices. □

6.3 Characterization of Transductions

Besides the simplified case of interpretation from Theorem 6.4, we can go much further with a bit of additional effort, as carried out in the following claim.

Theorem 6.5. *Let $\mathcal{D}_d$ be the class of (finite) graphs with maximum degree at most d and let τ be an FO transduction. Then there exist ℓ and q, depending only on d and τ, such that every graph $H \in \tau(\mathcal{D}_d)$ is (ℓ, q)-near-covered.*

Proof. Our strategy is to prove that there exist an integer d', an FO formula $\psi(x, y)$ and a class of labelled graphs $\mathcal{G} \subseteq \mathcal{D}_{d'}$, all depending on d and τ, such that the following holds: for every $H \in \tau(\mathcal{D}_d)$ there exists a graph $H' \in I_\psi(\mathcal{G})$ such that H' is obtained from H by adding isolated vertices.

Assuming the previous for a moment, we show how it implies our theorem. By Theorem 6.4, there exist ℓ_1 and q such that, for every graph $H' \in I_\psi(\mathcal{G})$, the following holds: there exists $S_1 = \{v_1, \ldots, v_q\} \subseteq V(H')$ such that every $u \in V(H')$ is a near-ℓ_1-twin of at least one element of S_1. We may assume that at most one element of S_1, say v_1, is among the added isolated vertices of H' (hence $S_1 \setminus V(H) \subseteq \{v_1\}$). If $v_1 \notin V(H)$ and there exists a vertex $w_1 \in V(H)$, which is a near-ℓ_1-twin of v_1, then w_1 is a near-$2\ell_1$-twin of every near-ℓ_1-twin of v_1, and so $S := (S_1 \cap V(H)) \cup \{w_1\}$ witnesses that H is $(2\ell_1, q)$-near-covered. Otherwise, H is (ℓ_1, q)-near-covered by $S := S_1 \cap V(H)$.

It remains to define an appropriate class $\mathcal{G}$ and the formula ψ as claimed above.

From the definition of FO transduction, let $\tau = \tau_0 \circ \gamma \circ \varepsilon$ where τ_0 is a basic transduction, γ is a m-copy operation for some m, and ε is a p-parameter expansion for some p. We start with setting $d' = \max(d + 1, m)$, and $\mathcal{G}_1 = \varepsilon(\mathcal{D}_d)$. For every $G_1 \in \mathcal{G}_1$, we take $G_2 = \gamma(G_1)$. Then, for every $v \in V(G_1)$, we add to G_2 a new vertex v_0 of a new label R (the same for each added v_0), and m edges from v_0 to the m copies of v in G_2 (making a star $K_{1,m}$ with the new centre v_0). The resulting graph G_3 has $(m + 1) \cdot |V(G)|$ vertices and maximum degree d'. Finally, after formally erasing the relation $\sim$ of γ, we add G_3 to $\mathcal{G}$.

Regarding the formula $\psi(x, y)$, we recall that the basic transduction τ_0 underlying τ is determined by a triple of FO formulas (χ, ν, μ), where the role of χ can be safely ignored for now. We make formulas $\nu'(x)$ from $\nu(x)$, and $\mu'(x, y)$ from $\mu(x, y)$, by restricting every quantifier to the vertices *not* of label R and replacing each occurrence of the predicate $u \sim v$ (recall that $\sim$ has been

erased from G_3) with $\exists t(R(t) \wedge edge(u,t) \wedge edge(v,t))$. Then, we set

$$\psi(x,y) \equiv \neg R(x) \wedge \nu'(x) \wedge \neg R(y) \wedge \nu'(y) \wedge \mu'(x,y). \tag{4}$$

Pick now any $H \in \tau(\mathcal{D}_d)$, and let $G_1 \in \mathcal{G}_1 = \varepsilon(\mathcal{D}_d)$ be the corresponding graph such that $G_2 = \gamma(G_1)$ and $H = \tau_0(G_2)$. Let $G_3 \in \mathcal{G}$ be constructed from G_2 as above. By τ_0 and Equation (4), every vertex of $I_\psi(G_3)$ not in $V(H)$ is isolated in $I_\psi(G_3)$. Moreover, by the construction of ψ' for every two vertices $u, v \in V(H)$ it holds $G_2 \models \mu(u,v) \iff G_3 \models \mu'(u,v)$. Hence, by Equation (4), $I_\psi(G_3)$ results from H by adding isolated vertices, and we can set $H' = I_\psi(G_3)$, as desired. The proof is finished. □

