First-Order Interpretations of Bounded Expansion Classes

Jakub Gajarský[1]
Technical University Berlin, Germany

Stephan Kreutzer[1]
Technical University Berlin, Germany

Jaroslav Nešetřil[2]
Charles University, Prague, Czech Republic

Patrice Ossona de Mendez[2]
CAMS (CNRS, UMR 8557), Paris, France

Michał Pilipczuk[3]
University of Warsaw, Warsaw, Poland

Sebastian Siebertz[3]
University of Warsaw, Warsaw, Poland

Szymon Toruńczyk[4]
University of Warsaw, Warsaw, Poland

Abstract

The notion of bounded expansion captures uniform sparsity of graph classes and renders various algorithmic problems that are hard in general tractable. In particular, the model-checking problem for first-order logic is fixed-parameter tractable over such graph classes. With the aim of generalizing such results to dense graphs, we introduce classes of graphs with *structurally bounded expansion*, defined as first-order interpretations of classes of bounded expansion. As a first step towards their algorithmic treatment, we provide their characterization analogous to the characterization of classes of bounded expansion via low treedepth decompositions, replacing treedepth by its dense analogue called shrubdepth.

1   Introduction

The interplay of methods from logic and graph theory has led to many important results in theoretical computer science, notably in algorithmics and complexity theory. The combination of logic and algorithmic graph theory is particularly fruitful in the area of *algorithmic meta-theorems*. Algorithmic meta-theorems are results of the form: *every computational problem definable in a logic $\mathcal{L}$ can be solved efficiently on any class of structures satisfying a property $\mathcal{P}$.* In other words, these theorems show that the *model-checking problem* for the logic $\mathcal{L}$ on any class $\mathcal{C}$ satisfying $\mathcal{P}$ can be solved efficiently, where *efficiency* usually means *fixed-parameter tractability.*

The archetypal example of an algorithmic meta-theorem is Courcelle's theorem [1, 2], which states that model-checking a formula φ of *monadic second-order logic* can be solved in time $f(\varphi) \cdot n$ on any graph with n vertices which comes from a fixed class of graphs of *bounded treewidth*, for some computable function f. Seese [27] proved an analogue of Courcelle's result for the model-checking problem of first-order logic on any class of graphs of bounded degree. Following this result, the complexity of first-order model-checking on specific classes of graphs has been studied extensively in the literature. See e.g. [17, 7, 19, 21, 4, 9, 10, 27, 8, 22, 28, 16, 5, 6, 13, 20]. One of the main goals of this line of research is to find a structural property $\mathcal{P}$ which precisely defines those graph classes $\mathcal{C}$ for which model checking of first-order logic is tractable.

So far, research on algorithmic meta-theorems has focused predominantly on *sparse* classes of graphs, such as classes of bounded *treewidth*, *excluding a minor* or which have *bounded expansion* or are *nowhere dense*. The concepts of *bounded expansion* and *nowhere denseness* were introduced by Nešetřil and Ossona de Mendez with the goal of capturing the intuitive notion of *sparseness*. See [25] for an extensive cover of these notions. The large number of equivalent ways in which they can be defined using either notions from combinatorics, theoretical computer science or logic, indicate that these two concepts capture some very natural limits of "well-behavedness" and algorithmic tractability. For instance, Grohe et al. [19] proved that if $\mathcal{C}$ is a class of graphs closed under taking subgraphs then model checking first-order logic on $\mathcal{C}$ is tractable if, and only if, $\mathcal{C}$ is nowhere dense (the lower bound was proved in [7]). As far as algorithmic meta-theorems for fixed-parameter tractability of first-order model-checking are concerned, this result completely solves the case for graph classes which are *closed under taking subgraphs*, which is a reasonable requirement for sparse but not for dense graph classes.

Consequently, research in this area has shifted towards studying the dense case, which is much less understood. While there are several examples of algorithmic meta-theorems on dense classes, such as for monadic second-order logic on classes of bounded *cliquewidth* [3] or for first-order logic on *interval graphs*, *partial orders*, classes of bounded *shrubdepth* and other classes, see e.g. [13, 11, 14, 12], a general theory of meta-theorems for dense classes is still missing. Moreover, unlike the sparse case, there is no canonical hierarchy of dense graph classes similar to the sparse case which could guide research on algorithmic meta-theorems in the dense world.

Hence, the main research challenge for dense model-checking is not only to prove tractability results and to develop the necessary logical and algorithmic tools. It is at least as important to define and analyze promising candidates for "structurally simple" classes of graph classes which are not necessarily sparse. This is the main motivation for the research in this paper. Since bounded expansion and nowhere denseness form the limits for tractability of certain problems in the sparse case, any extension of the theory should provide notions

which collapse to bounded expansion or nowhere denseness, under the additional assumption that the classes are closed under taking subgraphs. Therefore, a natural way of seeking such notions is to base them on the existing notions of bounded expansion or nowhere denseness.

In this paper, we take bounded expansion classes as a starting point and study two different ways of generalizing them towards dense graph classes preserving their good properties. In particular, we define and analyze classes of graphs obtained from bounded expansion classes by means of first-order interpretations and classes of graphs obtained by generalizing another, more combinatorial characterization of bounded expansion in terms of low treedepth colorings into the dense world. Our main structural result shows that these two very different ways of generalizing bounded expansion into the dense setting lead to the same classes of graphs. This is explained in greater detail below.

Interpretations and transductions.   One possible way of constructing "well-behaved" and "structurally simple" classes of graphs is to use logical *interpretations*, or the related concept of *transductions* studied in formal language and automata theory. For our purpose, transductions are more convenient and we will use them in this paper. Intuitively, a *transduction* is a logically defined operation I which takes a structure as input and nondeterministically produces as output a target structure. In this paper we use *first-order* transductions, which involve first-order formulas (see Section 2 for details). Two examples of such transductions are graph complementation, and the squaring operation which, given a graph G, adds an edge between every pair of vertices at distance 2 from each other.

