From Parikh's Theorem to Many-Sorted Spectra

J.A. Makowsky*

Department of Computer Science
Technion–Israel Institute of Technology, Haifa, Israel
{janos}@cs.technion.ac.il

Abstract. We discuss a generalization of Parikh's Theorem for context-free languages to classes of many-sorted relational structures which are both definable in Monadic Second Order Logic and which are of bounded patch-width. Patch-width is a generalization of both tree-width and clique-width. This gives a powerful unifying tool to prove that certain classes of graphs are of unbounded width.

For R. Parikh,
at the occasion of
his 70th birthday

1 Generalizing Parikh's Theorem

R. Parikh's celebrated theorem, first proved in [Par66], counts the number of occurrences of letters in words of a context-free languages L over an alphabet of k letters. For a given word w, the numbers of these occurrences is denoted by a vector $n(w) \in \mathbb{N}^k$, and the theorem states

Theorem 1 (Parikh 1966). *For a context-free language L, the set $Par(L) = \{n(w) \in \mathbb{N}^k : w \in L\}$ is semi-linear.*

A set $X \subseteq \mathbb{N}^s$ is *linear in* $\mathbb{N}^s$ iff there is vector $\bar{a} \in \mathbb{N}^s$ and a matrix $M \in \mathbb{N}^{s\times r}$ such that $X = A_{\bar{a},\bar{M}} = \{\bar{b} \in \mathbb{N}^s :$ there is $\bar{u} \in \mathbb{N}^r$ with $\bar{b} = \bar{a} + M \cdot \bar{u}\}$. Singletons are linear sets with $M = 0$. If $M \neq 0$ the series is *nontrivial.* $X \subseteq \mathbb{N}^s$ is *semi-linear in* $\mathbb{N}^s$ iff X is a finite union of linear sets $A_i \subseteq \mathbb{N}^s$. For $s = 1$ the semi-linear sets are exactly the ultimately periodic sets. The terminology is from [Par66], and has since become standard terminology in formal language theory. Several alternative proofs of Parikh's Theorem have appeared since. D. Pilling [Pil73] put it into a more algebraic form, and more recently, L. Aceto, Z. Esik and A. Ingolfsdottir [AÉI02] showed that it depends only on a few equational properties of least pre-fixed points. B. Courcelle [Cou95]. has generalized Theorem 1 further to certain graph grammars, and relational structures which are

* Partially supported by the Israel Science Foundation for the project "Model Theoretic Interpretations of Counting Functions" (2007-2010) and the Grant for Promotion of Research by the Technion–Israel Institute of Technology.

definable in Monadic Second Order Logic (MSOL) in labeled trees, counting not only letter occurrences but the size of MSOL-definable subsets. In [FM04] a similar generalization was proven, inspired by a theorem due to Gurevich and Shelah [GS03] on spectra of MSOL-sentences with one unary function symbol and a finite number of unary relation symbols. The results of [FM04] are formulated in a model theoretic framework rather than using the language of graph grammars. But it turned out that some of their result could have been obtained also using the techniques of [Cou95].

In this paper we explain discuss these generalizations of Parikh's Theorem without detailed proofs, but with the emphasis on the concepts involved and on applications. Like in the well known characterization of regular languages using Monadic Second Order Logic (MSOL) we also use MSOL, and an extension thereof, CMSOL, which allows for modular counting quantifiers. However, the languages in Parikh's Theorem are replaced by *arbitrary finite relational structures* which are of *bounded width*. The most general notion of width we shall use is *patch-width*, which was first introduced in [FM04]. It generalizes both tree-width and clique-width. The detailed discussion of these notions of width is given in Section 4. Finally, like in Courcelle's generalization of Parikh's Theorem, rather than counting occurrences of letters, we count cardinalities of CMSOL-definable unary relations. We shall first explain this for the case where these relations are given by unary predicates, where we speak of many-sorted spectra. The applications consist mostly in proving that certain classes of graphs and relational structures have unbounded tree-width, clique-width or patch-width. These are given in Section 5.

We assume the reader is familiar with the basics of Monadic Second Order Logic MSOL, cf. [Cou92,EF95]. For convenience, we collect the basics in Section 2. Otherwise this paper is rather self-contained.

2 Monadic Second Order Logic and its Extension

First Order Logic FOL restricts quantification to elements of the structure, Monadic Second Order Logic MSOL also allows for quantification over subsets, but not over binary relations,, or relations of arity $r \geq 2$. The logic MSOL can be extended by modular counting quantifiers $C_{k,m}$, where $C_{k,m}x\ \phi(x)$ is interpreted as "there are, modulo m, exactly k elements satisfying $\phi(x)$". We denote the extension of MSOL obtained by adding, for all $k, m \in \mathbb{N}$ the quantifiers $C_{k,m}$, by CMSOL. Over structures which have a linear order the quantifiers $C_{k,m}x\ \phi(x)$ can be eliminated without loss of expressive power.

