On the number of types in sparse graphs*

Michał Pilipczuk, Sebastian Siebertz, and Szymon Toruńczyk
Institute of Informatics, University of Warsaw, Poland
{michal.pilipczuk,siebertz,szymtor}@mimuw.edu.pl

Abstract

We prove that for every class of graphs $\mathscr{C}$ which is nowhere dense, as defined by Nešetřil and Ossona de Mendez [29, 30], and for every first order formula $\varphi(\bar{x}, \bar{y})$, whenever one draws a graph $G \in \mathscr{C}$ and a subset of its nodes A, the number of subsets of $A^{|\bar{y}|}$ which are of the form $\{\bar{v} \in A^{|\bar{y}|} : G \models \varphi(\bar{u}, \bar{v})\}$ for some valuation $\bar{u}$ of $\bar{x}$ in G is bounded by $O(|A|^{|\bar{x}|+\varepsilon})$, for every $\varepsilon > 0$. This provides optimal bounds on the VC-density of first-order definable set systems in nowhere dense graph classes. We also give two new proofs of upper bounds on quantities in nowhere dense classes which are relevant for their logical treatment. Firstly, we provide a new proof of the fact that nowhere dense classes are uniformly quasi-wide, implying explicit, polynomial upper bounds on the functions relating the two notions. Secondly, we give a new combinatorial proof of the result of Adler and Adler [1] stating that every nowhere dense class of graphs is stable. In contrast to the previous proofs of the above results, our proofs are completely finitistic and constructive, and yield explicit and computable upper bounds on quantities related to uniform quasi-wideness (margins) and stability (ladder indices).

Keywords Nowhere dense graphs, Stone space, first-order types, VC-density, stability, uniform quasi-wideness

1 Introduction

Nowhere dense classes of graphs were introduced by Nešetřil and Ossona de Mendez [29, 30] as a general and abstract model capturing uniform sparseness of graphs. These classes generalize many familiar classes of sparse graphs, such as planar graphs, graphs of bounded treewidth, graphs of bounded degree, and, in fact, all classes that exclude a fixed topological minor. Formally, a class $\mathscr{C}$ of graphs is *nowhere dense* if there is a function $t\colon \mathbb{N} \to \mathbb{N}$ such that for every $r \in \mathbb{N}$, no graph G in $\mathscr{C}$ contains the clique $K_{t(r)}$ on $t(r)$ vertices as *depth-r minor*, i.e., as a subgraph of a graph obtained from G by contracting mutually disjoint subgraphs of radius at most r to single vertices. We write The more restricted notion of *bounded expansion* requires in addition that for every fixed r, there is a constant (depending on r) upper bound on the ratio between the number of edges and the number of vertices in depth-r minors of graphs from $\mathscr{C}$.

The concept of nowhere denseness turns out to be very robust, as witnessed by the fact that it admits multiple different characterizations, uncovering intricate connections to seemingly distant branches of mathematics. For instance, nowhere dense graph classes can be characterized by upper bounds on the density of bounded-depth (topological) minors [29, 30], by uniform quasi-wideness [30] (a notion introduced by Dawar [10] in the context of homomorphism preservation properties), by low tree-depth colorings [28], by generalized coloring numbers [41], by sparse neighborhood covers [18, 19], by a game called the splitter game [18], and by the model-theoretic concepts of stability and independence [1]. For a broader discussion on graph theoretic sparsity we refer to the book of Nešetřil and Ossona de Mendez [31].

The combination of combinatorial and logical methods yields a powerful toolbox for the study of nowhere dense graph classes. In particular, the result of Grohe, Kreutzer and the second author [18] exploits this combination in order to prove that a given first order sentence φ can be evaluated in time $f(\varphi) \cdot n^{1+\varepsilon}$ on n-vertex graphs from a fixed nowhere dense class of graphs $\mathscr{C}$, for any fixed real $\varepsilon > 0$ and some function f. On the other hand, provided $\mathscr{C}$ is closed under taking subgraphs, it is known that if $\mathscr{C}$ is not nowhere dense, then there is no algorithm with running time of the form $f(\varphi) \cdot n^c$ for any constant c under plausible complexity assumptions [12]. In the terminology of parameterized complexity, these results show that the notion of nowhere denseness exactly characterizes subgraph-closed classes where model-checking first order logic is fixed-parameter tractable, and conclude a long line of research concerning the parameterized complexity of the model checking problem for sparse graph classes (see [16] for a survey).

Summary of contribution. In this paper, we continue the study of the interplay of combinatorial and logical properties of nowhere dense graph classes, and provide new upper bounds on several quantities appearing in their logical study. Our main focus is on the notion of *VC-density* for first order formulas. This concept originates from model theory and aims to measure the complexity of set systems definable by first order formulas, similarly to the better-known VC-dimension. We give optimal bounds on the VC-density in nowhere dense graph classes, and in particular we show that these bounds are "as good as one could hope for".

We also provide new upper bounds on quantities related to *stability* and *uniform quasi-wideness* of nowhere dense classes. For stability, we provide explicit and computable upper bounds on the *ladder index* of any first order formula on a given nowhere dense class. For uniform quasi-wideness, we give a new, purely combinatorial proof of polynomial upper bounds on *margins*, that is, functions governing this notion. We remark that the existence of upper bounds as above is known [1, 22], but the proofs are based on nonconstructive arguments, notably the compactness theorem for first order logic. Therefore, the upper bounds are given purely existentially and are not effectively computable. Contrary to these, our proofs are entirely combinatorial and effective, yielding computable upper bounds.

We now discuss the relevant background from logic and model theory, in order to motivate and state our results.

*The work of M. Pilipczuk and S. Siebertz is supported by the National Science Centre of Poland via POLONEZ grant agreement UMO-2015/19/P/ST6/03998, which has received funding from the European Union's Horizon 2020 research and innovation programme (Marie Skłodowska-Curie grant agreement No. 665778). The work of Sz. Toruńczyk is supported by the National Science Centre of Poland grant 2016/21/D/ST6/01485. M. Pilipczuk is supported by the Foundation for Polish Science (FNP) via the START stipend programme.

LICS 2018, July 9–12, 2018, Oxford, Great Britain
2018.

Model theory. Our work is inspired by ideas from model theory, more specifically, from *stability theory*. The goal of stability theory is to draw certain dividing lines specifying abstract properties of logical structures which allow the development of a good structure theory. There are many such dividing lines, depending on the specifics of the desired theory. One such dividing line encloses the class of *stable structures*, another encloses the larger class of *dependent structures* (also called *NIP*). A general theme is that the existence of a manageable structure is strongly related to the non-existence of certain forbidden patterns on one hand, and on the other hand, to bounds on cardinalities of certain *type sets*. Let us illustrate this phenomenon more concretely.

For a first order formula $\varphi(\bar{x}, \bar{y})$ with free variables split into $\bar{x}$ and $\bar{y}$, a *φ-ladder* of length n in a logical structure $\mathbb{A}$ is a sequence $\bar{u}_1, \ldots, \bar{u}_n, \bar{v}_1, \ldots, \bar{v}_n$ of tuples of elements of $\mathbb{A}$ such that

$$\mathbb{A} \models \varphi(\bar{u}_i, \bar{v}_j) \iff i \leqslant j \qquad \text{for all } 1 \leqslant i, j \leqslant n.$$

The least n for which there is no φ-ladder of length n is the *ladder index* of $\varphi(\bar{x}, \bar{y})$ in $\mathbb{A}$ (which may depend on the split of the variables, and may be ∞ for some infinite structures $\mathbb{A}$). A class of structures $\mathscr{C}$ is *stable* if the ladder index of every first order formula $\varphi(\bar{x}, \bar{y})$ over structures from $\mathscr{C}$ is bounded by a constant depending only on φ and $\mathscr{C}$. This notion can be applied to a single infinite structure $\mathbb{A}$, by considering the class consisting of $\mathbb{A}$ only. Examples of stable structures include $(\mathbb{N}, =)$, the field of complex numbers $(\mathbb{C}, +, \times, 0, 1)$, as well as any vector space V over the field of rationals, treated as a group with addition. On the other hand, $(\mathbb{Q}, \leqslant)$ and the field of reals $(\mathbb{R}, +, \times, 0, 1)$ are not stable, as they admit a linear ordering which is definable by a first order formula. Stable structures turn out to have more graspable structure than unstable ones, as they can be equipped with various notions useful for their study, such as *forking independence* (generalizing linear independence in vector spaces) and *rank* (generalizing dimension). We refer to the textbooks [32, 39] for an introduction to stability theory.

One of concepts studied in the early years of stability theory is a property of infinite graphs called *superflatness*, introduced by Podewski and Ziegler [33]. The definition of superflatness is the same as of nowhere denseness, but Podewski and Ziegler, instead of applying it to an infinite class of finite graphs, apply it to a single infinite graph. The main result of [33] is that every superflat graph is stable. As observed by Adler and Adler [1], this directly implies the following:

Theorem 1.1 ([1, 33])**.** *Every nowhere dense class of graphs is stable. Conversely, any stable class of finite graphs which is subgraph-closed is nowhere dense.*

Thus, the notion of nowhere denseness (or superflatness) coincides with stability if we restrict attention to subgraph-closed graph classes.

The proof of Adler and Adler does not yield effective or computable upper bound on the ladder index of a given formula for a given nowhere dense class of graphs, as it relies on the result of Podewski and Ziegler, which in turn invokes compactness for first order logic.

Cardinality bounds. One of the key insights provided by the work of Shelah is that stable classes can be characterized by admitting strong upper bounds on the cardinality of *Stone spaces*. For a first order formula $\varphi(\bar{x}, \bar{y})$ with free variables partitioned into *object variables* $\bar{x}$ and *parameter variables* $\bar{y}$, a logical structure $\mathbb{A}$, and a subset of its domain B, define the set of *φ-types* with parameters from B, which are realized in $\mathbb{A}$, as follows[1]:

$$S^{\varphi}(\mathbb{A}/B) = \left\{ \left\{ \bar{v} \in B^{|\bar{y}|} : \mathbb{A} \models \varphi(\bar{u}, \bar{v}) \right\} : \bar{u} \in V(\mathbb{A})^{|\bar{x}|} \right\} \subseteq \mathcal{P}(B^{|\bar{y}|}). \tag{1}$$

Here, $V(\mathbb{A})$ denotes the domain of $\mathbb{A}$ and $\mathcal{P}(X)$ denotes the powerset of X. Putting the above definition in words, every tuple $\bar{u} \in V(\mathbb{A})^{|\bar{x}|}$ defines the set of those tuples $\bar{v} \in B^{|\bar{y}|}$ for which $\varphi(\bar{u}, \bar{v})$ holds. The set $S^{\varphi}(\mathbb{A}/B)$ consists of all subsets of $B^{|\bar{y}|}$ that can be defined in this way.

