Connection Matrices and the Definability of Graph Parameters*

Tomer Kotek and Johann A. Makowsky

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

Abstract

In this paper we extend the Finite Rank Theorem for connection matrices of graph parameters definable in Monadic Second Order Logic with modular counting CMSOL of B. Godlin, T. Kotek and J.A. Makowsky, [16, 30], and demonstrate its vast applicability in simplifying known and new non-definability results of graph properties and finding new non-definability results for graph parameters. We also prove a Feferman-Vaught Theorem for the logic CFOL, First Order Logic with the modular counting quantifiers.

1998 ACM Subject Classification F.4.1 Mathematical Logic

Keywords and phrases Model theory, finite model theory, graph invariants

Digital Object Identifier 10.4230/LIPIcs.CSL.2012.411

1 Introduction

Difficulties in proving non-definability

Proving that a graph property P is not definable in first order logic FOL can be a challenging task, especially on graphs with an additional linear order on the vertices. Proving that a graph property such as 3-colorability, which is definable in monadic second order logic MSOL, is not definable in fixed point logic on ordered graphs amounts to solving the famous $\mathbf{P} \neq \mathbf{NP}$ problem.

In the case of FOL and MSOL properties the basic tools for proving non-definability are the various *Ehrenfeucht-Fraïssé games* also called *pebble games*. However, proving the existence of winning strategies for these games can be exasperating. Two additional tools can be used to make the construction of such winning strategies easier and more transparent: the *composition of winning strategies* and the use of *locality properties* such as Hanf locality and Gaifman locality. These techniques are by now well understood, even if not always simple to apply, and are described in monographs such as [11, 24]. However these techniques are not easily applicable for stronger logics, such as CMSOL, monadic second order logic with modular counting. Furthermore, the pebble game method or the locality method may be difficult to use when dealing with ordered structures or when proving non-definability for the case where the definition may use an order relation on the universe in an *order-invariant* way.

* The first author was partially supported by the Fein Foundation and the Graduate School of the Technion–Israel Institute of Technology. The second author was partially supported by the Israel Science Foundation for the project "Model Theoretic Interpretations of Counting Functions" (2007-2011) and the Grant for Promotion of Research by the Technion–Israel Institute of Technology.

Computer Science Logic 2012 (CSL'12).
Editors: Patrick Cégielski, Arnaud Durand; pp. 411–425
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The notion of definability was extended in [2, 3] to integer valued graph parameters, and in [9, 27, 29, 22, 23] to real or complex valued graph parameters and graph polynomials. In [27] and [22] graph polynomials definable in MSOL respectively SOL were introduced. The techniques of pebble games and locality do not lend themselves easily, or are not useful at all, for proving non-definability in these cases.

Connection matrices

Connection matrices were introduced in [15, 26] by M. Freedman, L. Lovász and A. Schrijver where they were used to characterize graph homomorphism functions. Let f be a real or complex valued graph parameter. A k-connection matrix $M(\sqcup_k, f)$ is an infinite matrix, where the rows and columns are indexed by finite k-labeled graphs G_i and the entry $M(\sqcup_k, f)_{i,j}$ is given by the value of $f(G_i \sqcup_k G_j)$. Here $\sqcup_k$ denotes the *k-sum* operation on G_i and G_j, i.e. the operation of taking the disjoint union of G_i and G_j and identifying vertices with the corresponding k-many labels.

In [16] connection matrices were used to show that certain graph parameters and polynomials are not MSOL-definable. The main result of [16] is the *Finite Rank Theorem*, which states that the connection matrices of CMSOL-definable graph polynomials have finite rank. Connection matrices and the Finite Rank Theorem were generalized in [30] to matrices $M(\Box, f)$ where $\Box$ is a binary operation on labeled graphs subject to a smoothness condition depending on the logic one wants to deal with. However, very few applications of the Finite Rank Theorem were given.

Properties not definable in CFOL and CMSOL

The purpose of this paper lies in the demonstration that the Finite Rank Theorem is a *truly manageable tool* for proving non-definability which leaves no room for hand-waving arguments. To make our point we discuss graph properties (not)-definable in CFOL and CMSOL, i.e., First Order Logic, respectively Monadic Second Order Logic with modular counting quantifiers $\mathrm{D}_{m,i}x\phi(x)$ which say that the number of elements satisfying ϕ equals i modulo m. We also discuss the corresponding (non)-definability questions in CMSOL for graph parameters and graph polynomials. Although one can derive pebble games for these two logics, see e.g. [21, 32], using them to prove non-definability may be very awkward.

Instead we use a Feferman-Vaught-type Theorem for CFOL for disjoint unions and Cartesian products, Theorem 3.3, which seems to be new for the case of products. The corresponding theorem for disjoint unions, Theorem 3.2(i), for CMSOL was proven by B. Courcelle [7, 8, 27].

The proof of the Finite Rank Theorem for these logics follows from the Feferman-Vaught-type theorems. The details will be spelled out in Section 3.

With the help of the Finite Rank Theorem we give new and uniform proofs for the following:

(i) Using connection matrices for various generalizations of the Cartesian product $\times_\Phi$ we prove non-definability of the following properties in CFOL with the vocabulary of graphs $\langle V, E, < \rangle$ with linear order:
- Forests, bipartite graphs, chordal graphs, perfect graphs, interval graphs, block graphs (every biconnected component, i.e., every block, is a clique), parity graphs (any two induced paths joining the same pair of vertices have the same parity);
- Trees, connected graphs;

- Planar graphs, cactus graphs (graphs in which any two cycles have at most one vertex in common) and pseudo-forests (graphs in which every connected component has at most one cycle).

The case of connected graphs was also shown undefinable in CFOL by. J. Nurmonen in [32] using his version of the pebble games for CFOL.

(ii) Using connection matrices for various generalizations of the disjoint union $\sqcup_\Phi$ we prove non-definability of the following properties in CMSOL with the vocabulary of graphs $\langle V, E, < \rangle$ with linear order:
- Hamiltonicity (via cycles or paths), graphs having a perfect matching, cage graphs (regular graphs with as few vertices as possible for their girth), well-covered graphs (where every minimal vertex cover has the same size as any other minimal vertex cover). Here $\sqcup_\Phi$ is the join operation $\bowtie$.
- The class of graphs which have a spanning tree of degree at most 3. Here $\sqcup_\Phi$ is a modified join operation.