Typical graph theoretic concepts expressible in FOL are the presence or absence (up to isomorphism) of a fixed (induced) subgraph H, and fixed lower or upper bounds on the degree of the vertices (hence also r-regularity). Typical graph theoretic concepts expressible in MSOL but not in FOL are connectivity, k-connectivity, reachability, k-colorability (of the vertices), and the presence or absence of a fixed (topological) minor. The latter includes planarity, and more generally, graphs of a fixed genus g. Typical graph theoretic concepts expressible in CMSOL but not in MSOL are the existence of an Eulerian circuit (path), the size of a connected component being a multiple of k, and the number of connected components is a multiple of k. All the non-definability statements above can be proved using Ehrenfeucht-Fraïssé games. The definability statements are straightforward.

3 Spectra of Sentences of Monadic Second Order Logic

Let ϕ be sentence of Monadic Second Order Logic MSOL over a vocabulary τ. The *spectrum* $\text{SPEC}(\phi)$ is the set of finite cardinalities $n \in \mathbb{N}$ of finite model of ϕ. We note that for unary languages, Parikh's Theorem looks at the spectrum of context-free languages. H. Scholz introduced spectra in [Sch52], where he asked to characterize spectra. Spectra of first order sentences have been studied ever since. For a history and survey of the study of spectra, cf. [DJMM08].

3.1 Spectra of sentences with one unary function symbol

Let $S \subseteq \mathbb{N}$ be an ultimately periodic set. It is not difficult to construct even a first order sentence ϕ such the $\text{SPEC}(\phi) = S$. Surprisingly, ultimately periodic sets are precisely the spectra of sentences with one unary function symbol, and a finite set of unary relation symbols. [DFL97], [GS03].

Theorem 2 (Durand, Fagin, Loescher (1997), Gurevich, Shelah (2003)). *Let ϕ be a sentence of* $\text{MSOL}(\tau)$ *where τ consists of*

- *finitely many unary relation symbols,*
- *one unary function and equality only.*

Then $\text{SPEC}(\phi)$ *is ultimately periodic,*

3.2 From one unary function to bounded tree-width

The finite structures which have only one unary function consist of disjoint unions of components of the form given in Figure 1. They look like directed

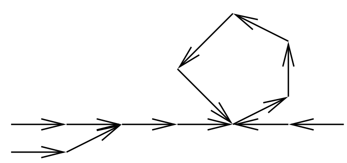

Fig. 1. Models of one unary function

forests where the roots are replaced by a directed cycle. The unary predicates are just colors attached to the nodes. The similarity of labeled graphs to labeled trees can be measured by the notion of *tree-width*, and in fact, these structures have tree width at most 2. The necessary background on tree-width will be given in Subsection 4.1. Inspired by Theorem 2, E. Fischer and J.A. Makowsky [FM04] generalized Theorem 2 by replacing the restriction on the vocabulary by a purely model theoretic condition involving the width of a relational structure.

Theorem 3 (E. Fischer and J.A. Makowsky (2004)).
Let ϕ be an MSOL *sentence and $k \in \mathbb{N}$. Assume that all the models of ϕ have tree-width at most k. Then* $\text{SPEC}(\phi)$ *is ultimately periodic.*

To prove Theorem 3, and its generalizations below, the authors use three tools:

(i) A generalization of the Feferman-Vaught Theorem for k-sums of labeled graphs to the logic CMSOL, due to B. Courcelle, [Cou90], and further refined by J.A. Makowsky in [Mak04].
(ii) A reduction of the problem to spectra of labeled trees by a technique first used by B. Courcelle in [Cou95] in his study of graph grammars.
(iii) An adaptation of the Pumping Lemma for labeled trees, cf. [GS97].

The proof of Theorem 3 is quite general. However, the details are rather technical. Its proof can be adapted to stronger logics, in particular to CMSOL introduced in Section 2. Second, one observes that very little of the definition of tree-width is used in the proof. The techniques used in the proof of Theorem 3 can be adapted to other notions of width of relational structures, such as *clique-width*, which was introduced first in [CER93] and studied more systematically in [CO00], and to *patch-width*, introduced in [FM04]. These notions will be discussed in detail in Section 4. Before we get to the various definitions of width we refer to any of these notions just as the *width* of a structure.

3.3 Many-sorted Spectra

We want to generalize Theorem 1 to spectra. Rather than counting occurrences of letters, we look at many-sorted structures and the sizes of the different sorts, which we call many-sorted spectra.

For our purposes, a *k-sorted τ-structure* is a structure of the form

$$\mathfrak{A} = \langle A, P_1, \ldots, P_k, R_1, \ldots R_m \rangle$$

where A is a finite set, $P_i \subseteq A$, $\bigcup_{i=1}^{k} P_i = A$ and for all $i, j \leq k, i \neq j$ we have that $P_i \cap P_j = \emptyset$. The sets P_i are the sorts of $\mathfrak{A}$. Furthermore, the relations $R_\ell, \ell \leq m$ are typed. τ now consists of the relation symbols $\mathbf{P}_i, \mathbf{R}_\ell$, which are interpreted in $\mathfrak{A}$ in the obvious way. For a finite k-sorted τ-structure $\mathfrak{A}$ we denote by $\text{MCARD}(\mathfrak{A})$ the k-tuple $(n_1, \ldots, n_k)$ of the cardinalities n_i of P_i. The many-sorted spectrum $\text{MSPEC}(\phi)$ of a k-sorted τ-sentence $\phi \in \text{CMSOL}(\tau)$ is the set

$$\text{MSPEC}(\phi) = \{\text{MCARD}(\mathfrak{A}) \in \mathbb{N}^k : \mathfrak{A} \models \phi\}.$$

These definitions can be extended to MSOL augmented by the modular counting quantifiers $C_{k,m}$, where $C_{k,m}x\ \phi(x)$ is interpreted as "there are, modulo m, exactly k elements satisfying $\phi(x)$". We denote the extension of MSOL obtained by adding, for all $k, m \in \mathbb{N}$ the quantifiers $C_{k,m}$, by CMSOL.