Note that in principle, $S^{\varphi}(\mathbb{A}/B)$ may be equal to $\mathcal{P}(B^{|\bar{x}|})$, and therefore have very large cardinality compared to B, even for very simple formulas. The following characterization due to Shelah (cf. [38, Theorem 2.2, Chapter II]) shows that for stable classes this does not happen. A class of structures C is stable if and only if there is an infinite cardinal κ such that the following holds for all structures $\mathbb{A}$ in the elementary closure[2] of C, and all $B \subseteq V(\mathbb{A})$: *if* $|B| \leqslant \kappa$, *then* $|S^{\varphi}(\mathbb{A}/B)| \leqslant \kappa$. Therefore Shelah's result provides an upper bound on the number of types, albeit using infinite cardinals, elementary limits, and infinite parameter sets. The cardinality bound provided by the above characterization, however, does not seem to immediately translate to a result of finitary nature. As we describe below, this can be done using the notions of *VC-dimension* and *VC-density*.

VC-dimension and VC-density. The notion of VC-dimension was introduced by Vapnik and Chervonenkis [9] as a measure of complexity of set systems, or equivalently of hypergraphs, and independently by Shelah [36] under the name of dependence (the equivalence of the two notions was observed by Laskowski [24]).

Formally, VC-dimension is defined as follows. Let X be a set and let $\mathcal{F} \subseteq \mathcal{P}(X)$ be a family of subsets of X. A subset $A \subseteq X$ is *shattered by* $\mathcal{F}$ if $\{A \cap F : F \in \mathcal{F}\} = \mathcal{P}(A)$; that is, every subset of A can be obtained as the intersection of some set from $\mathcal{F}$ with A. The *VC-dimension* of $\mathcal{F}$ is the maximum size of a subset $A \subseteq X$ that is shattered by $\mathcal{F}$ (or ∞ if there is no bound on the size of shattered subsets).

For a given structure $\mathbb{A}$, parameter set $B \subseteq V(\mathbb{A})$, and formula $\varphi(\bar{x}, \bar{y})$, we may consider the family $S^{\varphi}(\mathbb{A}/B)$ of subsets of $B^{|\bar{y}|}$ defined using equation (1). The *VC-dimension* of $\varphi(\bar{x}, \bar{y})$ on $\mathbb{A}$ is the VC-dimension of the family $S^{\varphi}(\mathbb{A}/V(\mathbb{A}))$. In other words, the VC-dimension of $\varphi(\bar{x}, \bar{y})$ on $\mathbb{A}$ is the largest cardinality of a finite set I for which there exist families of tuples $(\bar{u}_J)_{J \subseteq I}$ and $(\bar{v}_i)_{i \in I}$ of elements of $\mathbb{A}$ such that

$$\mathbb{A} \models \varphi(\bar{u}_J, \bar{v}_i) \iff i \in J \qquad \text{for all } i \in I \text{ and } J \subseteq I.$$

A formula $\varphi(\bar{x}, \bar{y})$ is *dependent* on a class of structures C if there is a bound $d \in \mathbb{N}$ such that the VC-dimension of $\varphi(\bar{x}, \bar{y})$ on $\mathbb{A}$ is at most d for all $\mathbb{A} \in C$. It is immediate from the definitions that if a formula $\varphi(\bar{x}, \bar{y})$ is stable over C, then it is also dependent on C (the bound being the ladder index). A class of structures C is *dependent* if every formula $\varphi(\bar{x}, \bar{y})$ is dependent over C. In particular, every

[1]Here, $S^{\varphi}(\mathbb{A}/B)$ is the set of types which are *realized* in $\mathbb{A}$. In model theory, one usually works with the larger class of *complete types*. This distinction will not be relevant here.

[2]The elementary closure of C is the class of all structures $\mathbb{A}$ such that every first order sentence φ which holds in $\mathbb{A}$ also holds in some $\mathbb{B} \in C$. Equivalently, it is the class of models of the theory of C.

stable class is dependent, and hence, by Theorem 1.1, every nowhere dense class of graphs is dependent. Examples of infinite dependent structures (treated as singleton classes) include $(\mathbb{Q}, \leqslant)$ and the field of reals $(\mathbb{R}, \times, +, 0, 1)$.

One of the main properties of VC-dimension is that it implies polynomial upper bounds on the number of different "traces" that a set system can have on a given parameter set. This is made precise by the well-known Sauer-Shelah Lemma, stated as follows.

Theorem 1.2 (Sauer-Shelah Lemma, [9, 35, 37]). *For any family $\mathcal{F}$ of subsets of a set X, if the VC-dimension of $\mathcal{F}$ is d, then for every finite $A \subseteq X$,*

$$|\{A \cap F \ : \ F \in \mathcal{F}\}| \leqslant \sum_{i \geqslant 0}^{d} \binom{|A|}{i} \leqslant |A|^d + 1.$$

In particular, this implies that in a dependent class of structures C, for every formula $\varphi(\bar{x}, \bar{y})$ there exists some constant $d \in \mathbb{N}$ such that

$$|S^{\varphi}(\mathbb{A}/B)| \leqslant |B|^d + 1, \tag{2}$$

for all $\mathbb{A} \in C$ and finite $B \subseteq V(\mathbb{A})$. Unlike Shelah's characterization theorem of stable classes, this result is of finitary nature: it provides quantitative upper bounds on the number of different definable subsets of a given finite parameter set. Together with Theorem 1.1, this implies that for every nowhere dense class of graphs and every first order formula $\varphi(\bar{x}, \bar{y})$, there exists a constant $d \in \mathbb{N}$ such that (2) holds.

For many structure classes C the combination of bounding VC-dimension and applying the Sauer-Shelah Lemma yields a suboptimal upper bound of the form (2). This motivates the notion of *VC-density*, a notion closely related to VC-dimension. The *VC-density* (also called the *VC-exponent*) of a set system $\mathcal{F}$ on an infinite set X is the infimum of all reals $\alpha > 0$ such that $|\{A \cap F \ : \ F \in \mathcal{F}\}| \in O(|A|^{\alpha})$, for all finite $A \subseteq X$ (where the constants hidden in the O notation may depend on α). Similarly, the VC-density of a formula $\varphi(\bar{x}, \bar{y})$ over a class of structures C is the infimum of all reals $\alpha > 0$ such that $|S^{\varphi}(\mathbb{A}/B)| \in O(|B|^{\alpha})$, for all $\mathbb{A} \in C$ and all finite $B \subseteq V(\mathbb{A})$. The Sauer-Shelah Lemma implies that the VC-density (of a set system, or of a formula over a class of structures) is bounded from above by the VC-dimension. Our main result, Theorem 1.3 below, provides optimal bounds on the VC-density in the setting of nowhere dense graph classes. In particular, we show that the VC-density of a formula $\varphi(\bar{x}, \bar{y})$ over a nowhere dense class C is bounded by $|\bar{x}|$.

Motivation. The motivation for finding bounds on the VC-density comes from the fact that it is this quantity, rather than VC-dimension, that is actually relevant in combinatorial and algorithmic applications [6, 8, 9, 26, 27]. For example in the framework of *probably approximately correct learning* (PAC learning, introduced by Valiant in [40]), the size of the random sample required as the training set is determined by the VC-density of the concept class rather than by its VC-dimension (see [6], Lemma 7). We refer to the work of Grohe and Turán in [17] and Adler and Adler [1] for more background on learning definable concepts. In Theorem 1.6 below we give another example where bounds on the VC-density yield bounds on certain combinatorial quantities. We refer to [4] for an overview of further applications of VC-dimension and VC-density in model theory and to the surveys [14, 26] on uses of VC-density in combinatorics.

The main result. Our main result, Theorem 1.3 stated below, improves the bound (2) for classes of sparse graphs by providing essentially the optimum exponent.

Theorem 1.3. *Let $\mathscr{C}$ be a nowhere dense class of graphs and let $\varphi(\bar{x}, \bar{y})$ be a first order formula with free variables partitioned into object variables $\bar{x}$ and parameter variables $\bar{y}$. Let $\ell = |\bar{x}|$. Then for every $\varepsilon > 0$ there exists a constant c such that for every $G \in \mathscr{C}$ and every nonempty $A \subseteq V(G)$, we have $|S^{\varphi}(G/A)| \leqslant c \cdot |A|^{\ell+\varepsilon}$.*

In particular, Theorem 1.3 implies that the VC-density of any fixed formula $\varphi(\bar{x}, \bar{y})$ over any nowhere dense class of graphs is $|\bar{x}|$, the number of object variables in φ.

To see that the bounds provided by Theorem 1.3 cannot be improved, consider a formula $\varphi(\bar{x}, y)$ (i.e. with one parameter variable) expressing that y is equal to one of the entries of $\bar{x}$. Then for each graph G and parameter set A, $S^{\varphi}(G/A)$ consists of all subsets of A of size at most $|\bar{x}|$, whose number is $\Theta(|A|^{|\bar{x}|})$. Note that this lower bound applies to any infinite class of graphs, even edgeless ones.

We moreover show that, as long as we consider only subgraph-closed graph classes, the result of Theorem 1.3 also cannot be improved in terms of generality. The following result is an easy corollary of known characterizations of obstructions to being nowhere dense.

Theorem 1.4. *Let $\mathscr{C}$ be a class of graphs which is closed under taking subgraphs. If $\mathscr{C}$ is not nowhere dense, then there is a formula $\varphi(x, y)$ such that for every $n \in \mathbb{N}$ there are $G \in \mathscr{C}$ and $A \subseteq V(G)$ with $|A| = n$ and $|S^{\varphi}(G/A)| = 2^{|A|}$.*

A similar characterization theorem can be proved for bounded expansion classes of graphs. We show that if $\mathscr{C}$ has bounded expansion, then there exists a constant c such that for every $G \in \mathscr{C}$ and every nonempty $A \subseteq V(G)$, we have $|S^{\varphi}(G/A)| \leqslant c \cdot |A|^{\ell}$. Conversely, if $\mathscr{C}$ has unbounded expansion, then there is a formula $\varphi(x, y)$ such that for every $c \in \mathbb{R}$ there exist $G \in \mathscr{C}$ and a nonempty $A \subseteq V(G)$ with $|S^{\varphi}(G/A)| > c|A|$.

