Efficient First-Order Model-Checking Using Short Labels*

Bruno Courcelle**, Cyril Gavoille, and Mamadou Moustapha Kanté

Université de Bordeaux, LaBRI, CNRS, France
{courcell,gavoille,mamadou.kante}@labri.fr

Abstract. We prove that there exists an $O(\log(n))$-labeling scheme for every first-order formula with free set variables in every class of graphs that is *nicely locally cwd-decomposable*, which contains in particular, the *nicely locally tree-decomposable classes*. For every class of graphs of *bounded expansion* we prove that every *bounded formula* has an $O(\log(n))$-labeling scheme. We also prove that every quantifier-free formula has an $O(\log(n))$-labeling scheme in graphs of bounded *arboricity*. Some of these results are extended to counting queries.

1 Introduction

The model-checking problem for a class of structures $\mathcal{C}$ and a logical language $\mathcal{L}$ consists in deciding, for given $S \in \mathcal{C}$ and for some fixed sentence $\varphi \in \mathcal{L}$ if $S \models \varphi$, i.e., if S satisfies the property expressed by φ. More generally, if φ is a formula with free variables $x_1, \ldots, x_m$ one asks whether $S \models \varphi(d_1, \ldots, d_m)$ where $d_1, \ldots, d_m$ are values given to $x_1, \ldots, x_m$. One may also wish to list the set of m-tuples $(d_1, \ldots, d_m)$ that satisfy φ in S, or simply count them.

Polynomial time algorithms for these problems (for fixed φ) exist for certain classes of structures and certain logical languages. In this sense graphs of bounded degree "fit" with first-order (FO for short) logic [17,7] and graphs of bounded tree-width or clique-width "fit" with monadic second-order (MSO for short) logic. Frick and Grohe [8,9,11] have defined *Fixed Parameter Tractable* (FPT for short) algorithms for FO model-checking problems on classes of graphs that may have unbounded degree and tree-width (Definitions and Examples are given in Section 4). We will also use graph classes introduced by Nešetřil and Ossona de Mendez [15].

We will use the same tools for the following labeling problem: let be given a class of graphs $\mathcal{C}$ and a property $P(x_1, \ldots, x_m, Y_1, \ldots, Y_q)$ of vertices $x_1, \ldots, x_m$ and of sets of vertices $Y_1, \ldots, Y_q$ of graphs in $\mathcal{C}$. We want two algorithms, an algorithm $\mathcal{A}$ that attaches to each vertex u of every n-vertex graph of $\mathcal{C}$ a label $L(u)$, defined as a sequence of 0's and 1's of length $O(\log(n))$ or $\log^{O(1)}(n)$, and an algorithm $\mathcal{B}$ that checks property $P(a_1, \ldots, a_m, W_1, \ldots, W_q)$ by using

* Research supported by the project "Graph decompositions and Algorithms (GRAAL)" of "Agence Nationale pour la Recherche".
** Member of "Institut Universitaire de France".

F.P. Preparata, X. Wu, and J. Yin (Eds.): FAW 2008, LNCS 5059, pp. 159–170, 2008.

the labels. This latter algorithm must take as input the labels $L(a_1), \ldots, L(a_m)$ and the sets of labels $L(W_1), \ldots, L(W_q)$ of the sets $W_1, \ldots, W_q$ and tell whether $P(a_1, \ldots, a_m, W_1, \ldots, W_q)$ is true. Moreover each label $L(u)$ identifies the vertex u in the graph, which is possible with a sequence of length $\lfloor \log(n) \rfloor$. An $f(n)$-*labeling scheme* for a class of structures $\mathcal{C}$ is a pair $(\mathcal{A}, \mathcal{B})$ of functions solving the labeling problem and using labels of size at most $f(n)$ for n-vertex graphs of $\mathcal{C}$. Results of this type have been established for MSO logic by Courcelle and Vanicat [5] and, for particular properties (connectivity queries, that are expressible in MSO logic) by Courcelle and Twigg in [4] and by Courcelle et al. in [2].

Let us review the motivations for looking for *compact labelings of graphs*. By *compact*, we mean of length of order less than $O(n)$, where n is the number of vertices of the graph, hence in particular of length $\log^{O(1)}(n)$.

In distributed computing over a communication network with underlying graph G, nodes must act according to their local knowledge only. This knowledge can be updated by message passing. Due to space constraints on the local memory of each node, and on the sizes of messages, a distributed task cannot be solved by representing the whole graph G in each node or in each message, but it must rather manipulate more compact representations of G. Typically, the routing task may use routing tables, that are sublinear in the size of G (preferably of poly-logarithmic size), and short addresses transmitted in the headers of messages (of poly-logarithmic size too). As surveyed in [12] many distributed tasks can be optimized by the use of labels attached to vertices. Such labels should be usable even when the network has node or link crashes. They are called *forbidden-set labeling* schemes in [4]. In this framework local informations can be updated just by transmitting to all surviving nodes the list of (short) labels of all defected nodes and links, so that the surviving nodes can update their local information, e.g., their routing tables.

Let us comment about using set arguments. The forbidden (or defective) parts of a network are handled as a set of vertices passed to a query as an argument. This means that algorithm $\mathcal{A}$ computes the labels once and for all, independently of the possible forbidden parts of the network. In other words the labeling supports node deletions from the given network. (Edge deletions are supported in the labelings of [2] and [4].) If the network is augmented with new nodes and links, the labels must be recomputed. We leave this incremental extension as a topic for future research. Set arguments can be used to handle deletions, but also constraints, or queries like "what are the nodes that are at distance at most 3 of X and Y" where X and Y are two specified sets of nodes.