In [FM04] the following theorem is proved:

Theorem 4 (E. Fischer and J.A. Makowsky (2006)).
Let ϕ be a many-sorted CMSOL(τ)*-sentence, such that all of its models have width at most k. Then the many-sorted spectrum* MSPEC(ϕ) *is a semi-linear set.*

Recall that width stands here for any of the notions, tree-width, clique-width, patch-width to be discussed next.

4 Structures of Bounded Width

4.1 Tree-width

In the eighties the notion of tree-width of a graph became a central focus of research in graph theory through the monumental work of Robertson and Seymour on graph minor closed classes of graphs, and its algorithmic consequences [RS86]. The literature is very rich, but good references and orientation may be found in [Die96,Bod93,Bod97]. Tree-width is a parameter that measures to what extent a graph is similar to a tree. Additional unary predicates do not affect the tree-width. Tree-width of directed graphs is defined as the tree-width of the underlying undirected graph[1].

Definition 1 (Tree-width). *A k-tree decomposition of a graph $G = (V, E)$ is a pair $(\{X_i \mid i \in I\}, T = (I, F))$ with $\{X_i \mid i \in I\}$ a family of subsets of V, one for each node of T, and T a tree such that*

(i) $\bigcup_{i \in I} X_i = V$.
(ii) for all edges $(v, w) \in E$ there exists an $i \in I$ with $v \in X_i$ and $w \in X_i$.
(iii) for all $i, j, k \in I$: if j is on the path from i to k in T, then $X_i \cap X_k \subseteq X_j$ in other words, the subset $\{t \mid v \in X_t\}$ is connected for all v.
(iv) for all $i \in I$, $|X_i| \leq k + 1$.

A graph G is of tree-width at most k *if there exists a k-tree decomposition of G. A class of graphs K is a $TW(k)$-class iff all its members have tree width at most k.*

Given a graph G and $k \in \mathbb{N}$ there are polynomial time, even linear time, algorithms, which determine whether G has tree-width k, and if the answer is yes, produce a tree decomposition, cf. [Bod97]. However, if k is part of the input, the problem is **NP**-complete [ACP87]

Trees have tree-width 1. The clique K_n has tree-width $n - 1$. Furthermore, for fixed k, the class of finite graphs of tree-width at most k denoted by $TW(k)$, is MSOL-definable.

[1] In [JRST01] a different definition is given, which attempts to capture the specific situation of directed graphs. But the original definition is the one which is used when dealing with hyper-graphs and general relational structures.

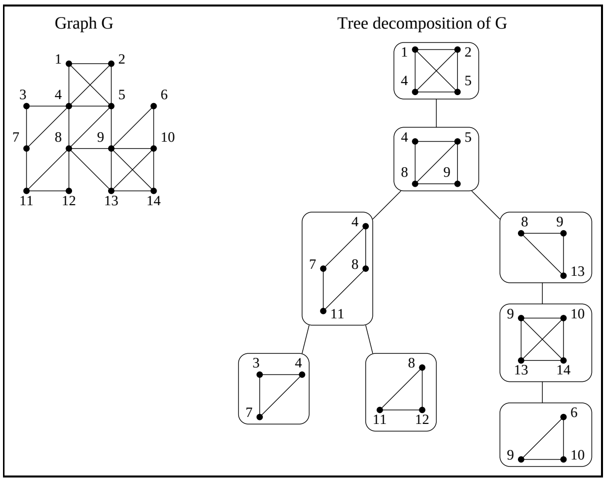

Fig. 2. A graph and one of its tree-decompositions

Example 1. The following graph classes are of tree-width at most k:

(i) Planar graphs of radius r with $k = 3r$.
(ii) Chordal graphs with maximal clique of size c with $k = c - 1$.
(iii) Interval graphs with maximal clique of size c with $k = c - 1$.

Example 2. The following graph classes have unbounded tree-width and are all MSOL-definable.

(i) All planar graphs and the class of all planar grids $G_{m,n}$.
Note that if $n \leq n_0$ for some fixed $n_0 \in \mathbb{N}$, then the tree-width of the grids $G_{m,n}, n \leq n_0$, is bounded by $2n_0$.
(ii) The regular graphs of degree $r, r \geq 3$ have unbounded tree-width.

Tree-width for labeled graphs can be generalized to arbitrary relational structures in a straightforward way. Clause (ii) in the above definition is replaced by

(ii-rel) For each r-ary relation R, if $\bar{v} \in R$, there exists an $i \in I$ with $\bar{v} \in X_i^r$.

This was first used in [FV99].

4.2 Clique-width

A k-colored τ-structures is a $\tau_k = \tau \cup \{P_1, \ldots, P_k\}$-structure where $P_i, i \leq k$ are unary predicate symbols the interpretation of which are disjoint (but can be empty).