Neighborhood complexity. To illustrate Theorem 1.3, consider the case when G is a graph and $\varphi(x, y)$ is the formula with two variables x and y expressing that the distance between x and y is at most r, for some fixed integer r. In this case, $S^{\varphi}(G/A)$ is the family consisting of all intersections $U \cap A$, for U ranging over all balls of radius r in G, and $|S^{\varphi}(G/A)|$ is called the *r-neighborhood complexity* of A. The concept of r-neighborhood complexity in sparse graph classes has already been studied before. In particular, it was proved by Reidl et al. [34] that in any graph class of bounded expansion, the r-neighborhood complexity of any set of vertices A is $O(|A|)$. Recently, Eickmeyer et al. [13] generalized this result to an upper bound of $O(|A|^{1+\varepsilon})$ in any nowhere dense class of graphs. Note that these results are special cases of Theorem 1.3. The study of r-neighborhood complexity on classes of bounded expansion and nowhere dense classes was motivated by algorithmic questions from the field of parameterized complexity. More precisely, the usage of this notion was crucial for the development of a linear kernel for the r-Dominating Set problem on any class of bounded expansion [11], and of an almost linear kernel for this problem on any nowhere dense class [13]. We will use the results of [11, 13] on r-neighborhood complexity in sparse graphs in our proof of Theorem 1.3.

Uniform quasi-wideness. One of the main tools used in our proof is the notion of *uniform quasi-wideness*, introduced by Dawar [10] in the context of homomorphism preservation theorems. Formally, a class of graphs $\mathscr{C}$ is *uniformly quasi-wide* if for each integer $r \in \mathbb{N}$ there is a function $N\colon \mathbb{N} \to \mathbb{N}$ and a constant $s \in \mathbb{N}$ such that for every $m \in \mathbb{N}$, graph $G \in \mathscr{C}$, and vertex subset $A \subseteq V(G)$ of size $|A| \geqslant N(m)$, there is a set $S \subseteq V(G)$ of size $|S| \leqslant s$ and a set $B \subseteq A \setminus S$ of size $|B| \geqslant m$ which is r-independent in $G - S$. Recall that a set $B \subseteq V(G)$ is *r-independent* in G if all distinct $u, v \in B$ are at distance larger than r in G.

Nešetřil and Ossona de Mendez proved that the notions of uniform quasi-wideness and nowhere denseness coincide for classes of finite graphs [29]. The proof of Nešetřil and Ossona de Mendez goes back to a construction of Kreidler and Seese [21] (see also Atserias et al. [5]), and uses iterated Ramsey arguments. Hence the original bounds on the function N_r are non-elementary. Recently, Kreutzer, Rabinovich and the second author proved that for each radius r, we may always choose the function N_r to be a polynomial [22]. However, the exact dependence of the degree of the polynomial on r and on the class $\mathscr{C}$ itself was not specified in [22], as the existence of a polynomial bound is derived from non-constructive arguments used by Adler and Adler in [1] when showing that every nowhere dense class of graphs is stable. We remark that polynomial bounds for uniform quasi-wideness are essential for some of its applications: the fact that N_r can be chosen to be polynomial was crucially used by Eickmeyer et al. [13] both to establish an almost linear upper bound on the r-neighborhood complexity in nowhere dense classes, and to develop an almost linear kernel for the r-Dominating Set problem. We use this in our proof of Theorem 1.3 as well.

In our quest for constructive arguments, we give a new construction giving polynomial bounds for uniform quasi-wideness. The new proof is considerably simpler than that of [22] and gives explicit and computable bounds on the degree of the polynomial. More precisely, we prove the following theorem; here, the notation $O_{r,t}(\cdot)$ hides computable factors depending on r and t. Below, we write $K_t \not\preccurlyeq_r G$ to denote that G does not contain K_t as a depth-r minor, i.e., as a subgraph of a graph obtained from G by contracting mutually disjoint subgraphs of radius at most r to single vertices.

Theorem 1.5. *For all $r, t \in \mathbb{N}$ there is a polynomial N with $N(m) = O_{r,t}(m^{(4t+1)^{2rt}})$, such that the following holds. Let G be a graph such that $K_t \not\preccurlyeq_{\lfloor 9r/2 \rfloor} G$, and let $A \subseteq V(G)$ be a vertex subset of size at least $N(m)$, for a given m. Then there exists a set $S \subseteq V(G)$ of size $|S| < t$ and a set $B \subseteq A \setminus S$ of size $|B| \geqslant m$ which is r-independent in $G - S$. Moreover, given G and A, such sets S and B can be computed in time $O_{r,t}(|A| \cdot |E(G)|)$.*

We remark that even though the techniques employed to prove Theorem 1.5 are inspired by methods from stability theory, at the end we conduct an elementary graph theoretic reasoning. In particular, as asserted in the statement, the proof can be turned into an efficient algorithm.

We also prove a result extending Theorem 1.5 to the case where $A \subseteq V(G)^d$ is a set of *tuples* of vertices, of any fixed length d. This result is essentially an adaptation of an analogous result due to Podewski and Ziegler [33] in the infinite case, but appears to be new in the context of finite structures. This more general result turns out to be necessary in the proof of Theorem 1.3.

Local separation. A simple, albeit important notion which permeates our proofs is a graph theoretic concept of *local separation*.

Let G be a graph, $S \subseteq V(G)$ a set of vertices, and let $r \in \mathbb{N}$ be a number. We say that two sets of vertices A and B are *r-separated* by S (in G) if every path from a vertex in A to a vertex in B of length at most r contains a vertex from S. Observe that taking $r = \infty$ in r-separation yields the familiar notion of a separation in graph theory. From the perspective of stability, separation (for $r = \infty$) characterizes *forking independence* in superflat graphs [20]. Therefore, r-separation can be thought of as a local analogue of forking independence, for nowhere dense graph classes.

A key lemma concerning r-separation (cf. Corollary 3.2) states that if A and B are r-separated by a set S of size s in G, then for any fixed formula $\varphi(\bar{x}, \bar{y})$ of quantifier rank $O(\log r)$, the set $\{\{\bar{v} \in B^{|\bar{y}|} : G \models \varphi(\bar{u}, \bar{v})\} : \bar{u} \in A^{|\bar{x}|}\}$ has cardinality bounded by a constant depending on s and φ only (and not on G, A, and B). This elementary result combines Gaifman's locality of first order logic (cf. [15]) and a Feferman-Vaught compositionality argument. This, in combination with the polynomial bounds for uniform quasi-wideness (Theorem 1.5, and its extension to tuples Theorem 2.9), as well as the previous results on neighborhood complexity [11, 13], are the main ingredients of our main result, Theorem 1.3.

A duality theorem. As an example application of our main result, Theorem 1.3, we state the following result. Below, $\tau(\mathcal{G})$ denotes the *transversality* of $\mathcal{G}$, i.e., the least number of elements of a set X which intersects every set in $\mathcal{G}$, and $\nu(\mathcal{G})$ denotes the *packing number* of $\mathcal{G}$, i.e., the largest number of pairwise-disjoint subsets of $\mathcal{G}$.

Theorem 1.6. *Fix a nowhere dense class of graphs $\mathscr{C}$ and a formula $\varphi(x, y)$ with two free variables x, y. Then there is a function $f\colon \mathbb{N} \to \mathbb{N}$ with the following property. Let $G \in \mathscr{C}$ be a graph and let $\mathcal{G}$ be a family of subsets of $V(G)$ consisting of sets of the form $\{v \in V(G) : \varphi(u, v)\}$, where u is some vertex of $V(G)$. Then $\tau(\mathcal{G}) \leqslant f(\nu(\mathcal{G}))$.*

Theorem 1.6 is an immediate consequence of the bound given by Theorem 1.3 and a result of Matoušek [27]. We remark that a similar, but incomparable result is proved by Bousquet and Thomassé [7]. In their result, the assumption on $\mathscr{C}$ is weaker, since they just require that it has *bounded distance VC-dimension*, but the assumption on $\mathcal{G}$ is stronger, as it is required to be the set of all balls of a fixed radius.

Stability. Finally, we observe that we can apply our tools to give a constructive proof of the result of Adler and Adler [1] that every nowhere dense class is stable, which yields computable upper bounds on ladder indices. More precisely, we translate the approach of Podewski and Ziegler [33] to the finite and replace the key non-constructive application of compactness with a combinatorial argument based on Gaifman's locality, in the flavor served by our observations on r-separation (Corollary 3.2). The following theorem summarizes our result.

Theorem 1.7. *There are computable functions $f\colon \mathbb{N}^3 \to \mathbb{N}$ and $g\colon \mathbb{N} \to \mathbb{N}$ with the following property. Suppose $\varphi(\bar{x}, \bar{y})$ is a formula of quantifier rank at most q and with d free variables. Suppose further that G is a graph excluding K_t as a depth-$g(q)$ minor. Then the ladder index of $\varphi(\bar{x}, \bar{y})$ in G is at most $f(q, d, t)$.*

Acknowledgments. We would like to thank Patrice Ossona de Mendez for pointing us to the question of studying VC-density of nowhere dense graph classes.

Due to space constraints not all proofs can be presented in the conference version of the paper, missing proofs are marked with ($\star$) The full version of the paper is available online [23].

2 Bounds for uniform quasi-wideness

In this section we prove Theorem 1.5, which strengthens the result of Kreutzer et al. [22] by providing an explicit polynomial N and bound s. We remark that the result of Kreutzer et al. is sufficient to prove our main result, Theorem 1.3. Our proof follows the same lines as the original proof of Nešetřil and Ossona de Mendez [29], with the difference that in the key technical lemma (Lemma 3.5 of [29]) we improve the bounds significantly by replacing a Ramsey argument with a refined combinatorial analysis (Lemma 2.1 below). The new argument essentially originates in the concept of *branching index* from stability theory. Due to space constraints we present only the proof of the key lemma and refer to the appended full version for a complete proof of the main theorem.

Lemma 2.1. *Let $\ell, m, t \in \mathbb{N}$ and assume $\ell \geqslant t^8$. If G is a graph and A is a 1-independent set in G with at least $(m+\ell)^{2t}$ elements, then at least one of the following conditions hold:*

- *$K_t \preccurlyeq_4 G$,*
- *A contains a 2-independent set of size m,*
- *some vertex v of G has at least $\ell^{1/4}$ neighbors in A.*

Moreover, if $K_t \not\preccurlyeq_4 G$, the structures described in the other two cases (a 2-independent set of size m, or a vertex v as above) can be computed in time $O_t(|A| \cdot |E(G)|)$.

The remainder of this section is devoted to the proof of Lemma 2.1. We will use the following bounds on the edge density of graphs with excluded shallow minors obtained by Alon et al. [3].