2 Notations and Definitions

All graphs and relational structures are finite. Let $\varphi(x_1, \ldots, x_m, Y_1, \ldots, Y_q)$ be a FO formula with free FO variables among $x_1, \ldots, x_m$ and free set variables among $Y_1, \ldots, Y_q$. Set variables are allowed in FO formulas but are not quantified. They occur in atomic formulas of the form "$y \in Y_i$". Gaifman's Theorem

[10] and its stronger versions are valid for such formulas because "$y \in Y_i$" is the same as "$R_i(y)$ holds" where R_i is a unary relation representing Y_i.

Let S be a relational structure of type relevant with signature $\mathcal{R}$, $S = \langle D_S, (R_S)_{R\in\mathcal{R}}\rangle$ with domain D_S. A *labeling* of S is an injective mapping $J : D_S \to \{0,1\}^*$ (or into some more convenient set Σ^* where Σ is a finite alphabet). If $Y \subseteq D_S$ we let $J(Y)$ be the family $(J(y))_{y\in Y}$. Clearly Y is defined from $J(Y)$.

For a formula $\varphi(x_1,\ldots,x_m,Y_1,\ldots,Y_q)$ and a class of structures $\mathcal{C}$ we are interested in the construction of two algorithms $\mathcal{A}$ and $\mathcal{B}$ doing the following:

1. $\mathcal{A}$ constructs for each $S \in \mathcal{C}$ a labeling J of S such that $|J(a)| = O(\log(n))$ for every $a \in D_S$, where $n = |D_S|$.
2. If J is computed from S by $\mathcal{A}$, then $\mathcal{B}$ takes as input an $(m+q)$-tuple $(J(a_1),\ldots,J(a_m),J(W_1),\ldots,J(W_q))$ and says correctly whether:

$$S \models \varphi(a_1,\ldots,a_m,W_1,\ldots,W_q).$$

In this case we say that the pair $(\mathcal{A},\mathcal{B})$ defines an $O(\log(n))$*-labeling supporting* the query defined by φ for the structures in $\mathcal{C}$.

Labelings based on logical descriptions of queries have been defined by Courcelle and Vanicat [5] for MSO queries and graphs of bounded clique-width (whence also of bounded tree-width). Applications to distance and connectivity queries in graphs of bounded clique-width and in planar graphs have been given by Courcelle and Twigg in [4] and by Courcelle, Gavoille, Kanté and Twigg in [2]. In the present article, we consider classes of graphs of unbounded clique-width and in particular, classes that are *locally decomposable* (Frick and Grohe [8,9]) and classes of *bounded expansion* (Nešetřil and Ossona de Mendez [15]). So, MSO logic cannot be achieved, we are thus obliged to consider FO logic.

In this extended abstract we only consider vertex-labeled graphs. The extension to structures can be done in a standard way through the so-called *Gaifman graphs.* A Σ-labeled graph is $G = \langle V_G, edg_G(\cdot,\cdot), (lab_{a,G})_{a\in\Sigma}\rangle$ (vertices, edge relations and unary relation for vertex labels).

By replacing everywhere "clique-width", "local clique-width", etc. by "tree-width", "local tree-width", etc., one can handle formulas with edge-set quantifications.

Definition 1 (Logic)
An FO formula $\varphi(x_1,\ldots,x_m,Y_1,\ldots,Y_q)$ is *basic bounded* if for some $p \in \mathbb{N}$ we have the following equivalence for all graphs G, all $a_1,\ldots,a_m \in V_G$ and all $W_1,\ldots,W_q \subseteq V_G$

$$G \models \varphi(a_1,\ldots,a_m,W_1,\ldots,W_q) \text{ iff } G[X] \models \varphi(a_1,\ldots,a_m,W_1\cap X,\ldots,W_q\cap X)$$

for some $X \subseteq V_G$ such that $|X| \le p$ and $a_1,\ldots,a_m \in X$. (If this is true for X, then $G[Y] \models \varphi(a_1,\ldots,a_m,W_1\cap Y,\ldots,W_q\cap Y)$ for every $Y \supseteq X$.)

An FO formula is *bounded* if it is a Boolean combination of basic bounded formulas. In particular, the negation of a basic bounded formula is not (in general) basic bounded, but it is bounded.

An FO formula $\varphi(x_1, \ldots, x_m, Y_1, \ldots, Y_q)$ is *t-local around* $(x_1, \ldots, x_m)$ if for every G and, every $a_1, \ldots, a_m \in V_G,\ W_1, \ldots, W_q \subseteq V_G$ we have

$$G \models \varphi(a_1, \ldots, a_m, W_1, \ldots, W_q) \text{ iff } G[N] \models \varphi(a_1, \ldots, a_m, W_1 \cap N, \ldots, W_q \cap N)$$

where $N = N_G^t(a_1, \ldots, a_m) = \{y \in V_G \mid d(y, a_i) \leq t \text{ for some } i = 1, \ldots, m\}$ and $d(u, v)$ is the length of a shortest undirected path between u and v in G.

An FO sentence is *basic* (t, s)*-local* if it is equivalent to a sentence of the form

$$\exists x_1 \cdots \exists x_s \left(\bigwedge_{1 \leq i < j \leq s} d(x_i, x_j) > 2t \ \wedge \ \bigwedge_{1 \leq i \leq s} \psi(x_i) \right)$$

where $\psi(x)$ is t-local around its unique free variable x.

Remark. The query $d(x, y) \leq r$ is basic bounded ($p = r + 1$) and t-local with $t = \lfloor r/2 \rfloor$. Its negation $d(x, y) > r$ is t-local and bounded (but not basic bounded).

