CONNECTION MATRICES AND THE DEFINABILITY OF GRAPH PARAMETERS

TOMER KOTEK [a] AND JOHANN A. MAKOWSKY [b]

[a] Faculty of Informatics, Vienna University of Technology, Vienna, Austria
   e-mail address: kotek@forsyte.at

[b] Faculty of Computer Science, Technion–Israel Institute of Technology, Haifa, Israel
   e-mail address: janos@cs.technion.ac.il

ABSTRACT. In this paper we extend and prove in detail the Finite Rank Theorem for connection matrices of graph parameters definable in Monadic Second Order Logic with counting (CMSOL) from B. Godlin, T. Kotek and J.A. Makowsky (2008) and J.A. Makowsky (2009). We 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.

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 [22, 40]. However these techniques are not easily applicable for stronger logics, such as CFOL and CMSOL, which extend FOL respectively MSOL with modular counting quantifiers $D_{m,i}x\phi(x)$ which say that the number of elements satisfying ϕ equals i modulo m. 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. Definability in MSOL and CMSOL depends on whether the vocabulary of graphs or hypergraphs is considered. The vocabulary of graphs is τ_G consisting of the edge relation symbol E. The vocabulary τ_{HG} of hypergraphs is two-sorted, with two unary relation symbols V and E whose interpretations partition the universe, and has the binary incidence relation R_{inc} between V elements and E elements. Structures of τ_{HG} can be not only graphs but hypergraphs. In [17] MSOL over τ_G is denoted MS_1, while MSOL over τ_{HG} is denoted MS_2. In the case of FOL the choice of τ_G or τ_{HG} does not have any effect on definablility.

The notion of definability was extended in [5, 6] to integer valued graph parameters, and in [18, 44, 46, 37, 38] to real or complex valued graph parameters and graph polynomials. In [44] and [37] 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.

We assume the reader is familiar with the basics of finite model theory [22, 40] and graph theory [9, 20].

Connection matrices. *Connection matrices* were introduced in [26, 42] by M. Freedman, L. Lovász and A. Schrijver where they were used to characterize to characterize various partition functions based on graph homomorphisms, cf. also [43]. Let f be a graph parameter whose image is in some field $\mathbb{F}$ such as the real or the complex numbers. 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 [27] connection matrices were used to show that certain graph parameters and polynomials are not MSOL-definable. The main result of [27] 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 [47] to matrices $M(\square, f)$ where $\square$ 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. 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. [35, 50], 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 [16, 17, 44].

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);
   - Bridgeless graphs, k-connected.

   The case of connected graphs was also shown undefinable in CFOL by. J. Nurmonen in [50] 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 [17, Remark 5.21, Page 350].

(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 7. 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. [45]. This led the second

author to express his belief in [45] that all "naturally occurring graph polynomials" are CMSOL-definable. However, in [27] 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. [23, 34]. That this is indeed a graph polynomial was shown in [38]. However, the main thrust of [27] consists in showing that certain graph parameters are not evaluations of the standard prominent graph polynomials.

In Section 9, 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 [38]. 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 in the language of graphs:

- The *chromatic polynomial* $\chi(G, k)$ which counts proper vertex k-colorings of G;
- For every fixed $t \in \mathbb{N}^+$, $\chi_{mcc(t)}(G, k)$ which counts vertex k-colorings $f : V(G) \to [k]$ for which no color induces a subgraph with a connected component of size larger than t;
- The *vertex-acyclic polynomial* $\chi_{v-acyclic}(G, k)$ which counts proper vertex k-colorings $f : V(G) \to [k]$ such that there is no two colored cycle in G.

We show that the following graph polynomials are not even CMSOL-definable in the language of hypergraphs:

- The *rainbow polynomial* $\chi_{rainbow}(G, k)$ which counts *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.
- $\chi_{convex}(G, k)$ 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.
- For every fixed $t \in \mathbb{N}^+$, $\chi_{t-improper}(G, k)$ which counts *t-improper colorings*. A t-improper coloring is function $f : V(G) \to [k]$ for which every color induces a graph in which no vertex has degree more than t.
- $\chi_{non-rep}(G, k)$ which counts *non-repetitive colorings*. A function f : E(G) [k] is a non-repetitive coloring if the sequence of colors on any path in G is non-repetitive. A sequence $a_1, \ldots, a_r$ is non-repetitive if there is no $i, j \geq 1$ such that $(a_i, \ldots, a_{i+j-1}) = (a_{i+j}, \ldots, a_{i+2j-1})$.
- The *harmonious polynomial* $\chi_{harm}(G, k)$ which counts proper vertex k-colorings $f : V(G) \to [k]$ such that for any two distinct edges (u_1, u_2) and (v_1, v_2), it holds that $\{f(u_1), f(u_2)\} \neq \{f(v_1), f(v_2)\}$. The underlying proof idea here is similar to that in [27]. We bring it here for completeness.

Path-rainbow connected colorings were introduced in [15] and their complexity was studied in [14]. $mcc(t)$-colorings were studied in [1], [41] and [25]. Note $\chi_{mcc(1)}(G, k)$ is the chromatic polynomial. Convex colorings were studied for their complexity e.g. in [49] and [28]. From [38] 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. Acyclic vertex colorings were introduced in [29] and A. V. Kostochka proved in 1978 in his thesis that it is NP-hard to decide for a given G and k if the there exists an acyclic vertex coloring with at most k colors, see [3]. Acyclic edge colorings were studied e.g. in [4, 51, 7]. It is NP-hard to determine whether G is t-improperly 2-colorable for any fixed positive t (even if G is planar), cf. [19]. Non-repetitive colorings were introduced in [2].

Their complexity was studied in [48]. The minimal number of colors needed to color G in a non-repetitive way is called the *Thue number* of G.

Section 9 contains more examples of graph polynomials and graph parameters not definable in CMSOL.

Outline of the paper. 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 give a proof of the Feferman-Vaught theorem for CFOL. In Section 6 we illustrate the use of connection matrices and the Finite Rank Theorem for proving non-definability of properties. In Section 7 we recall the framework of definable graph polynomials and τ-polynomials and the corresponding definable numeric parameters. In Section 8 we prove a Feferman-Vaught-type theorem for CMSOL-polynomials with respect to sum-like operations, which is the main ingredient in Finite Rank Theorems for CMSOL-polynomials, and in Section 9 we show how to prove non-definability of many numeric graph invariants.

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 [22, 40], 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 $u \circ v \in L$. [1]

The Myhill-Nerode Theorem, see [33, 32], 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[2] $M(\circ, L)^{[U_i, U_j]}$ 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 $m_L(u\bar{\circ}v) = \max\{m_L(u), m_L(v)\}$.

---

[1] Strictly speaking we should use the characteristic function of L rather than L. We allow this slight abuse of notation to achieve simpler notation.

[2] $M(\circ, L)^{[U_i, U_j]}$ denotes the submatrix of $M(\circ, L)$ with rows and columns corresponding to U_i and U_j respectively.

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: $\ell(w) = \sum_{u <_{in} w} 1$, 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 7. 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.*

Connection matrices for concatenation of words are known in Automata Theory as Hankel-matrices and were introduced in [13], see also [8] and [21].

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 the boolean operations $\wedge, \vee, \neg$ and under renaming of relation and constant symbols.

(ii) $\mathcal{L}$ is equipped with a notion of *quantifier rank* $qr : \mathcal{L} \to \mathbb{N}$ and we denote by $\mathcal{L}_q(\tau)$ the set of formulas of quantifier rank at most q. The quantifier rank of atomic formulas is 0. The quantifier rank of a Boolean combination of formulas $\alpha_1, \ldots, \alpha_t \in \mathcal{L}(\tau)$ is the maximum quantifier rank of $\alpha_1, \ldots, \alpha_t$. 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-distingui-shability*, and we write $\mathfrak{A} \equiv^{\mathcal{L}}_q \mathfrak{B}$, if they satisfy the same sentences from $\mathcal{L}_q(\tau)$.

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

$$\mathfrak{A}_1 \square \mathfrak{A}_2 \equiv^{\mathcal{L}}_q \mathfrak{B}_1 \square \mathfrak{B}_2.$$

A binary operation $\square$ between τ-structures is *$\mathcal{L}$-smooth* if it is $(0,\mathcal{L})$-smooth.

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[3]; $\mathfrak{A} \sqcup_{rich} \mathfrak{B}$ 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 obtained from the disjoint union of G and H by adding all possible edges between vertices of G and vertices of H.

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 [40, 44].

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 and containment of the form $\forall x(\varphi(x) \to \psi(x))$. 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 k-sum $\sqcup_k$ of τ-structures is MSOL-smooth and CMSOL-smooth, but not SOL-smooth, for $k \in \mathbb{N}$.*

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

(iv) *The join $\bowtie$ of τ-structures is FOL-smooth and MSOL- and CMSOL-smooth in the vocabulary of graphs, but is not MSOL- and CMSOL-smooth in the vocabulary of hypergraphs.*

(v) *Let Φ be a quantifier-free scalar transduction of τ-structures into τ-structures and let $\square$ be an $\mathcal{L}$-smooth operation. Then the operation $\square_{\Phi}(\mathfrak{A}, \mathfrak{B}) = \Phi^{\star}(\mathfrak{A} \square \mathfrak{B})$ is $\mathcal{L}$-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. [16, 17, 44].

---

[3] The standard definition of disjoint union e.g. in [22] is for relational structures, however we also allow contant symbols. For two relational τ-structures $\mathfrak{A}$ and $\mathfrak{B}$ and tuples $\bar{a}$ and $\bar{b}$ of A respectively B elements, we denote by $\langle \mathfrak{A}, \bar{a} \rangle \sqcup \langle \mathfrak{B}, \bar{b} \rangle$ the disjoint union $\mathfrak{A} \sqcup \mathfrak{B}$ extended with the tuples $\bar{a}$ and $\bar{b}$, i.e. the structure $\langle \mathfrak{A} \sqcup \mathfrak{B}, \bar{a}, \bar{b} \rangle$.

(iii) is again shown using the pebble game for FOL. (v) follows from Proposition 3.1. The negative statements are well-known, but also follow from the developments in the sequel. (ii) and (iv) follow from (i) and (v). □

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 [16, 17]. We will prove (ii) in Section 5. □

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

Remark 3.4. Theorem 3.3(ii) is proven using modifications of the reduction sequences from [44, Theorem 1.6]. Reduction sequences are tuples of formulas $\langle \psi_1^A, \ldots, \psi_m^A, \psi_1^B, \ldots, \psi_n^B \rangle$ obtained from a formula ϕ so that the truth-value of $\mathfrak{A}\square\mathfrak{B}$ on ϕ is a Boolean combination of the truth-values of $\mathfrak{A}$ on $\psi_1^A, \ldots, \psi_m^A$ and $\mathfrak{B}$ on $\psi_1^B, \ldots, \psi_m^B$. We call this a Feferman-Vaught Theorem, because our proof actually computes the reduction sequences explicitly. One might also try to prove the theorem using the pebble games defined in [50], but at least for the case of the Cartesian product, the proof would be rather complicated and less transparent.

Theorem 3.5 (Finite Rank Theorem for tame $\mathcal{L}$, [27, 47])**.**
Let $\mathcal{L}$ be a tame fragment of SOL. *Let $\square$ 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 $M(\square, \psi)$ be the connection matrix defined by*

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

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 $M(\square, \psi)$ to $M(\square, \psi)^{[U_i, U_j]}$ have constant entries.*

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

(iii) *$M(\square, \psi)$ 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) 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 6.9.

(ii) 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 6.

Limitations. The classical method of proving non-definability in FOL using pebble games is complete in the sense that a property is $\text{FOL}(\tau)_q$-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 $\square_\Phi$ be the binary operation on τ-structures:*

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

Then the connection matrix $M(\square_\Phi, \text{EVEN})$ satisfies the properties (i)-(iii) of Theorem 3.5.