Putting together the results of Theorems 5.5 and 6.5, we easily get also the following corollary, which is interesting on its own:

Corollary 6.6. *Let $\mathcal{C}$ be a near-uniform graph class, and τ be an FO transduction. Then the class $\tau(\mathcal{C})$ is again a near-uniform graph class.*

Proof. By Theorem 5.5, there exists an FO formula $\psi(x,y)$ such that $\mathcal{C} \subseteq I_\psi(\mathcal{D}_d)$ for suitable degree bound d depending on $\mathcal{C}$. Let τ_1 be the corresponding basic transduction (determined by $(true, true, \psi)$). Then $\tau(\mathcal{C}) \subseteq \tau(\tau_1(\mathcal{D}_d))$, and since $\tau \circ \tau_1$ is again a transduction by transitivity, by Theorem 6.5 every graph of $\tau(\tau_1(\mathcal{D}_d))$ is (ℓ, q)-near-covered. Consequently, using Lemma 4.4, $\tau(\mathcal{C})$ is also a near-uniform graph class. □

7 HARDNESS OF RECOGNIZING AN INTERPRETATION

Recall the aforementioned result [24] claiming that it is NP-hard to decide whether a given graph is a *square* of some graph. The square of a graph can be straightforwardly described by an FO interpretation with $\psi_s(x,y) \equiv edge(x,y) \vee [x \neq y \wedge \exists z(edge(x,z) \wedge edge(z,y))]$, expressing that edges of the square are original edges or pairs at distance exactly two.

In our context, Reference [24] hence means that there exist a graph class $\mathcal{C}$ and an FO formula $\psi(x,y)$ such that the problem, for a given graph $H \in I_\psi(\mathcal{C})$, to find $G \in \mathcal{C}$ such that $H = I_\psi(G)$ is not efficiently solvable (unless P=NP). Though, the reduction of Reference [24] requires a class $\mathcal{C}$ of unbounded maximum degree, while we are primarily interested in interpretations of the classes $\mathcal{D}_d$ of graphs of degrees at most d. Here, we show a straightforward alternative reduction working already with the class $\mathcal{D}_3$ of graphs of degree at most 3.

Notice that such a result is not in a contradiction with Theorem 5.5, since each of the two results speaks about a different particular formula(s) ψ.

Theorem 7.1. *Let $\mathcal{D}_3$ denote the class of graphs of degree at most* 3. *There exists an FO formula $\psi_0(x,y)$ such that the problem, for a given graph $H \in I_{\psi_0}(\mathcal{D}_3)$, to decide whether there exists a graph $G \in \mathcal{D}_3$ such that $H = I_{\psi_0}(G)$ is NP-hard.*

Proof. We reduce from the folklore NP-hard problem of deciding whether there exists a proper 3-colouring of a given 4-regular graph H_0. We construct a graph H from an arbitrary 4-regular graph H_0 as follows:

- Every vertex v of H_0 is replaced with a graph T_v, which is a copy of the graph in Figure 5 including the dashed edges.
- Every edge e of H_0 is replaced with a graph U_e, which is a copy of the graph in Figure 6 including the dashed edges.
- For every edge $e = \{u, v\}$ of H_0, the terminal e^1 of U_e is identified with u^i of T_u, and e^2 of U_e is identified with v^j of T_v, where e is the ith edge at u and the jth edge at v (for arbitrarily chosen orderings of edges incident to u, v).

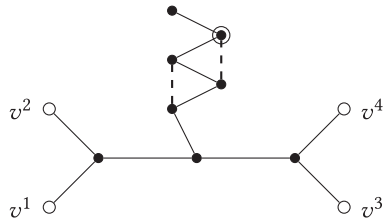

Fig. 5. The vertex gadget T_v in the proof of Theorem 7.1.

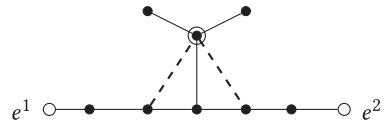

Fig. 6. The edge gadget U_e in the proof of Theorem 7.1.

The construction of H is independent of whether H_0 is 3-colourable. Note that since U_e contains a vertex of degree 5, we have $H \notin \mathcal{D}_3$.

Before defining the formula ψ_0, we briefly explain the underlying idea of the reduction. For a suitable subgraph G of H (on the same vertex set), we would like to have $H = I_{\psi_0}(G)$ if and only if every vertex gadget (of a vertex of H_0) restricted to G encodes one of three available colours (for this vertex in H_0), and every edge gadget in G "verifies" that the ends of the edge (in H_0) receive distinct colours.