(iii) Using connection matrices for various generalizations of the disjoint union $\sqcup_\Phi$ we prove non-definability of the following properties in CMSOL with the vocabulary of hypergraphs $\langle V, E; R, < \rangle$ with linear order:
- Regular graphs and bi-degree graphs.
- Graphs with average degree at most $\frac{|V|}{2}$.
- Aperiodic digraphs (where the greatest common divisor of the lengths of all cycles in the graph is 1).
- Asymmetric (also called rigid) graphs (i.e. graphs which have no non-trivial automorphisms).

Graph parameters and graph polynomials not definable in CMSOL

A graph parameter is CMSOL-definable if it is the evaluation of a CMSOL-definable graph polynomial. The precise definition of definability of graph polynomials is given in Section 6. Most prominent graph polynomials turned out to be definable in CMSOL, sometimes using a linear order on the vertices in an order-invariant way, among them the various Tutte polynomials, interlace polynomials, matching polynomials, and many more, cf. [28]. This led the second author to express his belief in [28] that all "naturally occurring graph polynomials" are CMSOL-definable. However, in [16] it was shown, using connection matrices, that the graph polynomial counting harmonious colorings is not CMSOL-definable. A vertex coloring is harmonious if each pair of colors appears at most once at the two end points of an edge, cf. [12, 20]. That this is indeed a graph polynomial was shown in [23]. However, the main thrust of [16] consists in showing that certain graph parameters are not evaluations of the standard prominent graph polynomials.

In Section 7, we use connection matrices to show that many "naturally occurring graph polynomials" are not CMSOL-definable. All these examples count various colorings and are graph polynomials by [23]. The corresponding notion of coloring is studied extensively in the literature.

To illustrate this we show that the following graph polynomials are not CMSOL-definable:
- $\chi_{rainbow}(G, k)$ is the number of *path-rainbow connected k-colorings*, which are functions $c : E(G) \to [k]$ such that between any two vertices $u, v \in V(G)$ there exists a path where all the edges have different colors.
- For every fixed $t \in \mathbb{N}$, $\chi_{mcc(t)}(G, k)$ is the number of vertex k-colorings $f : V(G) \to [k]$ for which no color induces a subgraph with a connected component of size larger than t.

- $\chi_{convex}(G,k)$ is the number of *convex colorings*, which are vertex k-colorings $f : V(G) \to [k]$ such that every color induces a connected subgraph of G.

Path-rainbow connected colorings were introduced in [6] and their complexity was studied in [5]. $mcc(t)$-colorings were studied in [1], [25] and [14]. Note $\chi_{mcc(1)}(G,k)$ is the chromatic polynomial. Convex colorings were studied for their complexity e.g. in [31] and [17]. From [23] we get that $\chi_{rainbow}(G,k)$, $\chi_{mcc(t)}(G,k)$, and $\chi_{convex}(G,k)$ are graph polynomials with k as the variable.

In Section 7 more examples of graph polynomials and graph parameters not definable in CMSOL are given.

Outline of the paper

We assume the reader is familiar with the basics of finite model theory [11, 24] and graph theory [4, 10].

In Section 2 we illustrate the use of connection matrices in the case of regular languages. This serves as a “warm-up” exercise. In Section 3 we introduce the general framework for connection matrices of graph properties, i.e., boolean graph parameters, and of properties of general τ-structures. In Section 4 we spell out the advantages and limitations of the method of connection matrices in proving non-definability. In Section 5 we illustrate the use of connection matrices and the Finite Rank Theorem for proving non-definability of properties. In Section 6 we recall the framework of definable graph polynomials and τ-polynomials and the corresponding definable numeric parameters, and in Section 7 we show how to prove non-definability of these.

2 Connection Matrices for Regular Languages

Our first motivating examples deal with regular languages and the operation of concatenation $\circ$. By the well-known Büchi-Elgot-Trakhtenbrot Theorem, see [11, 24], a language $L \subseteq \Sigma^*$ is regular if and only if the class S_L of ordered structures representing the words of L is definable in MSOL (or equivalently in CMSOL or $\exists$MSOL, the existential fragment of MSOL). The connection matrix $M(\circ, L)$ with columns and rows indexed by all words of Σ^* is defined by $M(\circ, L)_{u,v} = 1$ iff $u \circ v \in L$.

The Myhill-Nerode Theorem and the Pumping Lemma for regular languages, see [19, 18], can be used to derive the following properties of $M(\circ, L)$:

▶ Proposition 2.1. Let $L \subseteq \Sigma^*$ be a regular language.

(i) There is a finite partition $\{U_1, \ldots, U_k\}$ of Σ^* such that the sub-matrices obtained from restricting $M(\circ, L)$ to $M(\circ, L)^{[U_i,U_j]}$ have constant entries.

(ii) In particular, the infinite matrix $M(\circ, L)$ has finite rank over any field $\mathcal{F}$.

(iii) $M(\circ, L)$ has an infinite sub-matrix of rank at most 1.

Now we can also look at counting functions and numeric parameters of words, such as the length $\ell(w)$ of a word w or the number of words $s_L(w)$ in a language L which are (connected) sub-words of a given word w. The corresponding connection matrices $M(\circ, \ell)$ and $M(\circ, s_L)$ defined by $M(\circ, \ell)_{u,v} = \ell(u \circ v)$ and $M(\circ, s_L)_{u,v} = s_L(u \circ v)$ respectively do not satisfy (i) and (iii) above, but still have finite rank. On the other hand the function $m_L(w)$ which gives the maximal size of a word in L which occurs as a connected sub-word in w gives rise to connection matrix $M(\bar{\circ}, s_L)$ of infinite rank. Here $u\bar{\circ}v = u \circ a \circ v$ where $a \notin \Sigma$ and therefore $m_L(u\bar{\circ}v) = \max\{m_L(u), m_L(v)\}$.

We can use these connection matrices to show that $L_1 = \{0^n \circ 1^n : n \in \mathbb{N}\}$ is not regular, by noting that the sub-matrix $M(\circ, L_1)$ with columns indexed by 0^n and rows indexed by 1^n has 0 everywhere but in the diagonal, hence has infinite rank, contradicting (ii) of Proposition 2.1.

The numeric parameters on words ℓ, s_L are MSOL-definable as follows: $\ell(w) = \sum_{u <_{in} w} 1$, where $u <_{in} w$ means that u is a proper possibly empty initial segment of w. Similarly, $s_L(w) = \sum_{u <_{sw} w} 1$, where $u <_{sw} w$ denotes the relation u is a connected sub-word of w. We shall give a general definition of MSOL-definable numeric parameter in Section 6. But we state here already

▶ Proposition 2.2. The connection matrices $M(\circ, f)$ and $M(\bar{\circ}, f)$ have finite rank, provided f is MSOL-definable.

▶ Corollary 2.3. The function $m_L(w)$ is not MSOL-definable.