We postulate that if we start with a "structurally simple" class $\mathcal{C}$ of graphs, e.g. a class of bounded expansion or a nowhere dense class, and then study the graph classes $\mathcal{D}$ which can be obtained from $\mathcal{C}$ by first-order transductions, then the resulting classes should still have a simple structure and thus be well-behaved algorithmically as well as in terms of logic. In other words, the resulting classes are interesting graph classes with good algorithmic and logical properties, and which are certainly not sparse in general. For instance, a useful feature of transductions is that they provide a canonical way of reducing model-checking problems from the generated classes $\mathcal{D}$ to the original class $\mathcal{C}$, provided that given a graph $H \in \mathcal{D}$, we can effectively compute some graph $G \in \mathcal{C}$ that is mapped to H by the transduction. In general, this is a hard problem, requiring a *combinatorial* understanding of the structure of the resulting classes $\mathcal{D}$.

The above principle has so far been successfully applied in the setting of graph classes of bounded treewidth and monadic second-order transductions: it was shown by Courcelle, Makowsky and Rotics [3] that transductions of classes of bounded treewidth can be combinatorially characterized as classes of bounded cliquewidth. This, combined with Oum's result [26] gives a fixed-parameter algorithm for model-checking monadic second-order logic on classes of bounded cliquewidth. More recently, the same principle, but for first-order logic, has been applied to graphs of bounded degree [12], leading to a combinatorial characterization of first-order transductions of such classes, and to a model-checking algorithm.

Applying our postulate to bounded expansion classes yields the central notion of this paper: a class of graphs has *structurally bounded expansion* if it is the image of a class of bounded expansion under some fixed first-order transduction. This paper is a step towards a combinatorial, algorithmic, and logical understanding of such graph classes.

Low Shrubdepth Covers.   The method of transductions is one way of constructing complex graphs out of simple graphs. A more combinatorial approach is the method of *decompositions* (or *colorings*) [25], which we reformulate below in terms of *covers*. This method can be used

$$
\begin{array}{ccc}
\textit{bounded treedepth} & \xrightarrow{\text{transduction}} & \textit{transduction of bounded treedepth} \\
& & =_{(1)} \textit{ bounded shrubdepth} \\
\big\downarrow \text{cover} & & \big\downarrow \text{cover} \\
\textit{bounded treedepth covers} & \xrightarrow{\text{transduction}} & \textit{low shrubdepth covers} =_{(\star)} \\
=_{(2)} \textit{ bounded expansion} & & \textit{structurally bounded expansion}
\end{array}
$$
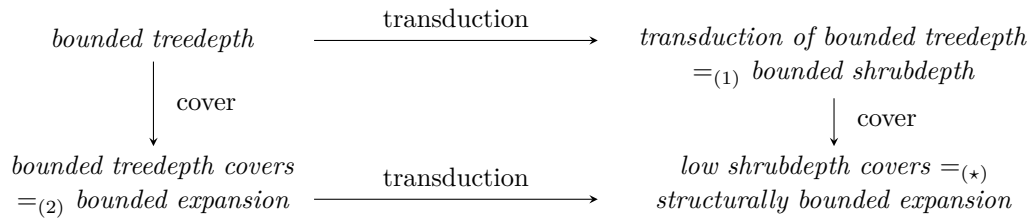
■ **Figure 1** The nodes in the diagram depict properties of graph classes, and the arrows depict operations on properties of graph classes. Equality (1) is by [14]. Equality (2) is by [24]. Equality ($\star$) is the main result of this paper, Theorem 5.1.

to provide a characterization of bounded expansion classes in terms of very simple graph classes, namely classes of *bounded treedepth*. A class of graphs has bounded treedepth if there is a bound on the length of simple paths in the graphs in the class (see Section 2 for a different but equivalent definition). A class $\mathcal{C}$ has *low treedepth covers* if for every number $p \in \mathbb{N}$ there is a number N and a class of bounded treedepth $\mathcal{T}$ such that for every $G \in \mathcal{C}$, the vertex set $V(G)$ can be covered by N sets $U_1, \ldots, U_N$ so that every set $X \subseteq V(G)$ of at most p vertices is contained in some U_i, and for each $i = 1, \ldots, N$, the subgraph of G induced by U_i belongs to $\mathcal{T}$. A consequence of a result by Nešetřil and Ossona de Mendez [24] on a related notion of *low treedepth colorings* is that a graph class has bounded expansion if, and only if, it has low treedepth covers.

The decomposition method allows to lift algorithmic, logical, and structural properties from classes of bounded treedepth to classes of bounded expansion. For instance, this was used to show tractability of first-order model-checking on bounded expansion classes [7, 18].

An analogue of treedepth in the dense world is the concept of *shrubdepth*, introduced in [14]. Shrubdepth shares many of the good algorithmic and logical properties of treedepth. This notion is defined combinatorially, in the spirit of the definition of cliquewidth, but can be also characterized by logical means, as first-order transductions of classes of bounded treedepth. Applying the method of decompositions to the notion of shrubdepth leads to the following definition. A class $\mathcal{C}$ of graphs has *low shrubdepth covers* if for every number $p \in \mathbb{N}$ there is a number N and a class $\mathcal{B}$ of bounded shrubdepth such that for every $G \in \mathcal{C}$, there is a *p-cover* of G consisting of N sets $U_1, \ldots, U_N \subseteq V(G)$, so that every set $X \subseteq V(G)$ of at most p vertices is contained in some U_i and for each $i = 1, \ldots, N$, the subgraph of G induced by U_i belongs to $\mathcal{B}$. Shrubdepth properly generalizes treedepth and consequently classes admitting low shrubdepth covers properly extend bounded expansion classes.

It was observed earlier [23] that for every fixed $r \in \mathbb{N}$ and every class $\mathcal{C}$ of bounded expansion, the class of rth power graphs G^r of graphs from $\mathcal{C}$ (the rth power of a graph is a simple first-order transduction) admits low shrubdepth colorings.

Our contributions.   Our main result, Theorem 5.1, states that the two notions introduced above are the same: a class of graphs $\mathcal{C}$ has structurally bounded expansion if, and only if, it has bounded shrubdepth covers. That is, transductions of classes of bounded expansion are the same as classes with low shrubdepth covers (cf. Figure 1). This gives a combinatorial characterization of structurally bounded expansion classes, which is an important step towards their algorithmic treatment.