5. Proof of the smoothness of Cartesian product in CFOL

Recall that we denote by $\text{D}_{m,i}$ the modular counting quantifiers $\text{D}_{m,i}x\phi(x)$ which says that the number of elements satisfying ϕ equals i modulo m.

The proof of Theorem 3.3(ii) follows exactly the proof of Theorem 1.6 in [44], in which an analogous statement was proven for the ordered sum of structures. We spell out the changes needed in the proof from [44] for the ordered product $\times$.

Proof of Theorem 3.3. Given $G_1 = \langle V_1, E_2, <_1 \rangle$ and $G_2 = \langle V_2, E_2, <_2 \rangle$, their ordered product is given by $\langle V_1, E_2 \rangle \times \langle V_2, E_2 \rangle$ together with the lexicographic order $<$ on $V_1 \times V_2$ induced by $<_1$ and $<_2$.

The proof proceeds by induction and computes the reduction sequences of CFOL-formulas for $\times$. Reduction sequences are tuples of formulas $\langle \psi_1^1, \ldots, \psi_m^1, \psi_1^2, \ldots, \psi_n^2 \rangle$ obtained from a formula ϕ so that the truth-value of $G_1 \times G_2$ on ϕ is a Boolean combination of the truth-values $b_1^1, \ldots, b_m^1$ of G_1 on $\psi_1^1, \ldots, \psi_m^1$ and the truth-values $b_1^2, \ldots, b_m^2$ of G_2 on $\psi_1^2, \ldots, \psi_m^2$. The existence of reduction sequences directly implies smoothness.

The reduction sequences and Boolean functions of the atomic relations are given as follows. The difficult cases are those of the quantifiers. We discuss $\text{D}_{2,0}$ in detail. For all other cases we just state the reduction sequences and the corresponding Boolean formulas. See the proof of Theorem 1.6 in [44] for a detailed discussion of $\exists x\phi$.

For $E(u,v)$:
    Reduction sequence: $\langle E_1(u,v), E_2(u,v) \rangle$
    Boolean function: $b_1^1 \wedge b_1^2$.

For $u \approx v$:
    Reduction sequence: $\langle u \approx_1 v, u \approx_2 v \rangle$
    Boolean function: $b_1^1 \wedge b_1^2$.

For $u < v$:
    Reduction sequence: $\langle u <_1 v, u \approx_1 v, u <_2 v, u \approx_2 v \rangle$
    Boolean function: $b_1^1 \vee \left(b_2^1 \wedge b_1^2 \right)$.

Let $\Phi = \left\langle \phi_1^1, \ldots, \phi_m^1, \phi_1^2, \ldots, \phi_m^2 \right\rangle$ and $\Psi = \left\langle \psi_1^1, \ldots, \psi_m^1, \psi_1^2, \ldots, \psi_m^2 \right\rangle$ be reduction sequences for ϕ and ψ respectively, and let $B_\phi(\bar{b})$ and $B_\psi(\bar{b'})$ be the corresponding Boolean functions with disjoint variables. .

For $(\phi \wedge \psi)$**:**

   Reduction sequences: $\left\langle \Phi, \Psi \right\rangle$

   Boolean function: $B_\phi(\bar{b}) \wedge B_\psi(\bar{b'})$.

For $\neg\phi$**:**

   Reduction sequences: Φ

   Boolean function: $\neg B_\phi(\bar{b})$.

Now we turn to $D_{2,0} x\phi$. We look at $B_\phi(\bar{b})$ in disjunctive normal form:

$$B_1 = \bigvee_{j \in J} C_j$$

with

$$C_j = C_j^A \wedge C_j^B$$

and

$$C_j^A = \left(\bigwedge_{i \in J(j,A,pos)} b_i^A \bigwedge_{i \in J(j,A,neg)} \neg b_i^A \right)$$

and

$$C_j^B = \left(\bigwedge_{i \in J(j,B,pos)} b_i^B \bigwedge_{i \in J(j,B,neg)} \neg b_i^B \right) .$$

B_1 has $2m$ Boolean variables, $b_1^A, \ldots, b_m^A, b_1^B, \ldots, b_m^B$. We assume without loss of generality that every C_j contains all of the variables. In other words, $\{1, \ldots, m\} \backslash J(j, A, pos) = J(j, A, neg)$ and $\{1, \ldots, m\} \backslash J(j, B, pos) = J(j, B, neg)$ for every $j \in J$.

Now let:

$$\alpha_j^A = \left(\bigwedge_{i \in J(j,A,pos)} \phi_i^A(x) \bigwedge_{i \in J(j,A,neg)} \neg\phi_i^A(x) \right)$$

and similarly

$$\alpha_j^B = \left(\bigwedge_{i \in J(j,B,pos)} \phi_i^B(x) \bigwedge_{i \in J(j,B,neg)} \neg\phi_i^B(x) \right)$$

Consider the formula $D_{2,0} x\phi$. $\mathfrak{A} \times \mathfrak{B} \models D_{2,0} x\phi$ iff the number of pairs $(a,b) \in A \times B$ such that $\langle \mathfrak{A} \times \mathfrak{B}, (a,b) \rangle \models \phi$ is even. $\langle \mathfrak{A} \times \mathfrak{B}, (a,b) \rangle \models \phi$ iff B_1 holds for $\langle \mathfrak{A}, a \rangle$ and $\langle \mathfrak{B}, b \rangle$. Note that $\langle \mathfrak{A}, a \rangle$ satisfies at most one of the α_j^A and similarly for $\langle \mathfrak{B}, b \rangle$.

The number of pairs (a,b) for which B_1 holds is even iff the number of C_j such that C_j holds for an odd number of pairs (a,b) is even. This holds iff the number of C_j such that C_j^A holds for an odd number of $a \in A$ and C_j^B holds for an odd number of $b \in B$, is even.

Let

$$\beta_j^A = D_{2,0} x\alpha_j^A \quad \text{and} \quad \beta_j^B = D_{2,0} x\alpha_j^B$$

and let

$$P = \{ T \subseteq J \mid |T| \text{ is even } \}.$$

Finally we put

$$B_{D_{2,0}}(\bar{c}) = \bigvee_{T \in P} \left(\bigwedge_{j \in T} (\neg c_j^A \wedge \neg c_j^B) \bigwedge_{j \notin T} (c_j^A \vee c_j^B) \right)$$

where $c_j^A = 1$ iff $\mathfrak{A} \models \beta_j^A$ and $c_j^B = 1$ iff $\mathfrak{B} \models \beta_j^B$.

So, we have:

For $D_{2,0} x \phi$:

Reduction sequence: $\left\langle \beta_1^A, \ldots, \beta_{m(J)}^A, \beta_1^B, \ldots, \beta_{m(J)}^B \right\rangle$

Boolean function: $B_{D_{2,0}}(\bar{c})$.

For $\exists x \phi$:

Reduction sequence: $\left\langle \theta_1^A, \ldots, \theta_{m(J)}^A, \theta_1^B, \ldots, \theta_{m(J)}^B \right\rangle$

Boolean function: $B_\exists(\bar{c}) = \bigvee_{j \in J} (c_j^A \wedge c_j^B)$.

$\square$

A similar proof covers all quantifiers $D_{a,b}$. E.g. consider the quantifiers $D_{3,0}, D_{3,1}, D_{3,2}$. Here we set, for each $i \in \{0, 1, 2\}$:

$$B_{D_{3,i}}(\bar{c}) = \bigvee_{T_{1,1}, T_{1,2}, T_{2,1}, T_{2,2}} \left(\bigwedge_{j \in T_{1,1}} (c_{j,1}^A \wedge c_{j,1}^B) \wedge \bigwedge_{j \in T_{1,2}} (c_{j,1}^A \wedge c_{j,2}^B) \wedge \right.$$
$$\bigwedge_{j \in T_{2,1}} (c_{j,2}^A \wedge c_{j,1}^B) \wedge \bigwedge_{j \in T_{2,2}} (c_{j,2}^A \wedge c_{j,2}^B) \wedge$$
$$\left. \bigwedge_{j \in J \setminus (T_{1,1} \cup T_{1,2} \cup T_{2,1} \cup T_{2,2})} (c_{j,0}^A \vee c_{j,0}^B) \right)$$

where the outer $\bigvee$ is over tuples $(T_{1,1}, T_{1,2}, T_{2,1}, T_{2,2})$ of disjoint subsets of J, which additionally satisfy that

$$|T_{1,1}| + |T_{2,2}| + 2|T_{1,2}| + 2|T_{2,1}| \equiv i \mod 3.$$

For every $i \in \{0, 1, 2\}$, $c_{j,i}^A = 1$ iff $\mathfrak{A} \models D_{3,i} x \alpha_j^A$ and $c_{j,i}^B = 1$ iff $\mathfrak{B} \models D_{3,i} x \alpha_j^A$. The reduction sequence of $D_{3,i} x \phi$ is [4]

$$\left\langle D_{3,0} x \alpha_1^A, D_{3,1} x \alpha_1^A, D_{3,2} x \alpha_1^A, \ldots, D_{3,0} x \alpha_{m(J)}^A, D_{3,1} x \alpha_{m(J)}^A, D_{3,2} x \alpha_{m(J)}^A, \right.$$
$$\left. D_{3,0} x \alpha_1^B, D_{3,1} x \alpha_1^B, D_{3,2} x \alpha_1^B, \ldots, D_{3,0} x \alpha_{m(J)}^B, D_{3,1} x \alpha_{m(J)}^B, D_{3,2} x \alpha_{m(J)}^B \right\rangle.$$

---

[4] If we were interested in making the reduction sequence shorter, we could have omitted $D_{3,2} x \alpha_j^A$ and $D_{3,2} x \alpha_j^B$ and express them using $D_{3,0} x \alpha_j^A$, $D_{3,1} x \alpha_j^A$, $D_{3,0} x \alpha_j^B$ and $D_{3,1} x \alpha_j^B$, in a similar way to our treatment of $D_{2,0}$ and $D_{2,1}$ in the proof above.

6. 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 $(v_{start}^1, v_{start}^2)$ from G^1 and G^2 respectively such that v_{start}^i 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 $G^i = (V_1, E_1, v_{start}^i, v_{end}^i)$ 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 transduction of the two input graphs G^1 and G^2:

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

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

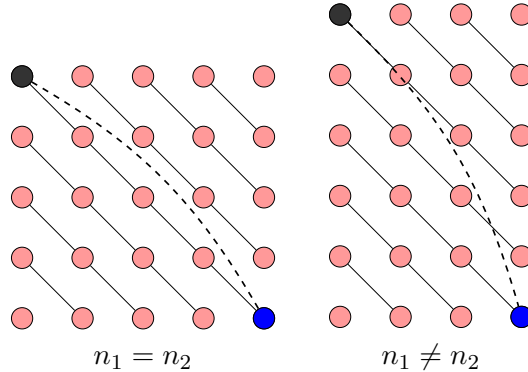The result of the application of the transduction is given in Figure 1. The result of the

Figure 1: The result of applying Φ_F on the two directed paths of length n_1 and n_2. There is a cycle iff the two paths are of the same length. The black vertex corresponds to the two initial vertices of the two directed paths.

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

Corollary 6.2. *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 Φ_T, obtained from Φ'_T below by additionally applying Φ_{sym}, transforms the two directed paths into the structures in Figure 2.