3 Connection Matrices for Properties: The Framework

Let τ be a purely relational finite vocabulary which may include constant symbols and may include a distinguished binary relation symbol for a linear order. A *τ-property* is a class of finite τ-structures closed under τ-isomorphisms. If the context is clear we just speak of properties and isomorphisms. We denote by $\mathrm{SOL}(\tau)$ the set of SOL formulas over τ. A sentence is a formula without free variables.

Let $\mathcal{L}$ be a subset of SOL. $\mathcal{L}$ is a *fragment of* SOL if the following conditions hold:

(i) For every finite relational vocabulary τ the set of $\mathcal{L}(\tau)$ formulas contains all the atomic τ-formulas and is closed under boolean operations and renaming of relation and constant symbols.

(ii) $\mathcal{L}$ is equipped with a notion of *quantifier rank* and we denote by $\mathcal{L}_q(\tau)$ the set of formulas of quantifier rank at most q. The quantifier rank is sub-additive under substitution of sub-formulas,

(iii) The set of formulas of $\mathcal{L}_q(\tau)$ with a fixed set of free variables is, up to logical equivalence, finite.

(iv) Furthermore, if $\phi(x)$ is a formula of $\mathcal{L}_q(\tau)$ with x a free variable of $\mathcal{L}$, then there is a formula ψ logically equivalent to $\exists x \phi(x)$ in $\mathcal{L}_{q'}(\tau)$ with $q' \geq q + 1$.

Typical fragments are FOL and MSOL. CMSOL and the fixed point logics IFPL and FPL and their corresponding finite variable subsets correspond to fragments of SOL if we replace the counting or fixed-point operators by their SOL-definitions.

For two τ-structures $\mathfrak{A}$ and $\mathfrak{B}$ we define the *equivalence relation of $\mathcal{L}_q(\tau)$- non-distinguishability*, and we write $\mathfrak{A} \equiv_q^{\mathcal{L}} \mathfrak{B}$, if they satisfy the same sentences from $\mathcal{L}_q(\tau)$.

Let $s : \mathbb{N} \to \mathbb{N}$ be a function. A binary operation $\Box$ between τ-structures is called *(s,$\mathcal{L}$)-smooth*, if for all $q \in \mathbb{N}$ whenever $\mathfrak{A}_1 \equiv_{q+s(q)}^{\mathcal{L}} \mathfrak{B}_1$ and $\mathfrak{A}_2 \equiv_{q+s(q)}^{\mathcal{L}} \mathfrak{B}_2$ then

$$\mathfrak{A}_1 \Box \mathfrak{A}_2 \equiv_q^{\mathcal{L}} \mathfrak{B}_1 \Box \mathfrak{B}_2.$$

If $s(q)$ is identically 0 we omit it.

For two τ-structures $\mathfrak{A}$ and $\mathfrak{B}$, we denote by $\mathfrak{A} \sqcup \mathfrak{B}$ the *disjoint union*, which is a τ-structure; $\mathfrak{A} \sqcup_{rich} \mathfrak{B}$ the *rich disjoint union* which is the disjoint union augmented with two unary predicates for the universes A and B respectively; $\mathfrak{A} \times \mathfrak{B}$ the *Cartesian product*, which is a τ-structure; and for graphs G, H by $G \bowtie H$ the *join* of two graphs, see [10].

A *$\mathcal{L}$-transduction* of τ-structures into σ-structures is given by defining a σ-structure inside a given τ-structure. The universe of the new structure may be a definable subset of an

m-fold Cartesian product of the old structure. If $m = 1$ we speak of *scalar* and otherwise of *vectorized* transductions. For every k-ary relation symbol $R \in \sigma$ we need a τ-formula in $k \cdot m$ free individual variables to define it. We denote by Φ a sequence of τ-formulas which defines a transduction. We denote by $\Phi^\star$ the map sending τ-structures into σ-structures induced by Φ. We denote by $\Phi^\sharp$ the map sending σ-formulas into τ-formulas induced by Φ. For a σ-formula $\Phi^\sharp(\theta)$ is the *backward translation* of θ into a τ-formula. Φ is *quantifier-free* if all its formulas are from $\mathrm{FOL}_0(\tau)$. We skip the details, and refer the reader to [24, 27].

A fragment $\mathcal{L}$ is *closed under scalar transductions*, if for Φ such that all the formulas of Φ are in $\mathcal{L}(\tau)$, Φ scalar, and $\theta \in \mathcal{L}(\sigma)$, the backward substitution $\Phi^\sharp(\theta)$ is also in $\mathcal{L}(\tau)$. A fragment of SOL is called *tame* if it is closed under scalar transductions. FOL, MSOL and CMSOL are all tame fragments. So are their finite variable versions.

FOL and SOL are also closed under vectorized transductions, but the monadic fragments MSOL and CMSOL are not.

We shall frequently use the following:

▶ Proposition 3.1. Let Φ define a $\mathcal{L}$-transduction from τ-structures to σ-structures where each formula is of quantifier rank at most q. Let θ be a $\mathcal{L}(\sigma)_r$-formula. Then

$$\Phi^\star(\mathfrak{A}) \models \theta \text{ iff } \mathfrak{A} \models \Phi^\sharp(\theta)$$

and $\Phi^\sharp(\theta)$ is in $\mathcal{L}(\tau)_{q+r}$.

▶ Proposition 3.2 (Smooth operations).

(i) The rich disjoint union $\sqcup_{rich}$ of τ-structures and therefore also the disjoint union are FOL-smooth, MSOL-smooth and CMSOL-smooth. They are not SOL-smooth.

(ii) The Cartesian product $\times$ of τ-structures is FOL-smooth, but not MSOL-smooth

(iii) Let Φ be a quantifier-free scalar transduction of τ-structures into τ-structures and let $\Box$ be an $\mathcal{L}$-smooth operation. Then the operation $\Box_\Phi(\mathfrak{A}, \mathfrak{B}) = \Phi^\star(\mathfrak{A} \Box \mathfrak{B})$ $\mathcal{L}$-smooth. If Φ has quantifier rank at most k, it is $(k, \mathcal{L})$-smooth.

Sketch of proof. (i) is shown for FOL and MSOL using the usual pebble games. For CMSOL one can use Courcelle's version of the Feferman-Vaught Theorem for CMSOL, cf. [7, 8, 27]. (ii) is again shown using the pebble game for FOL. (iii) follows from Proposition 3.1. The negative statements are well-known, but also follow from the developments in the sequel. ◀

▶ Theorem 3.3 (Feferman-Vaught Theorem for CFOL).