Lemma 2.2 (Theorem 2.2 in [3])**.** *Let H be a bipartite graph with maximum degree d on one side. Then there exists a constant c_H, depending only on H, such that every n-vertex graph G excluding H as a subgraph has at most $c_H \cdot n^{2-1/d}$ edges.*

Observe that if $K_t \not\preccurlyeq_1 G$, then in particular the 1-subdivision of K_t is excluded as a subgraph of G (the 1-subdivision of a graph H is obtained by replacing every edge of H by a path of length 2). Moreover, the 1-subdivision of K_t is a bipartite graph with maximum degree 2 on one side. Furthermore, it is easy to check in the proof of Theorem 2.2 in [3] that $c_H \leqslant |V(H)|$ in case $d = 2$. Since the 1-subdivision of K_t has $\binom{t+1}{2}$ vertices, we can choose $c_{K_t} = \binom{t+1}{2}$ and conclude the following.

Corollary 2.3. *Let G be an n-vertex graph such that $K_t \not\preccurlyeq_1 G$ for some constant $t \in \mathbb{N}$. Then G has at most $\binom{t+1}{2} \cdot n^{3/2}$ edges.*

We will use the following standard lemma saying that a shallow minor of a shallow minor is a shallow minor, where the parameters of shallowness are appropriately chosen.

Lemma 2.4 (adaptation of Proposition 4.11 in [31])**.** *Suppose J, H, G are graphs such that $H \preccurlyeq_a G$ and $J \preccurlyeq_b H$, for some $a, b \in \mathbb{N}$. Then $J \preccurlyeq_c G$, where $c = 2ab + a + b$.*

We will need one more technical lemma.

Lemma 2.5. *Let G be a graph such that $K_t \not\preccurlyeq_4 G$ for some $t \in \mathbb{N}$ and let $A \subseteq V(G)$ with $|A| \geqslant t^8$. Assume furthermore that every pair of elements of A has a common neighbor in $V(G) \setminus A$. Then there exists a vertex v in $V(G) \setminus A$ which has at least $|A|^{1/4}$ neighbors in A.*

Proof. Denote $k = \max\{\, |N(w) \cap A| \;:\; w \in V(G) - A \,\}$; our goal is to prove that $k \geqslant |A|^{1/4}$.

Let $B \subseteq V(G) - A$ be the set of those vertices outside of A that have a neighbor in A. Construct a function $f : B \to A$ by a random procedure as follows: for each vertex $v \in B$, choose $f(v)$ uniformly and independently at random from the set $N(v) \cap A$. Next, for each $u \in A$ define branch set $I_u = G[\{u\} \cup f^{-1}(u)]$. Observe that since, by construction, v and $f(v)$ are adjacent for all $v \in B$, each branch set I_u has radius at most 1, with u being the central vertex. Also, the branch sets $\{I_u\}_{u \in A}$ are pairwise disjoint. Finally, construct a graph H on vertex set A by making distinct $u, v \in A$ adjacent in H whenever there is an edge in G between I_u and I_v. Then the branch sets $\{I_u\}_{u \in A}$ witness that H is a 1-shallow minor of G.

For distinct $u, v \in A$, let us estimate the probability that the edge uv appears in H. By assumption, there is a vertex $w \in B$ that is adjacent both to u and to v. Observe that if $f(w) = u$ or $f(w) = v$, then uv for sure becomes an edge in H. Since w has at most k neighbors in A, the probability that $f(w) \in \{u, v\}$ is at least $\frac{2}{k}$.

By the linearity of expectation, the expected number of edges in H is at least $\binom{|A|}{2} \cdot \frac{2}{k} = \frac{|A|(|A|-1)}{k}$. Hence, for at least one run of the random experiment we have that H indeed has at least this many edges. On the other hand, observe that $K_t \not\preccurlyeq_1 H$; indeed, since $H \preccurlyeq_1 G$, by Lemma 2.4 we infer that $K_t \preccurlyeq_1 H$ would imply $K_t \preccurlyeq_4 G$, a contradiction with the assumptions on G. Then Corollary 2.3 implies H has at most $\binom{t+1}{2} \cdot |A|^{3/2}$ edges. Observe that $\binom{t+1}{2} \cdot |A|^{3/2} \leqslant 3t^2/4 \cdot |A|^{3/2} \leqslant \frac{3}{4}|A|^{7/4}$, where the first inequality holds due to $t \geqslant 2$, while the second holds by the assumption that $|A| \geqslant t^8$. By combining the above bounds, we obtain

$$\frac{|A|(|A|-1)}{k} \leqslant \frac{3}{4}|A|^{7/4},$$

which implies $k \geqslant |A|^{1/4}$ due to $|A| \geqslant t^8 \geqslant 64$. $\square$

We proceed with the proof of Lemma 2.1. The idea is to arrange the elements of A in a binary tree and prove that provided A is large, this tree contains a long path. From this path, we will extract the set B. In stability theory, similar trees are called *type trees* and they are used to extract long indiscernible sequences, see e.g. [25].

We will work with a two-symbol alphabet $\{\mathrm{D}, \mathrm{S}\}$, for *daughter* and *son*. We identify words in $\{\mathrm{D}, \mathrm{S}\}^*$ with *nodes* of the infinite rooted binary tree. The *depth* of a node w is the length of w. For $w \in \{\mathrm{D}, \mathrm{S}\}^*$, the nodes $w\mathrm{D}$ and $w\mathrm{S}$ are called, respectively, the *daughter* and the *son* of w, and w is the *parent* of both $w\mathrm{S}$ and $w\mathrm{D}$. A node w' is a *descendant* of a node w if w' is a prefix of w (possibly $w' = w$). We consider finite, labeled, rooted, binary trees, which are called simply trees below, and are defined as follows. For a set of labels U, a (U-labeled) *tree* is a partial function $\tau\colon \{\mathrm{D}, \mathrm{S}\}^* \to U$ whose domain is a finite set of nodes, called the *nodes of* τ, which is closed under taking parents. If v is a node of τ, then $\tau(v)$ its *label.*

Let G be a graph, $A \subseteq V(G)$ be a 1-independent set in G, and $\bar{a}$ be any enumeration of A, that is, a sequence of length $|A|$ in which every element of A appears exactly once. We define a binary tree τ which is labeled by vertices of G. The tree is defined by processing all elements of $\bar{a}$ sequentially.

We start with τ being the tree with empty domain, and for each element a of the sequence $\bar{a}$, processed in the order given by $\bar{a}$, execute the following procedure which results in adding a node with label a to τ. When processing the vertex a, do the following. Start with w being the empty word. While w is a node of τ, repeat the following step: if the distance from a to $\tau(w)$ in the graph G is at most 2, replace w by its son, otherwise, replace w by its daughter. Once w is not a node of τ, extend τ by setting $\tau(w) = a$. In this way, we have processed the element a, and now proceed to the next element of $\bar{a}$, until all elements are processed. This ends the construction of τ. Thus, τ is a tree labeled with vertices of A, and every vertex of A appears exactly once in τ.

Define the *depth* of τ as the maximal depth of a node of τ. For a word w, an *alternation* in w is any position α, $1 \leqslant \alpha \leqslant |w|$, such that $w_\alpha \neq w_{\alpha-1}$; here, w_α denotes the αth symbol of w, and w_0 is assumed to be D. The *alternation rank* of the tree τ is the maximum of the number of alternations in w, over all nodes w of τ.

Lemma 2.6. *Let $h, t \geqslant 2$. If τ has alternation rank at most $2t - 1$ and depth at most $h - 1$, then τ has fewer than h^{2t} nodes.*

Proof. With each node w of τ associate function $f_w \colon \{1, \dots, 2t\} \to \{1, \dots, h\}$ defined as follows: f_w maps each $i \in \{1, \dots, 2t\}$ to the ith alternation of w, provided i is at most the number of alternations of w, and otherwise we put $f_w(i) = |w| + 1$. It is clear that the mapping $w \mapsto f_w$ for nodes w of τ is injective and its image is contained in monotone functions from $\{1, \dots, 2t\}$ to $\{1, \dots, h\}$, whose number is less than h^{2t}. Hence, of τ has fewer than h^{2t} elements. □

Lemma 2.7. *Suppose that $K_t \not\preccurlyeq_2 G$. Then τ has alternation rank at most $2t - 1$.*

Proof. Let w be a node of τ with at least $2k$ alternations, for some $k \in \mathbb{N}$. Suppose $\alpha_1, \beta_1, \dots, \alpha_k, \beta_k$ be the first $2k$ alternations of w. By the assumption that $w_0 = \mathrm{D}$ we have that w contains symbol S at all positions α_i for $i = 1, \dots, k$, and symbol D at all positions β_i for $i = 1, \dots, k$. For each $i \in \{1, \dots, k\}$, define $a_i \in V(G)$ to be the label in τ of the prefix of w of length $\alpha_i - 1$, and similarly define $b_i \in V(G)$ to be the label in τ of the prefix of w of length $\beta_i - 1$. It follows that for each $i \in \{1, \dots, k\}$, the following assertions hold: the nodes in τ with labels $b_i, a_{i+1}, b_{i+1}, \dots, a_k, b_k$ are descendants of the son of the node with label a_i, and the nodes with labels $a_{i+1}, b_{i+1}, \dots, a_k, b_k$ are descendants of the daughter of the node with label b_i.

Claim 1. For every pair a_i, b_j with $1 \leqslant i \leqslant j \leqslant k$, there is a vertex $z_{ij} \notin A$ which is a common neighbor of a_i and b_j, and is not a neighbor of any b_s with $s \neq j$.

Proof. Note that since $i \leqslant j$, the node with label b_j is a descendant of the son of the node with label a_i, hence we have $\mathrm{dist}_G(a_i, b_j) \leqslant 2$ by the construction of τ. However, we also have $\mathrm{dist}_G(a_i, b_j) > 1$ since A is 1-independent. Therefore $\mathrm{dist}_G(a_i, b_j) = 2$, so there is a vertex z_{ij} which is a common neighbor of a_i and b_j. Suppose that z_{ij} was a neighbor of b_s, for some $s \neq j$. This would imply that $\mathrm{dist}_G(b_j, b_s) \leqslant 2$, which is impossible, because the nodes with labels b_s and b_j in τ are such that one is a descendant of the daughter of the other, implying that $\mathrm{dist}_G(b_s, b_j) > 2$. ⌟

Note that whenever $i \leqslant j$ and $i' \leqslant j'$ are such that $j \neq j'$, the vertices z_{ij} and $z_{i'j'}$ are different, because z_{ij} is adjacent to b_j but not to $b_{j'}$, and the converse holds for $z_{i'j'}$. However, it may happen that $z_{ij} = z_{i'j}$ even if $i \neq i'$. This will not affect our further reasoning.