The above rough sketch is made precise now. Considering colours 1,2,3, we define three *reduced vertex gadgets* of a vertex $v \in V(H_0)$ as $T_v^1 = T_v$ and T_v^2, T_v^3 obtained from T_v by removing one or the other dashed edge of T_v in Figure 5. Similarly, a *reduced edge gadget* U_e' of an edge $e \in E(H_0)$ is obtained from U_e in Figure 6 by removing both dashed edges. Assuming any 3-colouring $c : V(H_0) \to \{1, 2, 3\}$, we construct a graph $G \in \mathcal{D}_3$ analogously to the above construction of H, while replacing every vertex $v \in V(H_0)$ with $T_v^{c(v)}$ and every edge $e \in E(H_0)$ with U_e'.

Note that $G \subset H$. We call a vertex w a *v-marker* if w is adjacent to precisely one vertex of degree 1, and we call w an *e-marker* if w is adjacent to two vertices of degree 1 (see the circled vertices in Figures 5 and 6, respectively). Then every e-marker w of G belongs to some U_e' of $e = \{u, v\} \in E(H_0)$, and there are precisely two v-markers of G at distance 9 from w belonging to T_u^i and to T_v^j. We would now like to "verify" that the colouring c is proper, i.e., that $i \neq j$, in the formula ψ_0.

We define $\psi_0(x, y) \equiv edge(x, y) \vee v(x, y) \vee \eta(x, y)$, where

- $v(x, y)$ asserts that there exists z, which is a neighbour of x or y, such that z is a v-marker and the 5-neighbourhood of z is isomorphic to one of T_v^1, T_v^2, T_v^3, and that x, y are the ends of one of the dashed edges in Figure 5;
- $\eta(x, y)$ asserts that one of x, y, say x, is an e-marker, y is at distance two from x, and the following holds: there exist vertices z, z' at distance 9 from x such that z, z' are v-markers with their 5-neighbourhoods isomorphic to T_v^i and T_v^j where $i \neq j$.

It is routine to rewrite the above description into an FO formula.

Clearly, $H = I_{\psi_0}(G)$ if and only if the above colouring c is proper. Conversely, it remains to prove that if $H = I_{\psi_0}(G)$ for any $G \in \mathcal{D}_3$, then H_0 is 3-colourable. Notice that $G \subseteq H$ and that the formula ψ_0 does not "add" edges to degree-1 vertices, and so the degree-1 vertices of G must be in a one-to-one correspondence with the v-marker and e-marker vertices of H.

Fix an e-marker w belonging to $U_e \subseteq H$. Since w is of degree 5 in H and of degree ≤ 3 in $G \in \mathcal{D}_3$, it is $G \models \psi_0(w,t)$ for some (actually, at least two) neighbour t of w in H. In particular, by the definition of $\eta(w,t)$, this means there exist two v-markers w', w'' at distance 9 from w in G. From the construction of H, we know that w', w'' belong to T_u, T_v, respectively, where u, v are the ends of e in H_0. Again by $G \models \psi_0(w,t)$, the subgraph of G induced by $V(T_u)$ is one of T_u^1, T_u^2, T_u^3, say it is T_u^i. Similarly, the subgraph of G induced by $V(T_v)$ is, say, T_v^j and $i \neq j$. Since the same holds for any edge of H_0, an (arbitrary) graph $G \in \mathcal{D}_3$ such that $H = I_{\psi_0}(G)$ indeed encodes a proper 3-colouring of H_0. □

8 QUESTIONS AND OPEN PROBLEMS

Our interpretation approach and obtained results open several natural questions that we believe are worth further investigation. We list them in this last section of the article.