(i) The rich disjoint union $\sqcup_{rich}$ of τ-structures, and therefore the disjoint union, too, is CFOL-smooth.

(ii) The Cartesian product $\times$ of τ-structures is CFOL-smooth.

Sketch of proof. The proof does not use pebble games, but Feferman-Vaught-type reduction sequences. (i) can be proven using the same reduction sequences which are used in [7, 8]. (ii) is proven using modifications of the reduction sequences as given in detail in [27, Theorem 1.6]. ◀

To the best of our knowledge, (ii) of Theorem 3.3 has not been stated in the literature before.

▶ Remark. We call this a Feferman-Vaught Theorem, because our proof actually computes the reduction sequences explicitly. However, this is not needed here, so we refer the reader to [27] for the definition of reduction sequences. One might also try to prove the theorem using the pebble games defined in [32], but at least for the case of the Cartesian product, the proof would be rather complicated and less transparent.

▶ Theorem 3.4 (Finite Rank Theorem for tame $\mathcal{L}$, [16, 30]).
Let $\mathcal{L}$ be a tame fragment of SOL. Let $\Box$ be a binary operation between τ-structures which is $\mathcal{L}$-smooth. Let $\mathcal{P}$ be a τ-property which is definable by a $\mathcal{L}$-formula ψ and $M(\Box, \psi)$ be the connection matrix defined by

$$M(\Box, \psi)_{\mathfrak{A},\mathfrak{B}} = 1 \text{ iff } \mathfrak{A}\Box\mathfrak{B} \models \psi \text{ and } 0 \text{ otherwise} .$$

Then

- **(i)** There is a finite partition $\{U_1, \ldots, U_k\}$ of the (finite) τ-structures such that the sub-matrices obtained from restricting $M(\Box, \psi)$ to $M(\Box, \psi)^{[U_i,U_j]}$ have constant entries.
- **(ii)** In particular, the infinite matrix $M(\Box, \psi)$ has finite rank over any field $\mathcal{F}$.
- **(iii)** $M(\Box, \psi)$ has an infinite sub-matrix of rank at most 1.

Sketch of proof. (i) follows from the definition of a tame fragment and of smoothness and the fact that there are only finitely many formulas (up to logical equivalence) in $\mathcal{L}(\tau)_q$. (ii) and (iii) follow from (i). ◀

4 Merits and Limitations of Connection Matrices

Merits

The advantages of the Finite Rank Theorem for tame $\mathcal{L}$ in proving that a property is not definable in $\mathcal{L}$ are the following:

- **(i)** It suffices to prove that certain binary operations on graphs (τ-structures) are $\mathcal{L}$-smooth operations.
- **(ii)** Once the $\mathcal{L}$-smoothness of a binary operation has been established, proofs of non-definability become surprisingly simple and transparent. One of the most striking examples is the fact that asymmetric (rigid) graphs are not definable in CMSOL, cf. Corollary 5.7.
- **(iii)** Many properties can be proven to be non-definable using the same or similar sub-matrices, i.e., matrices with the same row and column indices. This is well illustrated in the examples of Section 5.

Limitations

The classical method of proving non-definability in FOL using pebble games is complete in the sense that a property is $\mathrm{FOL}(\tau)_q$-definable iff the class of its models is closed under game equivalence of length q. Using pebble games one proves easily that the class of structures without any relations of even cardinality, EVEN, is not FOL-definable. This cannot be proven using connection matrices in the following sense:

▶ Proposition 4.1. Let Φ be a quantifier-free transduction between τ-structures and let $\Box_\Phi$ be the binary operation on τ-structures:

$$\Box_\Phi(\mathfrak{A}, \mathfrak{B}) = \Phi^\star(\mathfrak{A} \sqcup_{rich} \mathfrak{B})$$

Then the connection matrix $M(\Box_\Phi, \mathrm{EVEN})$ satisfies the properties (i)-(iii) of Theorem 3.4.

5 Proving Non-definability of Properties

Non-definability on CFOL

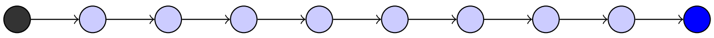
We will prove non-definability in CFOL using Theorem 3.3 for Cartesian products combined with FOL transductions. It is useful to consider a slight generalization of the Cartesian product as follows. We add two constant symbols *start* and *end* to our graphs. In $G^1 \times G^2$ the symbol *start* is interpreted as the pair of vertices $(v^1_{start}, v^2_{start})$ from G^1 and G^2 respectively such that v^i_{start} is the interpretation of $start_i$ (i.e. *start* in G^i) for $i = 1, 2$.

The transduction $\Phi_{sym}(x, y) = E_D(x, y) \vee E_D(y, x)$ transforms a digraph $D = (V_D, E_D)$ into an undirected graph whose edge relation is the symmetric closure of the edge relation of the digraph.

The following transduction Φ_F transforms the Cartesian product of two directed graphs $G^i = (V_1, E_1, v^i_{start}, v^i_{end})$ with the two constants $start_i$ and end_i, $i = 1, 2$ into a certain digraph. It is convenient to describe Φ_F as a tranduction of the two input graphs G^1 and G^2:

$$\begin{aligned}\Phi_F\left((v_1, v_2), (u_1, u_2)\right) \quad = \quad & (E_1(v_1, u_1) \wedge E_2(v_2, u_2)) \vee \\ & ((v_1, v_2), (u_1, u_2)) = ((start_1, start_2), (end_1, end_2))\end{aligned}$$

Consider the transduction obtained from Φ_F by applying Φ_{sym} when the input graphs are directed paths $P^i_{n_i}$ of length n_i. The input graphs look like this:



The result of the application of the transduction is given in Figure 1. The result of the

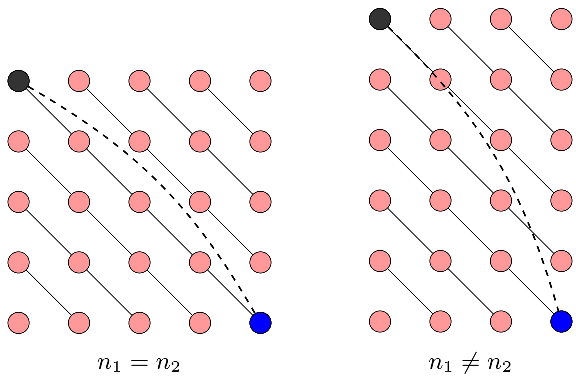


Figure 1 The result of applying Φ_F and then Φ_{sym} on the two directed paths. There is a cycle iff the two directed paths are of the same length.

transduction has a cycle iff $n_1 = n_2$. The length of this cycle is n_1. Hence, the connection sub-matrix with rows and columns labeled by directed paths of odd (even) length has ones on the main diagonal and zeros everywhere else, so it has infinite rank. Thus we have shown:

▶ **Theorem 5.1.** The graphs without cycles of odd (even) length are not CFOL-definable even in the presence of a linear order.