3 Graphs

We are interested in on-line checking properties of networks in case of (reported) failures. Hence for each property of interest $\varphi(x_1, \ldots, x_m)$ we are not only interested in checking if $G \models \varphi(a_1, \ldots, a_m)$ by using $J(a_1), \ldots, J(a_m)$ where $a_1, \ldots, a_m \in V_G$ but also in checking $G \backslash X \models \varphi(a_1, \ldots, a_m)$ by using $J(a_1), \ldots, J(a_m)$ and $J(X)$ where $X \subseteq V_G - \{a_1, \ldots, a_m\}$ and $G \backslash X$ is the subgraph of G induced on $V_G - X$.

However, $G \backslash X \models \varphi(a_1, \ldots, a_m)$ for a FO formula $\varphi(x_1, \ldots, x_m)$ is equivalent to $G \models \varphi'(a_1, \ldots, a_m, X)$ and to $G_X \models \varphi''(a_1, \ldots, a_m)$ for FO formulas $\varphi'(x_1, \ldots, x_m, Y)$ and $\varphi''(x_1, \ldots, x_m)$ that are easy to write. We denote by G_X the graph G equipped with an additional vertex-label. Hence, we consider G_X as the structure G augmented with a unary relation lab such that $lab_{G_X}(u)$ holds iff $u \in X$. We will handle "holes" in graphs by means of set variables.

A graph has *arboricity at most k* if it is the union of k-edge disjoint forests (independently of the orientations of its edges).

Classes with *bounded expansion*, defined in [15] have several equivalent characterizations. We will use the following one: a class $\mathcal{C}$ has *bounded expansion* if for every integer p, there exists a constant $N(\mathcal{C}, p)$ such that for every $G \in \mathcal{C}$, one can partition V_G in at most $N(\mathcal{C}, p)$ parts such that any $i \leq p$ of them induce a subgraph of tree-width at most $i - 1$. (For $i = 1$ this implies that each part is a stable set, hence the partition can be seen as a *proper vertex-coloring.*)

4 Locally Decomposable Classes

We refer to [16] and to [3,5] for the definitions of *tree-width* and of *clique-width* respectively. (We denote by $cwd(G)$ the clique-width of a graph G). We will use the same notations as in [8,9]. Definition 2 is analogous to [9, Definition 5.1].

Definition 2

1. The *local clique-width* of a graph G is the function $lcw^G : \mathbb{N} \to \mathbb{N}$ defined by $lcw^G(t) := \max\{cwd(G[N_G^t(a)]) \mid a \in V_G\}$.
2. A class $\mathcal{C}$ of graphs has *bounded local clique-width* if there is a function $f : \mathbb{N} \to \mathbb{N}$ such that $lcw^G(t) \leq f(t)$ for every $G \in \mathcal{C}$ and $t \in \mathbb{N}$.

Examples

1. Every class of graphs of bounded clique-width has also bounded local clique-width since $cwd(G[A]) \leq cwd(G)$ for every $A \subseteq V_G$ (see [3]).
2. The classes of graphs of bounded local tree-width have bounded local clique-width since every class of graphs of bounded tree-width has bounded clique-width (see [3]). We can cite graphs of bounded degree and minor-closed classes of graphs that do not contain all apex-graphs (see [8,9]) as examples of classes of bounded local tree-width.
3. The class of unit-interval graphs has bounded local clique-width (using results from [14]) but neither bounded clique-width nor bounded local tree-width.
4. The class of interval graphs has not bounded local clique-width.

In order to obtain an $O(\log(n))$-labeling for certain classes of graphs of bounded local clique-width, we cover them as in [8,9], by graphs of small clique-width in a suitable way. In [8] a notion of *nicely locally tree-decomposable* class of structures was introduced. We will define a slightly more general notion.

Definition 3. Let $r, l \geq 1$ and $g : \mathbb{N} \to \mathbb{N}$. An (r, l, g)-*cwd cover* of a graph G is a family $\mathcal{T}$ of subsets of V_G such that:

1. For every $a \in V_G$ there exists a $U \in \mathcal{T}$ such that $N_G^r(a) \subseteq U$.
2. For each $U \in \mathcal{T}$ there exist less than l many $V \in \mathcal{T}$ such that $U \cap V \neq \emptyset$.
3. For each $U \in \mathcal{T}$ we have $cwd(G[U]) \leq g(1)$.

An (r, l, g)-cwd cover is *nice* if condition 3 is replaced by condition 3' below:

3'. For all $U_1, \ldots, U_q \in \mathcal{T}$ and $q \geq 1$ we have $cwd(G[U_1 \cup \cdots \cup U_q]) \leq g(q)$.

A class $\mathcal{C}$ of graphs is *locally cwd-decomposable* (resp. *nicely locally cwd-decomposable*) if there is a polynomial time algorithm that given a graph $G \in \mathcal{C}$ and $r \geq 1$, computes an (r, l, g)-cwd cover (resp. a nice (r, l, g)-cwd cover) of G for suitable l, g depending on r. (These two definitions are the same as in [9,8] where we substitute clique-width to tree-width.)

Examples

1. It is clear that every nicely locally cwd-decomposable class is locally cwd-decomposable and the converse is not true.
2. Each class of nicely locally tree-decomposable structures [8] is nicely locally cwd-decomposable.

3. Let G be a unit-interval graph. Using results from [14, Theorems 1,3 and Corollary 5] one can prove that G has an $(r, r, f(2r+1))$-cwd cover where f is the function that bounds local clique-width of unit-interval graphs. Thus every class of unit-interval graphs is locally cwd-decomposable.
4. Figure 1 shows inclusion relations between the many classes defined in Sections 3 and 4. It completes the diagram [9, Figure 2].