1. Can one characterize under which conditions on a formula $\psi(x,y)$ and a graph class C, the following holds? Given a graph $H \in I_\psi(C)$ as an input, it would be possible to compute in polynomial (or in FPT with respect to ψ and C) time a graph $G \in C$ such that $H = \psi(G)$. We know both of positive and negative examples (Theorems 5.5 and 7.1), but any plausible conjecture seems now out of reach.
2. It is easy to generalize the notion of near-k-twins u, v in such a way that it would measure not the size of the symmetric difference between the neighbourhoods, $|N(u) \triangle N(v)|$, but structural properties of the subgraph induced on $N(u) \triangle N(v)$. For example, we may define a *near-sd_k-twin relation*, in which two vertices u, v would be near-sd_k-twins if the subgraph induced on $N(u) \triangle N(v)$ has shrub-depth at most k (see Reference [18] for the definition of shrub-depth). One may then consider graph classes where the near-sd_k-twin relation is an equivalence. Is there an FPT algorithm for FO model checking on such graph classes?
3. What are the characterizations of graph classes interpretable in other sparse graph classes? Are there FPT algorithms for FO model checking on such classes? A recent result of Gajarský et al. [16] provides a structural characterization of graph classes interpretable in graph classes of bounded expansion. Even though many commonly studied classes of sparse graphs (such as planar graphs, graphs with excluded minor, etc.) have bounded expansion, it would still be very interesting to have a precise characterization of graph classes interpretable in these special instances of sparse graph classes. Moreover, it is not known how to efficiently compute the decomposition defined in Reference [16], and so the existence of an efficient FO model checking algorithm for graph classes interpretable in sparse graph classes remains open.
4. In relation to the previous point, we know from Corollary 6.6 that the notion of near-uniform graph classes is robust under FO interpretations and transductions. We know of (at least) three other examples of such behaviour—the graph classes of bounded clique-width [6] and those of bounded shrub-depth [18] (which are both robust even under MSO transductions), and stable classes of graphs [1]. Can one come up with other natural and interesting graph properties defining graph classes robust under FO transductions?
5. Inspired by the classification of sparse graph classes by Nešetřil and Ossona de Mendez [25], we may investigate graph classes $\mathcal{D}$ with the property that, for every FO formula $\psi(x,y)$ there exists a graph F_ψ (as "forbidden") such that F_ψ is not present as an induced subgraph in any member of $I_\psi(\mathcal{D})$. This logical definition may be considered in analogy to the structural definition(s) of nowhere dense classes [25] (as "nowhere FO dense"). What can we say about complexity of FO model checking on such classes $\mathcal{D}$?

To conclude, we make the following two explicit conjectures related to points 3 and 5 of the discussion.

Conjecture 8.1. *Let $\mathcal{C}$ be a nowhere dense graph class and $\mathcal{D}$ a graph class FO interpretable in $\mathcal{C}$. Then $\mathcal{D}$ has an FPT algorithm for FO model checking.*

Conjecture 8.2 ("Nowhere FO Dense"). *Let $\mathcal{D}$ be a graph class with the following property: for every FO formula $\psi(x, y)$ there exists a graph F_ψ such that F_ψ is not an induced subgraph of any member of $I_\psi(\mathcal{D})$. Then $\mathcal{D}$ has an FPT algorithm for FO model checking.*

Regarding Conjecture 8.2, it is tempting to strengthen its conclusion to; "then $\mathcal{D}$ is FO interpretable in some nowhere dense graph class," but that actually fails. For example, take the class $\mathcal{D}$ of graphs of clique-width 2. Then, for every FO formula $\psi(x, y)$, the interpreted class $I_\psi(\mathcal{D})$ is of bounded clique-width, too, and so a forbidden graph F_ψ always exists in this case. However, the class $\mathcal{D}$ is not interpretable in any nowhere dense graph class.

REFERENCES

[1] H. Adler and I. Adler. 2014. Interpreting nowhere dense graph classes as a classical notion of model theory. *Eur. J. Comb.* 36 (2014), 322–330.

[2] A. Blumensath and B. Courcelle. 2010. On the Monadic second-order transduction hierarchy. *Log. Methods Comput. Sci.* 6, 2:2 (2010), 1–28.

[3] S. Bova, R. Ganian, and S. Szeider. 2014. Model checking existential logic on partially ordered sets. In *Proceedings of the CSL-LICS'14*. ACM, 1–10. Article No. 21.

[4] B. Courcelle and J. Engelfriet. 2012. *Graph Structure and Monadic Second-Order Logic: A Language-Theoretic Approach*. Encyclopedia of Mathematics and Its Applications, Vol. 138. Cambridge University Press.

[5] B. Courcelle, J. A. Makowsky, and U. Rotics. 2000. Linear time solvable optimization problems on graphs of bounded clique-width. *Theory Comput. Syst.* 33, 2 (2000), 125–150.

[6] B. Courcelle and S. Olariu. 2000. Upper bounds to the clique width of graphs. *Discrete Appl. Math.* 101, 1–3 (2000), 77–114.

[7] A. Dawar, M. Grohe, and S. Kreutzer. 2007. Locally excluding a minor. In *Proceedings of the LICS'07*. IEEE Computer Society, 270–279.