▶ **Corollary 5.2.** Not definable in CFOL with order are:
(i) Forests, bipartite graphs, chordal graphs, perfect graphs
(ii) interval graphs (cycles are not interval graphs)
(iii) Block graphs (every biconnected component is a clique)

(iv) Parity graphs (any two induced paths joining the same pair of vertices have the same parity)

The transduction

$$\begin{aligned}\Phi_T\left((v_1, v_2), (u_1, u_2)\right) &= (E_1(v_1, u_1) \wedge E_2(v_2, u_2)) \vee \\ &\quad (v_1 = u_1 = start_1 \wedge E(v_2, u_2)) \vee \\ &\quad (v_1 = u_1 = end_1 \wedge E(v_2, u_2)),\end{aligned}$$

combined with Φ_{sym} transforms the two directed paths into the structures in Figure 2.

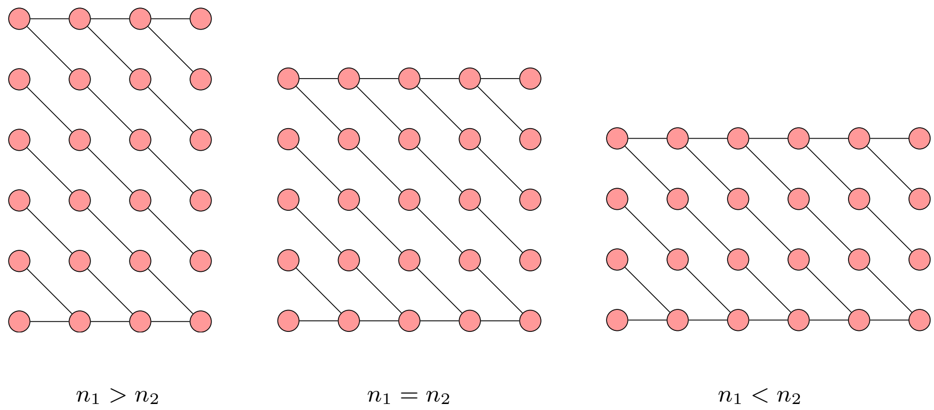


Figure 2 The result of applying Φ_T and then Φ_{sym} on two directed paths. We get a tree iff the two directed paths are of equal length.

So, the result of the transduction is a tree iff $n_1 = n_2$. It is connected iff $n_1 \leq n_2$. Hence, both the connection matrices with directed paths as row and column labels of the property of being a tree and of connectivity have infinite rank.

▶ Theorem 5.3. The properties of being a tree or a connected graph are not CFOL-definable even in the presence of linear order.

For our next connection matrix we use the 2-sum of the following two 2-graphs:

(i) the 2-graph (G, a, b) obtained from K_5 by choosing two vertices a and b and removing the edge between them

(ii) the symmetric closure of the Cartesian product of the two digraphs $P^1_{n_1}$ and $P^2_{n_2}$:

We denote this transduction by Φ_P, see Figure 3.

So, the result of this construction has a clique of size 5 as a minor iff $n_1 = n_2$. It can never have a $K_{3,3}$ as a minor.

▶ Theorem 5.4. The class of planar graphs is not CFOL-definable on ordered graphs.

If we modify the above construction by taking K_3 instead of K_5 and making $(start_1, start_2)$ and (end_1, end_2) adjacent, we get

▶ Corollary 5.5. The following classes of graphs are not CFOL-definable on ordered graphs.

(i) Cactus graphs, i.e. graphs in which any two cycles have at most one vertex in common.

(ii) Pseudo-forests, i.e. graphs in which every connected component has at most one cycle.

The case of connected graphs was also shown non-definable in CFOL by. J. Nurmonen in [32] using his version of the pebble games for CFOL.

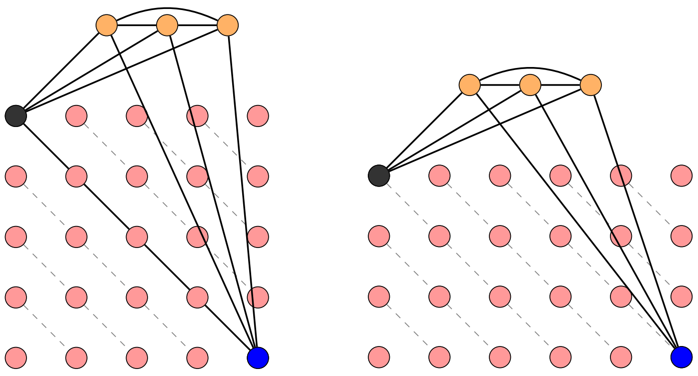


Figure 3 The result of Φ_P on two directed paths. The graph obtained here is planar iff the two directed paths are of equal length.

Non-definability in CMSOL

Considering the connection matrix where the rows and columns are labeled by the graphs on n vertices but without edges E_n, the graph $E_i \bowtie E_j = K_{i,j}$ is

(i) Hamiltonian iff $i = j$;
(ii) has a perfect matching iff $i = j$;
(iii) is a cage graph (a regular graph with as few vertices as possible for its girth) iff $i = j$;
(iv) is a well-covered graph (every minimal vertex cover has the same size as any other minimal vertex cover) iff $i = j$.

All of these connection matrices have infinite rank, so we get

▶ Corollary 5.6. None of the properties above are CMSOL-definable as graphs even in the presence of an order.

Using a modification $\boxtimes$ of the join operation used in [8, Remark 5.21] one can show the same for the class of graphs which have a spanning tree of degree at most 3. For any fixed natural number $d > 3$, by performing a transduction on $G \boxtimes H$ which attaches $d - 3$ new vertices as pendants to each vertex of $G \boxtimes H$, one can extend the non-definability result to the class of graphs which have a spanning tree of degree at most d.