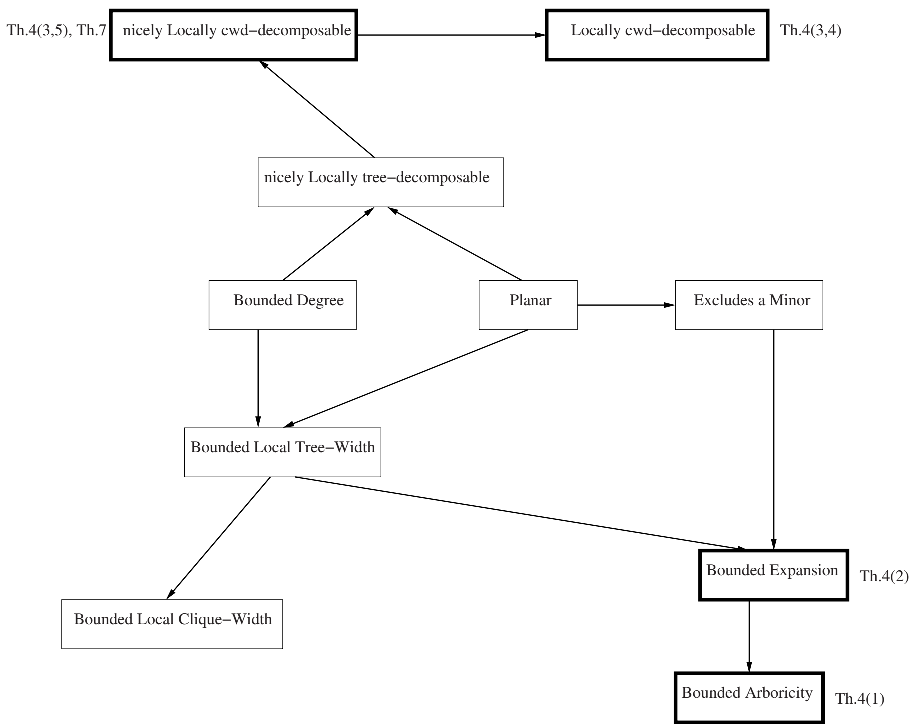

Fig. 1. Inclusion diagram indicating which results apply to which classes. An arrow means an inclusion of classes. Bold boxes are proved in this paper.

5 Results

The main results are as follows. In each case we consider labeled graphs over a finite set Σ of vertex-labels.

Theorem 4 (First Main Theorem). *There exist* $O(\log(n))$*-labeling schemes for the following queries and graph classes:*

1. *Quantifier-free queries in graphs of bounded arboricity.*
2. *Bounded FO queries for each class of graphs of bounded expansion.*
3. *Local queries with set arguments on locally cwd-decomposable classes.*

4. FO queries without set arguments on locally cwd-decomposable classes.
5. FO queries with set arguments on nicely locally cwd-decomposable classes.

We recall that for graphs G of clique-width at most k, there exists a cubic time algorithm that computes a cwd-term that defines G without being optimal [13]. (It uses $2^{k+1}-1$ labels, hence does not witness $cwd(G) \leq k$; however this term is enough for using [5].) And if a graph G has tree-width at most k, there exists a linear time algorithm that computes a tree-decomposition of width k of G [1]. We will also use results by Gaifman [10], Frick and Grohe [9,8] recalled below.

Theorem 5 ([10]). *Let $\varphi(\bar{x})$ be a FO formula where $\bar{x} = (x_1, \ldots, x_m)$. Then φ is logically equivalent to a Boolean combination $B(\varphi_1(\bar{u_1}), \ldots, \varphi_p(\bar{u_p}), \psi_1, \ldots, \psi_h)$ where:*

- *each φ_i is a t-local formula around $\bar{u_i} \subseteq \bar{x}$.*
- *each ψ_i is a basic (t', s)-local sentence.*

Moreover B can be computed effectively and, t, t' and s can be bounded in terms of m and the quantifier-rank of φ.

We will use a stronger form from [8]. Let $m, t \geq 1$. The *t-distance type* of an m-tuple $\bar{a}$ is the undirected graph $\epsilon = ([m], edg_\epsilon)$ where $edg_\epsilon(i,j)$ iff $d(a_i, a_j) \leq 2t+1$. The satisfaction of a t-distance type by an m-tuple can be expressed by a t-local formula:

$$\rho_{t,\epsilon}(x_1, \ldots, x_m) := \bigwedge_{(i,j) \in edg_\epsilon} d(x_i, x_j) \leq 2t+1 \ \wedge \bigwedge_{(i,j) \notin edg_\epsilon} d(x_i, x_j) > 2t+1.$$

We recall that Gaifman's Theorem and its variants extend to FO formulas with set variables.

Lemma 1 ([8]). *Let $\varphi(\bar{x}, Y_1, \ldots, Y_q)$ be a t-local formula around $\bar{x} = (x_1, \ldots, x_m)$, $m \geq 1$. For each t-distance type ϵ with $\epsilon_1, \ldots, \epsilon_p$ as connected components, one can compute a Boolean combination $F^{t,\epsilon}(\varphi_{1,1}, \ldots, \varphi_{1,j_1}, \ldots, \varphi_{p,1}, \ldots, \varphi_{p,j_p})$ of formulas $\varphi_{i,j}$ with FO free variables among those of $\bar{x}$ and set arguments in $\{Y_1, \ldots, Y_q\}$ such that:*

- *The FO free variables of each $\varphi_{i,j}$ are among $\bar{x} \mid \epsilon_i$ ($\bar{x} \mid \epsilon_i$ is the restriction of $\bar{x}$ to ϵ_i).*
- *$\varphi_{i,j}$ is t-local around $\bar{x} \mid \epsilon_i$.*
- *For each m-tuple $\bar{a}$, each q-tuple of sets $W_1, \ldots, W_q$, $G \models \rho_{t,\epsilon}(\bar{a}) \wedge \varphi(\bar{a}, W_1, \ldots, W_q)$ iff $G \models \rho_{t,\epsilon}(\bar{a}) \wedge F^{t,\epsilon}(\ldots, \varphi_{i,j}(\bar{a} \mid \epsilon_i, W_1, \ldots, W_q), \ldots)$.*

The lemma below is an easy adaptation of the results in [9].