One of the key ingredients of our proof is a quantifier-elimination result (Lemma 5.5) for transductions on classes of structurally bounded expansion. This result strengthens

in several ways similar results for bounded expansion classes due to Dvořák, Kráľ, and Thomas [7] and Grohe and Kreutzer [18]. Our assumption is more general, as they assume that $\mathcal{C}$ has bounded expansion, and here $\mathcal{C}$ is only required to have low shrubdepth covers. Also, our conclusion is stronger, as their results provide quantifier-free formulas involving some unary functions and unary predicates which are computable algorithmically, whereas our result shows that these functions can be defined using very restricted transductions. Quantifier-elimination results of this type proved to be useful for the model-checking problem on bounded expansion classes [7, 18], and this is also the case here.

As explained earlier, the transduction method allows to reduce the model-checking problem to the problem of finding inverse images under transductions, which is a hard problem in general and depends very much on the specific transduction. On the other hand, as we show, the cover method allows to reduce the model-checking problem for classes with low shrubdepth covers to the problem of computing a bounded shrubdepth cover of a given graph. In fact, as a consequence of our proof, in Theorem 6.1 we show that it is enough to compute a 2-cover of a given graph G from a structurally bounded expansion class, in order to obtain an algorithm for the model-checking problem for such classes. We strongly conjecture that such an algorithm exists and that therefore first-order model-checking is fixed-parameter tractable on any class of graphs of structurally bounded expansion. We leave this problem for future work.

2 Transductions

In this section, apart from recalling some background notions from logic and graph theory, we introduce the notion of transductions which we use in this paper.

Structures. We use standard logical notation and terminology with the following exceptions. A *signature* Σ is a finite set of relation and function symbols. We allow relations of any finite arity but only unary functions. We use boldface letters $\mathbf{A}$ for logical structures and denote their domains by $V(\mathbf{A})$. A structure $\mathbf{A}$ over Σ is defined as usual with the exception that each function symbol $f \in \Sigma$ is interpreted as a *partial* function $f^{\mathbf{A}} \colon V(\mathbf{A}) \rightharpoonup V(\mathbf{A})$. If $\mathbf{A}$ is a structure and $X \subseteq V(\mathbf{A})$ then we define the *substructure* of $\mathbf{A}$ induced by X in the usual way except that a unary function $f(x)$ in $\mathbf{A}$ becomes undefined on all $x \in X$ for which $f(x) \notin X$. The semantics of first-order logic is defined as usual, with the proviso that an atomic formula evaluates to false if any of the terms involved in it is undefined.

Graphs and colored graphs. We consider finite, simple and undirected graphs. These can be viewed as finite structures over the signature consisting of a binary relation symbol E, interpreted as the edge relation, in the usual way. For a finite label set Λ, by a Λ-*colored* graph we mean a graph enriched by a unary predicate U_λ for every $\lambda \in \Lambda$. We will follow the convention that if $\mathcal{C}$ is a class of colored graphs, then we implicitly assume that all graphs in $\mathcal{C}$ are over the same fixed finite signature. A rooted forest is an acyclic graph F together with a unary predicate $R \subseteq V(F)$ selecting one root in each connected component of F.

Transductions. We now define the notion of transduction used in the sequel. A *transduction* is a special type of first-order interpretation with set parameters, which we see here (from a computational point of view) as a nondeterministic operation that maps input structures to output structures. Transductions are defined as compositions of *atomic operations* listed below.

An **extension** operation is parameterized by a first-order formula $\varphi(x_1, \ldots, x_k)$ and a relation symbol R. Given an input structure $\mathbf{A}$, it outputs the structure $\mathbf{A}$ extended by the relation R interpreted as the set of k-tuples of elements satisfying φ in $\mathbf{A}$. A **restriction** operation is parameterized by a unary formula $\psi(x)$. Applied to a structure $\mathbf{A}$ it outputs the substructure of $\mathbf{A}$ induced by all elements satisfying ψ. A **reduct** operation is parameterized by a relation symbol R, and results in removing the relation R from the input structure. **Copying** is an operation which, given a structure $\mathbf{A}$ outputs a disjoint union of two copies of $\mathbf{A}$ extended with a new unary function which maps both copies of a vertex v back to v. A **function extension** operation is parameterized by a binary formula $\varphi(x, y)$ and a function symbol f, and extends a given input structure by a partial function f defined as follows: $f(x) = y$ if y is the unique vertex such that $\varphi(x, y)$ holds. Note that if there is no such y or more than one such y, then $f(x)$ is undefined. Finally, suppose σ is function that maps each structure $\mathbf{A}$ to a nonempty family $\sigma(\mathbf{A})$ of subsets of its universe. A **unary lift** operation, parametrized by σ, takes as input a structure $\mathbf{A}$ and outputs the structure $\mathbf{A}$ enriched by a unary predicate X interpreted by a nondeterministically chosen set $U \in \sigma(\mathbf{A})$.

We remark that a function extension operations can be simulated by extension operations, defining the graphs of the functions in the obvious way. They are, however, useful as a means of extending the expressive power of transductions in which only quantifier-free formulas are allowed, as defined below.

▶ **Definition 2.1.** *Transductions* are defined inductively: every atomic transduction is a transduction, and the composition of two transductions I and J is the transduction $\mathsf{I}; \mathsf{J}$ that, given a structure $\mathbf{A}$, first applies J to $\mathbf{A}$ and then I to the output $\mathsf{J}(\mathbf{A})$. A transduction is *deterministic* if it does not use unary lifts. In this case, for every input structure there is exactly one output structure. A transduction is *almost quantifier-free* if all formulas that parameterize atomic operations comprising it are quantifier-free (observe that such transductions still can access elements that are not among its free variables via functions, hence, to avoid confusion we do not speak of quantifier-free transductions), and is *deterministic almost quantifier-free* if it additionally does not use unary lifts.

If $\mathcal{C}$ is a class of structures, we write $\mathsf{I}(\mathcal{C})$ for the class which contains all possible outputs $\mathsf{I}(\mathbf{A})$ for $\mathbf{A} \in \mathcal{C}$. We say that two transductions I and J are *equivalent* on a class $\mathcal{C}$ of structures if every possible output of $\mathsf{I}(\mathbf{A})$ is also a possible output of $\mathsf{J}(\mathbf{A})$, and vice versa, for every $\mathbf{A} \in \mathcal{C}$.