For the language of hypergraphs we cannot use the join operation, since it is not smooth. Note also that Hamiltonian and having a perfect matching are both definable in CMSOL in the language of hypergraphs. But using the connection sub-matrices of the disjoint union we still get:

(i) Regular: $K_i \sqcup K_j$ is regular iff $i = j$;
(ii) A generalization of regular graphs are *bi-degree* graphs, i.e., graphs where every vertex has one of two possible degrees. $K_i \sqcup (K_j \sqcup K_1)$ is a bi-degree graph iff $i = j$.
(iii) The average degree of $K_i \sqcup K_j$ is at most $\frac{|V|}{2}$ iff $i = j$;
(iv) A digraph is *aperiodic* if the common denominator of the lengths of all cycles in the graph is 1. We denote by C_i^d the directed cycle with i vertices. For prime numbers p, q the digraphs $C_p \sqcup C_q$ is aperiodic iff $p \neq q$.
(v) A graph is *asymmetric (or rigid)* if it has no non-trivial automorphisms. It was shown by P. Erdös and A. Rényi [13] that almost all finite graphs are asymmetric. So there is

an infinite set $I \subseteq \mathbb{N}$ such that for $i \in I$ there is an asymmetric graph R_i of cardinality i. $R_i \sqcup R_j$ is asymmetric iff $i \neq j$.

▶ Corollary 5.7. None of the properties above are CMSOL-definable as hypergraphs even in the presence of an order.

▶ Remark. The case of asymmetric graphs illustrates that it is not always necessary to find explicit infinite families of graphs whose connection matrices are of infinite rank in order to show that such a family exists.

6 $\mathcal{L}$-Definable Graph Polynomials and Graph Parameters

$\mathcal{L}(\tau)$-polynomials

Here we follow closely the exposition from [23]. Let $\mathcal{L}$ be a tame fragment of SOL. We are now ready to introduce the $\mathcal{L}$-definable polynomials. They are defined for τ-structures and generally are called $\mathcal{L}(\tau)$ *invariants* as they map τ-structures into some commutative semi-ring $\mathcal{R}$, which contains the semi-ring of the integers $\mathbb{N}$, and are invariant under τ-isomorphisms. If τ is the vocabulary of graphs or hypergraphs, we speak of graph invariants and graph polynomials.

For our discussion $\mathcal{R} = \mathbb{N}$ or $\mathcal{R} = \mathbb{Z}$ suffices, but the definitions generalize. Our polynomials have a fixed finite set of variables (indeterminates, if we distinguish them from the variables of $\mathcal{L}$), $\mathbf{X}$.

▶ **Definition 1** ($\mathcal{L}$-monomials). Let $\mathcal{M}$ be a τ-structure. We first define the $\mathcal{L}$-definable $\mathcal{M}$*-monomials.* inductively.

(i) Elements of $\mathbb{N}$ are $\mathcal{L}$-definable $\mathcal{M}$-monomials.
(ii) Elements of $\mathbf{X}$ are $\mathcal{L}$-definable $\mathcal{M}$-monomials.
(iii) Finite products of monomials are $\mathcal{L}$-definable $\mathcal{M}$-monomials.
(iv) Let $\phi(a)$ be a $\tau \cup \{a\}$-formula in $\mathcal{L}$, where a is a constant symbol not in τ. Let t be a $\mathcal{M}$-monomial. Then $\prod_{a:\langle \mathcal{M},a\rangle \models \phi(a)} t$ is a $\mathcal{L}$-definable $\mathcal{M}$-monomial.

The monomial t may depend on relation or function symbols occurring in ϕ.

Note the degree of a monomial is polynomially bounded by the cardinality of $\mathcal{M}$.

▶ **Definition 2** ($\mathcal{L}$-polynomials). The $\mathcal{M}$-polynomials definable in $\mathcal{L}$ are defined inductively:

(i) $\mathcal{M}$-monomials are $\mathcal{L}$-definable $\mathcal{M}$-polynomials.
(ii) Let $\phi(\overline{a})$ be a $\tau \cup \{\overline{a}\}$-formula in $\mathcal{L}$ where $\overline{a} = (a_1, \ldots, a_m)$ is a finite sequence of constant symbols not in τ. Let t be a $\mathcal{M}$-polynomial. Then $\sum_{\overline{a}:\langle \mathcal{M},\overline{a}\rangle \models \phi(\overline{a})} t$ is a $\mathcal{L}$-definable $\mathcal{M}$-polynomial.
(iii) Let $\phi(\overline{R})$ be a $\tau \cup \{\overline{R}\}$-formula in $\mathcal{L}$ where $\overline{R} = (R_1, \ldots, R_m)$ is a finite sequence of relation symbols not in τ. Let t be a $\mathcal{M}$-polynomial definable in $\mathcal{L}$. Then $\sum_{\overline{R}:\langle \mathcal{M},\overline{R}\rangle \models \phi(\overline{R})} t$ is a $\mathcal{L}$-definable $\mathcal{M}$-polynomial.

The polynomial t may depend on relation or function symbols occurring in ϕ.

An $\mathcal{M}$-polynomial $p_{\mathcal{M}}(\mathbf{X})$ is an expression with parameter $\mathcal{M}$. The family of polynomials, which we obtain from this expression by letting $\mathcal{M}$ vary over all τ-structures, is called, by abuse of terminology, a $\mathcal{L}(\tau)$-polynomial.

Among the $\mathcal{L}$-definable polynomials we find most of the known graph polynomials from the literature, cf. [28, 23]. $\mathcal{L}$*-definable numeric graph parameters* are evaluations of $\mathcal{L}$-definable polynomials and take values in $\mathcal{R}$. $\mathcal{L}$*-definable properties* are special cases of numeric parameters which have boolean values.

Some simple graph parameters are even FOL-definable, e.g. $|V|$, the number of vertices and $|E|$, the number of edges. However, we leave the discussion of FOL-definable parameters for the journal version of this paper, and concentrate on tame fragments of SOL which do have second order variables.

Sum-like operations

For the proof of the The Finite Rank Theorem for $\mathcal{L}$-polynomials which involve second order variables it is not enough that the binary operation $\Box$ on τ-structures be $\mathcal{L}$-smooth. We need a way to uniquely decompose the relation over which we perform summation in $\mathfrak{A}\Box\mathfrak{B}$ into relations in $\mathfrak{A}$ and $\mathfrak{B}$ respectively, from which we can reconstruct the relation in $\mathfrak{A}\Box\mathfrak{B}$. For our discussion here it suffices to restrict $\Box$ to *$\mathcal{L}$-sum-like operations*. $\mathfrak{A}\Box\mathfrak{B}$ is *$\mathcal{L}$-sum-like* if there is a *scalar* $\mathcal{L}$-transduction Φ such that

$$\mathfrak{A}\Box\mathfrak{B} = \Phi^\star(\mathfrak{A} \sqcup_{rich} \mathfrak{B}).$$

An operation is *$\mathcal{L}$-product-like* if instead of scalar transductions we also allow *vectorized* transductions. Typically, the Cartesian product is FOL-product-like, but not sum-like. The k-sum and the join operation on graphs are FOL-sum-like (but, in the case of join, not on hypergraphs).