Definition 2. *Let $\mathfrak{A}$ be a k-colored τ-structure.*

(i) (Adding hyper-edges) Let $R \in \tau$ *be an* r*-ary relation symbol.* $\eta_{R,P_{j_1},\ldots,P_{j_r}}(\mathfrak{A})$ *denotes the* k*-colored* τ *structure* $\mathfrak{B}$ *with the same universe as* $\mathfrak{A}$*, and for each* $S \in \tau_k$*,* $S \neq R$ *the interpretation is also unchanged. Only for* R *we put*

$$R^B = R^A \cup \{\bar{a} \in A^r : a_i \in P^A_{j_i}\}.$$

We call the operation η hyper edge creation*, or simply* edge creation *in the case of directed graphs. In the case of undirected graphs we denote by* $\eta_{P_{j_1},P_{j_2}}$ *the operation of adding the corresponding undirected edges.*

(ii) (Recoloring) $\rho_{i,j}(\mathfrak{A})$ *denotes the* k*-colored* τ *structure* $\mathfrak{B}$ *with the same universe as* $\mathfrak{A}$*, and all the relations unchanged but for* P^A_i *and* P^A_j*. We put*

$$P^B_i = \emptyset \text{ and } P^B_j = P^A_j \cup P^A_i.$$

We call this operation recoloring.

(iii) (modification via quantifier free translation) More generally, for $S \in \tau_k$ *of arity* r *and* $B(x_1,\ldots,x_r)$ *a quantifier free* τ_k*-formula,* $\delta_{S,B}(\mathfrak{A})$ *denotes the* k*-colored* τ *structure* $\mathfrak{B}$ *with the same universe as* $\mathfrak{A}$*, and for each* $S' \in \tau_k$*,* $S' \neq S$ *the interpretation is also unchanged. Only for* S *we put*

$$S^B = \{\bar{a} \in A^r : \bar{a} \in B^A\}.$$

where B^A *denotes the interpretation of* B *in* $\mathfrak{A}$*.*

Note that the operations of type ρ *and* η *are special cases of the operation of type* δ*.*

Definition 3 (Clique-Width, [CO00,Mak04]).

(i) Here $\tau = \{R_E\}$ *consist of the symbol for the edge relation. Given a graph* $G = (V, E)$*, the* clique-width of G *(cwd(G)) is the minimal number of colors required to obtain the given graph as an* $\{R_E\}$*-reduct from a* k*-colored graph constructed inductively from colored singletons and closure under the following operations:*
  - *(i.a) disjoint union* ($\sqcup$)
  - *(i.b) recoloring* ($\rho_{i\to j}$)
  - *(i.c) edge creation* (η_{E,P_i,P_j})

(ii) For τ *containing more than one binary relation symbol, we replace the edge creation by the corresponding hyper edge creation* $\eta_{R,P_{j_1},\ldots,P_{j_r}}$ *for each* $R \in \tau$*.*

(iii) A class of τ*-structures is a* $CW(k)$*-class if all its members have clique-width at most* k*.*

If τ contains a unary predicate symbol U, the interpretation of U is not affected by the operations recoloring or edge creation. Only the disjoint union affects it.

A description of a graph or a structure using these operations is called a *clique-width parse term* (or *parse term*, if no confusion arises). Every structure of size n has clique-width at most n. The simplest class of graphs of unbounded tree-width but of clique-width at most 2 are the cliques. Given a graph G and

$k \in \mathbb{N}$, determining whether G has clique-width k is in **NP**. A polynomial time algorithm was presented for $k \leq 3$ in [CHL+00]. It was shown in [FRRS05,FRRS06] that for fixed $k \geq 4$ the problem is **NP**-complete. The recognition problem for clique-width of relational structures has not been studied so far even for $k = 2$. The relationship between tree-width and clique-width was studied in [CO00], and for relational structures in [GM03].

Theorem 5 (Courcelle, Olariu (2000), Glikson, Makowsky (2003)). *Let K be a $TW(k)$-class. Then*

(i) If K is a class of graphs, then K is a $CW(m)$-class with $m \leq 2^{k+1} + 1$.
(ii) For arbitrary τ-structures, K is a $CW_\tau(m')$-class with $m' \leq f(k)$ for some function $f(k) = O(2^{p(k)})$ where p is a polynomial in k.

The following examples are from [MR99,GR00].

Example 3 (Classes of clique-width at most k).

(i) The cographs with $k = 2$.
(ii) The distance-hereditary graphs with $k = 3$.
(iii) The cycles C_n with $k = 4$.
(iv) The complement graphs $\bar{C_n}$ of the cycles C_n with $k = 4$.

The cycles C_n have tree-width at most 2, but the other examples have unbounded tree-width.

Example 4 (Classes of unbounded clique-width).

(i) The class of all finite graphs.
(ii) The class of unit interval graphs.
(iii) The class of permutation graphs.
(iv) The regular graphs of degree 4 have unbounded clique-width.
(v) The class grids *Grid*, consisting of the graphs $Grid_{n \times n}$.

For more non-trivial examples, cf. [MR99,GR00]. In contrast to $TW(k)$, we do not know whether the class of all $CW(k)$-graphs is MSOL-definable.