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

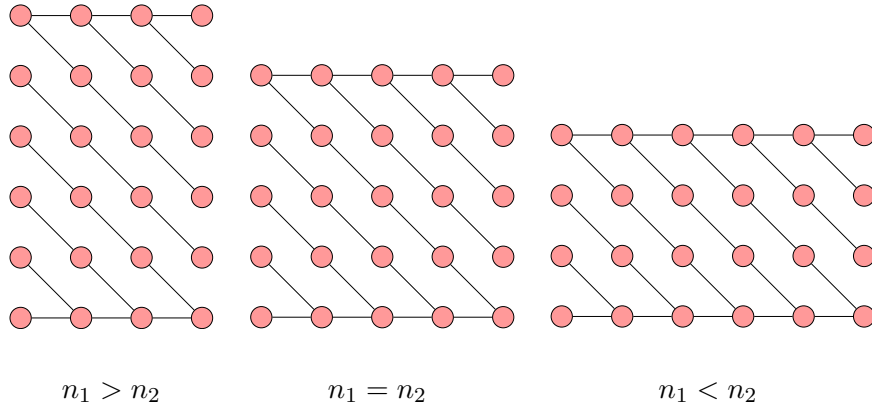$$n_1 > n_2 \qquad\qquad n_1 = n_2 \qquad\qquad n_1 < n_2$$

Figure 2: The result of applying Φ_T 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 $n_1 \le n_2$. 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 6.3. *The properties of being a tree or a connected graph are not* CFOL-*definable even in the presence of linear order.*

The case of connected graphs was also shown non-definable in CFOL by. J. Nurmonen in [50] using his version of the pebble games for CFOL.

For our next connection matrix we use the 2-sum of the following two 2-graphs. Recall that the 2-sum $G \sqcup_2 H$ of two 2-graphs G and H is obtained by taking the disjoint union of G and H and identifying the corresponding labels. In this case the two labels are the constant symbols $start$ and end.

(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 $P_{n_1}^1$ and $P_{n_2}^2$:

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 6.4. *The class of planar graphs is not* CFOL-*definable on ordered graphs.*
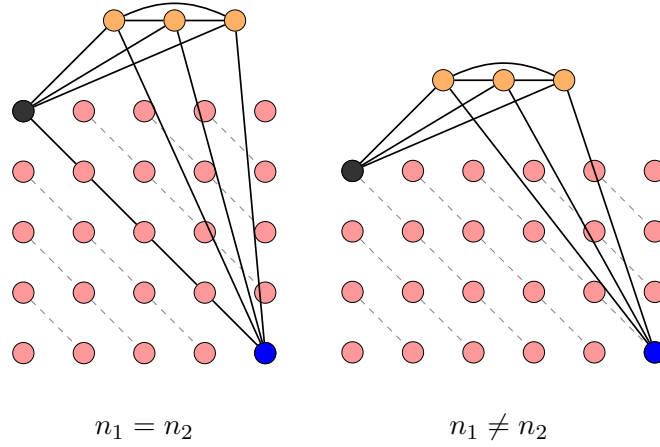
$$n_1 = n_2 \qquad\qquad\qquad n_1 \neq n_2$$

Figure 3: The result of Φ_P on two directed paths. The graph obtained here is planar iff the two directed paths are not of equal length.

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 6.5. *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.*

For the next transduction we need to refer to the second and second to last vertices of G^1, both of which are definable using E_1 and $start_1$ or end_1. The transduction Φ_B adds to the Cartesian product of G^1 and G^2 an edge between the second vertex in the first column and the second to last vertex in the last column. Moreover, Φ_B adds all the edges in the first and last rows and columns, except for two edges, the first edge in the first column and the last edge in the last column.

The transduction Φ_B, obtained from Φ'_B below by additionally applying Φ_{sym}, transforms the two directed paths into the structures in Figure 4.

$$
\begin{aligned}
\Phi'_B\left((v_1, v_2), (u_1, u_2)\right) \;=\; & (E_1(start_1, v_1) \wedge E_1(u_1, end_1) \wedge (start_2 = v_2) \wedge (end_2 = u_2)) \;\vee \\
& (E_1(v_1, u_1) \wedge E_2(v_2, u_2)) \;\vee \\
& (v_1 = u_1 = start_1 \wedge E(v_2, u_2)) \;\vee \\
& (v_2 = u_2 = start_2 \wedge E(v_1, u_1) \wedge (v_1 \neq start_1)) \;\vee \\
& (v_1 = u_1 = end_1 \wedge E(v_2, u_2)) \;\vee \\
& (v_2 = u_2 = end_2 \wedge E(v_1, u_1) \wedge (u_1 \neq end_1)),
\end{aligned}
$$

The result of the transduction has a bridge iff $n_1 \geq n_2 + 1$. It is 2-connected iff $n_1 \leq n_2$. Hence, the connection matrices of bridgelessness and of 2-connectivity with directed paths as row and column labels have infinite rank.

Theorem 6.6. *The properties of being bridgeless or 2-connected are not* CFOL-*definable even in the presence of linear order.*

For any $\ell > 0$, if we join K_ℓ to the result of Φ_B we get non-definability of $(\ell + 2)$-connectivity, so we have:

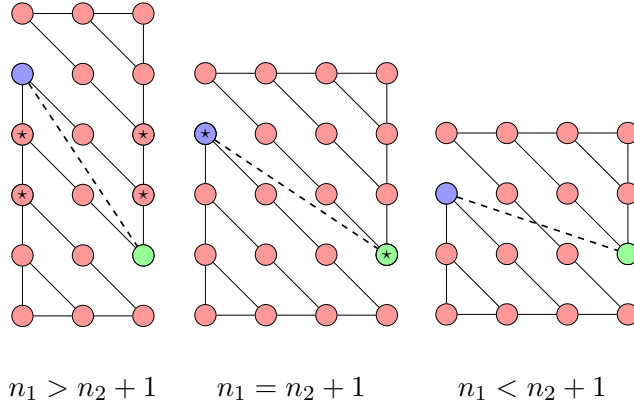$$n_1 > n_2 + 1 \qquad n_1 = n_2 + 1 \qquad n_1 < n_2 + 1$$

Figure 4: The result of applying Φ_B on two directed paths. We always get a connected graph as a result of this transduction. The resulting graph has articulation vertices (vertices whose removal leaves the graph disconnected) iff $n_1 \geq n_2 + 1$. The same is true for the existence of bridges (edges whose removal leaves the graph disconnected). We marked the articulation vertices with $\star$. The bridges are the edges between adjacent vertices marked with $\star$. Note that if $n_1 = n_2 + 1$ then there are two articulation vertices and one bridge, whereas if $n_1 > n_2 + 1$ there are four articulation vertices and two bridges.

Corollary 6.7. *For every $\ell \geq 0$, the property of being $(\ell+2)$-connected is not* CFOL-*definable even in the presence of linear order.*

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 6.8. *None of the properties above are* CMSOL-*definable as graphs even in the presence of an order.*

Using a modification $\bowtie$ of the join operation from [17, Remark 5.21, Page 350] one can show the same for the class of graphs which have a spanning tree of degree at most 3. Let G and H be graphs. Let $u_1, \ldots, u_m$ be the vertices of H. $G \bowtie H$ is obtained from $G \bowtie H$ by taking two additional copies H_1 and H_2 of H with vertices of the form u_i^b, $b = 1$ or $b = 2$ respectively, and adding, for every $u_i \in V(H)$, the edges (u_i, u_i^1) and (u_i, u_i^2). If $G = \overline{K_n}$ and $H = \overline{K_m}$ are edgeless graphs with n respectively m vertices, then $G \bowtie H$ has a spanning tree of degree at most 3 iff $|V(G)| + 2 \geq |V(H)|$. The operation $\bowtie$ is MSOL-smooth over the vocabulary of graphs since it can be defined as a transduction of the rich disjoint union of G and H. The connection matrix of $G \bowtie H$ and the property of having a spanning tree of

degree at most 3 has infinite rank, implying that the property is not MSOL-definable in the vocabulary of graphs.

For any fixed natural number $d > 3$, by performing a transduction on $G \bowtie H$ which attaches $d - 3$ new vertices as pendants to each vertex of $G \bowtie H$, the non-definability result extends 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 [24] 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 6.9. *None of the properties above are* CMSOL-*definable as hypergraphs even in the presence of an order.*

Remark 6.10. 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.

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

$\mathcal{L}(\tau)$-**polynomials.** Here we follow closely the exposition from [38]. 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 ($\mathcal{L}$-simple monomials)**.** *Let $\mathcal{M}$ be a τ-structure. Elements of $\mathbb{N} \cup \mathbf{X}$ are $\mathcal{L}(\tau)$-definable simple $\mathcal{M}$-monomials.*

Definition 2 ($\mathcal{L}$-monomials)**.** *Let $\mathcal{M}$ be a τ-structure. We define the $\mathcal{L}(\tau)$-definable $\mathcal{M}$-monomials inductively.*

   (i) *$\mathcal{L}(\tau)$-definable simple monomials are $\mathcal{L}(\tau)$-definable monomials.*

(ii) *Let $\phi(a)$ be a $\tau \cup \{a\}$-formula in $\mathcal{L}$, where a is a constant symbol not in τ and $U \in \tau$ is a unary relation symbol. Let t be a simple monomial. Then*

$$\prod_{a:\langle \mathcal{M}, a \rangle \models \phi(a)} t$$

is a $\mathcal{L}(\tau)$-definable $\mathcal{M}$-monomial.

(iii) *Finite products of $\mathcal{L}(\tau)$-definable monomials and simple monomials are $\mathcal{L}(\tau)$-definable monomials.*

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

Definition 3 ($\mathcal{L}$-polynomials). *The $\mathcal{M}$-polynomials definable in $\mathcal{L}(\tau)$ are defined inductively:*

(i) *$\mathcal{L}(\tau)$-definable monomials are $\mathcal{L}(\tau)$-definable polynomials.*
(ii) *Let $\phi(\bar{a})$ be a $\tau \cup \{\bar{a}\}$-formula in $\mathcal{L}$ where $\bar{a} = (a_1, \ldots, a_m)$ be a finite sequence of constant symbols not in τ. Let t be a $\mathcal{L}(\tau \cup \{\bar{a}\})$-polynomial. Then*

$$\sum_{\bar{a}:\langle \mathcal{M}, \bar{a} \rangle \models \phi(\bar{a})} t$$

is a $\mathcal{L}(\tau)$-definable polynomial.

(iii) *Let $\phi(\overline{R})$ be a $\tau \cup \{\overline{R}\}$-formula in $\mathcal{L}$ where $\overline{R} = (R_1, \ldots, R_m)$ be a finite sequence of relation symbols not in τ. Let t be a $\mathcal{L}(\tau \cup \{\overline{R}\})$-polynomial definable in $\mathcal{L}$. Then*

$$\sum_{\overline{R}:\langle \mathcal{M}, \overline{R} \rangle \models \phi(\overline{R})} t$$

is a $\mathcal{L}(\tau)$-definable 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.

The quantifier rank of a $\mathcal{L}(\tau)$-polynomial p is the maximal quantifier rank of the formulas defining p.

Among the $\mathcal{L}$-definable polynomials we find most of the known graph polynomials from the literature, cf. [45, 38].

Example 7.1 (Matching polynomial). The matching generating polynomial of a graph G is the generating function of matchings in G:

$$m(G) = \sum_{i=0}^{|E(G)|} m_i \mathbf{X}^i$$

where m_i is the number of matchings of size i of G. $m(G)$ is an MSOL-polynomial given by

$$\sum_{M \subseteq E(G):\langle G, M \rangle \models \varphi_{match}} \prod_{v:\langle G, M, v \rangle \models T} \mathbf{X}$$

where T is any tautology, $\varphi_{sym}(R) = \forall x \forall y (R(x,y) \rightarrow R(y,x))$, and $\varphi_{match}(R) = \varphi_{sym} \wedge \forall x \forall y \forall z \neg (R(x,y) \wedge R(x,z))$. The monomial $t = \prod_{v:\langle G, M, v \rangle \models T} \mathbf{X}$ depends on M. We simplify

the notation by writing

$$m(G, \mathbf{X}) = \sum_{M \subseteq E(G): \varphi_{match}(M)} \mathbf{X}^{|M|}. \tag{7.1}$$

Definition 4 ($\mathcal{L}$-parameters and $\mathcal{L}$-properties). $\mathcal{L}$-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.*

Remark 7.2 (FOL- and CFOL-parameters). For FOL- and CFOL-parameters, the formulas must not contain any second order variables. Therefore, sums of the form $\sum_{\overline{R}: \langle \mathcal{M}, \overline{R} \rangle \models \phi(\overline{R})} t$ are not allowed, while sums of the form $\sum_{\overline{a}: \langle \mathcal{M}, \overline{a} \rangle \models \phi(\overline{a})} t$ are allowed. Moreover, the monomials are required to be simple monomials.