Lemma 2 ([9]). *Let G be in a locally cwd-decomposable class. Every basic (t,s)-local sentence can be decided in polynomial time.*

We now give the proofs of each statement of Theorem 4 (except statement *1* because of space constraints). For clarity, we give them separately.

Proof (of Theorem 4 (2)). Let φ be a basic bounded formula with bound p and at least one free FO variable. We let $N = N(\mathcal{C}, p)$ and we partition V_G into $V_1 \uplus V_2 \uplus \cdots \uplus V_N$ as in the definition, $V_i \neq \emptyset$.

For every $\alpha \subseteq [N]$ of size p we let $V_\alpha = \bigcup_{i\in\alpha} V_i$ so that the tree-width of $G[V_\alpha]$ is at most $p-1$. Each vertex u belongs to less than $(N-1)^{p-1}$ sets V_α.

Hence a basic bounded formula $\varphi(x_1, \ldots, x_m, Y_1, \ldots, Y_q)$ is true in G iff it is true in some $G[X]$ with $|X| \leq p$, hence in some $G[V_\alpha]$ such that $x_1, \ldots, x_m \in V_\alpha$. For each α we construct a labeling J_α of $G[V_\alpha]$ (of tree-width at most $p-1$) supporting query φ by using [5]. We let $J(x) = \big(\ulcorner x \urcorner, \{(\ulcorner \alpha \urcorner, J_\alpha(x)) \mid x \in V_\alpha\}\big)$. We have $|J(x)| = O(\log(n))$.

We now explain how to decide φ by using the labels only. Given $J(a_1)$, $\ldots, J(a_m)$ we can determine all those sets α such that V_α contains $a_1, \ldots, a_m$. Using the components $J_\alpha(\cdot)$ of $J(a_1), \ldots, J(a_m)$ and the labels in $J(W_1), \ldots, J(W_q)$ we can determine if for some α, $G[V_\alpha] \models \varphi(a_1, \ldots, a_m, W_1 \cap V_\alpha, \ldots, W_q \cap V_\alpha)$ hence whether $G \models \varphi(a_1, \ldots, a_m, W_1, \ldots, W_q)$.

It remains to consider the case of a basic bounded formula of the form $\varphi(Y_1, \ldots, Y_q)$. For each α we determine the truth value t_α of $\varphi(\emptyset, \ldots, \emptyset)$ in $G[V_\alpha]$. The family of pairs (α, t_α) is of fixed size (depending on p) and is appended to $J(x)$ defined as above. From $J(W_1), \ldots, J(W_q)$ we get $D = \{\alpha \mid V_\alpha \cap (W_1 \cup \cdots \cup W_q) \neq \emptyset\}$.

By using the $J_\alpha(\cdot)$ components of the labels in $J(W_1) \cup \cdots \cup J(W_q)$ we can determine if for some $\alpha \in D$ we have $G[V_\alpha] \models \varphi(W_1 \cap V_\alpha, \ldots, W_q \cap V_\alpha)$. If one is found we conclude positively. Otherwise we look for some $t_\beta = True$ where $\beta \notin D$. This gives the final answer.

For a Boolean combination of basic bounded formulas $\varphi_1, \ldots, \varphi_t$ with associated labelings $J_1, \ldots, J_t$ we take the concatenation $J_1(x) \bullet J_2(x) \bullet \cdots \bullet J_t(x)$. It is of size $O(\log(n))$ and gives the desired result. □

Proof (of Theorem 4 (3)). Let $\varphi(\bar{x}, Y_1, \ldots, Y_q)$ be a t-local formula around $\bar{x} = (x_1, \ldots, x_m)$, $m \geq 1$. Then $G \models \varphi(\bar{a}, W_1, \ldots, W_q)$ iff $G[N_G^t(\bar{a})] \models \varphi(\bar{a}, W_1 \cap N_G^t(\bar{a}), \ldots, W_q \cap N_G^t(\bar{a}))$. Let ϵ be a t-distance type with $\epsilon_1, \ldots, \epsilon_p$ as connected components. By Lemma 1, $G \models \rho_{t,\epsilon}(\bar{a}) \wedge \varphi(\bar{a}, W_1, \ldots, W_q)$ iff $G \models \rho_{t,\epsilon}(\bar{a}) \wedge F^{t,\epsilon}(\varphi_{1,1}(\bar{a} \mid \epsilon_1, W_1, \ldots, W_q), \ldots, \varphi_{p,j_p}(\bar{a} \mid \epsilon_p, W_1, \ldots, W_q))$.

We let $\mathcal{T}$ be an (r, l, g)-cwd cover of G where $r = m(2t+1)$. We use such an r in order to warranty that if $a_1, \ldots, a_m$ are in a connected component of a t-distance type, there exists a $U \in \mathcal{T}$ such that $N_G^t(a_1, \ldots, a_m) \subseteq U$. For each vertex x there exist less than l many $V \in \mathcal{T}$ such that $x \in V$. We assume that each $U \in \mathcal{T}$ has an index encoded as a bit string $\ulcorner U \urcorner$. There are at most $n \cdot l$ sets in $\mathcal{T}$. Hence $\ulcorner U \urcorner$ has length $O(\log(n))$.