For each $j \in \{1, \dots, k\}$, let B_j be the subgraph of G induced by the set $\{a_j, b_j\} \cup \{z_{ij} \colon 1 \leqslant i \leqslant j\}$. Observe that B_j is connected and has radius at most 2, with b_j being the central vertex. By Claim 1 and the discussion from the previous paragraph, the graphs B_j for $j \in \{1, \dots, k\}$ are pairwise disjoint. Moreover, for all $1 \leqslant i \leqslant j \leqslant k$, there is an edge between B_i and B_j, namely, the edge between $z_{ij} \in B_j$ and $a_i \in B_i$. Hence, the graphs B_j, for $j \in \{1, \dots, k\}$, define a depth-2 minor model of K_k in G. Since $K_t \not\preccurlyeq_2 G$, this implies that $k < t$, proving Lemma 2.7. □

We continue with the proof of Lemma 2.1. Fix integers $\ell \geqslant t^8$ and m, and define $h = m + \ell$. Let A be a 1-independent set in G of size at least h^{2t}. Suppose that the first case of Lemma 2.1 does not hold. In particular $K_t \not\preccurlyeq_2 G$, so by Lemma 2.7, τ has alternation rank at most $2t - 1$. From Lemma 2.6 we conclude that τ has depth at least h. As $h = m + \ell$, it follows that either τ has a node w which contains at least m letters D, or τ has a node w which contains at least ℓ letters S.

Consider the first case, i.e., there is a node w of τ which contains at least m letters D, and let X be the set of all vertices $\tau(u)$ such that $u\mathrm{D}$ is a prefix of w. Then, by construction, X is a 2-independent set in G of size at least m, so the second case of the lemma holds.

Finally, consider the second case, i.e., there is a node w in τ which contains at least ℓ letters S. Let Y be the set of all vertices $\tau(u)$ such that $u\mathrm{S}$ is a prefix of w. Then, by construction, $Y \subseteq A$ is a set of at least ℓ vertices which are mutually at distance exactly 2 in G. Since $K_t \not\preccurlyeq_4 G$ and $\ell \geqslant t^8$, by Lemma 2.5 we infer that there is a vertex $v \in G$ with at least $\ell^{1/4}$ neighbors in Y. This finishes the proof of the existential part of Lemma 2.1. We defer the proof of the algorithmic statement of the lemma to the full version.

2.1 Uniform quasi-widness for tuples

We now formulate and prove an extension of Theorem 1.5 which applies to sets of tuples of vertices, rather than sets of vertices. This more general result will be used later on in the paper. The result and its proof are essentially adaptations to the finite of their infinite analogues introduced by Podewski and Ziegler (cf. [33], Corollary 3), modulo the numerical bounds.

Fix a graph G and a number $r \in \mathbb{N}$, and let $S \subseteq V(G)$ be a subset of vertices of G. We say that vertices u and v are *r-separated* by S in G if every path of length at most r connecting u and v in G passes through a vertex of S. We extend this notion to tuples: two tuples $\bar{u}, \bar{v}$ of vertices of G are *r-separated* by S every vertex appearing in $\bar{u}$ is r-separated by S from every vertex appearing in $\bar{v}$. Finally, if $A \subseteq V(G)^d$ is a set of d-tuples of vertices, for some $d \in \mathbb{N}$, then we say that A is *mutually r-separated* by S in G if any two distinct $\bar{u}, \bar{v} \in A$ are r-separated by S in G.

With these definitions set, we may introduce the notion of uniform quasi-wideness for tuples.

Definition 2.8. Fix a class C and numbers $r, d \in \mathbb{N}$. For a function $N \colon \mathbb{N} \to \mathbb{N}$ and number $s \in \mathbb{N}$, we say that C satisfies property $\mathrm{UQW}^d_r(N, s)$ if the following condition holds:

> *for every $m \in \mathbb{N}$ and every subset $A \subseteq V(G)^d$ with $|A| \geqslant N(m)$, there is a set $S \subseteq V(G)$ with $|S| \leqslant s$ and a subset $B \subseteq A$ with $|B| \geqslant m$ which is mutually r-separated by S in G.*

We say that C satisfies property UQW_r^d if C satisfies property $\mathrm{UQW}_r^d(N,s)$ for some $N\colon \mathbb{N} \to \mathbb{N}$ and $s \in \mathbb{N}$. If moreover one can take N to be a polynomial, then we say that C satisfies property PUQW_r^d.

The following result provides a generalization of Theorem 1.5 to higher dimensions.

Theorem 2.9. *If C is a nowhere dense class of graphs, then for all $r, d \in \mathbb{N}$, the class C satisfies PUQW_r^d. More precisely, for any class of graphs C and numbers $r, t \in \mathbb{N}$, if $K_t \not\preccurlyeq_{18r} G$ for all $G \in C$, then for all $d \in \mathbb{N}$ the class C satisfies $\mathrm{UQW}_r^d(N_r^d, s_r^d,)$ for some number $s_r^d \in \mathbb{N}$ and polynomial $N_r^d\colon \mathbb{N} \to \mathbb{N}$ that can be computed given r, t, and d.*

Theorem 2.9 is an immediate consequence of Theorem 1.5 and of the following result.

Lemma 2.10. *For all $r, d \in \mathbb{N}$, if C satisfies $\mathrm{UQW}_{2r}(N_{2r}, s_{2r})$ for some $s_{2r} \in \mathbb{N}$ and $N_{2r}\colon \mathbb{N} \to \mathbb{N}$, then C satisfies $\mathrm{UQW}_r^d(N_r^d, s_r^d)$ for $s_r^d = d \cdot s_{2r}$ and function $N_r^d\colon \mathbb{N} \to \mathbb{N}$ defined as $N_r^d(m) = f^d((d^2+1)\cdot m)$, where $f(m') = m' \cdot N_{2r}(m')$ and f^d is the d-fold composition of f with itself.*

The rest of Section 2.1 is devoted to the proof of Lemma 2.10. Fix a class C such that $\mathrm{UQW}_{2r}(N_{2r}, s_{2r})$ holds for some number $s_{2r} \in \mathbb{N}$ and function $N_{2r}\colon \mathbb{N} \to \mathbb{N}$. We also fix the function f defined in the statement of Lemma 2.10.

Let us fix dimension $d \in \mathbb{N}$, radius $r \in \mathbb{N}$, and graph $G \in C$. For a coordinate $i \in \{1, \ldots, d\}$, by $\pi_i\colon V(G)^d \to V(G)$ we denote the *projection* onto the ith coordinate; that is, for $\bar{x} \in V(G)^d$ by $\pi_i(\bar{x})$ we denote the ith coordinate of $\bar{x}$.

Our first goal is to find a large subset of tuples that are mutually $2r$-separated by some small S on each coordinate separately. Note that in the following statement we ask for $2r$-separation, instead of r-separation.

Lemma 2.11. *For all $r, m \in \mathbb{N}$ and $A \subseteq V(G)^d$ with $|A| \geqslant f^d(m)$, there is a set $B \subseteq A$ with $|B| \geqslant m$ and a set $S \subseteq V(G)$ with $|S| \leqslant d \cdot s_{2r}$ such that for each coordinate $i \in \{1, \ldots, d\}$ and all distinct $\bar{x}, \bar{y} \in B$, the vertices $\pi_i(\bar{x})$ and $\pi_i(\bar{y})$ are $2r$-separated by S.*

Proof. We will iteratively apply the following claim.

Claim 2. Fix a coordinate $i \in \{1, \ldots, d\}$, an integer $m' \in \mathbb{N}$, and a set $A' \subseteq V(G)^d$ with $|A'| \geqslant f(m')$. Then there is a set $B' \subseteq A'$ with $|B'| \geqslant m'$ and a set $S' \subseteq V(G)$ with $|S'| \leqslant s_{2r}$, such that for all distinct $\bar{x}, \bar{y} \in B$, the vertices $\pi_i(\bar{x})$ and $\pi_i(\bar{y})$ are $2r$-separated by S.

Proof. We consider two cases, depending on whether $|\pi_i(A')| \geqslant N_{2r}(m')$.

Suppose first that $\pi_i(A')$ contains at least $N_{2r}(m')$ distinct vertices. Then we may apply the property UQW_{2r} to $\pi_i(A')$, yielding sets $S' \subseteq V(G)$ and $X \subseteq \pi_i(A')$ such that $|X| \geqslant m'$, $|S'| \leqslant s_{2r}$, and X is mutually $2r$-separated by S' in G. Let $B' \subseteq A'$ be a subset of tuples constructed as follows: for each $u \in X$, include in B' one arbitrarily chosen tuple $\bar{x} \in A'$ such that the ith coordinate of $\bar{x}$ is u. Clearly $|B'| = |X| \geqslant m'$ and for all distinct $\bar{x}, \bar{y} \in B'$, we have that $\pi_i(\bar{x})$ and $\pi_i(\bar{y})$ are different and $2r$-separated by S' in G; this is because X is mutually $2r$-separated by S' in G. Hence B' and S' satisfy all the required properties.

Suppose now that $|\pi_i(A')| < N_{2r}(m')$. Then choose a vertex $a \in \pi_i(A')$ for which the pre-image $\pi_i^{-1}(a)$ has the largest cardinality. Since $|A'| \geqslant f(m') = m' \cdot N_{2r}(m')$, we have that

$$|\pi_i^{-1}(a)| \geqslant \frac{|A'|}{|\pi_i(A')|} \geqslant \frac{m' \cdot N_{2r}(m')}{N_{2r}(m')} = m'.$$

Hence, provided we set $S' = \{a\}$ and $B' = \pi_i^{-1}(a)$, we have that B' is mutually $2r$-separated by S', $|B'| \geqslant m$, and $|S'| = 1$.

We proceed with the proof of Theorem 2.11. Let $A \subseteq V(G)^d$ be such that $|A| \geqslant f^d(m)$. We inductively define subsets $B_0 \supseteq B_1 \supseteq \ldots \supseteq B_d$ of A and sets $S_1, \ldots, S_d \subseteq V(G)$ as follows. First put $B_0 = A$. Then, for each $i = 1, \ldots, d$, let B_i and S_i be the B' and S' obtained from Claim 2 applied to the set of tuples $B_{i-1} \subseteq V(G)^d$, the coordinate i, and $m' = f^{d-i}(m)$. It is straightforward to see that the following invariant holds for each $i \in \{1, \ldots, d\}$: $|B_i| \geqslant f^{d-i}(m)$ and for all $j \leqslant i$ and distinct $\bar{x}, \bar{y} \in B_i$, the vertices $\pi_j(\bar{x})$ and $\pi_j(\bar{y})$ are $2r$-separated by $S_1 \cup \ldots \cup S_i$ in G. In particular, by taking $B = B_d$ and $S = S_1 \cup \ldots \cup S_d$, we obtain that $|B| \geqslant m$, $|S| \leqslant d \cdot s_{2r}$, and B and S satisfy the condition requested in the lemma statement. □

The next lemma will be used to turn mutual $2r$-separation on each coordinate to mutual r-separation of the whole tuple set.