The Finite Rank Theorem for $\mathcal{L}$-polynomials

Now we can state the Finite Rank Theorem for $\mathcal{L}$-polynomials. The proof uses the same techniques as in [9, 27].

▶ Theorem 6.1 (The Finite Rank Theorem for $\mathcal{L}$-polynomials).
Let $\mathcal{L}$ be a tame fragment of SOL and $\Box$ be an $\mathcal{L}$-sum-like operation between τ-structures which is $(s, \mathcal{L})$-smooth. Let P be an $\mathcal{L}(\tau)$-polynomial. Then the connection matrix $M(\Box, P)$ has finite rank.

▶ Remark. In [16] the theorem was only formulated for k-sums, and the join operation and for the logic CMSOL.

7 Non-definability of $\mathcal{L}(\tau)$-invariants

7.1 Numeric $\mathcal{L}(\tau)$-parameters

Theorem 6.1 can be used to show that many τ-parameters are not $\mathcal{L}$-definable.

$\Box$-maximizing and $\Box$-minimizing parameters

We say a τ-parameter f is *$\Box$-maximizing* (*$\Box$-minimizing*) if there exist an infinite sequence of non-isomorphic τ-structures $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \ldots, \mathcal{A}_i, \ldots$ such that for any $i \neq j$,

$$f(\mathcal{A}_i \Box \mathcal{A}_j) = \max\{f(\mathcal{A}_i), f(\mathcal{A}_j)\}\,.$$

Furthermore, if f is unbounded on $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \ldots$ then f is *unboundedly $\Box$-maximizing*. Analogously we define *(unboundedly) $\Box$-minimizing*.

▶ Proposition 7.1. If f is a unboundedly $\Box$-maximizing ($\Box$-minimizing) τ-parameter, then $M(f, \Box)$ has infinite rank.

Using Proposition 7.1 we show that many τ-parameters are not CMSOL-definable:

▶ Proposition 7.2. Let $\mathcal{L}$ be tame and $\sqcup$-smooth. The following graph parameters are not $\mathcal{L}$-definable in the language of hypergraphs. In particular they are not CMSOL-definable. *Spectral radius, chromatic number, acyclic chromatic number, arboricity, star chromatic number, clique number, Hadwiger number, Hajós number, tree-width, path-width, clique-width, edge chromatic number, Thue number, maximum valency, circumference, longest path, maximal connected planar (bipartite) induced subgraph, boxicity, minimal eigenvalue, spectral gap, girth, degeneracy,* and *minimum valency*.

Proof. All these graph parameters g are unboundedly $\sqcup$-maximizing or $\sqcup$-minimizing. ◀

Variations of the notions of $\Box$-maximizing or $\Box$-minimizing τ-parameters can also lead to non-definability results, e.g.:

▶ Proposition 7.3. Under the same assumption on $\mathcal{L}$ as before, the number of connected components (blocks, simple cycles, induced paths) of maximum (minimum) size is not $\mathcal{L}$-definable in the language of hypergraphs.

Proof. Consider the connection matrix of graphs $i \cdot K_i$ which consist of the disjoint union of i cliques of size i with the operation of disjoint union. We denote the number of connected components of maximum size in a graph G by $\#_{\max-\mathrm{cc}}(G)$. Then

$$\#_{\max-\mathrm{cc}}(nK_n \sqcup mK_m) = \begin{cases} \max\{n,m\} & n \neq m \\ n+m & n = m \end{cases}$$

So $M(\#_{\max-\mathrm{cc}}, \sqcup)$ is of infinite rank. The other cases are proved similarly. ◀

7.2 τ-polynomials

Here we use the method of connection matrices for showing that (hyper)graph polynomials are not MSOL-definable. Some of the material here is taken from the first author's thesis [33]. As examples we consider the polynomials $\chi_{rainbow}(G,k)$, $\chi_{mcc(t)}(G,k)$, and $\chi_{convex}(G,k)$, which were defined in the introduction.

To show that none of $\chi_{rainbow}(G,k)$, $\chi_{mcc(t)}(G,k)$, or $\chi_{convex}(G,k)$ are CMSOL-polynomials in the language of graphs, and that neither $\chi_{rainbow}(G,k)$ nor $\chi_{convex}(G,k)$ are CMSOL-polynomials in the language of hypergraphs, we prove the following general proposition:

▶ Lemma 7.4. Given a τ-parameter p, a binary operation $\Box$ on τ-structures and an infinite sequence of non-isomorphic τ-structures $\mathcal{A}_i, i \in \mathbb{N}$, let $f : \mathbb{N} \to \mathbb{N}$ be an unbounded function such that for every $\lambda \in \mathbb{N}$, $p(\mathcal{A}_i \Box \mathcal{A}_j, \lambda) = 0$ iff $i + j > f(\lambda)$. Then the connection matrix $M(\Box, p)$ has infinite rank.

Proof. Let $\lambda \in \mathbb{N}$ and let p_λ be the graph parameter given by $p_\lambda(G) = p(G, \lambda)$. The restriction of the connection matrix $M(p_\lambda, \Box)$ to the rows and columns corresponding to $\mathcal{A}_i$, $0 \leq i \leq f(\lambda) - 1$, yields a finite triangular matrix with non-zero diagonal. Hence the rank of $M(p_\lambda, \Box)$ is at least $f(\lambda) - 1$.

Using that f is unbounded, we get that $M(p, \Box)$ contains infinitely many finite sub-matrices with ranks which tend to infinity. Hence, the rank of $M(p, \Box)$ is infinite, ◀

We now use Lemma 7.4 to compute connection matrices where $\Box$ is the disjoint union $\sqcup$, the 1-sum $\sqcup_1$ or the join $\bowtie$.

▶ Proposition 7.5. The following connection matrices have infinite rank:

(i) $M(\sqcup_1, \chi_{rainbow}(G,k))$;
(ii) $M(\sqcup_1, \chi_{convex}(G,k))$;
(iii) For every $t > 0$ the matrix $M(\bowtie, \chi_{mcc(t)}(G,k))$;

Proof.
(i) For $\chi_{rainbow}(G,k)$, we use that the 1-sum of paths with one end labeled is again a path with $P_i \sqcup_1 P_j = P_{i+j-1}$ and that $\chi_{rainbow}(P_r,k) = 0$ iff $r > k+3$.
(ii) For $\chi_{convex}(G,k)$, we use edgeless graphs and disjoint union $E_i \sqcup E_j = E_{i+j}$ and that $\chi_{convex}(E_r,k) = 0$ iff $r > k$.
(iii) For $\chi_{mcc(t)}(G,k)$ we use the join and cliques, $K_i \bowtie K_j = K_{i+j}$ and that $\chi_{mcc(t)}(K_r,k) = 0$ iff $r > kt$.