To find more examples it is useful to note, cf. [MM03]:

Proposition 1. *If a graph is of clique-width at most k and G' is an induced subgraph of G, then the clique-width of G' is at most k.*

In [FM04] the following is shown:

Theorem 6 (E. Fischer and J.A. Makowsky (2004)).
Let $\phi \in$ CMSOL(τ) be such that all its finite models have clique-width at most k. Then there are $m_0, n_0 \in \mathbb{N}$ such that if ϕ has a model of size $n \geq n_0$ then ϕ has also a model of size $n + m_0$.

From this we get immediately a further generalization of Theorem 3.

Corollary 1. *Let $\phi \in$ CMSOL(τ) be such that all its finite models have clique-width at most k. Then spec(ϕ) is ultimately periodic.*

4.3 Patch-width

Here is a further generalization of clique-width for which our theorem still works. The choice of operation is discussed in detail in [CM02].

Definition 4. *Given a τ-structure $\mathfrak{A}$, the* patch-width of G *($pwd(G)$) is the minimal number of colors required to obtain $\mathfrak{S}$ as a $\{\tau\}$-reduct from a k-colored τ-structure inductively from fixed finite number of τ_k-structures and closure under the following operations:*

(i) disjoint union ($\sqcup$),
(ii) recoloring ($\rho_{i \to j}$) and
(iii) modifications ($\delta_{S,B}$).

A class of τ-structures is a $PW_\tau(k)$-class if all its members have patch-width at most k.

A description of a τ-structure using these operations is called a *patch term.*

Example 5.

(i) In [CO00] it is shown that if a graph G has clique-width at most k then its complement graph $\bar{G}$ has clique-width at most $2k$. However, its patch-width is also k as $\bar{G}$ can be obtained from G by $\delta_{E,\neg E}$.
(ii) The clique K_n has clique-width 2. However if we consider graphs as structures on a two-sorted universe (respectively for vertices and edges), then K_n has clique-width $c(n)$ and patch-width $p(n)$ where $c(n)$ and $p(n)$ are functions which tend to infinity. This will easily follow from Theorem 4. For the clique-width of K_n as a two-sorted-structure this was already shown in [Rot98].

In [CM02] it is shown that a class of graphs of patch-width at most k is of clique-width at most $f(k)$ for some function f. It is shown in [FM00] that this is not true for relational structures in general.

In the definition of patch-width we allowed only unary predicates as auxiliary predicates (colors). We could also allow r-ary predicates and speak of r-ary patch-width. The theorems where bounded patch-width is required are also true for this more general case. The relative strength of clique-width and the various forms of patch-width are discussed in [FM00].

In [FM04] the following is shown:

Theorem 7 (E. Fischer and J.A. Makowsky (2004)).
Let $\phi \in \mathrm{CMSOL}(\tau)$ be such that all its finite models have patch-width at most k. Then there are $m_0, n_0 \in \mathbb{N}$ such that if ϕ has a model of size $n \geq n_0$ then ϕ has also a model of size $n + m_0$.

From this we get yet another generalization of Theorem 3.

Corollary 2. *Let $\phi \in \mathrm{CMSOL}(\tau)$ be such that all its finite models have patch-width at most k. Then $spec(\phi)$ is ultimately periodic.*

More recent work on spectra and patch-width may be found in [She04,DS06].

5 Applications of Theorem 7

In this paper we give several applications of Theorem 7 showing that certain classes of graphs have unbounded clique-width or patch-width.

5.1 Classes of unbounded patch-width

Theorem 7 gives a new method to show that certain classes K of graphs have unbounded tree-width, clique-width or patch-width.

To see this we look at the class $Grid$ of all grids $Grid_{n\times n}$. They are known to have unbounded tree-width, cf. [Die96], and in fact, every minor closed class of graphs of unbounded tree-width contains these grids. They were shown to have unbounded clique-width in [CO00,GR00]. However, for patch-width these arguments do not work. On the other hand $Grid$ is MSOL-definable, and its spectrum consists of the numbers n^2, so by Theorems 6 and 7, the unboundedness follows directly.

In particular, as this is also true for every $K' \supseteq K$, the class of all graphs is of unbounded patch-width.

Without Theorem 7, there was only a conditional proof of unbounded patch-width available. It depends on the assumption that the polynomial hierarchy Σ^{P} does not collapse to **NP**. The argument then proceeds as follows:

(i) Checking patch-width at most k of a structure $\mathfrak{A}$, for k fixed, is in **NP**. Given a structure $\mathfrak{A}$, one just has to guess a patch-term of size polynomial in the size of $\mathfrak{A}$.
(ii) Using the results of [Mak04] one gets that checking a CMSOL(τ)-property ϕ on the class $PW_\tau(k)$ is in **NP**, whereas, by [MP96], there are Σ_n^{P}-hard problems definable in MSOL for every level Σ_n^{P} of the polynomial hierarchy.
(iii) Hence, if the polynomial hierarchy does not collapse to **NP**, the class of all τ-structures is of unbounded patch-width, provided τ is large enough.