▶ **Example 2.2.** Let $\mathcal{C}$ be the class of rooted forests of depth at most d, for some fixed $d \in \mathbb{N}$. We describe an almost quantifier-free transduction which defines the *parent function* in $\mathcal{C}$. First, using unary lifts introduce $d + 1$ unary predicates $D_0, \ldots, D_d$, where D_i marks the vertices of the input tree which are at distance i from a root. Next, using a function extension, define a function f which maps a vertex v in the input tree to its parent, or is undefined in case of a root. This can be done by a quantifier-free formula, which selects those pairs x, y such that x and y are adjacent and $D_i(x)$ implies $D_{i-1}(y)$.

It will sometimes be convenient to work with the encoding of bounded-depth trees and forests as node sets endowed with the parent function, rather than graphs with prescribed roots. As seen in Example 2.2, these two encodings can be translated to each other by means of almost quantifier-free transductions, which render them essentially equivalent.

3   Treedepth, shrubdepth and bounded expansion

Treedepth.   The *depth* of a node x in a rooted forest F is the number of edges on the root-to-x path in F. The *depth* of F is the maximal depth of any of its nodes. The *treedepth* of a graph G is the minimal depth of a rooted forest F with the same vertex set as G, such that for every edge vw of G, v is an ancestor of w, or w is an ancestor of v in F. A class $\mathcal{C}$ of graphs has *bounded treedepth* if there is a bound $d \in \mathbb{N}$ such that every graph in $\mathcal{C}$ has treedepth at most d. Equivalently, there is some number k such that no graph in $\mathcal{C}$ has a simple path of length k [25].

Shrubdepth.   The following notion of *shrubdepth* has been proposed in [14] as a dense analogue of treedepth. Originally, shrubdepth was defined using the notion of *tree-models*. We present an equivalent definition based on the notion of *connection models*, introduced in [14] under the name of *m-partite cographs* of bounded depth.

▶ **Definition 3.1.** A *connection model* with labels from Λ is a rooted labeled tree T in which every leaf x is assigned a label $\lambda(x) \in \Lambda$, and every non-leaf node v is labeled by a (symmetric) binary relation $C(v) \subseteq \Lambda \times \Lambda$. Such a model defines a graph G on the leaves of T, in which two distinct leaves x and y are connected by an edge if and only if $(\lambda(x), \lambda(y)) \in C(v)$, where v is the common ancestor of x and y of largest depth. We say that T is a *connection model* of the resulting graph G.

▶ **Example 3.2.** Fix $n \in \mathbb{N}$, and let G_n be the bi-complement of a matching of order n, i.e., the bipartite graph with nodes $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$, such that a_i is adjacent to b_j if, and only if, $i \neq j$. A connection model for G_n is shown below:
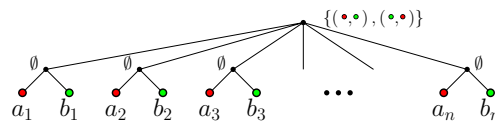
▶ **Definition 3.3.** A class $\mathcal{C}$ of graphs has *bounded shrubdepth* if there is an $h \in \mathbb{N}$ and a finite set Λ such that every $G \in \mathcal{C}$ has a connection model of depth at most h using labels from Λ. A class of colored graphs has bounded shrubdepth if the class of underlying graphs has bounded shrubdepth.

Note that contrary to other graph parameters, it is meaningless to speak about the shrubdepth of a single graph. Rather, shrubdepth measures directly the complexity of a class.

Shrubdepth can be equivalently defined in terms of another graph parameter, as follows.

▶ **Definition 3.4.** Given a graph G and a set of vertices $W \subseteq V(G)$, the graph obtained by *flipping the adjacency within W* is the graph G' with vertices $V(G)$ and edge set which is the symmetric difference of the edge set of G and the edge set of the clique on W. The *subset-complementation depth*, or *SC-depth*, of a graph is defined inductively as follows: (1) a graph with one vertex has SC-depth 0, and (2) for $d \geqslant 1$ a graph G has SC-depth at most d if there is a set $W \subseteq V(G)$ of vertices such that in the graph obtained from G by flipping the adjacency within W all connected components have SC-depth at most $d - 1$.

The notion of SC-depth leads to a natural notion of decompositions. An *SC-decomposition* of a graph G of SC-depth at most d is a rooted tree T of depth d with leaf set $V(G)$, equipped with unary predicates $W_0, \ldots, W_d$ on the leaves. Each child s of the root in T corresponds to a connected component C_s of the graph G' obtained from G by flipping the adjacency within W_0, such that the subtree of T rooted at s, together with the unary predicates $W_1, \ldots, W_d$ restricted to $V(C_s)$, form an SC-decomposition of C_s.

We will make use of the following properties, where the first one follows from the definition of shrubdepth, and the remaining ones follow from [14].

▶ **Proposition 3.5.** *Let* $\mathcal{C}$ *be a class of graphs. Then:*
1. *If* $\mathcal{C}$ *has bounded shrubdepth, then the class of all induced subgraphs of graphs from* $\mathcal{C}$ *also has bounded shrubdepth.*
2. $\mathcal{C}$ *has bounded shrubdepth if and only if for some* $d \in \mathbb{N}$ *all graphs in* $\mathcal{C}$ *have SC-depth at most* d.
3. *If* $\mathcal{C}$ *has bounded treedepth, then* $\mathcal{C}$ *has bounded shrubdepth.*
4. *If* $\mathcal{C}$ *has bounded shrubdepth and* I *is a transduction that outputs colored graphs, then* $\mathsf{I}(\mathcal{C})$ *has bounded shrubdepth.*

Bounded expansion. A graph H is a *depth-r minor* of a graph G if H can be obtained from a subgraph of G by contracting mutually disjoint connected subgraphs of radius at most r. A class $\mathcal{C}$ of graphs has *bounded expansion* if there is a function $f : \mathbb{N} \to \mathbb{N}$ such that $\frac{|E(H)|}{|V(H)|} \leqslant f(r)$ for every $r \in \mathbb{N}$ and every depth-r minor H of a graph from $\mathcal{C}$. Examples include the class of planar graphs, or any class of graphs with bounded maximum degree.