[8] G. Dong, L. Libkin, and L. Wong. 1997. Local properties of query languages. In *Proceedings of the ICDT'97 (LNCS)*, Vol. 1186. Springer, 140–154.

[9] Z. Dvořák, D. Král', and R. Thomas. 2010. Deciding first-order properties for sparse graphs. In *Proceedings of the FOCS'10*. IEEE Computer Society, 133–142.

[10] K. Eickmeyer and K. Kawarabayashi. 2016. Successor-invariant first-order logic on graphs with excluded topological subgraphs. In *Proceedings of the CSL'16 (LIPIcs)*, Vol. 62. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 18:1–18:15.

[11] V. Engelmann, S. Kreutzer, and S. Siebertz. 2012. First-order and Monadic second-order model-checking on ordered structures. In *Proceedings of the LICS'12*. IEEE Computer Society, 275–284.

[12] M. Frick and M. Grohe. 2001. Deciding first-order properties of locally tree-decomposable structures. *J. ACM* 48, 6 (2001), 1184–1206.

[13] Haim Gaifman. 1982. On local and non-local properties. In *Proceedings of the Herbrand Symposium*. Stud. Logic Found. Math., Vol. 107. Elsevier, 105–135.

[14] J. Gajarský, P. Hliněný, D. Lokshtanov, J. Obdržálek, S. Ordyniak, M. S. Ramanujan, and S. Saurabh. 2015. FO model checking on posets of bounded width. In *Proceedings of the FOCS'15*. IEEE Computer Society, 963–974.

[15] J. Gajarský, P. Hliněný, J. Obdržálek, and S. Ordyniak. 2014. Faster existential FO model checking on posets. In *Proceedings of the ISAAC'14 (LNCS)*, Vol. 8889. Springer, 441–451.

[16] J. Gajarský, S. Kreutzer, J. Nešetřil, P. Ossona de Mendez, M. Pilipczuk, S. Siebertz, and S. Toruńczyk. 2018. First-order interpretations of bounded expansion classes. In *Proceedings of the ICALP'18 (LIPIcs)*, Vol. 107. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 126:1–126:14.

[17] R. Ganian, P. Hliněný, D. Král', J. Obdržálek, J. Schwartz, and J. Teska. 2015. FO model checking of interval graphs. *Log. Methods Comput. Sci.* 11, 4:11 (2015), 1–20.

[18] R. Ganian, P. Hliněný, J. Nešetřil, J. Obdržálek, P. Ossona de Mendez, and R. Ramadurai. 2012. When trees grow low: Shrubs and fast MSO_1. In *Proceedings of the MFCS'12 (LNCS)*, Vol. 7464. Springer, 419–430.

[19] M. Grohe. 2001. Generalized model-checking problems for first-order logic. In *Proceedings of the STACS'01 (Lecture Notes in Computer Science)*, Vol. 2010. Springer, 12–26.
[20] M. Grohe, S. Kreutzer, and S. Siebertz. 2014. Deciding first-order properties of nowhere dense graphs. In *Proceedings of the STOC'14*. ACM, 89–98.
[21] L. Hella, L. Libkin, and J. Nurmonen. 1999. Notions of locality and their logical characterizations over finite models. *J. Symb. Log.* 64, 4 (1999), 1751–1773.
[22] M. Lampis. 2010. Algorithmic meta-theorems for restrictions of treewidth. In *Proceedings of the ESA'10*. 549–560.
[23] L. Libkin. 2004. *Elements of Finite Model Theory*. Springer.
[24] R. Motwani and Madhu Sudan. 1994. Computing roots of graphs is hard. *Discrete Appl. Math.* 54, 1 (1994), 81–88.
[25] J. Nešetřil and P. Ossona de Mendez. 2012. *Sparsity: Graphs, Structures, and Algorithms*. Algorithms Combin., Vol. 28. Springer.
[26] B. Rossman. 2007. Successor-invariant first-order logic on finite structures. *J. Symb. Log.* 72, 2 (2007), 601–618.
[27] D. Seese. 1996. Linear time computable problems and first-order descriptions. *Math. Struct. Comput. Sci.* 6, 6 (1996), 505–526.
[28] J. van den Heuvel, S. Kreutzer, M. Pilipczuk, D. A. Quiroz, R. Rabinovich, and S. Siebertz. 2017. Model-checking for Successor-Invariant First-Order Formulas on Graph Classes of Bounded Expansion. Retrieved from arXiv:1701.08516.

Received April 2018; accepted December 2019