5.2 The patch-width of incidence graphs

We look at two presentation a graph G. The first consists of a set of vertices $V(G)$ and a binary symmetric relation $E(G)$. The second is the incidence graph $I(G)$, which consists of two sorts, $V(G)$ and $E(G)$ and an incidence relation $R(G) \subseteq V(G) \times E(G)$ with $(u,e) \in R(G)$ iff there is $v \in V(G)$ with $(u,v) = e$.

It is well known that if G has tree-width at most k, so has $I(G)$. This is not so for clique-width. The cliques K_n have clique-width 2, but the clique-width of $I(K_n)$ grows with n. This follows immediately from Theorem 4. The cliques are definable by an MSOL-sentence ϕ_{clique} whose spectrum $\textsc{mspec}(\phi_{clique})$ consists of the pairs $(n, \frac{n(n-1)}{2})$, which is not semi-linear.

Similarly, the class of square grids $Grid_{n\times n}$ and the class of its incidence graphs $I(Grid_{n\times n})$ are known to be definable by MSOL-sentences $\phi_{sqgrids}$ and $\psi_{sqgrids}$, respectively. But their spectra are of the form n^2 and $(n^2, 2n^2+2n)$, respectively, which are not semi-linear. Hence they are not of bounded patch-width.

Our main result in this section is a new proof of the following theorem, implicit in papers by B. Courcelle and J. Engelfriet [CE95,EvO97,Cou03].

Theorem 8. *Let K be a class of graphs and $I(K)$ be the class of its incidence graphs. Assume $I(K)$ is defined by an* CMSOL*-sentence ψ and has bounded clique-width. Then K and $I(K)$ have bounded tree-width.*

5.3 Proof of Theorem 8

The proof of Theorem 8 combines Theorem 4 with the fact, taken from [GW00]:

Proposition 2. *Graphs of clique-width at most k which do not have the complete bipartite graph $K_{n,n}$ as a subgraph, have tree-width at most $3k(n-1)-1$.*

The clique-width of $I(K)$ and K are related as follows:

Proposition 3. *If $I(K)$ has bounded clique-width, so has K.*

This is a consequence of results in [EvO97], formulated for graph grammars.

Furthermore, when we expand a τ-structure $\mathfrak{A}$ by adding new unary predicates $P_1, \ldots, P_m$, whether definable or not, the width (tree-width, clique-width, patch-width) does not change.

Proposition 4. *Let $\mathfrak{A}$ be a τ-structure of width k. Then $\langle \mathfrak{A}, P_1, \ldots, P_m \rangle$ as a $\tau \cup \{P_1, \ldots P_m\}$ has also width k.*

Now assume $I(K)$ is defined by an CMSOL-sentence ψ and has bounded clique-width. Then, by Proposition 3, K has bounded clique-width. Assume further, for contradiction, that K has unbounded tree-width. Then, by Proposition 2, K contains all the complete bipartite graphs $K_{n,n}$ is subgraphs of graphs in K, and hence all the graphs $I(K_{n,n})$ as subgraphs of graphs of $I(K)$. Let $K' \subseteq K$ be the subclass of K of graphs which do contain a copy of $K_{n,n}$. Clearly, $I(K')$ is CMSOL-definable. We now show that $I(K') \subseteq I(K)$ has unbounded clique-width. We expand each graph of $I(K')$ by three disjoint unary predicates $P_i : i = 1, 2, 3$, one for each vertex set of $I(K_{n,n})$ and one for the edge set of $I(K_{n,n})$. The many-sorted spectrum of $I(K')$ with respect to these predicates is of the from $(n, n, \binom{n}{2})$, which is not semi-linear. Using Theorem 7, we conclude that $I(K')$ and hence $I(K)$ has unbounded clique-width.

6 Conclusions and Open problems

We have seen how Parikh's Theorem can be extended to arbitrary relational structures, provided they are of bounded width. Classes of structures of bounded patch-width share all of the interesting algorithmic properties which were proven for classes of bounded clique-width. In particular, membership in CMSOL-definable classes of structures of bounded patch-width can be decided in polynomial time, whereas for classes of unbounded patch-width this can arbitrarily hard within the polynomial hierarchy. As a matter of fact all the theorems proven for clique-width in [Mak04] are also valid for patch-width. But this is only interesting, if the structures carry more relations than graphs. For graphs, it was shown in [CM02] that any class K of graphs bounded patch-width is also of bounded clique-width.

The true usefulness of patch-width as a structural parameter of relational structures still has to be explored. A first step is done in [FM00], where it is shown that there are classes of relational structures with unbounded clique-width but of bounded patch-width.

Open Question 1 *What is the complexity of checking whether a τ-structure $\mathfrak{A}$ has patch-width at most k, for a fixed k? What is the complexity, if k is part of the input?*

We conclude with some directions of further research. Proposition 3 actually is a special case of a more general theorem dealing with MSOL-transductions of graphs. A *transduction* T is, roughly speaking, a (possibly partial) map which, in a uniform way, associates with a τ-structure a τ_1-structure where the new atomic relations are replaced by MSOL(τ)-definable relations. The transductions here are scalar, i.e., the universe may be duplicated by taking disjoint copies of it, but the arities of the relations may not be changed. For details, the reader may consult [Cou94,Mak04]. In [EvO97] it the following is shown:

Theorem 9. *Let K be a class of graphs of bounded clique-width, and let T be a* MSOL*-transduction of graphs into graphs. Then $T(K) = \{T(G) : G \in K\}$ is also of bounded clique-width.*