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

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

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).

Remark 7.3 (CFOL-polynomials). In Remark 7.2 we required that CFOL-parameters only have simple monomials. We did so because the Feferman-Vaught theorem and the Finite Rank theorem for $\times$ do not hold when allowing general monomials. Let CFOL-polynomials be the extension of CFOL-parameters by allowing monomials which are not simple.

For example, consider the following CFOL-polynomial, which is not a CFOL-parameter:

$$\prod_{a \in M} \mathbf{X} = \mathbf{X}^{|M|}$$

The connection matrix $M(\times, \mathbf{X}^{|M|})$ restricted to E_n, the edgeless graphs, has entries $m_{i,j} = \mathbf{X}^{ij}$, is a Vandermonde matrix and therefore has infinite rank.

The graph polynomial $\mathbf{X}^{|M|}$ does not satisfy a bilinear reduction theorem such as Theorem 8.7. Assume there is a polynomial Q, and CFOL-polynomials $p_1, \ldots, p_t$ such that

$$p(G_1 \times G_2) = Q(p_1(G_1), \ldots, p_t(G_1), p_1(G_2), \ldots, p_t(G_2))$$

Let $G_1 = G_2 = E_n$ where E_n is the edgeless graph with n vertices. By definition of CFOL-polynomials, $p_i(E_n)$ has degree at most n in $\mathbf{X}$. Therefore, $Q(p_1(G_1), \ldots, p_t(G_1), p_1(G_2), \ldots, p_t(G_2))$ has degree at most cn in $\mathbf{X}$ for some $c \in \mathbb{N}$. In contrast, $p(E_n \times E_n) = \mathbf{X}^{n^2}$. Therefore, it cannot be the case that there is such Q.

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 [18, 44].

Theorem 7.4 (The Finite Rank Theorem for $\mathcal{L}$-polynomials)**.**
Let $\mathcal{L}$ be a tame fragment of SOL *such that $\sqcup_{rich}$ is $\mathcal{L}$-smooth and $\square$ be an $\mathcal{L}$-sum-like operation between τ-structures. Let P be a $\mathcal{L}(\tau)$-polynomial. Then the connection matrix $M(\square, P)$ has finite rank.*

The proof is given in Section 8 In [27] the theorem was only formulated for k-sums, and the join operation and for the logic CMSOL.

Theorem 7.5 (The Finite Rank Theorem for CFOL-parameters)**.**
Let $\boxtimes$ be an CFOL*-product-like operation between τ-structures. Let P be an* CFOL(τ)*-parameter. Then the connection matrix $M(\boxtimes, P)$ has finite rank.*

The proof of Theorem 7.5 is similar to that of Theorem 7.4.

8. A FEFERMAN-VAUGHT-TYPE THEOREM FOR $\mathcal{L}$-POLYNOMIALS AND SMOOTH OPERATIONS

In this section we prove an analogue of the Feferman-Vaught theorem for sum-like operations, Theorem 7.4. We start by explicitly giving a Feferman-Vaught theorem for the matching polynomial, proceed with the case of rich disjoint union, and end with all sum-like operations.

8.1. **The matching polynomial.** As an example, we prove here a Feferman-Vaught theorem for the matching generating polynomial with the 1-sum operation. $m(G, \mathbf{X})$ is a polynomial in $\mathbb{Z}[\mathbf{X}]$. As we need a field we work in the field of rational functions $\mathbb{Q}(\mathbf{X})$.

We will need the following auxiliary graph polynomials whose inputs are 1-labeled graphs (G, v_G):

$$m^+(G, v_G, \mathbf{X}) \quad = \sum_{M \subseteq E_G : \varphi_{match}(M) \wedge \varphi_+(M, v_G)} \mathbf{X}^{|M|} \tag{8.1}$$

$$m^-(G, v_G, \mathbf{X}) \quad = \sum_{M \subseteq E_G : \varphi_{match}(M) \wedge \neg \varphi_+(M, v_G)} \mathbf{X}^{|M|} \tag{8.2}$$

where φ_+ says that v_G is incident to an edge of M. The definition of $m(G, \mathbf{X})$ in Equation (7.1) extends to 1-labeled graphs (G, v_G) naturally by ignoring the label on v_G. For a 1-labeled graph (G, v_G), $m(G, \mathbf{X}) = m(G, v_G, \mathbf{X})$ and

$$m(G, v_G, \mathbf{X}) = m^+(G, v_G, \mathbf{X}) + m^-(G, v_G, \mathbf{X}). \tag{8.3}$$

Theorem 8.1. *Let $\mathfrak{m}(G, v_G; \mathbf{X}) = (m^+(G, v_G, \mathbf{X}), m^-(G, v_G, \mathbf{X}), m(G, v_G, \mathbf{X}))$.*

(i) *There exists a polynomial $Q \in \mathbb{Z}[\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2]$ such that for any two 1-labeled graphs (G, v_G) and (H, v_H),*

$$m((G, v_G) \sqcup_1 (H, v_H), \mathbf{X}) \quad = \quad Q\Big(m^+(G, v_G, \mathbf{X}), m^-(G, v_G, \mathbf{X}),$$
$$m^+(H, v_H, \mathbf{X}), m^-(H, v_H, \mathbf{X})\Big). \tag{8.4}$$

(ii) *There exist a matrix $A \in \mathbb{Q}(\mathbf{X})^{(3 \times 3)}$ such that Q can be written as the bilinear form*

$$m((G, v_G) \sqcup_1 (H, v_H), \mathbf{X}) = \mathfrak{m}(G, v_G, \mathbf{X}) \cdot A \cdot \mathfrak{m}(H, v_H, \mathbf{X})^{tr}. \tag{8.5}$$

Proof. (i): A matching M of $(G, v_G) \sqcup_1 (H, v_H)$ is a disjoint union of two matchings M_G and M_H of G and H respectively such that at most one of M_G and M_H is incident to v_G respectively v_H. $m^+((G, v_G) \sqcup_1 (H, v_H))$ is the generating function of matchings $M_G \sqcup M_H$ such that one of M_G and M_H is incident to v_G respectively V_H. Analogously, $m^-((G, v_G) \sqcup_1 (H, v_H))$ is the generating function of matchings $M_G \sqcup M_H$ such that both M_G and M_H are not incident to v_G respectively v_H. So we have:

$$
\begin{aligned}
m((G, v_G) \sqcup_1 (H, v_H), \mathbf{X}) \;=\; & m^+(G, v_G, \mathbf{X}) \cdot m^-(H, v_H, \mathbf{X}) + \\
& m^-(G, v_G, \mathbf{X}) \cdot m^+(H, v_H, \mathbf{X}) + \\
& m^-(G, v_G, \mathbf{X}) \cdot m^-(H, v_H, \mathbf{X}) \, .
\end{aligned}
$$

The theorem holds with the polynomial

$$
Q = \mathbf{X}_1 \mathbf{Y}_2 + \mathbf{X}_2 \mathbf{Y}_1 + \mathbf{X}_2 \mathbf{Y}_2 \, . \tag{8.6}
$$

(ii): The matrix A can be taken to be

$$
A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{8.7}
$$

$\square$

Corollary 8.2. *The connection matrix $M(\sqcup_1, m)$ has rank 2.*

Proof. The connection matrix $M(\sqcup_1,)$ is spanned by two rows which correspond to m^+ and m^-. Hence, the rank of $M(\sqcup_1, m)$ is a t most 2. $\square$

Remark 8.3. The matrix A turns out to be a $(0,1)$-matrix due to a good choice of auxiliary graph polynomials. If we replace $\mathfrak{m}(G, v_G, \mathbf{X})$ by

$$
\mathfrak{m}_1(G, v_G, \mathbf{X}) = \mathfrak{m}(G, v_G, \mathbf{X}) \cdot B, \tag{8.8}
$$

where B is any (3×3)-matrix over $\mathbb{Q}(\mathbf{X})$, we get

$$
m((G, v_G) \sqcup_1 (H, v_H), \mathbf{X}) = \mathfrak{m}_1(G, v_G, \mathbf{X}) \cdot A_1 \cdot \mathfrak{m}_1(H, v_H, \mathbf{X})^{tr} \tag{8.9}
$$

with $A_1 = B \cdot A \cdot B^{tr}$ where A_1 can have arbitrary entries from $\mathbb{Q}(\mathbf{X})$.

8.2. **Rich disjoint union.** Here we prove a Feferman-Vaught-type theorem for rich disjoint union and structures of some vocabulary τ. We do this for $\mathcal{L}(\tau)$-polynomials of the form

$$
p(\mathfrak{M}) = \sum_{U \subseteq M : \Omega(U)} \left(\prod_{c \in M : \phi_{\mathbf{X}}(c, U)} \mathbf{X} \right) \left(\prod_{c \in M : \phi_{\mathbf{Y}}(c, U)} \mathbf{Y} \right) \tag{8.10}
$$

where $\Omega, \phi_{\mathbf{X}}, \phi_{\mathbf{Y}}$ are $\mathcal{L}$-sentences over the appropriate expansion of τ. Generalization to all $\mathcal{L}(\tau)$-polynomials is not hard using the normal form lemma for $\mathcal{L}$-polynomials, Lemma 8.4:

Lemma 8.4 (Normal Form Lemma, [38]). *Let $\mathcal{L}$ be a fragment of* SOL. *Let p be a $\mathcal{L}(\tau)$-polynomial. Then there exist $s, t \in \mathbb{N}$, $\mathcal{L}$-formulas Ω, and $\phi_1, \ldots, \phi_t$, and $X_1, \ldots, X_t \in \mathbf{X} \cup \mathbb{N}$ such that*

$$
p(\mathfrak{M}) = \sum_{U_1, \ldots, U_s, a_1, \ldots, a_r : \Omega(\overline{U}, \overline{a})} \left(\prod_{c \in M : \phi_1(c, \overline{U}, \overline{a})} X_1 \right) \cdots \left(\prod_{c \in M : \phi_t(c, \overline{U}, \overline{a})} X_t \right) \tag{8.11}
$$

where $\overline{U} = U_1, \ldots, U_s$ and $\overline{a} = a_1, \ldots, a_r$.

The monomials of $\mathcal{L}$-polynomials are already in the desired form. To prove Lemma 8.4, the main property of $\mathcal{L}$ that we use here is closure under conjunction which allows us to eliminate nested sums.

The following is well known [22]:

Proposition 8.5 (Hintikka sentences). *Let $\mathcal{L}$ be a fragment of* SOL. *Let τ be a vocabulary. For every $q \in \mathbb{N}$ there is a finite set $\Theta_{\tau,q}(\tau)$ of $\mathcal{L}(\tau)$-sentences of quantifier rank q such that:*

  (i) *every $\theta \in \Theta_{\tau,q}(\tau)$ has a model;*
 (ii) *the conjunction of any two sentences $\theta_1, \theta_2 \in \Theta_{\tau,q}(\tau)$ is not satisfiable;*
(iii) *every $\mathcal{L}(\tau)$ sentence ψ of quantifier rank at most q is equivalent to exactly one finite disjunction of sentences in $\Theta_{\tau,q}(\tau)$;*
 (iv) *every finite τ-structure $\mathfrak{A}$ satisfies exactly one sentence $\theta_{\tau,q}(\mathfrak{A})$ of $\Theta_{\tau,q}(\tau)$.*

Proposition 8.5 uses that $\mathcal{L}$ is fragment and therefore $\mathcal{L}_q(\tau)$ is finite up to equivalence and is closed under Boolean operations.