The connection between treedepth and graph classes of bounded expansion can be established via *treedepth-p colorings*. For an integer p, a function $c : V(G) \to C$ is a treedepth-p coloring if, for every $i \leqslant p$ and set $X \subseteq V(G)$ with $|c(X)| = i$, the induced graph $G[X]$ has treedepth at most i. A graph class $\mathcal{C}$ has *low treedepth colorings* if for every $p \in \mathbb{N}$ there is a number N_p such that for every $G \in \mathcal{C}$ there exists a treedepth-p coloring $c \colon V(G) \to C$ with $|C| \leqslant N_p$.

▶ **Theorem 3.6** ([24])**.** *A class* $\mathcal{C}$ *of graphs has bounded expansion if, and only if, it has low treedepth colorings.*

4 Structurally bounded expansion and low shrubdepth covers

In this section we introduce two notions which generalize the concept of bounded expansion. First, we introduce classes of *structurally bounded expansion*. This notion arises from closing bounded expansion graph classes under transductions.

▶ **Definition 4.1.** A class $\mathcal{C}$ of graphs has *structurally bounded expansion* if there exists a class of graphs $\mathcal{D}$ of bounded expansion and a transduction I such that $\mathcal{C} \subseteq \mathsf{I}(\mathcal{D})$.

The second notion, *low shrubdepth covers*, arises from the low treedepth coloring characterization of bounded expansion (see Theorem 3.6) by replacing treedepth by its dense counterpart, shrubdepth. For convenience, we formally define this in terms of *covers*.

▶ **Definition 4.2.** A *cover* of a graph G is a family $\mathcal{U}_G$ of subsets of $V(G)$ such that $\bigcup \mathcal{U}_G = V(G)$. A cover $\mathcal{U}_G$ is a *p-cover*, where $p \in \mathbb{N}$, if every set of at most p vertices is contained in some $U \in \mathcal{U}_G$. If $\mathcal{C}$ is a class of graphs, then a (p-)cover of $\mathcal{C}$ is a family $\mathcal{U} = (\mathcal{U}_G)_{G \in \mathcal{C}}$, where $\mathcal{U}_G$ is a (p-)cover of G. The cover $\mathcal{U}$ is *finite* if $\sup\{|\mathcal{U}_G| \colon G \in \mathcal{C}\}$ is finite. Let $\mathcal{C}[\mathcal{U}]$ denote the class of graphs $\{G[U] \colon G \in \mathcal{C}, U \in \mathcal{U}_G\}$. We say that the cover $\mathcal{U}$ has *bounded treedepth* (respectively, bounded shrubdepth) if the class $\mathcal{C}[\mathcal{U}]$ has bounded treedepth (respectively, shrubdepth).

▶ **Example 4.3.** Let $\mathcal{T}$ be the class of trees and $p \in \mathbb{N}$ be a number. We construct a finite p-cover $\mathcal{U}$ of $\mathcal{T}$ which has bounded treedepth. Given a rooted tree T, let $\mathcal{U}_T = \{U_0, \dots, U_p\}$, where U_i is the set of vertices of T whose depth is not congruent to i modulo $p + 1$. Note that $T[U_i]$ is a forest of height p, and that $\mathcal{U}_T$ is a p-cover of T. Hence $\mathcal{U} = (\mathcal{U}_T)_{T \in \mathcal{T}}$ is a finite p-cover of $\mathcal{T}$ of bounded treedepth.

In analogy to low treedepth colorings, we can now characterize graph classes of bounded expansion using covers. We say that a class $\mathcal{C}$ of graphs has *low treedepth covers* if for every $p \in \mathbb{N}$ there is a finite p-cover $\mathcal{U}$ of $\mathcal{C}$ with bounded treedepth. The following lemma follows easily from Theorem 3.6.

▶ **Lemma 4.4.** *A class of graphs has bounded expansion if, and only if, it has low treedepth covers.*

We now define the second notion generalizing the concept of bounded expansion. The idea is to use low shrubdepth covers instead of low treedepth covers.

▶ **Definition 4.5.** A class $\mathcal{C}$ of structures has *low shrubdepth covers* if, and only if, for every $p \in \mathbb{N}$ there is a finite p-cover $\mathcal{U}$ of $\mathcal{C}$ with bounded shrubdepth.

It is easily seen that Lemma 4.4 together with Proposition 3.5 (3) imply that every class of bounded expansion has low shrubdepth covers.

5   Main Result

The main result of our paper is the followiong.

▶ **Theorem 5.1.** *A class of graphs has structurally bounded expansion if, and only if, it has low shrubdepth covers.*

We prove the left-to-right direction of Theorem 5.1 in Section 5.1 and the right-to-left direction in Section 5.2. But before we need some lemmas on classes of bounded shrubdepth.

The first lemma is an analogue of a known fact that every class of bounded cliquewidth which excludes a fixed bi-clique as a subgraph has bounded treewidth. The lemma is proved by an easy induction on the depth of the connection models.

▶ **Lemma 5.2.** *A class $\mathcal{C}$ of graphs has bounded treedepth if, and only if, $\mathcal{C}$ has bounded shrubdepth and all $G \in \mathcal{C}$ exclude some fixed bi-clique as a subgraph.*

The second lemma, Lemma 5.3 below, is substantially more involved, and is the combinatorial cornerstone of our approach. Essentially, it says that if $\mathcal{C}$ is a class of bounded shrubdepth, then a bounded-depth SC-decomposition of a graph $G \in \mathcal{C}$ can be computed using an *almost quantifier-free* transduction B. In other words, an SC-decomposition of a graph $G \in \mathcal{C}$ can be encoded in G using a finite number of unary predicates and reconstructed using function extension operations involving deterministic almost quantifier-free formulas.

▶ **Lemma 5.3.** *Let $\mathcal{B}$ be a class of graphs of bounded shrubdepth. Then there is a class $\mathcal{T}$ of colored trees of bounded height and a pair of transductions T and B such that T is almost quantifier-free, B is deterministic almost quantifier-free, $\mathsf{T}(\mathcal{B}) \subseteq \mathcal{T}$, $\mathsf{B}(\mathcal{T}) \subseteq \mathcal{B}$, and*

$$\mathsf{B}(\mathsf{T}(G)) = \{G\} \quad \text{for all } G \in \mathcal{B} \quad \text{and} \quad \mathsf{T}(\mathsf{B}(t)) \ni t \quad \text{for all } t \in \mathcal{T}.$$

Moreover, for any $G \in \mathcal{B}$, every $t \in \mathsf{T}(G)$ is an SC-decomposition of G.