◀

▶ **Corollary 3.**
(i) $\chi_{rainbow}(G,k)$ *and* $\chi_{convex}(G,k)$ *are not* CMSOL-*definable in the language of graphs and hypergraphs.*
(ii) $\chi_{mcc(t)}(G,k)$ *(for any fixed* $t > 0$*) is not* CMSOL-*definable in the language of graphs.*

Proof. (i) The 1-sum and the disjoint union are CMSOL-sum-like and CMSOL-smooth for hypergraphs. (ii) The join is only CMSOL-sum-like and CMSOL-smooth for graphs. ◀

The same method yields non-definability results for other graph polynomials which arise by counting other graph colorings from the literature, such as acyclic colorings, non-repetitive colorings, t-improper colorings, co-colorings, sub-colorings and G-free colorings.

References

1 N. Alon, G. Ding, B. Oporowski, and D. Vertigan. Partitioning into graphs with only small components. *Journal of Combinatorial Theory, Series B*, 87(2):231–243, 2003.
2 S. Arnborg, J. Lagergren, and D. Seese. Problems easy for tree-decomposable graphs (extended abstract). In *ICALP*, pages 38–51, 1988.
3 S. Arnborg, J. Lagergren, and D. Seese. Easy problems for tree decomposable graphs. *Journal of Algorithms*, 12:308–340, 1991.
4 B. Bollobás. *Modern Graph Theory.* Springer, 1999.
5 S. Chakraborty, E. Fischer, A. Matsliah, and R. Yuster. Hardness and algorithms for rainbow connection. *Journal of Combinatorial Optimization*, 21(3):330–347, 2011.
6 G. Chartrand, G.L. Johns, K.A. McKeon, and P. Zhang. Rainbow connection in graphs. *Mathematica Bohemica*, 133(1):85–98, 2008.
7 B. Courcelle. Graph rewriting: An algebraic and logic approach. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume 2, chapter 5, pages 193–242. Elsevier Science Publishers, 1990.
8 B. Courcelle and J. Engelfriet. *Graph Structure and Monadic Second-order Logic, a Language Theoretic Approach.* Cambridge University Press, 2012.
9 B. Courcelle, J.A. Makowsky, and U. Rotics. On the fixed parameter complexity of graph enumeration problems definable in monadic second order logic. *Discrete Applied Mathematics*, 108(1-2):23–52, 2001.
10 R. Diestel. *Graph Theory.* Graduate Texts in Mathematics. Springer, 3 edition, 2005.
11 H.-D. Ebbinghaus and J. Flum. *Finite Model Theory.* Springer Verlag, 1995.
12 K. Edwards. The harmonious chromatic number and the achromatic number. In R. A. Bailey, editor, *Survey in Combinatorics*, volume 241 of *London Math. Soc. Lecture Note Ser.*, pages 13–47. Cambridge Univ. Press, 1997.

13 P. Erdös and A. Rényi. Asymmetric graphs. *Acta Mathematica Hungaric*, 14(3):295–315, 1963.
14 A. Farrugia. Vertex-partitioning into fixed additive induced-hereditary properties is NP-hard. *Electric Journal of Combinatorics*, 11(1), 2004.
15 M. Freedman, László Lovász, and A. Schrijver. Reflection positivity, rank connectivity, and homomorphisms of graphs. *Journal of AMS*, 20:37–51, 2007.
16 B. Godlin, T. Kotek, and J.A. Makowsky. Evaluation of graph polynomials. In *34th International Workshop on Graph-Theoretic Concepts in Computer Science, WG08*, volume 5344 of *Lecture Notes in Computer Science*, pages 183–194, 2008.
17 A.J. Goodall and S.D. Noble. Counting cocircuits and convex two-colourings is #P-complete, 2008. http://arxiv.org/abs/0810.2042.
18 M. A. Harrison. *Introduction to Formal Language Theory.* Addison-Wesley, 1978.
19 J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages and Computation.* Addison-Wesley Series in Computer Science. Addison-Wesley, 1980.
20 J.E. Hopcroft and M.S. Krishnamoorthy. On the harmonious coloring of graphs. *SIAM J. Algebraic Discrete Methods*, 4:306–311, 1983.
21 Phokion G. Kolaitis and Jouko A. Väänänen. Generalized quantifiers and pebble games on finite structures. *Ann. Pure Appl. Logic*, 74(1):23–75, 1995.
22 T. Kotek, J.A. Makowsky, and B. Zilber. On counting generalized colorings. In *Computer Science Logic, CSL'08*, volume 5213 of *Lecture Notes in Computer Science*, page 339–353, 2008.
23 T. Kotek, J.A. Makowsky, and B. Zilber. On counting generalized colorings. In M. Grohe and J.A. Makowsky, editors, *Model Theoretic Methods in Finite Combinatorics*, volume 558 of *Contemporary Mathematics*, pages 207–242. American Mathematical Society, 2011.
24 L. Libkin. *Elements of Finite Model Theory.* Springer, 2004.
25 N. Linial, J. Matousek, O. Sheffet, and G. Tardos. Graph coloring with no large monochromatic components. *Combinatorics, Probability, and Computing*, 17.4:577–589, 2008.
26 L. Lovász. Connection matrics. In G. Grimmet and C. McDiarmid, editors, *Combinatorics, Complexity and Chance, A Tribute to Dominic Welsh*, pages 179–190. Oxford University Press, 2007.
27 J.A. Makowsky. Algorithmic uses of the Feferman-Vaught theorem. *Annals of Pure and Applied Logic*, 126.1-3:159–213, 2004.
28 J.A. Makowsky. From a zoo to a zoology: Towards a general theory of graph polynomials. *Theory of Computing Systems*, 43:542–562, 2008.
29 J.A. Makowsky and B. Zilber. Polynomial invariants of graphs and totally categorical theories. MODNET Preprint No. 21,
`http://www.logique.jussieu.fr/`
`modnet/Publications/Preprint%20 server`, 2006.
30 Johann A. Makowsky. Connection matrices for msol-definable structural invariants. In *ICLA*, pages 51–64, 2009.
31 S. Moran and S. Snir. Efficient approximation of convex recolorings. *Journal of Computer and System Sciences*, 73(7):1078–1089, 2007.
32 Juha Nurmonen. Counting modulo quantifiers on finite structures. *Inf. Comput.*, 160(1-2):62–87, 2000.
33 T. Kotek. *Definability of Combinatorial Functions.* PhD thesis, Technion—Israel Institute of Technology, 2012.