The following proposition reformulating $\mathcal{L}$-smoothness follows directly from the definition of $\mathcal{L}$-smoothness in Section 3 and Proposition 8.5:

Proposition 8.6. *Let $\mathcal{L}$ be a fragment of* SOL *such that $\sqcup_{rich}$ is $\mathcal{L}$-smooth. Let τ_1 and τ_2 be a vocabularies and let σ be the vocabulary of the rich disjoint union of a τ_1-structure and τ_2-structure. There exists a function $g_{\tau_1,\tau_2} : \Theta_{\tau_1,q} \times \Theta_{\tau_2,q} \to \Theta_{\sigma,q}$ such that for every two τ_1-structures $\mathfrak{A}$ and $\mathfrak{B}$,*

$$\theta_{\sigma,q}(\mathfrak{A} \sqcup_{rich} \mathfrak{B}) = g_{\tau_1,\tau_2}(\theta_{\tau,q}(\mathfrak{A}), \theta_{\tau_2}(\mathfrak{B})).$$

We write g_τ rather than g_{τ_1,τ_2} when $\tau = \tau_1 = \tau_2$.

In the proof of Theorem 8.7 we will use σ with various subscripts to denote vocabularies which correspond to structures obtained as rich disjoint unions, and τ with various subscripts to denote vocabularies of pre-union structures).

Theorem 8.7 (Bilinear Reduction Theorem). *Let $\mathcal{L}$ be a tame fragment of* SOL *such that $\sqcup_{rich}$ is $\mathcal{L}$-smooth. Let τ be a vocabulary and let σ be the vocabulary of the rich disjoint union of τ-structures. Let p be a $\mathcal{L}(\sigma)$-polynomial in the form of Equation (8.10). There exist $t \in \mathbb{N}$, $\mathcal{L}(\tau)$-polynomials $p_1, \ldots, p_{2t}$ and a polynomial $Q \in \mathbb{Z}[w_1, \ldots, w_t, v_1, \ldots, v_t]$ such that*

  (i) *for any two τ-structures $\mathfrak{A}$ and $\mathfrak{B}$*

$$p(\mathfrak{A} \sqcup_{rich} \mathfrak{B}) = Q(p_1(\mathfrak{A}), \ldots, p_t(\mathfrak{A}), p_{t+1}(\mathfrak{B}), \ldots, p_{2t}(\mathfrak{B})).$$

 (ii) *$Q(\overline{w}, \overline{v})$ can be written as a bilinear form*

$$\overline{w}^{tr} M \overline{v}$$

  where the matrix M has as entries polynomials with coefficients from $\mathbb{Z}$ and is independent of the τ-structures.

Remark 8.8. In the following proof, the graph polynomials p_i as well as p all have the same quantifier rank $qr(p)$ and hence t can be taken as the number $\beta(qr(p))$ of $\mathcal{L}(\sigma)$-polynomials of quantifier rank at most $qr(p)$, which is finite.

Proof. Let q be the maximum quantifier rank of Ω, $\phi_{\mathbf{X}}$ and $\phi_{\mathbf{Y}}$.

By definition of p, $p(\mathfrak{A} \sqcup_{rich} \mathfrak{B})$ is given by

$$\sum_{U \subseteq A \sqcup B : \Omega(U)} \left(\prod_{c \in A \sqcup B : \phi_{\mathbf{X}}(c,U)} \mathbf{X} \right) \left(\prod_{c \in A \sqcup B : \phi_{\mathbf{Y}}(c,U)} \mathbf{Y} \right) = \tag{8.12}$$

$$\sum_{U \subseteq A \sqcup B : \Omega(U)} \left(\prod_{c \in A : \phi_{\mathbf{X}}(c,U)} \mathbf{X} \right) \left(\prod_{c \in B : \phi_{\mathbf{X}}(c,U)} \mathbf{X} \right) \left(\prod_{c \in A : \phi_{\mathbf{Y}}(c,U)} \mathbf{Y} \right) \left(\prod_{c \in B : \phi_{\mathbf{Y}}(c,U)} \mathbf{Y} \right)$$

Consider the summation in Equation (8.12). It is a sum over all $U \subseteq A \sqcup B$ such that $\langle \mathfrak{A} \sqcup_{rich} \mathfrak{B}, U \rangle \models \Omega$. There is a unique partition $(U \cap A) \sqcup (U \cap B) = U$. We have that $\langle \mathfrak{A} \sqcup_{rich} \mathfrak{B}, U \rangle \models \Omega$ iff $\langle \mathfrak{A}, U \cap A \rangle \sqcup_{rich} \langle \mathfrak{B}, U \cap B \rangle \models \Omega$ by the definition of rich disjoint union.

Let τ_U be the vocabulary of $\langle \mathfrak{A}, U \cap A \rangle$ and $\langle \mathfrak{B}, U \cap B \rangle$, and let σ_U be the vocabulary of $\langle \mathfrak{A}, U \cap A \rangle \sqcup_{rich} \langle \mathfrak{B}, U \cap B \rangle$. By Proposition 8.6, there exists g_{τ_U} such that

$$\theta_{\sigma_U,q}(\langle \mathfrak{A}, U \cap A \rangle \sqcup_{rich} \langle \mathfrak{B}, U \cap B \rangle) = g_{\tau_U}(\theta_{\tau_U,q}(\langle \mathfrak{A}, U \cap A \rangle), \theta_{\tau_U,q}(\langle \mathfrak{B}, U \cap B \rangle))$$

Hence we can write: [5]

$$p(\mathfrak{A} \sqcup_{rich} \mathfrak{B}) = \sum_{\substack{(\theta_1,\theta_2) \in \Theta_{\tau_U,q} \times \Theta_{\tau_U,q} : \\ g_{\tau_U}(\theta_1,\theta_2) \models \Omega}} \sum_{\substack{U = U_1 \cup U_2 \\ U_1 \subseteq A : \langle \mathfrak{A}, U_1 \rangle \models \theta_1 \\ U_2 \subseteq B : \langle \mathfrak{B}, U_2 \rangle \models \theta_2}} \left(\prod_{c \in A : \phi_{\mathbf{X}}(c,U)} \mathbf{X} \right)$$
$$\left(\prod_{c \in B : \phi_{\mathbf{X}}(c,U)} \mathbf{X} \right) \left(\prod_{c \in A : \phi_{\mathbf{Y}}(c,U)} \mathbf{Y} \right) \left(\prod_{c \in B : \phi_{\mathbf{Y}}(c,U)} \mathbf{Y} \right) \tag{8.13}$$

Now consider e.g. the product $\prod_{c \in A : \phi_{\mathbf{X}}(c,U)} \mathbf{X}$. It is a product over all $c \in A$ such that

$$\langle \mathfrak{A} \sqcup_{rich} \mathfrak{B}, c, U \rangle \models \phi_{\mathbf{X}} .$$

Again this condition can be written as a condition on a rich disjoint union of two structures: $\langle \mathfrak{A} \sqcup_{rich} \mathfrak{B}, c, U \rangle \models \phi_{\mathbf{X}}$ iff $\langle \mathfrak{A}, U_1, c \rangle \sqcup_{rich} \langle \mathfrak{B}, U_2 \rangle \models \phi_{\mathbf{X}}$.

Let τ_{Uc} and σ_{Uc} be the vocabularies of $\langle \mathfrak{A}, U_1, c \rangle$ and $\langle \mathfrak{A}, U_1, c \rangle \sqcup_{rich} \langle \mathfrak{B}, U_2 \rangle$ respectively. The vocabulary of $\langle \mathfrak{B}, U_2 \rangle$ is τ_U.

Using Proposition 8.6 there exists a function g_{τ_{Uc},τ_U} such that

$$\theta_{\sigma_{Uc},q}(\langle \mathfrak{A}, U_1, c \rangle \sqcup_{rich} \langle \mathfrak{B}, U_2 \rangle) = g_{\tau_{Uc},\tau_U}(\theta_{\tau_{Uc},q}(\langle \mathfrak{A}, U_1, c \rangle), \theta_{\tau_U,q}(\langle \mathfrak{B}, U_2 \rangle)) .$$

For every $\alpha \in \mathcal{L}_q(\tau_U)$, let

$$\psi_{\mathbf{X},1,\alpha} = \bigvee_{\substack{\theta \in \Theta_{\tau_{Uc},q} : \\ g_{\tau_{Uc},\tau_U}(\theta,\alpha) \models \phi_{\mathbf{X}}}} \theta$$

$\psi_{\mathbf{X},1,\alpha}$ is an $\mathcal{L}_q$-formula.

---

[5]The notation $\models$ in $g_{\tau_U}(\theta_1,\theta_2) \models \Omega$ in the following formula denotes that $g_{\tau_U}(\theta_1,\theta_2)$ entails Ω. (Note that the notation $\models$ is usually used in this paper in expressions such as $\mathfrak{A} \models \phi$, where $\models$ denotes that the structure $\mathfrak{A}$ satisfies ϕ.)

If $\langle \mathfrak{B}, U_2 \rangle \models \alpha$, then

$$\prod_{\substack{c \in A: \\ \langle \mathfrak{A} \sqcup_{rich} \mathfrak{B}, c, U \rangle \models \phi_{\mathbf{X}}}} \mathbf{X} = \prod_{\substack{c \in A: \\ \langle \mathfrak{A}, U_1, c \rangle \models \psi_{\mathbf{X},1,\alpha}}} \mathbf{X}$$

Similarly, there exist $\mathcal{L}_q$-formulas $\psi_{\mathbf{Y},1,\alpha}$, $\psi_{\mathbf{X},2,\alpha}$ and $\psi_{\mathbf{Y},2,\alpha}$ such that if $\langle \mathfrak{B}, U_2 \rangle \models \alpha$, then

$$\prod_{\substack{c \in A: \\ \langle \mathfrak{A} \sqcup_{rich} \mathfrak{B}, c, U \rangle \models \phi_{\mathbf{Y}}}} \mathbf{Y} = \prod_{\substack{c \in A: \\ \langle \mathfrak{A}, U_1, c \rangle \models \psi_{\mathbf{Y},1,\alpha}}} \mathbf{Y}$$

and if $\langle \mathfrak{A}, U_1 \rangle \models \alpha$, then

$$\prod_{\substack{c \in B: \\ \langle \mathfrak{A} \sqcup_{rich} \mathfrak{B}, c, U \rangle \models \phi_{\mathbf{X}}}} \mathbf{X} = \prod_{\substack{c \in B: \\ \langle \mathfrak{B}, U_2, c \rangle \models \psi_{\mathbf{X},2,\alpha}}} \mathbf{X}$$

$$\prod_{\substack{c \in B: \\ \langle \mathfrak{A} \sqcup_{rich} \mathfrak{B}, c, U \rangle \models \phi_{\mathbf{Y}}}} \mathbf{Y} = \prod_{\substack{c \in B: \\ \langle \mathfrak{B}, U_2, c \rangle \models \psi_{\mathbf{Y},2,\alpha}}} \mathbf{Y}.$$