By the results of [5] we can label each vertex with a label $K(x)$ of length $O(\log(n))$ and decide in $O(\log(n))$-time if $d(u, v) \leq 2t+1$ or not by using $K(u)$ and $K(v)$[1]. We build a labeling K_U for each $U \in \mathcal{T}$; then for each x we let $K(x) = \Big(\ulcorner x \urcorner,\ \{\big(\ulcorner U \urcorner, K_U(x)\big) \mid N(x) \subseteq U\},\ \{\big(\ulcorner U \urcorner, K_U(x)\big) \mid N(x) \nsubseteq U\}\Big)$. where $N(x) = N_G^{2t+1}(x)$. (We always assume that $x \in N_G^t(x)$ for all $t \in \mathbb{N}$.)

[1] For checking if $d(u, v) \leq 2t+1$, an (r', l', g')-cwd cover suffices, with $r' = 2t+1$.

By [5] for each $\varphi_{i,j}(\bar{x} \mid \epsilon_i, Y_1, \ldots, Y_q)$ and each $U \in \mathcal{T}$ we can label each vertex $x \in U$ with $J^{\epsilon}_{i,j,U}(x)$ of length $O(\log(n))$ and decide $\varphi_{i,j}(\bar{a} \mid \epsilon_i, W_1, \ldots, W_q)$ in $G[U]$ by using $\left(J^{\epsilon}_{i,j,U}(b)\right)_{b \in \bar{a} \ |\epsilon_i}$ and $J^{\epsilon}_{i,j,U}(W_1 \cap U), \ldots, J^{\epsilon}_{i,j,U}(W_q \cap U)$. For each x we let

$$J_\epsilon(x) := \left\{ \left(\ulcorner U \urcorner, J^{\epsilon}_{1,1,U}(x), \ldots, J^{\epsilon}_{1,j_1,U}(x), \ldots, J^{\epsilon}_{p,1,U}(x), \ldots, J^{\epsilon}_{p,j_p,U}(x)\right) \mid N^t_G(x) \subseteq U \right\}.$$

It is clear that $|J_\epsilon(x)| = O(\log(n))$ since each x is in less than l many $V \in \mathcal{T}$. There exist at most $k' = 2^{k(k-1)/2}$ t-distance type graphs; we enumerate them by $\epsilon^1, \ldots, \epsilon^{k'}$. For each x we let $J(x) := (\ulcorner x \urcorner, K(x), J_{\epsilon^1}(x), \ldots, J_{\epsilon^{k'}}(x))$.

From the labels $K(x)$, we can determine $\{\ulcorner U \urcorner \mid U \in \mathcal{T},\ x \in U\}$, hence the sets $U \in \mathcal{T}$ such that $W \cap U \neq \emptyset,\ W \subseteq V_G$, where W is a set argument. It is clear that $J(x)$ is of length $O(\log(n))$ and is computed in polynomial time since $\mathcal{T}$ is computed in polynomial time and each J_ϵ is computed in polynomial time. We now explain how to decide whether $G \models \varphi(a_1, \ldots, a_m, W_1, \ldots, W_q)$ by using $J(a_1), \ldots, J(a_m)$ and $J(W_1), \ldots, J(W_q)$.

By using $K(a_1), \ldots, K(a_m)$ from $J(a_1), \ldots, J(a_m)$ we can construct the t-distance type ϵ satisfied by $a_1, \ldots, a_m$; let $\epsilon_1, \ldots, \epsilon_p$ be the connected components of ϵ. From each $J(a_i)$ we can recover $J_\epsilon(a_i)$. For each $\bar{a} \mid \epsilon_i$ there exists at least one $U \in \mathcal{T}$ such that $N^t_G(\bar{a} \mid \epsilon_i) \subseteq U$. We can recover them (there are less than l) from the $J(b),\ b \in \bar{a} \mid \epsilon_i$. We can now decide whether $G \models F^{t,\epsilon}(\varphi_{1,1}(\bar{a} \mid \epsilon_1, W_1 \cap U_1, \ldots, W_q \cap U_1), \ldots, \varphi_{p,j_p}(\bar{a} \mid \epsilon_p, W_1 \cap U_p, \ldots, W_q \cap U_p))$ for some $U_1, \ldots, U_p$ determined from $J(a_1), \ldots, J(a_m)$. By using also $J(W_1), \ldots, J(W_q)$ we can determine the sets $W_i \cap U_j$ and this is sufficient by Lemma 1. □

Proof (of Theorem 4 (4)). Let $\varphi(x_1, \ldots, x_m)$ be a FO formula without set arguments. By Theorem 5 φ is equivalent to a Boolean combination $B(\varphi_1(\bar{x}), \ldots, \varphi_p(\bar{x}), \psi_1, \ldots, \psi_h)$ where φ_i is t-local and ψ_i is a basic (t', s)-local sentence for suitable t, t', s.

By Lemma 2 one can decide in polynomial time each sentence ψ_i. Let $b = (b_1, \ldots, b_h)$ where $b_i = 1$ if G satisfies ψ_i and 0 otherwise. For each $1 \leq i \leq p$ we construct a labeling J_i supporting query φ_i by Theorem 4 (3) (G belongs to a locally cwd-decomposable class and φ_i is a t-local formula around $\bar{x}$). For each x we let $J(x) := (\ulcorner x \urcorner, J_1(x), \ldots, J_p(x), b)$.