By Proposition 4 this holds also for unary expansions of graphs. Therefore we have the following corollary to Theorem 7:

Corollary 3. *Let K be a many-sorted class of graphs of bounded clique-width, and let T be a* MSOL*-transduction of graphs into graphs. Let ϕ be a* CMSOL*-sentence such that all its models are in $T(K)$. Then $mspec(\phi)$ is semi-linear.*

This suggest the following questions:

Open Question 2 *Are the grids $Grid_{n,n}$ distinctive for unbounded patch-width in the following sense:*

> *(S-1): Let K be a class of many-sorted τ-structures of unbounded patch-width. Does there exist a* CMSOL*-transduction T of τ-structures into graphs, such for all $n \in \mathbb{N}$ the grid $G_{n,n} \in T(K)$?*

One could also ask for a weaker version of (S-1) where we only require in the conclusion that

> *(S-2): (...) for infinitely many $n \in \mathbb{N}$ the grid $G_{n,n} \in T(K)$?*

Open Question 3 *Can we generalize Theorem 9 to the following:*

(S-2): Let K be a class of many-sorted τ-structures of bounded patch-width, and let T be a CMSOL*-transduction of τ-structures into τ_1-structures. Then the class of τ_1-structures $T(K)$ is also of bounded patch-width.*

Open Question 4 *Can we generalize Corollary 3 to*

(S-3): Let K be a many-sorted class of τ-structures bounded patch-width, and let T be a CMSOL*-transduction of τ-structures into τ_1-structures. Let ϕ be a* CMSOL(τ_1)*-sentence such that all its models are in $T(K)$. Then $mspec(\phi)$ is semi-linear.*

Open Question 5 *Is there a converse to Corollary 3? More precisely, is the following true?*

(S-4): Let K be a many-sorted class of τ-structures such that for every CMSOL*-transduction T of τ-structures into τ_1-structures, and for every* CMSOL(τ_1)*-sentence ϕ such that all its models are in $T(K)$, $mspec(\phi)$ is semi-linear. Then K is of bounded patch-width.*

Note that trivially, the class of all graphs has unbounded patch-width, but it spectrum consists of $\mathbb{N}$ which is semi-linear. So the closure under definable classes in $T(K)$ is essential here. The statements (S-i) (i=1,2,3,4,5) are not independent. (S-3) together with Theorem 7 implies (S-4). Also (S-1) and (S-3) implies (S-5). Note that (S-2) and (S-3) are not enough to get (S-5), unless we know that $T(K)$ is CMSOL-definable.

Acknowledgments Some passages are taken verbatim from [FM04] and from [DJMM08]. I would like to thank my co-authors A. Durand, E. Fischer, M. More and N. Jones for kindly allowing me to do so. I would like to thank B. Courcelle for valuable comments and references.

References

[ACP87] S. Arnborg, D.G. Corneil, and A. Proskurowski. Complexity of finding embedding in a k–tree. *SIAM. J. Algebraic Discrete Methods*, 8:277–284, 1987.

[AÉI02] Luca Aceto, Zoltán Ésik, and Anna Ingólfsdóttir. A fully equational proof of Parikh's theorem. *Informatique theéorique et applications*, 36(2):129–153, 2002.

[Bod93] H. Bodlaender. A tourist guide through tree width. *Acta Cybernetica*, 11:1–23, 1993.

[Bod97] H. Bodlaender. Treewidth: Algorithmic techniques and results. In I. Privara and P. Ruzicka, editors, *Proceedings of the 22th International Symposium on the Mathematical Foundation of Computer Science, MFCS'97*, volume 1295 of *Lecture Notes in Computer Science*, pages 29–36. Springer, 1997.

[CE95] B. Courcelle and J. Engelfriet. A logical characterization of the sets of hypergraphs defined by hyperedge replacement grammars. *Mathematical Systems Theory*, 28:515–552, 1995.

[CER93] B. Courcelle, J. Engelfriet, and G. Rozenberg. Handle-rewriting hypergraph grammars. *J. Comput. System Sci.*, 46:218–270, 1993.

[CHL+00] D. G. Corneil, M. Habib, J.-M. Lanlignel, B. Reed, and U. Rotics. Polynomial time recognition of clique-width ≤ 3 graphs. In *Proceedings of LATIN'2000*, volume 1776 of *Lecture Notes in Computer Science*, pages 126–134. Springer, 2000.

[CM02] B. Courcelle and J.A. Makowsky. Fusion on relational structures and the verification of monadic second order properties. *Mathematical Structures in Computer Science*, 12.2:203–235, 2002.

[CO00] B. Courcelle and S. Olariu. Upper bounds to the clique–width of graphs. *Discrete Applied Mathematics*, 101:77–114, 2000.

[Cou90] B. Courcelle. The monadic second–order logic of graphs I: Recognizable sets of finite graphs. *Information and Computation*, 85:12–75, 1990.

[Cou92] B. Courcelle. Monadic second-order logic of graphs VII: Graphs as relational structures. *Theoretical Computer Science*, 101:3–33, 1992.

[Cou94] B. Courcelle. Monadic second order graph transductions: A survey. *Theoretical Computer Science*, 126:53–75, 1994.