For every two $\mathcal{L}(\tau_{Uc})$-sentence $\varphi_{\mathbf{X}}$ and $\varphi_{\mathbf{Y}}$ and every Hintikka sentence $\varphi \in \Theta_{\tau_U,q}$, let $\widetilde{p}_{\varphi_{\mathbf{X}},\varphi_{\mathbf{Y}},\varphi}$ be the following $\mathcal{L}(\tau)$-polynomial:

$$\widetilde{p}_{\varphi_{\mathbf{X}},\varphi_{\mathbf{Y}},\varphi}(\mathfrak{M}) = \sum_{U_{\mathfrak{M}} \subseteq M: \langle \mathfrak{A}, U_{\mathfrak{M}} \rangle \models \varphi} \prod_{\substack{c \in M: \\ \langle \mathfrak{M}, U_{\mathfrak{M}}, c \rangle \models \varphi_{\mathbf{X}}}} \mathbf{X} \prod_{\substack{c \in M: \\ \langle \mathfrak{M}, U_{\mathfrak{M}}, c \rangle \models \varphi_{\mathbf{Y}}}} \mathbf{Y}$$

Note that $\widetilde{p}_{\varphi_{\mathbf{X}},\varphi_{\mathbf{Y}},\varphi}$ is a $\mathcal{L}$-polynomial over the vocabulary τ with quantifier rank q. Then using Eq. (8.13) we have

$$p(\mathfrak{A} \sqcup_{rich} \mathfrak{B}) = \sum_{\substack{(\theta_1,\theta_2) \in \Theta_{\tau_U,q} \times \Theta_{\tau_U,q}: \\ g_{\tau_U}(\theta_1,\theta_2) \models \Omega}} \widetilde{p}_{\psi_{\mathbf{X},1,\theta_2},\psi_{\mathbf{Y},1,\theta_2},\theta_1}(\mathfrak{A}) \cdot \widetilde{p}_{\psi_{\mathbf{X},2,\theta_1},\psi_{\mathbf{Y},2,\theta_1},\theta_2}(\mathfrak{B}).$$

Let $t = |\Theta_{\tau_U,q} \times \Theta_{\tau_U,q}|$ and let $\pi : \{1,\ldots,t\} \to \Theta_{\tau_U,q} \times \Theta_{\tau_U,q}$ be a bijection. For each $i = 1,\ldots,t$, denote $\pi(i) = (\pi_1(i), \pi_2(i))$ and let $p_i = \widetilde{p}_{\psi_{\mathbf{X},1,\pi_2(i)},\psi_{\mathbf{Y},1,\pi_2(i)},\pi_1(i)}$ and $p_{i+t} = \widetilde{p}_{\psi_{\mathbf{X},2,\pi_1(i)},\psi_{\mathbf{Y},2,\pi_1},\pi_2(i)}$. We set

$$Q(w_1,\ldots,w_t,v_1,\ldots,v_t) = \sum_{\substack{(\theta_1,\theta_2) \in \Theta_{\tau_U,q} \times \Theta_{\tau_U,q}: \\ g_{\tau_U}(\theta_1,\theta_2) \models \Omega}} w_{\pi^{-1}(\theta_1,\theta_2)} \cdot v_{\pi^{-1}(\theta_2,\theta_1)}$$

and the theorem follows. The matrix $M = (m_{ij})$ is given explicitly as

$$m_{ij} = \begin{cases} 1, & \pi_1(i) = \pi_2(j), \ \pi_2(i) = \pi_1(j), \text{ and } g_{\tau_U}(\pi(i)) \models \Omega \\ 0, & \text{otherwise.} \end{cases} \qquad \square$$

Remark 8.9. If the statement of Theorem 8.7 is changed such that instead of rich disjoint union we take simple disjoint union, and correspondingly, p has vocabulary τ, then the theorem holds as well. Additionally, there are polynomials $Q_1,\ldots,Q_{2t} \in \mathbb{Z}[w_1,\ldots,w_t,v_1,\ldots,v_t]$ such that

$$p_i(\mathfrak{A} \sqcup \mathfrak{B}) = Q_i(p_1(\mathfrak{A}),\ldots,p_t(\mathfrak{A}),p_{t+1}(\mathfrak{B}),\ldots,p_{2t}(\mathfrak{B}))$$

and Q_i can be written in bilinear form. Here we use the fact that the number of $\mathcal{L}(\tau)$-polynomials with a fixed quantifier rank bound on their formulas and any fixed set of

indeterminates is finite. This follows from the fact that the same is true for the number of formulas of any fixed vocabulary and quantifier rank.

As a consequence of Theorem 8.7 we have:

Theorem 8.10. *Let $\mathcal{L}$ be a tame fragment of* SOL *such that $\sqcup_{rich}$ is $\mathcal{L}$-smooth. Let p be a $\mathcal{L}(\sigma)$-polynomial in the form of Equation (8.10). Then the connection matrix $M(\sqcup_{rich}, p)$ has finite rank over the field of rational functions with indeterminates $\mathbf{X}$ and $\mathbf{Y}$.*

Remark 8.11. The entries of $M(\sqcup_{rich}, p)$ are polynomial, and hence we consider the rank of $M(\sqcup_{rich}, p)$ over the field of rational functions. Note that for every evaluation $p_{\mathbf{X}_0, \mathbf{Y}_0}(G) = p(G, \mathbf{X}_0, \mathbf{Y}_0)$ of p such that $\mathbf{X}_0, \mathbf{Y}_0$ belong to some subfield $\mathbb{F}$ of $\mathbb{C}$, the matrix $M(\square, p_{\mathbf{X}_0, \mathbf{Y}_0})$ has finite rank over $\mathbb{F}$.

8.3. **Replacing $\mathbb{Z}$ by arbitrary semirings.** Theorem 8.7 remains true when interpret the $\mathcal{L}$-polynomials not over the ring $\mathbb{Z}$ with the standard $+$ and $\cdot$, but over other rings and semirings. Stated explicitly, the following is true:

Theorem 8.12 (Bilinear Reduction Theorem for Semirings)**.** *Let $\mathcal{L}$ be a tame fragment of* SOL *such that $\sqcup_{rich}$ is $\mathcal{L}$-smooth. Let $\mathcal{S}$ be a semiring. Let p be a $\mathcal{L}(\sigma)$-polynomial in the form of Equation (8.10). There exist $t \in \mathbb{N}$, $\mathcal{L}(\tau)$-polynomials $p_1, \ldots, p_{2t}$ and a polynomial $Q \in \mathcal{S}[w_1, \ldots, w_t, v_1, \ldots, v_t]$ such that*
   (i) *for any two τ-structures $\mathfrak{A}$ and $\mathfrak{B}$*
$$p(\mathfrak{A} \sqcup_{rich} \mathfrak{B}) = Q\left(p_1(\mathfrak{A}), \ldots, p_t(\mathfrak{A}), p_{t+1}(\mathfrak{B}), \ldots, p_{2t}(\mathfrak{B})\right) .$$
   (ii) *$Q(\overline{w}, \overline{v})$ can be written as a bilinear form*
$$\overline{w}^{tr} M \overline{v}$$
   where the matrix M has as entries polynomials with coefficients from in $\mathcal{S}$ and is independent of the τ-structures.

For instance, the above holds for $\mathcal{S} = \mathbb{R} \cup \{-\infty\}$ with max and $+$ as the semiring's addition and multiplication, respectively, cf. [12]. Note, however, that it is not immediately clear how to extend the finite rank theorem to this context, since ranks of matrices over semirings can be defined in various ways.

The Bilinear Reduction Theorem for Semirings has found applications in the theory of weighted (aka multiplicity) automata, cf. [39].

8.4. **Sum-like operations.** Here we prove a Feferman-Vaught-type theorem for sum-like operations. Let ρ and τ be vocabularies and let σ be the vocabulary of the disjoint union of two τ-structures.

Theorem 8.13. *Let $\mathcal{L}$ be a tame fragment of* SOL *such that $\sqcup_{rich}$ is $\mathcal{L}$-smooth and $\square$ be an $\mathcal{L}$-sum-like operation between τ-structures. Let p be a $\mathcal{L}(\rho)$-polynomial in the form of Equation (8.10). There exists a polynomial $Q \in \mathbb{Z}[w_1, \ldots, w_t, v_1, \ldots, v_t]$ and $\mathcal{L}(\tau)$-polynomials $p_1, \ldots, p_{2t}$ such that*
   (i) *for any two τ-structures $\mathfrak{A}$ and $\mathfrak{B}$*
$$p(\mathfrak{A} \square \mathfrak{B}) = Q\left(p_1(\mathfrak{A}), \ldots, p_t(\mathfrak{A}), p_{t+1}(\mathfrak{B}), \ldots, p_{2t}(\mathfrak{B})\right) .$$
   (ii) *$Q(\overline{w}, \overline{v})$ can be written as a bilinear form.*

Proof. Let Φ be a scalar $\mathcal{L}$-transduction from σ_1 structures to ρ structures such that

$$\mathfrak{A}\square\mathfrak{B} = \Phi^*(\mathfrak{A} \sqcup_{rich} \mathfrak{B})\,.$$

Using Theorem 8.7 it is enough to show that there exists a $\mathcal{L}(\sigma)$-polynomial r such that

$$p(\Phi^*(\mathfrak{A} \sqcup_{rich} \mathfrak{B})) = r(\mathfrak{A} \sqcup_{rich} \mathfrak{B})\,.$$

Let $\phi_{univ}(v)$ be the formula of Φ which defines the universe of the transduction. Such r is given by

$$r(\mathfrak{M}) = \sum_{\substack{U\subseteq M: \\ \forall v(v\in U \to \phi_{univ}(v)) \wedge \Phi^\sharp(\Omega)}} \left(\prod_{c\in M:\Phi^\sharp(\phi_{\mathbf{X}}(c,U))} \mathbf{X}\right)\left(\prod_{c\in M:\Phi^\sharp(\phi_{\mathbf{Y}}(c,U))} \mathbf{Y}\right),$$

where we consider U and c in Ω, $\phi_{\mathbf{X}}$ and $\phi_{\mathbf{Y}}$ as free variables for the purpose of applying the map $\Phi^\sharp$. We use here that $\mathcal{L}$ is tame and therefore closed under transductions and containments. $\square$

Theorem 8.14. *Let p be a $\mathcal{L}(\rho)$-polynomial in the form of Equation (8.10). Then the connection matrix $M(\square, p)$ has finite rank over the field of rational functions with indeterminates $\mathbf{X}$ and $\mathbf{Y}$.*

Again evaluations of p have finite rank over the relevant subfield of $\mathbb{C}$.

9. Proving Non-definability of $\mathcal{L}(\tau)$-invariants

Using the Finite Rank Theorems requires showing that connection matrices have infinite rank. In the case of numeric τ-invariants and τ-polynomials, a simple technique for obtaining infinite rank is by showing that f is $\square$-maximizing or $\square$-minimizing. We will also use other techniques.

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

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

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

Proposition 9.1. *If f is a unboundedly $\square$-maximizing ($\square$-minimizing) τ-parameter, then $M(\square, f)$ has infinite rank.*

9.1. **Non-definability: numeric** CMSOL(τ)**-parameters.** Using Proposition 9.1, Theorem 7.4 and Proposition 3.2 we show that many τ-parameters are not CMSOL-definable:

Proposition 9.2. *The following graph parameters are not* CMSOL-*definable in the language of hypergraphs.*
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, total coloring number, Thue number, maximum degree, circumference, longest path, maximal connected planar (bipartite) induced subgraph, boxicity, minimal eigenvalue, spectral gap, girth, degeneracy, and minimum degree.

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

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

Proposition 9.3. *The number of connected components (blocks, simple cycles, induced paths) of maximum (minimum) size is not* CMSOL-*definable in the language of hypergraphs.*

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

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

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

In Section 9.3 we discuss the rank of connection matrices of various graph parameters based on averages, such as the average degree, and many others. These are examples where the computation of the rank is a bit more sophisticated.