Lemma 2.12. *Let $B \subseteq V(G)^d$ and $S \subseteq V(G)$ be such that for each $i \in \{1, \ldots, d\}$ and all distinct $\bar{x}, \bar{y} \in B$, the vertices $\pi_i(\bar{x})$ and $\pi_i(\bar{y})$ are $2r$-separated by S in G. Then there is a set C with $C \subseteq B$ and $|C| \geqslant \frac{|B|}{d^2+1}$ such that C is mutually r-separated by S in G.*

Proof. Let C be a maximal subset of B that is mutually r-separated by S in G. By the maximality of C, with each tuple $\bar{a} \in B - C$ we may associate a tuple $\bar{b} \in C$ and a pair of indices $(i, j) \in \{1, \ldots, d\}^2$ that witness that a cannot be added to C, namely $\pi_i(\bar{a})$ and $\pi_j(\bar{b})$ are not r-separated by S in G. Observe that two different tuples $\bar{a}, \bar{a}' \in B - C$ cannot be associated with exactly the same $\bar{b} \in C$ and same pair of indices (i, j). Indeed, then both $\pi_i(\bar{a})$ and $\pi_i(\bar{a}')$ would not be r-separated from $\pi_j(\bar{b})$ by S in G, which would imply that $\pi_i(\bar{a})$ and $\pi_i(\bar{a}')$ would not be $2r$-separated from each other by S, a contradiction with the assumption on B. Hence, $|B - C|$ is upper bounded by the number of tuples of the form $(\bar{b}, i, j) \in C \times \{1, \ldots, d\}^2$, which is $d^2|C|$. We conclude that $|B - C| \leqslant d^2|C|$, which implies $|C| \geqslant \frac{|B|}{d^2+1}$. □

To finish the proof of Lemma 2.10, given a set $A \subseteq V(G)^d$ and integer $m \in \mathbb{N}$, first apply Theorem 2.11 with $m' = m \cdot (d^2+1)$. Assuming that $|A| \geqslant f^d(m')$, we obtain a set $B \subseteq A$ with $|B| \geqslant m \cdot (d^2+1)$ and a set $S \subseteq V(G)$ with $|S| \leqslant d \cdot s_{2r}$, such that for each $i \in \{1, \ldots, d\}$ and all distinct $\bar{x}, \bar{y} \in B$, the vertices $\pi_i(\bar{x})$ and $\pi_i(\bar{y})$ are $2r$-separated by S in G. Then, apply Theorem 2.12 to B and S, yielding a set $C \subseteq B$ which is mutually r-separated by S and has size at least m. This concludes the proof of Lemma 2.10.

3 Bounds on the number of types

In this section we prove Theorem 1.3. We start with formulating a standard Gaifman-type result.

3.1 Locality

We will use the following intuitive notion of functional determination. Suppose X, A, B are sets and we have two functions: $f\colon X \to A$ and $g\colon X \to B$. We say that $f(x)$ *determines* $g(x)$ for $x \in X$ if for every pair of elements $x, x' \in X$ the following implication holds: $f(x) = f(x')$ implies $g(x) = g(x')$. Equivalently, there is a function $h\colon A \to B$ such that $g = h \circ f$.

Recall that if A, B, S are subsets of vertices of a graph G and $r \in \mathbb{N}$, then A and B are r-separated by S in G if every path from A to B of length at most r contains a vertex from S.

The following lemma is a consequence of Gaifman locality and a Feferman-Vaught lemma. If $q \in \mathbb{N}$ is a number and $\bar{u} \in V^d$ is a tuple of some length d, then by $\mathrm{tp}^q_G(\bar{u}/A)$ we denote the set of all formulas $\varphi(\bar{x}, \bar{v})$ of quantifier rank at most q, with parameters $\bar{v}$ from A, and with $|\bar{x}| = d$, such that $G, \bar{u} \models \varphi(\bar{y}, \bar{v})$.

Lemma 3.1 (⋆). *For any given numbers q and d one can compute numbers p and r with the following properties. Let $G = (V, E)$ be a fixed graph and let $A, B, S \subseteq V$ be fixed subsets of its vertices such that A and B are r-separated by S in G. Then, for tuples $\bar{u} \in A^d$, the type $\mathrm{tp}^q(\bar{u}/B)$ is determined by the type $\mathrm{tp}^p(\bar{u}/S)$.*

We will use the following consequence of the above lemma.

Corollary 3.2 (⋆). *For every formula $\varphi(\bar{x}, \bar{y})$ and number $s \in \mathbb{N}$ there exist numbers $T, r \in \mathbb{N}$, where r is computable from φ and T is computable from φ and s, such that the following holds. For every graph G and vertex subsets $A, B, S \subseteq V(G)$ where S has at most s vertices and r-separates A from B, we have $|S^\varphi(A/B)| \leqslant T$.*

3.2 Bounds on the number of types

We now come to the proof of Theorem 1.3. In the proof, we will will first enlarge the set A to a set B, called an *r-closure of A* (where r is chosen depending on φ), such that the connections of elements from $V(G) - B$ toward B are well controlled. This approach was first used in Drange et al. [11] in the context of classes of bounded expansion, and then for nowhere dense classes in Eickmeyer et al. [13]. We start by recalling these notions.

Let G be a graph and let $B \subseteq V(G)$ be a subset of vertices. For vertices $v \in B$ and $u \in V(G)$, a path P leading from u to v is called *B-avoiding* if all its vertices apart from v do not belong to B. Note that if $u \in B$, then there is only one B-avoiding path leading from u, namely the one-vertex path where $u = v$. For a positive integer r and $u \in V(G)$, the *r-projection* of u on B, denoted $M^G_r(u, B)$, is the set of all vertices $v \in B$ such that there is a B-avoiding path of length at most r leading from u to v. Note that for $u \in B$, we have $M^G_r(u, B) = \{b\}$. Equivalently, $M^G_r(u, B)$ is the unique inclusion-minimal subset of B which r-separates u from B. We will use the following result from [13].

Lemma 3.3 ([13]). *Let $\mathscr{C}$ be a nowhere dense class. Then for every $r \in \mathbb{N}$ and $\delta > 0$ there is a constant $c \in \mathbb{N}$ such that for every $G \in \mathscr{C}$ and $A \subseteq V(G)$ there exists a set B, called an r-*closure *of A, with the following properties:*

1. *$A \subseteq B \subseteq V(G)$;*
2. *$|B| \leqslant c \cdot |A|^{1+\delta}$; and*
3. *$|M^G_r(u, B)| \leqslant c \cdot |A|^\delta$ for each $u \in V(G)$.*

Moreover, for every set $X \subseteq V(G)$, it holds that

4. *$|\{M^G_r(u, X) \;:\; u \in V(G)\}| \leqslant c \cdot |X|^{1+\delta}$.*

We note that in [11, 13] projections on B are defined only for vertices outside of B. However, adding singleton projections for vertices of B to the definition only adds $|B|$ possible projections of size 1 each, so this does not influence the validity of the above results.

We proceed with the proof of Theorem 1.3. Let us fix: a nowhere dense class of graphs $\mathscr{C}$, a graph $G \in \mathscr{C}$, a vertex set $A \subseteq V(G)$, a real $\varepsilon > 0$, and a first order formula $\varphi(\bar{x}, \bar{y})$, where $\bar{x}$ is the distinguished ℓ-tuple of object variables. Our goal is to show that $|S^\varphi(G/A)| = O(|A|^{\ell+\varepsilon})$.

In the sequel, d denotes a positive integer depending on C, ℓ, φ only (and not on G, A and ε), and will be specified later. We may choose positive reals δ, ε_1 such that $(\ell + \varepsilon_1)(1 + \delta) \leqslant \ell + \varepsilon$ and $\varepsilon_1 > \delta(d + \ell) > \delta\ell$, for instance as follows: $\varepsilon_1 = \varepsilon/2$ and $\delta = \frac{\varepsilon}{4d+4\ell}$. The constants hidden in the $O(\cdot)$ notation below depend on $\varepsilon, \delta, \varepsilon_1, C, \ell$ and φ, but not on G and A. By *tuples* below we refer to tuples of length ℓ.

Let q be the quantifier rank of φ and let p, r be the numbers obtained by applying Lemma 3.1 to q and ℓ. Let B be an r-closure of A, given by Lemma 3.3. By Lemma 3.3, the total number of distinct r-projections onto B is at most $O(|B|^{1+\delta})$, and each of these projections has size $O(|B|^\delta)$. Figure 1 serves as an illustration to the steps of the proof in the case $\ell = 1$.

The first step is to reduce the statement to the following claim.

Claim 1. If X is a set of tuples with pairwise different φ-types over B, then $|X| = O(|B|^{\ell+\varepsilon_1})$.

Claim 1 implies that $|S^\varphi(G/B)| = O(|B|^{\ell+\varepsilon_1})$, which is bounded by $O(|A|^{(\ell+\varepsilon_1)(1+\delta)})$ since $|B| = O(|A|^{1+\delta})$.

As $(\ell + \varepsilon_1)(1 + \delta) \leqslant \ell + \varepsilon$, this shows that $|S^\varphi(G/B)| = O(|A|^{\ell+\varepsilon})$. Since $A \subseteq B$, we have $|S^\varphi(G/A)| \leqslant |S^\varphi(G/B)|$ so $|S^\varphi(G/A)| = O(|A|^{\ell+\varepsilon})$, and we are done. Therefore, it remains to prove Claim 1.

For a tuple $\bar{w} = w_1 \ldots w_\ell \in V(G)^\ell$, define its *projection* to be the set $C_1 \cup \ldots \cup C_\ell \subseteq B$ where $C_i = M^G_r(w_i, B)$. Note that there are at most $O(|B|^{\ell(1+\delta)})$ different projections of tuples in total, and each projection has size $O(|B|^\delta)$. To prove Claim 1, we consider the special case when all the tuples have the same projection, say $C \subseteq B$, and obtain a stronger conclusion, for $\varepsilon_2 \coloneqq \varepsilon_1 - \delta\ell > 0$.

Claim 2. If Y is a set of tuples with pairwise different φ-types over B, and each $u \in Y$ has the same projection $C \subseteq B$, then $|Y| = O(|B|^{\varepsilon_2})$.