It is clear that $|J(x)| = O(\log(n))$. Since from b one can recover the truth value of each sentence ψ_i, we can decide whether $G \models \varphi(a_1, \ldots, a_m)$ by using $J(a_1), \ldots, J(a_m)$, the truth values of $\varphi_i(\bar{a})$ and b. □

Proof (of Theorem 4 (5)). By Theorem 4 (3) it is sufficient to consider FO formulas $\varphi(Y_1, \ldots, Y_q)$ of the form:

$$\exists x_1 \cdots \exists x_m \left(\bigwedge_{1 \leq i < j \leq m} d(x_i, x_j) > 2t \ \wedge \bigwedge_{1 \leq i \leq m} \psi(x_i, Y_1, \ldots, Y_q) \right)$$

where $\psi(x, Y_1, \ldots, Y_q)$ is t-local around x. We show how to check their validity by means of $O(\log(n))$-labelings.

We consider for purpose of clarity the particular case of $m = 2$. Let $\mathcal{T}$ be a nice (r, l, g)-cwd cover of G where $r = 2t + 1$. We let $K(U) = \{x \in U \mid N_G^{2t}(x) \subseteq U\}$ (the $2t$-kernel of U (see [8])).

We let γ be a distance-2 coloring of the intersection graph of $\mathcal{T}$ (vertices at distance 1 or 2 have different colors). For every 2 colors i, j we let $G_{i,j}$ be the graph induced by the union of the blocks $U \in \mathcal{T}$ of colors i and j.

Claim 1. *$cwd(G_{i,j}) \leq g(2)$.*

Proof (of Claim 1). $G_{i,j}$ is a disjoint union of sets U in $\mathcal{T}$ and of unions $U \cup U'$ with $U \cap U' \neq \emptyset$ for $U, U' \in \mathcal{T}$. This union is disjoint because if $U \cup U'$ with $U \cap U' \neq \emptyset$ would meet some $U'' \in \mathcal{T}$, $U'' \neq U$, $U'' \neq U'$, then we would have $\gamma(U) = i$, $\gamma(U') = j$ and U'' meets U or U'. It cannot have color i or j because γ is a distance-2 coloring. Since $cwd(G[U \cup U']) \leq g(2)$, we are done. □

Claim 2. *Let $x \in K(U)$ and $y \in K(U')$ for some $U, U' \in \mathcal{T}$. Then $d_G(x, y) > 2t$ iff $d_{G[U \cup U']}(x, y) > 2t$.*

Proof (of Claim 2). The "if direction" is clear since distance increases if we go to induced subgraphs.

For the "converse direction", we let $d_G(x, y) \leq 2t$; there exists a path of length $\leq 2t$ from x to y. This path is in $U \cup U'$ since $x \in K(U)$ and $y \in K(U')$. Hence it is also in $G[U \cup U']$, hence $d_{G[U \cup U']} \leq 2t$. □

Let us now give to each vertex x of G the smallest color i such that $x \in K(U)$ and $\gamma(U) = i$. Hence a vertex has one and only one color. For each pair i, j we consider the formula $\psi_{i,j}$ (possibly $j = i$):

$$\exists x, y\Big(d(x, y) > 2t \ \wedge \ \psi(x, Y_1, \ldots, Y_q) \ \wedge \\ \psi(y, Y_1, \ldots, Y_q) \ \wedge \text{ “}x \text{ has color } i\text{”} \ \wedge \text{ “}y \text{ has color } j\text{”}\Big)$$

We use [5] to construct a labeling $J_{i,j}$ for the formula $\psi_{i,j}$ in the graph $G_{i,j}$ (with vertices colored by i or j, that is, we use new unary "color" predicates). We compute the truth value $b_{i,j}$ of $\psi_{i,j}(\emptyset, \ldots, \emptyset)$ in $G_{i,j}$; we get a vector $\boldsymbol{b}$ of fixed length. We also label each vertex x by its color. We concatenate to that b and the $J_{i,j}(x)$ for $x \in V_{G_{i,j}}$, giving $J(x)$.

From $J(W_1), \ldots, J(W_q)$ we can determine those $G_{i,j}$ such that $V_{G_{i,j}} \cap (W_1 \cup \cdots \cup W_q) \neq \emptyset$, and check if for one of them $G_{i,j} \models \psi_{i,j}(W_1, \ldots, W_q)$. If one is found we are done. Otherwise we use the $b_{i,j}$'s to look for $G_{i,j}$ such that $G_{i,j} \models \psi_{i,j}(\emptyset, \ldots, \emptyset)$ and $(W_1 \cup \cdots \cup W_q) \cap V_{G_{i,j}} = \emptyset$. This gives the correct results because of the following facts:

- If x, y satisfy the formula φ, then $x \in K(U)$, $y \in K(U')$ (possibly $U = U'$) and $d_G(x, y) > 2t$ implies $d_{G_{i,j}}(x, y) > 2t$, hence $G_{i,j} \models \psi_{i,j}(W_1, \ldots, W_q)$ where $i = \gamma(U)$ and $j = \gamma(U')$.
- If $G_{i,j} \models \psi_{i,j}(W_1, \ldots, W_q)$ then we get $G \models \varphi(W_1, \ldots, W_q)$ by similar argument (in particular $d_{G_{i,j}}(x, y) > 2t$ implies $d_{G[U \cup U']}(x, y) > 2t$ which implies that $d_G(x, y) > 2t$ by Claim 2).

For $m = 1$, the proof is similar with γ a proper (distance-1) coloring and we use G_i instead of $G_{i,j}$.