9.2. **Non-definability:** CMSOL(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 [36]. 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 9.4. *Given a τ-parameter p, a binary operation $\square$ 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 one of the following occurs:*
  (i) *for every $\lambda \in \mathbb{N}$, $p(\mathcal{A}_i \square \mathcal{A}_j, \lambda) = 0$ iff $i + j > f(\lambda)$.*
  (ii) *for every $\lambda \in \mathbb{N}$, $p(\mathcal{A}_{2i} \square \mathcal{A}_{2j+1}, \lambda) = 0$ iff $i + j > f(\lambda)$.*
Then the connection matrix $M(\square, p)$ has infinite rank.

Proof. Let $\lambda \in \mathbb{N}$ and let p_λ be the graph parameter given by $p_\lambda(G) = p(G, \lambda)$. If ((i)) holds, consider the restriction N of the connection matrix $M(\square, p_\lambda)$ to the rows and columns corresponding to $\mathcal{A}_i$, $0 \leq i \leq f(\lambda) - 1$. If ((ii)) holds, consider the restriction N of $M(\square, p_\lambda)$ to to the rows corresponding to $\mathcal{A}_{2i}$ and the columns corresponding to $\mathcal{A}_{2i+1}$, $0 \leq i \leq f(\lambda) - 1$. In both cases N is a finite triangular matrix with non-zero diagonal. Hence the rank of $M(\square, p_\lambda)$ is at least $f(\lambda) - 1$.

Using that f is unbounded, we get that $M(\square, p)$ contains infinitely many finite submatrices with ranks which tend to infinity. Hence, the rank of $M(\square, p)$ is infinite. $\qquad\square$

We now use Lemma 9.4 to compute connection matrices where $\square$ is the disjoint union $\sqcup$, the 1-sum $\sqcup_1$ or the join $\bowtie$. In addition to the graph polynomials introduced in the introduction, we consider the definability of some $\mathcal{P}$-*polynomials* denoted $\chi_\mathcal{P}(G, k)$ where $\mathcal{P}$ is a graph property. $\chi_\mathcal{P}(G, k)$ counts vertex k-colorings $f : V(G) \to [k]$ such that for every color $c \in [k]$, the graph $f^{-1}(c)$ induced by the vertices colored c under f belongs to $\mathcal{P}$. Such colorings were introduced by F. Harary, cf. [30, 31, 10, 11]. Several of the colorings presented so far are $\mathcal{P}$- colorings for appropriate choices of $\mathcal{P}$.

Proposition 9.5. *The following connection matrices have infinite rank:*

   (i) $M(\bowtie, \chi(G, k))$;
   (ii) *For every $t > 0$ the matrix $M(\bowtie, \chi_{mcc(t)}(G, k))$;*
  (iii) $M(\bowtie, \chi_{v-acyclic}(G, k))$;
  (iv) $M(\bowtie, \chi_{\mathcal{P}_{Bipartite}}(G, k))$;
   (v) $M(\bowtie, \chi_{\mathcal{P}_{Forest}}(G, k))$;
  (vi) $M(\bowtie, \chi_{\mathcal{P}_{Tree}}(G, k))$;
 (vii) $M(\bowtie, \chi_{\mathcal{P}_{Planar}}(G, k))$;
(viii) $M(\bowtie, \chi_{\mathcal{P}_{3-regular}}(G, k))$;
  (ix) $M(\sqcup_1, \chi_{rainbow}(G, k))$;
   (x) $M(\sqcup, \chi_{convex}(G, k))$;
  (xi) *For every $t > 0$ the matrix $M(\sqcup_1, \chi_{t-improper}(G, k))$;*
 (xii) $M(\sqcup_1, \chi_{non-rep}(G, k))$;
(xiii) $M(\sqcup, \chi_{harm}(G, k))$.

Proof.

   (i) For $\chi(G, k)$ we use the fact that join of cliques is again a clique, $K_i \bowtie K_j = K_{i+j}$ and that $\chi(K_r, k) = 0$ iff $r > k$.
   (ii) In a similar fashion, for $\chi_{mcc(t)}(G, k)$ we use cliques and that $\chi_{mcc(t)}(K_r, k) = 0$ iff $r > kt$.
  (iii) For $\chi_{v-acyclic}(G, k)$ we use cliques. We have $\chi_{v-acyclic}(K_n, k) = 0$ iff $n > k$.
  (iv) For $\chi_{\mathcal{P}_{Bipartite}}(G, k)$ we use cliques and the fact that $\chi_{\mathcal{P}_{Bipartite}}(K_i, k) = 0$ iff $i > 2k$.
   (v) For $\chi_{\mathcal{P}_{Forest}}(G, k)$ we use cliques and the fact that $\chi_{\mathcal{P}_{Forest}}(K_i, k) = 0$ iff $i > 2k$.
  (vi) For $\chi_{\mathcal{P}_{Tree}}(G, k)$ we use cliques and the fact that $\chi_{\mathcal{P}_{Tree}}(K_i, k) = 0$ iff $i > 2k$.
 (vii) For $\chi_{\mathcal{P}_{Planar}}(G, k)$ we use cliques and the fact that $\chi_{\mathcal{P}_{Planar}}(K_i, k) = 0$ iff $i > 4k$.
(viii) For $\chi_{\mathcal{P}_{3-regular}}(G, k)$ we use cliques and the fact that $\chi_{\mathcal{P}_{3-regular}}(K_i, k) = 0$ iff $i > 3k$.
  (ix) 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 $r > k + 3$.
   (x) 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 $r > k$.

(xi) For $\chi_{t-improper}(G, k)$ we use cliques and that $\chi_{t-improper}(K_i \sqcup_1 K_j, k) = 0$ iff $\left\lceil \frac{i+j-2}{k} \right\rceil >$ t

(xii) For $\chi_{non-rep}(G, k)$ we use that 1-sum of stars S_n is again a star. We have $\chi_{non-rep}(S_n, k) = 0$ iff $n > k$.

(xiii) For $\chi_{harm}(G, k)$ we use the graphs nK_2 consisting of n disjoint edges. We have $\chi_{harm}(nK_2, k) = 0$ iff $n > \binom{k}{2}$.

$\square$

Corollary 9.6.

(i) $\chi(G, k)$, $\chi_{mcc(t)}(G, k)$ *(for any fixed $t > 0$)*, $\chi_{v-acyclic}(G, k)$, $\chi_{\mathcal{P}_{Bipartite}}(G, k)$, $\chi_{\mathcal{P}_{Forest}}(G, k)$, $\chi_{\mathcal{P}_{Tree}}(G, k)$, $\chi_{\mathcal{P}_{Planar}}(G, k)$ *and* $\chi_{\mathcal{P}_{3-regular}}(G, k)$ *are not* CMSOL-*definable in the language of graphs.*

(ii) $\chi_{rainbow}(G, k)$, $\chi_{convex}(G, k)$, $\chi_{non-rep}(G, k)$, *and* $\chi_{t-improper}(G, k)$ *(for any fixed $t > 0$) are not* CMSOL-*definable in the languages of graphs and hypergraphs.*

Proof. (i) The join is CMSOL-sum-like and CMSOL-smooth for graphs (but not for hypergraphs). (ii) The 1-sum and the disjoint union are CMSOL-sum-like and CMSOL-smooth for graphs and hypergraphs. $\square$

9.3. **Non-definability: Means and** CMSOL(τ). Maximization and minimization can sometimes be thought of as a type of mean. For example, the maximal (minimal) degree of a graph is obtained from the generalized mean

$$\left(\frac{\sum_{v \in V(G)} \text{degree}(v)^p}{|V(G)|} \right)^{\frac{1}{p}}$$

when p tends to ∞ $(-\infty)$.

Other instances of the generalized mean also lead to infinite connection matrices ranks. In particular, below we show examples for $p = 1$, $p = 2$ and $p = -1$. The following Lemma will be useful:

Lemma 9.7. *Let $\square$ be any binary operation between τ-structures.*

(i) *The $\square$-connection matrix of a linear combination of τ-invariants with $\square$-connection matrices of finite rank is of finite rank.*

(ii) *The $\square$-connection matrix of a finite product of τ-invariants with $\square$-connection matrices of finite rank is of finite rank.*

Proof.

(i) Follows from the sub-additivity of the rank of matrices.

(ii) It is enough to prove the claim for the product of two graph invariants, f and g. Denote the $\square$-connection matrices of f and g by M and N respectively. Since M and N are of finite rank, there exists $t \in \mathbb{N}$ such that for every $i \in \mathbb{N}$, the row M_i is a linear combination of the rows $M_1, \ldots, M_t$ and N_i is a linear combination of the rows $N_1, \ldots, N_t$. Hence, for every $i \in \mathbb{N}$ there exist $c_1, \ldots, c_t, d_1, \ldots, d_t$ such that for every $j \in \mathbb{N}$,

$$M_{i,j} = \sum_{r \leq t} c_r M_{r,j} \quad \text{and} \quad N_{i,j} = \sum_{s \leq t} d_s N_{s,j} \, .$$

Hence,

$$M_{i,j}N_{i,j} = \sum_{r,s \leq t} c_r d_s M_{r,j} N_{s,j} .$$

So, $M(\Box, f \cdot g)$ is spanned by t^2 rows.

$\square$

Let $\operatorname{avg val}(G)$ denote the average degree:

$$\operatorname{avg val}(G) = \frac{\sum_{v \in V(G)} \operatorname{degree}(v)}{|V(G)|} .$$

Proposition 9.8 (Arithmetic mean). *The rank of $M(\sqcup, \operatorname{avg val})$ is infinite.*

Proof. To see this, look at the connection matrix $M(\sqcup, \operatorname{avg val})$, whose entries can be expressed as follows:

$$M(\sqcup, \operatorname{avg val})_{i,j} = 2\frac{|E_i| + |E_j|}{|V_i| + |V_j|}.$$

The sub-matrix of $M(\sqcup, \operatorname{avg val})$ which consists only of rows and columns corresponding to graphs with exactly one edge is the Cauchy matrix $(\frac{2}{i+j})_{i,j}$, hence the rank of $M(\sqcup, \operatorname{avg val})$ is infinite. $\square$

We can prove a similar statement for other graph parameters given as arithmetic means:

Proposition 9.9. *The following arithmetic means have $\sqcup$-connection matrices of infinite rank:*
  (i) *For every $i \in \mathbb{N}$, the average size of the i-th neighborhood of vertices $v \in V(G)$, $|\{u \mid 0 \leq \operatorname{distance}(u,v) \leq i\}|$.*
 (ii) *Average number of simple cycles in which vertices $v \in V(G)$ occur.*
(iii) *Average number of triangles in which vertices $v \in V(G)$ occur.*
 (iv) *Average size of connected component in which vertices $v \in V(G)$ occur.*
  (v) *Average number of edges incident to an edge $e \in E(G)$.*
 (vi) *Average number of cycles which include $e \in E(G)$.*
(vii) *For every $i \geq 2$, the average number of i-paths which include $e \in E(G)$.*

Let the quadratic mean of the degrees of vertices $v \in V(G)$ be

$$\operatorname{q avg val}(G) = \left(\frac{\sum_{v \in V(G)} \operatorname{degree}(v)^2}{|V(G)|} \right)^{\frac{1}{2}} .$$

Proposition 9.10 (Quadratic mean). *The rank of $M(\sqcup, \operatorname{q avg val})$ is infinite.*

Proof. $(\operatorname{q avg val}(G))^2$ has infinite connection matrix rank with respect to $\sqcup$ by looking again at graphs with exactly one edge. Again, $M(\sqcup, \operatorname{q avg val}^2)$ has entries $M_{i,j} = \frac{2}{i+j}$. Hence, by Lemma 9.7, $M(\sqcup, \operatorname{q avg val}(G))$ has infinite rank. $\square$

With similar proofs, we have:

Proposition 9.11. *The corresponding quadratic means to those in Proposition 9.9 have $\sqcup$-connection matrices of infinite rank.*

Let the harmonic mean of the degrees of vertices $v \in V(G)$ be

$$\mathrm{h\,avg\,val}(G) = \frac{|V(G)|}{\sum_{v \in V(G)} \frac{1}{\mathrm{degree}(v)}} \, .$$

We only consider graphs with no isolated vertices in order to avoid devision by zero.