Since there are at most $O(|B|^{\ell(1+\delta)})$ different projections in total and $\ell(1 + \delta) + \varepsilon_2 = \ell + \varepsilon_1$, Claim 1 can be proved by summing the bound given by Claim 2 through all different projections C. It therefore remains to prove Claim 2.

We apply Theorem 2.9 to the set of ℓ-tuples Y, for m being the largest integer such that $|Y| \geqslant N^\ell_{2r}(m)$. As a conclusion, we obtain a set $Z \subseteq Y$ of m tuples that is mutually $2r$-separated by S in G, for some set of vertices $S \subseteq V(G)$ of size $s \coloneqq s^\ell_{2r}$. Let d be the degree of the polynomial $N^\ell_{2r}(\cdot)$ obtained from Theorem 2.9. Note that $s = O(1)$ and $|Y| = O(m^d)$.

Claim 3. It holds that $|Z| = O(|C|)$.

We first show how Claim 3 implies Claim 2. Since $m = |Z| = O(|C|)$, and $|C| = O(|B|^\delta)$, it follows that $|Y| = O(m^d) = O(|B|^{d\delta})$. As $\delta(d + \ell) > \varepsilon_1$, this implies that $d\delta < \varepsilon_2$, yielding Claim 2. We now prove Claim 3.

Let $Z_0 \subseteq Z$ be the set of those tuples in Z which are r-separated by S from B in G, and let $Z_1 = Z - Z_0$ be the remaining tuples. Since tuples from Z_0 have pairwise different φ-types over B, and each of them is r-separated by S from B in G, by Corollary 3.2 we infer that $|Z_0| = O(1)$. On the other hand, by the definition of Z_1,

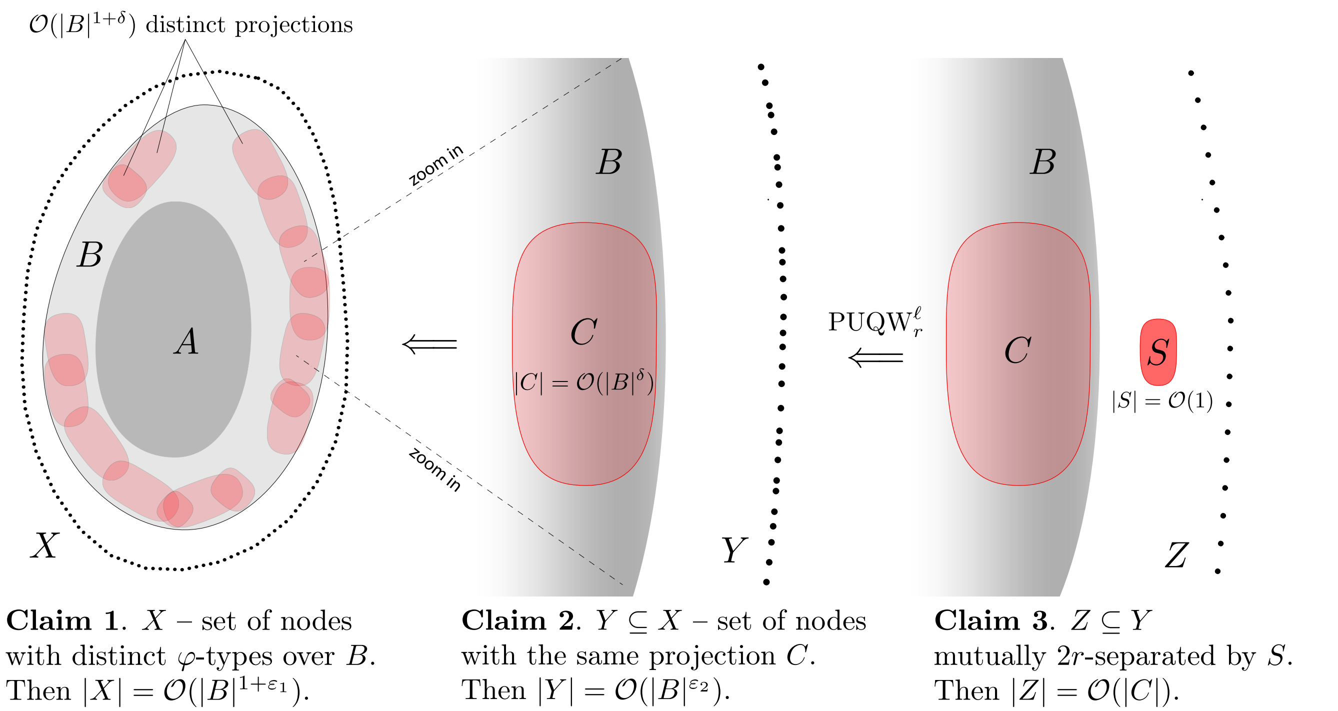

Figure 1. The proof of Theorem 1.3 in case $\ell = 1$. The logical implications flow from right to left, but our description below proceeds in the other direction.

with each tuple $\bar{u} \in Z_1$ we may associate a vertex $b(\bar{u}) \in C$ which is not r-separated from $\bar{u}$ by S in G. Since the set Z is mutually $2r$-separated by S in G, it follows that for any two different tuples $\bar{u}, \bar{v} \in Z_1$ we have $b(\bar{u}) \neq b(\bar{v})$. Hence $b(\cdot)$ is an injection from Z_1 to C, which proves that $|Z_1| \leqslant |C|$. To conclude, we have $|Z| = |Z_0| + |Z_1| = O(1) + O(|C|) = O(|C|)$. This finishes the proof of Claim 3 and ends the proof of Theorem 1.3.

4 Packing and traversal numbers for nowhere dense graphs

In this section, we give an application ofTheorem 1.3, proving a duality result, Theorem 1.6, for nowhere dense graph classes.

A *set system* is a family $\mathcal{F}$ of subsets of a set X. Its *packing* is a subfamily of $\mathcal{F}$ of pairwise disjoint subsets, and its *traversal* (or *hitting set*) is a subset of X which intersects every member of $\mathcal{F}$. The *packing number* of $\mathcal{F}$, denoted $\nu(\mathcal{F})$, is the largest cardinality of a packing in $\mathcal{F}$, and the *transversality* of $\mathcal{F}$, denoted $\tau(\mathcal{F})$, is the smallest cardinality of a traversal of $\mathcal{F}$. Note that if $\mathcal{G}$ is a finite set system, then $\nu(\mathcal{G}) \leqslant \tau(\mathcal{G})$. The set system $\mathcal{F}$ has the *Erdős-Pósa property* if there is a function $f : \mathbb{N} \to \mathbb{N}$ such that every finite subfamily $\mathcal{G}$ of $\mathcal{F}$ satisfies $\tau(\mathcal{G}) \leqslant f(\nu(\mathcal{G}))$.

Theorem 1.6 states that set systems defined by first order formulas in nowhere dense graph classes have the Erdős-Pósa property.

We will apply the following result of Matoušek [27], which relies on the proof of Alon and Kleitman [2] of the conjecture of Hardwiger and Debrunner. In the result of Matoušek, the set system $\mathcal{F}$ is infinite. For $m \in \mathbb{N}$, by $\pi^*_{\mathcal{F}}(m)$ we denote the *dual shatter function* of $\mathcal{F}$, which is defined as the maximal number of occupied cells in the Venn diagram of m sets in $\mathcal{F}$.

Theorem 4.1 (Matoušek, [27]). *Let $\mathcal{F}$ be a set system with $\pi^*_{\mathcal{F}}(m) = o(m^k)$, for some integer k, and let $p \geqslant k$. Then there is a constant T such that the following holds for every finite family $\mathcal{G} \subseteq \mathcal{F}$: if $\mathcal{G}$ has the (p, k)-property, meaning that among every p sets in $\mathcal{G}$ some k have a non-empty intersection, then $\tau(\mathcal{G}) \leqslant T$.*

Proof of Theorem 1.6. For a graph G, define the set system $\mathcal{F}_G$ on the ground set $V(G)$ as

$$\mathcal{F}_G = \{\{v \in V(G) \ : \ \varphi(u, v)\} \ : \ u \in V(G)\}.$$

Let then $\mathcal{F}$ be the disjoint union of set systems $\mathcal{F}_G$ for $G \in \mathscr{C}$. That is, the ground set of $\mathcal{F}$ is the disjoint union of the vertex sets $V(G)$ for $G \in \mathscr{C}$, and for each $G \in \mathscr{C}$ we add to $\mathcal{F}$ a copy of $\mathcal{F}_G$ over the copy of relevant $V(G)$. Then the following claim follows directly from Theorem 1.3.

Claim 4. The dual shatter function of $\mathcal{F}$ satisfies $\pi^*_{\mathcal{F}}(m) = O(m^{1+\varepsilon})$, for every fixed $\varepsilon > 0$. In particular, $\pi^*_{\mathcal{F}}(m) = o(m^2)$.

Consider the function $f : \mathbb{N} \to \mathbb{N}$ defined so that $f(\nu)$ is the value T obtained fromTheorem 4.1 applied to $\mathcal{F}$, $k = 2$, and $p = \nu+1$. Suppose now that $G \in \mathscr{C}$ is a graph and $\mathcal{G} \subseteq \mathcal{F}_G$ is a family of subsets of $V(G)$ consisting of sets of the form $\{v \in V(G) : \varphi(u, v)\}$, where u is some vertex of G. We identify $\mathcal{G}$ with a subfamily of $\mathcal{F}$ in the natural way, following the embedding of $\mathcal{F}_G$ into $\mathcal{F}$ used in the construction of the latter. Let ν be the packing number of $\mathcal{G}$. In particular, for every $\nu + 1$ subsets of $\mathcal{G}$ there is a vertex $v \in V(G)$ which is contained in two elements of $\mathcal{G}$. Hence, $\mathcal{G}$ is a $(p, 2)$-family for $p = \nu + 1$. By Theorem 4.1, $\tau(\mathcal{G}) \leqslant T = f(\nu) = f(\nu(\mathcal{G}))$, as required. ◻

5 Bounds for stability

Adler and Adler [1], proved that every nowhere dense class of graphs is stable. In this section, we prove its effective variant, Theorem 1.7.

Proof of Theorem 1.7. Fix a formula $\varphi(\bar{x}, \bar{y})$ of quantifier rank q and a partitioning of its free variables into $\bar{x}$ and $\bar{y}$. Let $d = |\bar{x}| + |\bar{y}|$ be the total number of free variables of φ. Let $r \in \mathbb{N}$ be the number given by Corollary 3.2, which depends on φ only. Let C be the class of all graphs such that $K_t \not\preccurlyeq_{18r} G$. Then, by Theorem 2.9, C satisfies $\mathrm{UQW}_r^d(N_r^d, s_r^d)$, for some polynomial $N_r^d \colon \mathbb{N} \to \mathbb{N}$ and number $s = s_r^d \in \mathbb{N}$ computable from d, t, r. Let T be the number given by Corollary 3.2 for φ and s. Finally, let $\ell = N_r^d(2T+1)$. We show that every φ-ladder in a graph $G \in C$ has length smaller than ℓ.