[Cou95] B. Courcelle. Structural properties of context-free sets of graphs generated by vertex replacement. *ic*, 116:275–293, 1995.

[Cou03] B. Courcelle. The monadic second order logic of graphs, XIV: Uniformly sparse graphs and edge set quantification. *Theoretical Computer Science*, 299(1):1–36, 2003.

[DFL97] A. Durand, R. Fagin, and B. Loescher. Spectra with only unary function symbols. In M. Nielsen and W. Thomas, editors, *CSL'97*, volume 1414 of *Lecture Notes in Computer Science*, pages 189–202. Springer, 1997.

[Die96] R. Diestel. *Graph Theory.* Graduate Texts in Mathematics. Springer, 1996.

[DJMM08] A. Durand, N. Jones, J.A. Makowsky, and M. More. 50 years of the spectrum problem: Survey and new results. in preparation, 2008.

[DS06] M. Doron and S. Shelah. Bounded m-ary patch-width are equivalent for $m > 2$. Electronically available at arXiv:math/0607375v1, 2006.

[EF95] H. Ebbinghaus and J. Flum. *Finite Model Theory.* Springer Verlag, 1995.

[EvO97] J. Engelfriet and V. van Oostrom. Logical description of context-free graph-languages. *Journal of Computer and System Sciences*, 55:489–503, 1997.

[FM00] E. Fischer and J.A. Makowsky. Patch-width, a generalization of clique-width for relational structures. in preparation, 200?

[FM04] E. Fischer and J.A. Makowsky. On spectra of sentences of monadic second order logic with counting. *Journal of Symbolic Logic*, 69(3):617–640, 2004.

[FRRS05] M.R. Fellows, F.A. Rosamond, U. Rotics, and S. Szeider. Proving NP-hardness for clique width ii: Non-approximability of clique-width. Electronic Colloquium on Computational Complexity, 2005.

[FRRS06] Michael R. Fellows, Frances A. Rosamond, Udi Rotics, and Stefan Szeider. Clique-width minimization is np-hard. In *STOC '06: Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 354–362. ACM, 2006.

[FV99] T. Feder and M. Vardi. The computational structure of monotone monadic SNP and constraint satisfaction: A study through Datalog and group theory. *SIAM Journal on Computing*, 28:57–104, 1999.

[GM03] A. Glikson and J.A. Makowsky. NCE graph grammars and clique-width. In H.L. Bodlaender, editor, *Proceedings of the 29th International Workshop on Graph-Theoretic Concepts in Computer Science (WG 2003), Elspeet (The Netherlands)*, volume 2880 of *Lecture Notes in Computer Science*, pages 237–248. Springer, 2003.

[GR00] M. C. Golumbic and U. Rotics. On the clique-width of some perfect graph classes. *Internation Journal of Foundations of Computer Science*, 11:423–443, 2000.

[GS97] F. Gécseg and M. Steinby. Tree languages. In G. Rozenberg and A. Salomaa, editors, *Handbook of formal languages, Vol. 3 : Beyond words*, pages 1–68. Springer Verlag, Berlin, 1997.

[GS03] Y. Gurevich and S. Shelah. Spectra of monadic second-order formulas with one unary function. In *LICS'03*, pages 291–300. IEEE, 2003.

[GW00] F. Gurski and E. Wanke. The tree-width of clique-width bounded graphs without $K_{n,n}$. *Lecture Notes in Computer Science*, 1928:196–205, 2000.

[JRST01] T. Johnson, N. Robertson, P. Seymour, and R. Thomas. Directed tree–width. *Journal of Combinatorial Theory, Serie B*, 81(1):138–154, 2001.

[Mak04] J.A. Makowsky. Algorithmic uses of the Feferman-Vaught theorem. *Annals of Pure and Applied Logic*, 126.1-3:159–213, 2004.

[MM03] J.A. Makowsky and J.P. Mariño. Tree-width and the monadic quantifier hierarchy. *Theoretical Computer Science*, 303:157–170, 2003.

[MP96] J.A. Makowsky and Y. Pnueli. Arity vs. alternation in second order logic. *Annals of Pure and Applied Logic*, 78(2):189–202, 1996.

[MR99] J.A. Makowsky and U. Rotics. On the cliquewidth of graphs with few P_4's. *International Journal on Foundations of Computer Science*, 10:329–348, 1999.

[Par66] R. Parikh. On context-free languages. *jacm*, 13:570–581, 1966.

[Pil73] D.L. Pilling. Commutative regular equations and Parikh's theorem. *J. London Math. Soc.*, 6:663–666, 1973.

[Rot98] U. Rotics. *Efficient Algorithms for Generally Intractable Graph Problems Restricted to Specific Classes of Graphs*. PhD thesis, Technion- Israel Institute of Technology, 1998.

[RS86] N. Robertson and P. D. Seymour. Graph minors. ii. algorithmic aspects of tree–width. *Journal of Algorithms*, 7:309–322, 1986.

[Sch52] H. Scholz. Problem # 1: Ein ungelöstes Problem in der symbolischen Logik. *Journal of Symbolic Logic*, 17:160, 1952.

[She04] S. Shelah. Spectra of monadic second order sentences. Paper No. 817, Electronically available at arXiv:math/0405158, 2004.