Proposition 9.12 (Harmonious mean). *The rank of $M(\bowtie, \mathrm{h\,avg\,val})$ is infinite.*

Proof. Note that $K_{n,m} = E_n \bowtie E_m$ and consider

$$\begin{aligned}
\mathrm{h\,avg\,val}(K_{n,m}) &= \frac{n+m}{\frac{n}{m} + \frac{m}{n}} \\
&= \frac{1}{n^2 + m^2} \cdot nm(n+m)
\end{aligned}$$

For every function $f : \mathbb{N} \times \mathbb{N} \to \mathbb{Q}$, let $N(f)$ be the matrix such that the entry in row i and column j is $f(i,j)$. Assume $M(\mathrm{h\,avg\,val}, \bowtie)$ is of finite rank. Then $N_1 = N\left(\frac{ij(i+j)}{i^2+j^2}\right)$ is of finite rank. Clearly, $N_2 = N\left(\frac{j(i+j)}{i^2+j^2}\right)$ is also of finite rank, since N_2 is obtained from N_1 by multiplying each row i of N_1 by a scalar $\frac{1}{i}$. The matrix $N\left(\frac{1}{j}\right)$ is of row rank 1 because all of its rows are equal. So, $N_3 = N\left(\frac{i+j}{i^2+j^2}\right)$ is of finite rank by Lemma 9.7. The matrix $N(i+j)$ is of row rank 2, since it is spanned by the vectors $(1\ 2\ 3\ 4\ \cdots)$ and $\mathbf{1}$. So, $N_4 = N\left(\frac{(i+j)^2}{i^2+j^2}\right)$ is of finite rank again by Lemma 9.7. Now notice

$$\frac{(i+j)^2}{i^2+j^2} = 1 + \frac{2ij}{i^2+j^2} \, .$$

Hence, we have that $N_5 = N\left(\frac{1}{i^2+j^2}\right)$ is of finite rank, but N_5 is a Cauchy matrix and is therefore of infinite rank, in contradiction. □

9.4. **Non definability: CFOL(τ)-parameters.** Here we discuss definability and non-definability of τ-parameters in CFOL. Recall from Section 7 that formulas in CFOL-parameters may not have second order variables. Examples of CFOL-parameters $p(G)$, $G = (V, E)$, in the vocabulary of graphs:

$$\begin{aligned}
|V| &= \sum_{v \in V} 1 & |Apex(G)| &= \sum_{v \in V : \forall x((x \approx v) \vee E(x,v))} 1 \\
|E| &= \sum_{v,u \in V : E(u,v) \wedge (u<v)} 1 & odd - deg(G) &= \sum_{x \in V : \mathrm{D}_{2,1} x E(y,x)} 1
\end{aligned}$$

where $Apex(G)$ is the set of vertices of G adjacent to all other vertices, $dgen(G)$ is the generating function of the degrees of vertices in G, and $odd - deg(G)$ is the number of vertices of odd degree.

On the other hand, we can use Theorem 7.5 to show non-definability of CFOL-parameters:

Proposition 9.13. *The following are not* CFOL-*definable:*
   (i) *The number* $spn - f(G)$ *of spanning forests*
   (ii) *The number* $spn - t(G)$ *of spanning trees.*
   (iii) *The number* $cyc(G)$ *of cycles in* G.
   (iv) *The number* $k(G)$ *of connected components in* G.
   (v) *The tree-width* $tw(G)$ *of* G.
   (vi) *The number* $blk(G)$ *of blocks of* G.

Proof. We use the constructions from Section 3.
   (i) The number of spanning forests $spn - f$ in the graph from Figure 1, the graph obtained
      by applying Φ_F to two directed paths of length n_1 and n_2, is 1 if $n_1 \neq n_2$, and is n_1 if
      $n_1 = n_2$. Hence, $M(\Phi_F, spn - f)$ has infinite rank. Since Φ_F is a CFOL-product-like
      operation, $spn - f$ is not CFOL-definable.

The other cases are similar. We describe them shortly.
   (ii) We use Φ_T from Figure 2. The graph is connected iff $n_1 \leq n_2$, implying that the
      connection matrix of the number of spanning trees is zero below the diagonal and
      non-zero otherwise, so has infinite rank.
   (iii) We use Φ_F from Figure 1 again. We have $cyc(G) = 1$ iff $n_1 = n_2$, otherwise $cyc(G) =$
      0. Hence the connection matrix of $cyc(G)$ has 2 on the diagonal and 1 otherwise.
      Consequently it has infinite rank.
   (iv) We use Φ_T from Figure 2. We have $k(G) = 1$ if $n_1 \leq n_2$ and $k(G) = n_1 - n_2 + 1$
      otherwise. Hence, the connection matrix of $k(G)$ so has infinite rank.
   (v) We use Φ_P from Figure 3. We have $tw(G) = 4$ iff $n_1 \neq n_2$, and otherwise $tw(G) = 5$.
   (vi) We use Φ_B from Figure 4. We have $blk(G) = 1$ if $n_1 < n_2 + 1$, $blk(G) = 2$ if $n_1 = n_2 + 1$
      and otherwise $blk(G) = 3$.

$\square$

Acknowledgements

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] N. Alon, J. Grytczuk, M. Hauszczak, and O. Riordan. Nonrepetitive colorings of graphs. In M. Káronski, J. Spencer, and A. Ruciński, editors, *Proceedings of the 10th International Conference on Random Structures and Algorithms*, volume 21, 3–4 of *Random Structures and Algorithms*, pages 336–346, Danvers, MA, August 6–10 2002. Wiley Periodicals.
[3] N. Alon, C. McDiarmid, and B.A. Reed. Acyclic coloring of graphs. *Random Structures and Algorithms*, 2(3):277–288, 1991.
[4] N. Alon and A. Zaks. Algorithmic aspects of acyclic edge colorings. *Algorithmica*, 32(4):611–614, 2002.
[5] S. Arnborg, J. Lagergren, and D. Seese. Problems easy for tree-decomposable graphs (extended abstract). In *ICALP*, pages 38–51, 1988.
[6] S. Arnborg, J. Lagergren, and D. Seese. Easy problems for tree decomposable graphs. *Journal of Algorithms*, 12:308–340, 1991.
[7] M. Basavaraju and L.S. Chandran. Acyclic edge coloring of subcubic graphs. *Discrete Mathematics*, 308(24):6650–6653, 2008.

[8] J. Berstel and C. Reutenauer. *Rational Series and their languages*, volume 12 of *EATCS Monographs on Theoretical Computer Science*. Springer, 1984.

[9] B. Bollobás. *Modern Graph Theory*. Springer, 1999.

[10] J.I. Brown and D.G. Corneil. On generalized graph colorings. *Journal of Graph Theory*, 11(1):87–99, 1987.

[11] J.I. Brown and D.G. Corneil. Graph properties and hypergraph colourings. *Discrete Mathematics*, 98, 1991.

[12] P. Butkovič. *Max-linear Systems: Theory and Algorithms*. Springer Monographs in Mathematics. Springer, 2010.

[13] J.W. Carlyle and A. Paz. Realizations by stochastic finite automata. *J. Comp. Syst. Sc.*, 5:26–40, 1971.

[14] S. Chakraborty, E. Fischer, A. Matsliah, and R. Yuster. Hardness and algorithms for rainbow connection. *Journal of Combinatorial Optimization*, 21(3):330–347, 2011.

[15] G. Chartrand, G.L. Johns, K.A. McKeon, and P. Zhang. Rainbow connection in graphs. *Mathematica Bohemica*, 133(1):85–98, 2008.

[16] 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.

[17] B. Courcelle and J. Engelfriet. *Graph Structure and Monadic Second-order Logic, a Language Theoretic Approach*. Cambridge University Press, 2012.

[18] 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.

[19] L. Cowen, W. Goddard, and C.E. Jesurum. Defective coloring revisited. *Journal of Graph Theory*, 24(3):205–219, 1997.

[20] R. Diestel. *Graph Theory*. Graduate Texts in Mathematics. Springer, 3 edition, 2005.

[21] M. Droste, W. Kuich, and H. Vogler, editors. *Handbook of Weighted Automata*. Springer, 2009.

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

[23] 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.

[24] P. Erdös and A. Rényi. Asymmetric graphs. *Acta Mathematica Hungaric*, 14(3):295–315, 1963.

[25] A. Farrugia. Vertex-partitioning into fixed additive induced-hereditary properties is NP-hard. *Electric Journal of Combinatorics*, 11(1), 2004.

[26] M. Freedman, László Lovász, and A. Schrijver. Reflection positivity, rank connectivity, and homomorphisms of graphs. *Journal of AMS*, 20:37–51, 2007.

[27] 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.

[28] A.J. Goodall and S.D. Noble. Counting cocircuits and convex two-colourings is #P-complete. arXiv:0810.2042, 2008.

[29] B. Grünbaum. Acyclic colorings of planar graphs. *Israel Journal of Mathematics*, 14:390–412, 1973.

[30] F. Harary. Conditional colorability in graphs. In *Graphs and applications: proceedings of the First Colorado Symposium on Graph Theory (Boulder, Colo., 1982)*, pages 127–136. Wiley-Intersci. Publ., Wiley, New York, 1985.

[31] F. Harary and L.H. Hsu. Conditional chromatic number and graph operations. *Bulletin of Institute of the Mathematics Academia Sinica*, 19:125–134, 1991.

[32] M. A. Harrison. *Introduction to Formal Language Theory*. Addison-Wesley, 1978.

[33] J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley Series in Computer Science. Addison-Wesley, 1980.

[34] J.E. Hopcroft and M.S. Krishnamoorthy. On the harmonious coloring of graphs. *SIAM J. Algebraic Discrete Methods*, 4:306–311, 1983.

[35] P. G. Kolaitis and Jouko A. Väänänen. Generalized quantifiers and pebble games on finite structures. *Ann. Pure Appl. Logic*, 74(1):23–75, 1995.

[36] T. Kotek. *Definability of Combinatorial Functions*. PhD thesis, Technion—Israel Institute of Technology, 2012.

[37] 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 339353, 2008.

[38] 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.

[39] N. Labai and J. A. Makowsky. Weighted automata and monadic second order logic. In *GandALF*, pages 122–135, 2013.

[40] L. Libkin. *Elements of Finite Model Theory*. Springer, 2004.

[41] 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.

[42] L. Lovász. Connection matrices. In G. Grimmet and C. McDiarmid, editors, *Combinatorics, Complexity and Chance, A Tribute to Dominic Welsh*, pages 179–190. Oxford University Press, 2007.

[43] L. Lovász. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.

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

[45] J.A. Makowsky. From a zoo to a zoology: Towards a general theory of graph polynomials. *Theory of Computing Systems*, 43:542–562, 2008.

[46] 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.

[47] J. A. Makowsky. Connection matrices for msol-definable structural invariants. In *ICLA*, pages 51–64, 2009.

[48] F. Manin. The complexity of nonrepetitive edge coloring of graphs. *CoRR*, abs/0709.4497, 2007. informal publication.

[49] S. Moran and S. Snir. Efficient approximation of convex recolorings. *Journal of Computer and System Sciences*, 73(7):1078–1089, 2007.

[50] J. Nurmonen. Counting modulo quantifiers on finite structures. *Inf. Comput.*, 160(1-2):62–87, 2000.

[51] S. Skulrattanakulchai. Acyclic colorings of subcubic graphs. *Information Processing Letters*, 92(4):161–167, 2004.