The key ingredient in the proof of Lemma 5.3 is that we can provide an almost quantifier-free transduction that defines connected components in graphs of bounded shrubdepth, as formulated in the following lemma.

▶ **Lemma 5.4.** *Let $\mathcal{C}$ be a class of graphs of bounded shrubdepth. There is an almost quantifier-free transduction F such that for a given $G \in \mathcal{C}$, every output of F on G is equal to G enriched by a function $g \colon V(G) \to V(G)$ such that $g(v) = g(w)$ if, and only if, v and w are in the same connected component of G.*

In the proof of Lemma 5.4 we use the fact that bounded shrubdepth implies bounded diameter [14], and the connection between bounded shrubdepth and (bi-)cographs [15].

5.1   Structurally bounded expansion implies low shrubdepth covers

The left-to-right implication of Theorem 5.1 – that every graph class with structurally bounded expansion has low shrub-depth covers – follows from the two lemmas below, whose proofs are outlined later in this section.

▶ **Lemma 5.5.** *Let $\mathcal{C}$ be a class of colored graphs which has low shrubdepth covers. Then every transduction I is equivalent on $\mathcal{C}$ to an almost quantifier-free transduction J.*

▶ **Lemma 5.6.** *If a class of graphs $\mathcal{C}$ has low shrubdepth covers and J is an almost quantifier-free transduction which outputs graphs, then $\mathsf{J}(\mathcal{C})$ also has low shrubdepth covers.*

We first show how to conclude the left-to-right implication of Theorem 5.1 from Lemma 5.5 and Lemma 5.6. Let $\mathcal{C}$ be a class of bounded expansion and I be a transduction which outputs graphs. By Lemma 4.4, $\mathcal{C}$ has low treedepth covers and hence, by Lemma 3.5, also has low shrubdepth covers. Applying Lemma 5.5, we obtain an almost quantifier-free transduction J such that $\mathsf{I}(\mathcal{C}) = \mathsf{J}(\mathcal{C})$. By Lemma 5.6, we deduce that $\mathsf{J}(\mathcal{C}) = \mathsf{I}(\mathcal{C})$ has low shrubdepth covers, proving the left-to-right implication of Theorem 5.1.

In the remainder of this section we sketch the proofs of Lemma 5.5 and Lemma 5.6.

Proof outline for Lemma 5.5. Our proof of Lemma 5.5 is a quantifier elimination procedure similar to that of Dvořák et al. [7] and Grohe and Kreutzer [18]. First, we prove the statement for classes of colored trees of bounded depth, just as in [7, 18]. Lemma 5.3 allows us to immediately lift the statement to classes of bounded shrubdepth. Finally, this is lifted to classes with low shrubdepth covers.   ◀

Proof outline for Lemma 5.6. To prove the lemma, we first observe that every almost quantifier-free transduction is equivalent to a transduction which consists of a sequence of unary lifts followed by a deterministic almost quantifier-free transduction. Since, by Proposition 3.5, adding colors (the result of applying unary lifts) to a graph class of bounded shrubdepth will again result in a graph class of bounded shrubdepth, it suffices to prove the lemma for a deterministic almost quantifier-free transduction and a class of colored graphs $\mathcal{C}$.

Consider a p-cover $\mathcal{U}$ of $\mathcal{C}$ of bounded shrubdepth. As a very special case, suppose that I is a single quantifier-free extension operation that does not use any functions in the formula governing it. Then $\mathsf{I}(G[W]) = \mathsf{I}(G)[W]$ for $G \in \mathcal{C}$ and any $W \subseteq V(G)$. As a consequence, $\mathcal{U}_G$ is a p-cover of $\mathsf{I}(G)$ and hence $\mathsf{I}(\mathcal{C}[\mathcal{U}]) = \mathsf{I}(\mathcal{C})[\mathcal{U}]$ is a p-cover of $\mathsf{I}(\mathcal{C})$. Moreover, by Proposition 3.5, $\mathsf{I}(\mathcal{C})[\mathcal{U}]$ has bounded shrubdepth, so $\mathcal{U}$ is a finite cover of $\mathcal{C}$ of bounded shrubdepth. As p is arbitrary, this proves that $\mathsf{I}(\mathcal{C})$ has bounded shrubdepth covers.

For a general deterministic almost quantifier-free transduction I, the equality $\mathsf{I}(G[W]) = \mathsf{I}(G)[W]$ may fail, due to the fact that I may involve functions which reach outside of W. However, the value of any term involved in I depends only on a bounded number of vertices of G. This leads to the following lemma.

▶ **Lemma 5.7.** *For every deterministic almost quantifier-free transduction I there is a number c such that the following holds. For every graph G and vertex v of $\mathsf{I}(G)$ there is a set $S_v \subseteq V(G)$ of size at most c such that for any sets U, W with $W \subseteq V(\mathsf{I}(G))$ and $U \subseteq V(G)$, if $U \supseteq \bigcup_{v \in W} S_v$, then*

$$\mathsf{I}(G)[W] = \mathsf{I}(G[U])[W].$$

We now proceed to the general case. Choose any $p \in \mathbb{N}$. Let c be as in Lemma 5.7. Let $\mathcal{U}$ be a finite $(p \cdot c)$-cover of $\mathcal{U}$ of bounded shrubdepth. For a graph $G \in \mathcal{C}$ and a set $U \in \mathcal{U}_G$, define $W_U \subseteq \mathsf{I}(G)$ as the set of all vertices $v \in \mathsf{I}(G)$ such that $S_v \subseteq U$, where S_v is as in Lemma 5.7. From the lemma and the fact that classes of bounded shrubdepth are closed under transductions and induced subgraphs it follows easily that letting $\mathcal{W}_{\mathsf{I}(G)} = \{W_U : U \in \mathcal{U}_G\}$ yields a p-cover of $\mathsf{I}(G)$, which collectively, for all $G \in \mathcal{C}$, form a p-cover of $\mathsf{I}(\mathcal{C})$ which has bounded shrubdepth. Since p is arbitrary, this proves the lemma. ◀

5.2 Low shrubdepth covers imply structurally bounded expansion

For the right-to-left implication of Theorem 5.1, we prove the following statement.