For the case $m > 2$, the proof is the same: one takes for γ a distance-m proper coloring of the intersection graph, one considers graphs $G_{i_1,\ldots,i_m}$ defined as (disjoint) unions of sets $U_1 \cup \cdots \cup U_m$ for $U_1, \ldots, U_m$ in $\mathcal{T}$, of respective colors $i_1, \ldots, i_m$ and $cwd(G[U_1 \cup \cdots \cup U_m]) \leq g(m)$. This terminates the proof of Theorem 4. □

Let us ask a very general question: what can be done with $O(\log(n))$ labels ? Here is a fact that limits the extension of these results.

Proposition 1. *There exists a constant $c > 0$ such that one cannot handle all local or bounded FO queries for n-vertex graphs of arboricity at most 2 with labels of size $c\sqrt{n}$.*

We now discuss extension to counting queries. Let $\varphi(x_1, \ldots, x_m, Y_1, \ldots, Y_q)$ be a MSO formula and S be a finite structure. For $W_1, \ldots, W_q \subseteq D_S$ we define

$$\#_S\varphi(W_1, \ldots, W_q) := \left| \left\{ (a_1, \ldots, a_m) \in D_S^m \mid S \models \varphi(a_1, \ldots, a_m, W_1, \ldots, W_q) \right\} \right|$$

A *counting query* consists in determining $\#_S\varphi(W_1, \ldots, W_q)$ for given $(W_1, \ldots, W_q)$. We will need the following extension of the results of [5].

Theorem 6. *Let $\varphi(x_1, \ldots, x_m, Y_1, \ldots, Y_q)$ be a MSO formula over labeled graphs and $k \in \mathbb{N}$. There exists an $O(\log^2(n))$-labeling scheme for n-vertex graphs of clique-width or tree-width at most k supporting the counting query $\#_G\varphi$. For computing $\#_G\varphi(W_1, \ldots, W_q)$ modulo some fixed integer s, or up to s (threshold counting) we need only labels of size $O(\log(n))$.*

We now state our second main theorem. The proof is omitted because of space constraints.

Theorem 7 (Second Main Theorem). *There exists an $O(\log^2(n))$-labeling scheme for counting queries based on FO formulas for nicely locally cwd-decomposable classes. $O(\log(n))$ is enough for modulo counting.*

We conjecture that the results of Theorem 4 (3,4,5) extend to classes of graphs excluding, or locally excluding a minor [6,11].

Question. Does there exist an $O(\log(n))$-labeling scheme for FO formulas with set arguments on locally cwd-decomposable classes ?

References

1. Bodlaender, H.L.: A Linear-Time Algorithm for Finding Tree-Decompositions of Small Tree-width. SIAM J. Comput. 25(6), 1305–1317 (1996)
2. Courcelle, B., Gavoille, C., Kanté, M.M., Twigg, A.: Optimal Labeling for Connectivity Checking in Planar Networks with Obstacles. (manuscript, 2008); An extended abstract will appear in Electronic Notes in Discrete Mathematics. In: Proceedings of the first Conference Topological and Geometric Graph Theory, Paris (2008)

3. Courcelle, B., Olariu, S.: Upper Bounds to the Clique-Width of Graphs. Discrete Applied Mathematics 101(1-3), 77–114 (2000)
4. Courcelle, B., Twigg, A.: Compact Forbidden-Set Routing. In: Thomas, W., Weil, P. (eds.) STACS 2007. LNCS, vol. 4393, pp. 37–48. Springer, Heidelberg (2007)
5. Courcelle, B., Vanicat, R.: Query Efficient Implementation of Graphs of Bounded Clique-Width. Discrete Applied Mathematics 131(1), 129–150 (2003)
6. Dawar, A., Grohe, M., Kreutzer, S.: Locally Excluding a Minor. In: 22^{nd} IEEE Symposium on Logic in Computer Science (LICS), pp. 270–279. IEEE Computer Society, Los Alamitos (2007)
7. Durand, A., Grandjean, E.: First-Order Queries on Structures of Bounded Degree are Computable with Constant Delay. ACM Trans. Comput. Log 8(4) (2007)
8. Frick, M.: Generalized Model-Checking over Locally Tree-Decomposable Classes. Theory Comput. Syst. 37(1), 157–191 (2004)
9. Frick, M., Grohe, M.: Deciding First-Order Properties of Locally Tree-Decomposable Structures. J. ACM 48(1), 1184–1206 (2001)
10. Gaifman, H.: On Local and Non-Local Properties. In: Proceedings of the Herbrand Symposium Logic Colloquium 1981, pp. 105–135 (1982)
11. Grohe, M.: Logic, Graphs and Algorithms. In: Flum, Grädel, Wilke (eds.) Logic, Automata, History and Perspectives, pp. 357–422. Amsterdam University Press (2007)
12. Gavoille, C., Peleg, D.: Compact and Localized Distributed Data Structures. Distributed Computing 16(2-3), 111–120 (2003)
13. Hliněný, P., Oum, S.: Finding Branch-Decompositions and Rank-Decompositions. In: Arge, L., Hoffmann, M., Welzl, E. (eds.) ESA 2007. LNCS, vol. 4698, pp. 163–174. Springer, Heidelberg (2007)
14. Lozin, V.: Clique-Width of Unit Interval Graphs. arXiv:0709.1935 (manuscript, 2007)
15. Nešetřil, J., Ossona de Mendez, P.: Linear Time Low Tree-Width Partitions and Algorithmic Consequences. In: Kleinberg, J.M. (ed.) 38^{th} Annual ACM Symposium on Theory of Computing (STOC), pp. 391–400. ACM, New York (2006)
16. Robertson, N., Seymour, P.: Graph Minors V: Excluding a Planar Graph. J. Combin. Theory Ser. B 41(1), 92–114 (1986)
17. Seese, D.: Linear Time Computable Problems and First-Order Descriptions. Mathematical Structures in Computer Science 6(6), 505–526 (1996)