For the sake of contradiction, assume that there is a graph $G \in C$ and tuples $\bar{u}_1, \ldots, \bar{u}_\ell \in V(G)^{|\bar{x}|}$ and $\bar{v}_1, \ldots, \bar{v}_\ell \in V(G)^{|\bar{y}|}$ which form a φ-ladder in G, i.e., $\varphi(\bar{u}_i, \bar{v}_j)$ holds in G if and only if $i \leqslant j$. Let $A = \{\bar{u}_i\bar{v}_i \;:\; i = 1, \ldots, \ell\} \subseteq V(G)^d$. Note that $|A| = \ell \geqslant N_r^d(2T+1)$, since tuples $\bar{u}_i$ have to be pairwise different.

Applying property $\mathrm{UQW}_r^d(N_r^d, s_r^d)$ to the set A, radius r, and target size $m = 2T+1$ yields a set $S \subseteq V(G)$ with $|S| \leqslant s$ and a set $B \subseteq A$ with $|B| \geqslant 2T+1$ of tuples which are mutually r-separated by S in G. Let $J \subseteq \{1, \ldots, \ell\}$ be the set of indices corresponding to B, i.e., $J = \{j \colon \bar{u}_j\bar{v}_j \in B\}$.

Since $|J| = 2T+1$, we may partition J into $J_1 \uplus J_2$ with $|J_1| = T+1$ so that the following condition holds: for each $i, k \in J_1$ satisfying $i < k$, there exists $j \in J_2$ with $i < j < k$. Indeed, it suffices to order the indices of J and put every second index to J_1, and every other to J_2. Let X be the set of vertices appearing in the tuples $\bar{u}_i$ with $i \in J_1$, and let Y be the set of vertices appearing in the tuples $\bar{v}_j$ with $j \in J_2$. Since the tuples of B are mutually r-separated by S in G, it follows that X and Y are r-separated by S. As $|J_1| = T+1$, by Corollary 3.2 we infer that there are distinct indices $i, k \in J_1$, say $i < k$, such that $\mathrm{tp}^\varphi(\bar{u}_i/Y) = \mathrm{tp}^\varphi(\bar{u}_k/Y)$. This implies that for each $j \in J_2$, we have $G, \bar{u}_i, \bar{v}_j \models \varphi(\bar{x}, \bar{y})$ if and only if $G, \bar{u}_k, \bar{v}_j \models \varphi(\bar{x}, \bar{y})$. However, there is an index $j \in J_2$ such that $i < j < k$, and for this index we should have $G, \bar{u}_i, \bar{v}_j \models \varphi(\bar{x}, \bar{y})$ and $G, \bar{u}_k, \bar{v}_j \not\models \varphi(\bar{x}, \bar{y})$ by the definition of a ladder. This contradiction concludes the proof. ◻

References

[1] H. Adler and I. Adler. Interpreting nowhere dense graph classes as a classical notion of model theory. *European Journal of Combinatorics*, 36:322–330, 2014.

[2] N. Alon and D. J. Kleitman. Piercing convex sets and the Hadwiger-Debrunner (p, q)-problem. *Advances in Mathematics*, 96(1):103 – 112, 1992.

[3] N. Alon, M. Krivelevich, and B. Sudakov. Turán numbers of bipartite graphs and related Ramsey-type questions. *Combinatorics, Probability and Computing*, 12(5+6):477–494, 2003.

[4] M. Aschenbrenner, A. Dolich, D. Haskell, D. Macpherson, and S. Starchenko. Vapnik-Chervonenkis density in some theories without the independence property, I. *Transactions of the American Mathematical Society*, 368(8):5889–5949, 2016.

[5] A. Atserias, A. Dawar, and P. G. Kolaitis. On preservation under homomorphisms and unions of conjunctive queries. *Journal of the ACM (JACM)*, 53(2):208–237, 2006.

[6] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM (JACM)*, 36(4):929–965, 1989.

[7] N. Bousquet and S. Thomassé. VC-dimension and Erdős-Pósa property. *Discrete Mathematics*, 338(12):2302–2317, 2015.

[8] H. Brönnimann and M. T. Goodrich. Almost optimal set covers in finite VC-dimension. *Discrete & Computational Geometry*, 14(4):463–479, 1995.

[9] A. Chervonenkis and V. Vapnik. Theory of uniform convergence of frequencies of events to their probabilities and problems of search for an optimal solution from empirical data. *Automation and Remote Control*, 32:207–217, 1971.

[10] A. Dawar. Homomorphism preservation on quasi-wide classes. *Journal of Computer and System Sciences*, 76(5):324–332, 2010.

[11] P. G. Drange, M. S. Dregi, F. V. Fomin, S. Kreutzer, D. Lokshtanov, M. Pilipczuk, M. Pilipczuk, F. Reidl, F. Sánchez Villaamil, S. Saurabh, S. Siebertz, and S. Sikdar. Kernelization and sparseness: the case of Dominating Set. In *STACS 2016*, volume 47 of *LIPIcs*, pages 31:1–31:14. Schloss Dagstuhl—Leibniz-Zentrum für Informatik, 2016. See https://arxiv.org/abs/1411.4575 for full proofs.

[12] Z. Dvořák, D. Král, and R. Thomas. Testing first-order properties for subclasses of sparse graphs. *Journal of the ACM (JACM)*, 60(5):36, 2013.

[13] K. Eickmeyer, A. C. Giannopoulou, S. Kreutzer, O. Kwon, M. Pilipczuk, R. Rabinovich, and S. Siebertz. Neighborhood complexity and kernelization for nowhere dense classes of graphs. In *ICALP 2017*, volume 80 of *LIPIcs*, pages 63:1–63:14. Schloss Dagstuhl—Leibniz-Zentrum für Informatik, 2017. See https://arxiv.org/abs/1612.08197 for full proofs.

[14] Z. Füredi and J. Pach. Traces of finite sets: extremal problems and geometric applications. *Extremal problems for finite sets*, 3:255–282, 1991.

[15] H. Gaifman. On local and non-local properties. *Studies in Logic and the Foundations of Mathematics*, 107:105–135, 1982.

[16] M. Grohe and S. Kreutzer. Methods for algorithmic meta theorems. In M. Grohe and J. Makowsky, editors, *Model Theoretic Methods in Finite Combinatorics*, volume 558 of *Contemporary Mathematics*, pages 181–206. American Mathematical Society, 2011.

[17] M. Grohe and G. Turán. Learnability and definability in trees and similar structures. *Theory of Computing Systems*, 37(1):193–220, 2004.

[18] M. Grohe, S. Kreutzer, and S. Siebertz. Deciding first-order properties of nowhere dense graphs. In *STOC 2014*, pages 89–98. ACM, 2014.

[19] M. Grohe, S. Kreutzer, R. Rabinovich, S. Siebertz, and K. Stavropoulos. Colouring and covering nowhere dense graphs. In *WG 2015*, volume 9224 of *Lecture Notes in Computer Science*, pages 325–338. Springer, 2015.

[20] A. A. Ivanov. The structure of superflat graphs. *Fundamenta Mathematicae*, 143:107–117, 1993.

[21] M. Kreidler and D. Seese. Monadic NP and graph minors. In *CSL 1998*, volume 1584 of *Lecture Notes in Computer Science*, pages 126–141. Springer, 1998.

[22] S. Kreutzer, R. Rabinovich, and S. Siebertz. Polynomial kernels and wideness properties of nowhere dense graph classes. In *SODA 2017*, pages 1533–1545. SIAM, 2017.

[23] M. l Pilipczuk, S. Siebertz, and S. Torunczyk. On the number of types in sparse graphs. *arXiv preprint arXiv:1705.09336*, 2017.

[24] M. C. Laskowski. Vapnik-Chervonenkis classes of definable sets. *Journal of the London Mathematical Society*, 2(2):377–384, 1992.

[25] M. Malliaris and S. Shelah. Regularity lemmas for stable graphs. *Transactions of the American Mathematical Society*, 366(3):1551–1585, 2014.

[26] J. Matoušek. Geometric set systems. In *European Congress of Mathematics*, volume 2, page 23. Birkhäuser, Basel, 1998.

[27] J. Matoušek. Bounded VC-dimension implies a fractional Helly theorem. *Discrete & Computational Geometry*, 31(2):251–255, 2004.

[28] J. Nešetřil and P. Ossona de Mendez. Grad and classes with bounded expansion I. Decompositions. *European Journal of Combinatorics*, 29(3):760–776, 2008.

[29] J. Nešetřil and P. Ossona de Mendez. First order properties on nowhere dense structures. *The Journal of Symbolic Logic*, 75(03):868–887, 2010.

[30] J. Nešetřil and P. Ossona de Mendez. On nowhere dense graphs. *European Journal of Combinatorics*, 32(4):600–617, 2011.

[31] J. Nešetřil and P. Ossona de Mendez. *Sparsity — Graphs, Structures, and Algorithms*, volume 28 of *Algorithms and combinatorics*. Springer, 2012.

[32] A. Pillay. *Introduction to Stability Theory*. Dover Books on Mathematics. Dover Publications, 2008.

[33] K.-P. Podewski and M. Ziegler. Stable graphs. *Fundamenta Mathematicae*, 100(2):101–107, 1978.

[34] F. Reidl, F. Sánchez Villaamil, and K. Stavropoulos. Characterising bounded expansion by neighbourhood complexity. *CoRR*, abs/1603.09532, 2016.

[35] N. Sauer. On the density of families of sets. *Journal of Combinatorial Theory, Series A*, 13(1):145–147, 1972.

[36] S. Shelah. Stability, the f.c.p., and superstability; model theoretic properties of formulas in first order theory. *Annals of Mathematical Logic*, 3(3):271–362, 1971.

[37] S. Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific Journal of Mathematics*, 41(1):247–261, 1972.

[38] S. Shelah. *Classification theory: and the number of non-isomorphic models*, volume 92 of *Studies in Logic and the Foundations of Mathematics*. Elsevier, 1990.

[39] K. Tent and M. Ziegler. *A Course in Model Theory*. Lecture Notes in Logic. Cambridge University Press, 2012.

[40] L. G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[41] X. Zhu. Colouring graphs with bounded generalized colouring number. *Discrete Mathematics*, 309(18):5562–5568, 2009.