▶ **Lemma 5.8.** *Let $\mathcal{C}$ be a class of graphs with low shrubdepth covers. There is a pair of transductions S and I, where S is almost quantifier-free and I is deterministic almost quantifier-free, such that $\mathsf{S}(\mathcal{C})$ is a class of colored graphs of bounded expansion and $\mathsf{I}(\mathsf{S}(G)) = \{G\}$ for each $G \in \mathcal{C}$.*

Clearly, the lemma implies that $\mathcal{C}$ has structurally bounded expansion, since it can be obtained as a result of a transduction I to a class $\mathsf{S}(\mathcal{C})$ of bounded expansion. Thus, the right-to-left implication of Theorem 5.1 is a consequence of the lemma.

Proof outline. Let $\mathcal{U}$ be a finite 2-cover of $\mathcal{C}$, and let $N = \sup\{|\mathcal{U}_G| : G \in \mathcal{C}\}$. We define a transduction S as follows. For a given $G \in \mathcal{C}$, let $\{U_1, \ldots, U_N\} = \mathcal{U}_G$. Given G, the transduction S introduces unary predicates $U_1, \ldots, U_N$, and then, using the transduction T from Lemma 5.3, computes the union $\bigcup_{i=1}^{N} \mathsf{T}(G[U_i])$, which is a union of N trees glued along the leaves. It is easy to see that the resulting graph excludes the bi-clique $K_{N+1,N+1}$ as a subgraph. By Lemma 5.6, the class $\mathsf{S}(\mathcal{C})$ has low shrubdepth covers. Any p-cover of $\mathsf{S}(\mathcal{C})$ of bounded shrubdepth has bounded treedepth by Lemma 5.2. Therefore, $\mathsf{S}(\mathcal{C})$ has low treedepth covers, so has bounded expansion.

The transduction I is easily constructed from the transduction B from Lemma 5.3, allowing to reconstruct the graph G from the colored union of the trees produced by S. ◀

6 Algorithmic aspects

In this section we give a partial result about efficient computability of transductions on classes with structurally bounded expansion. When we refer to the size of a structure in the algorithmic context, we refer to its total size, i.e., the sum of its universe size and the total sum of sizes of tuples in its relations.

Call a class $\mathcal{C}$ of graphs of structurally bounded expansion *efficiently decomposable* if there is a finite 2-cover $\mathcal{U}$ of $\mathcal{C}$ and an algorithm that, given a graph $G \in \mathcal{C}$, in linear time computes the cover $\mathcal{U}_G$ and for each $U \in \mathcal{U}_G$, an SC-decomposition S_U of depth at most d of the graph $G[U]$, for some constant d depending only on $\mathcal{C}$. Our result is as follows.

▶ **Theorem 6.1.** *Suppose J is a deterministic transduction and $\mathcal{C}$ is a class of graphs that has structurally bounded expansion and is efficiently decomposable. Then given a graph $G \in \mathcal{C}$, one may compute $\mathsf{J}(G)$ in time linear in the size of the input plus the size of the output.*

We remark that instead of efficient decomposability we could assume that the 2-cover $\mathcal{U}_G$ of a graph G and corresponding SC-decompositions for all $U \in \mathcal{U}_G$ is given together with G as input. If only the cover is given but not the SC-decompositions, we would obtain cubic running time because bounded shrubdepth implies bounded cliquewidth and we can compute

an approximate clique decomposition in cubic time [26]. Then, SC-decompositions of small height are definable in monadic second-order logic, and hence they can be computed in linear time using the result of Courcelle, Makowski and Rotics [3].

Observe that the theorem implies that we can efficiently evaluate a first-order sentence and enumerate all tuples satisfying a formula $\varphi(x_1, \ldots, x_k)$ on the given input graph, since this amounts to applying the theorem to a transduction consisting of a single extension operation. This strengthens the analogous result of Kazana and Segoufin [21] for classes of bounded expansion.

Proof outline. We will make use of transductions S and I constructed in the proof of Lemma 5.8. Recall that $\mathsf{S}(\mathcal{C})$ is a class of colored graphs of bounded expansion, I is deterministic, and $\mathsf{I}(\mathsf{S}(G)) = \{G\}$ for each $G \in \mathcal{C}$. Observe that J is equivalent to $\mathsf{S}; \mathsf{I}; \mathsf{J}$ on $\mathcal{C}$. Defining K as $\mathsf{I}; \mathsf{J}$, we get that $\mathsf{J}(G) = \mathsf{K}(\mathsf{S}(G))$ for $G \in \mathcal{C}$. Moreover, since I is deterministic, it follows that K is deterministic.

Let $G \in \mathcal{C}$ be an input graph. By efficient decomposability of $\mathcal{C}$, in linear time we can compute a cover $\mathcal{U}_G$ of G together with an SC-decomposition S_U of depth at most d of $G[U]$, for $U \in \mathcal{U}_G$. Each S_U is a colored tree, and by the construction described in the proof of Lemma 5.8, the trees S_U for $U \in \mathcal{U}_G$, glued along the leaves form a structure belonging to $\mathsf{S}(G)$. As $\mathsf{J}(G) = \mathsf{K}(\mathsf{S}(G))$, it suffices to apply the enumeration result of Kazana and Segoufin for classes of bounded expansion [21] to the colored graph $\mathsf{S}(G)$ and to all formulas occurring in the transduction K. ◀

7   Conclusion

In this paper we have provided a natural combinatorial characterization of graph classes that are first-order transductions of bounded expansion classes of graphs. Our characterization parallels the known characterization of bounded expansion classes by the existence of low treedepth decompositions, by replacing the notion of treedepth by shrubdepth. We believe that we have thereby taken a big step towards solving the model-checking problem for first-order logic on classes of structurally bounded expansion.

On the structural side we remark that transductions of bounded expansion graph classes are just the same as transductions of classes of structures of bounded expansion (i.e., classes whose Gaifman graphs or whose incidence encodings have bounded expansion). On the other hand, it remains an open question to characterize classes of relational structures, rather than just graphs, which are transductions of bounded expansion classes. We are lacking the analogue of Lemma 5.3; the problem is that within the proof we crucially use the characterization of shrubdepth via SC-depth, which works well for graphs but is unclear for structures of higher arity.

Finally, observe that classes of bounded expansion can be characterized among classes with structurally bounded expansion as those which are bi-clique free. It follows, that every monotone (i.e., subgraph closed) class of structurally bounded expansion has bounded expansion. Exactly the same statement holds characterizing bounded treedepth among bounded shrubdepth, and the second item holds for treewidth vs cliquewidth. In particular, for monotone graph classes all pairs of notions collapse.

We do not know how to extend our results to nowhere dense classes of graphs, mainly due to the fact that we do not know whether there exists a robust quantifier-elimination procedure for these graph classes. Obtaining such a procedure remains an open problem of prime importance in this field of research.

References

1 B. Courcelle. Graph rewriting: an algebraic and logic approach. In *Handbook of Theoretical Computer Science*, volume 2, chapter 5, pages 142–193. Elsevier, Amsterdam, 1990.

2 B. Courcelle. The monadic second-order logic of graphs I: recognizable sets of finite graphs. *Inform. and Comput.*, 85:12–75, 1990.

3 B. Courcelle, J. A. Makowsky, and U. Rotics. Linear time solvable optimization problems on graphs of bounded clique-width. *Theory of Computing Systems*, 33:125–150, 2000.

4 A. Dawar, M. Grohe, and S. Kreutzer. Locally excluding a minor. In *22nd Annual IEEE Symposium on Logic in Computer Science*, pages 270–279, 2007.

5 A. Durand and E. Grandjean. First-order queries on structures of bounded degree are computable with constant delay. *ACM Transactions on Computational Logic (TOCL)*, 8(4):21, 2007.

6 A. Durand, N. Schweikardt, and L. Segoufin. Enumerating answers to first-order queries over databases of low degree. In *Proceedings of the 33rd ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 121–131. ACM, 2014.

7 Z. Dvořák, D. Kráľ, and R. Thomas. Testing first-order properties for subclasses of sparse graphs. *Journal of the ACM*, 60:5 Article 36, 2013.

8 K. Eickmeyer and K. Kawarabayashi. FO model checking on map graphs. In *Proceedings of the 21st International Symposium on Fundamentals of Computation Theory, FCT 2017*, pages 204–216, 2017.

9 J. Flum and M. Grohe. Fixed-parameter tractability, definability, and model checking. *SIAM Journal of Computing*, 31:113–145, 2001.

10 M. Frick and M. Grohe. Deciding first-order properties of locally tree-decomposable structures. *Journal of the ACM*, 48:1148–1206, 2001.

11 J. Gajarský, P. Hliněný, D. Lokshtanov, J. Obdržálek, S. Ordyniak, M.S. Ramanujan, and S. Saurabh. Fo model checking on posets of bounded width. In *56th Annual Symposium on Foundations of Computer Science (FOCS), 2015*, pages 963–974. IEEE, 2015.

12 J. Gajarský, P. Hliněný, J. Obdržálek, D. Lokshtanov, and M. S. Ramanujan. A new perspective on fo model checking of dense graph classes. In *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science*, pages 176–184. ACM, 2016.

13 R. Ganian, P. Hliněný, J. Obdržálek, J. Schwartz, J. Teska, and D. Kráľ. Fo model checking of interval graphs. In *International Colloquium on Automata, Languages, and Programming*, pages 250–262. Springer, 2013.

14 R. Ganian, P. Hliněný, J. Nešetřil, J. Obdržálek, P. Ossona de Mendez, and R. Ramadurai. When trees grow low: Shrubs and fast MSO$_1$. In *MFCS 2012*, volume 7464 of *Lecture Notes in Computer Science*, pages 419–430. Springer-Verlag, 2012.

15 V. Giakoumakis and J.-M. Vanherpe. Bi-complement reducible graphs. *Adv. Appl. Math.*, 18:389–402, 1997.

16 M. Grohe. Generalized model-checking problems for first-order logic. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 12–26. Springer, 2001.

17 M. Grohe and S. Kreutzer. Methods for algorithmic meta-theorems. *Contemporary Mathematics*, 588, American Mathematical Society 2011.

18 M. Grohe and S. Kreutzer. Methods for algorithmic meta theorems. In *Model Theoretic Methods in Finite Combinatorics*, Contemporary mathematics, pages 181–206, 2011.

19 M. Grohe, S. Kreutzer, and S. Siebertz. Deciding first-order properties of nowhere dense graphs. *Journal of the ACM*, 64(3):17:1–17:32, 2017.

20 Petr Hlinený, Filip Pokrývka, and Bodhayan Roy. FO model checking of geometric graphs. In *12th International Symposium on Parameterized and Exact Computation, IPEC 2017*, pages 19:1–19:12, 2017. `doi:10.4230/LIPIcs.IPEC.2017.19`.

21 W. Kazana and L. Segoufin. Enumeration of first-order queries on classes of structures with bounded expansion. In *Proceedings of the 16th International Conference on Database Theory*, pages 10–20, 2013.

22 W. Kazana and L. Segoufin. Enumeration of first-order queries on classes of structures with bounded expansion. In *Proceedings of the 32nd ACM SIGMOD-SIGACT-SIGAI symposium on Principles of database systems*, pages 297–308. ACM, 2013.

23 O. Kwon, Mi. Pilipczuk, and S. Siebertz. On low rank-width colorings. In *43rd International Workshop on Graph-Theoretic Concepts in Computer Science, WG 2017*, pages 372–385, 2017.

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

25 J. Nešetřil and P. Ossona de Mendez. *Sparsity (Graphs, Structures, and Algorithms)*, volume 28 of *Algorithms and Combinatorics*. Springer, 2012. 465 pages.

26 S. Oum. Approximating rank-width and clique-width quickly. *ACM Transactions on Algorithms (TALG)*, 5(1):10, 2008.

27 D. Seese. Linear time computable problems and first-order descriptions. *Mathematical Structures in Computer Science*, 5:505–526, 1996.

28 L. Segoufin and W. Kazana. First-order query evaluation on structures of bounded degree. *Logical Methods in Computer Science*, 7, 2011.