Connection Matrices for MSOL-definable Structural Invariants

J.A. Makowsky*

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

Abstract. Connection matrices of graph parameters were first introduced by M. Freedman, L. Lovász and A. Schrijver (2007) to study the question which graph parameters can be represented as counting functions of weighted homomorphisms. The rows and columns of a connection matrix $M(f, \Box)$ of a graph parameter f and a binary operation $\Box$ are indexed by all finite (labeled) graphs G_i and the entry at (G_i, G_j) is given by the value of $f(G_i \Box G_j)$. Connection matrices turned out to be a very powerful tool for studying graph parameters in general.
B. Godlin, T. Kotek and J.A. Makowsky (2008) noticed that connection matrices can be defined for general relational structures and binary operations between them, and for general structural parameters. They proved that for structural parameters f definable in Monadic Second Order Logic, (*MSOL*) and binary operations compatible with *MSOL*, the connection matrix $M(f, \Box)$ has always finite rank. In this talk we discuss several applications of this Finite Rank Theorem, and outline ideas for further research.

1 Introduction

Graph Parameters and Graph Polynomials. A graph parameter (also called a numeric graph invariant) f is a function from the class of all finite graphs $\mathcal{G}$ to some numeric domain which is an ordered commutative ring $\mathcal{R}$ or an ordered field $\mathcal{F}$ with 0 and 1, usually the integers $\mathbb{Z}$, the rational numbers $\mathbb{Q}$ or the reals $\mathbb{R}$. Graph properties are the special case where the values of f are 0 or 1. In the case of graph properties the ring can be taken to be the two-element boolean algebra, or, alternatively the field $\mathbb{Z}_2$. We shall use the latter, to make our use of linear algebra uniform.

Graph polynomials are functions p from $\mathcal{G}$ into a polynomial ring, usually $\mathbb{Z}[\bar{X}]$, where $\bar{X}$ is a fixed finite set of indeterminates. Graph polynomials are a way to encode infinitely many graph parameters. Every evaluation of the polynomial $p(G; \bar{X})$ at some point $\bar{X} = \bar{x_0}$ is a graph parameter. So are the coefficients

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

of $p(G;\bar{X})$, the total degree or the degree of monomials where the coefficient satisfies certain properties, and the zeros of $p(G;\bar{X})$.

Instead of graphs one can also consider hypergraphs or relational structures over some fixed finite vocabulary τ, a set consisting of relation symbols and constants. In this case we speak of *structural invariants for τ-structures*, or just of *τ-invariants* and *τ-polynomials*. We include here the empty τ-structure, which we denote by $\emptyset_\tau$.

Connection Marices. Let $\Box$ be a binary operation on τ-structures (which respects τ-isomorphisms. A τ-structure $\mathcal{I}$ is $\Box$-neutral if for every τ-structure $\mathcal{A}$ we have $\mathcal{A}\Box\mathcal{I} \simeq \mathcal{I}\Box\mathcal{A}$. For the disjoint union of τ-structures, denoted by $\sqcup$, the empty structure is $\sqcup$-neutral. For the cartesian product of τ-structures, denoted by $\times$, the one-element structure with full relations is $\times$-neutral.

Let f be a τ-invariant and $\Box$ be a binary operation on τ-structures which respects τ-isomorphisms. Let $\{\mathcal{A}_i : i \in \mathbb{N}\}$ be an enumeration of all finite τ-structures (up to isomorphisms). We define the infinite matrix $M(f,\Box) = (m_{i,j}(f,\Box))$ by $m_{i,j}(f,\Box)) = f(\mathcal{A}_i\Box\mathcal{A}_j$. $M(f,\Box) = (m_{i,j}(f,\Box))$ is called the *connection matrix of f and $\Box$*. We denote by $r_{\mathcal{R}}(f,\Box)$ the rank over $\mathcal{R}$ of the matrix $M(f,\Box)$. We usually omit the subscript in $r_{\mathcal{R}}(f,\Box)$, when no confusion arises.

Multiplicative τ-invariants. A τ-invariant f is called $\Box$-*multiplicative* if it satisfies $f(\mathcal{A}\Box\mathcal{B}) = f(\mathcal{A})\cdot f(\mathcal{B})$ for all finite τ-structures $\mathcal{A}$ and $\mathcal{B}$. With respect to the disjoint union, $\sqcup$, typical examples of $\sqcup$-multiplicative graph parameters are $\chi(G,k)$, the number of proper vertex colorings with k colors, $pm(G)$, the number of perfect matchings, or the number of acyclic orientations. With respect to the join of graphs, denoted by $\bowtie$, counting the number of covers by independent sets is $\bowtie$-multiplicative.

In [FLS07] the following characterization of graph parameters, which are multiplicative with respect to the disjoint union, was given. We state it here in the general context. The proof is verbatim the same.

Proposition 1 *Let f be a τ-invariant with values in an ordered field $\mathcal{R}$, and which is not identically 0, and let $\Box$ be a binary operation on τ-structures, with $\mathcal{I}$ being a unique $\Box$-neutral structure. Then f is $\Box$-multiplicative if and only if $f(\mathcal{I}) = 1$, and the matrix $M(f,\Box)$ has rank 1 and is positive semi-definite.*

Connection matrices for various operations on labeled graphs are studied in [FLS07,Sze07,Lov07,Sch08]. In these papers they are used to characterize graph invariants arising from various vertex-coloring and edge-colorings models. In [GKM08] connection matrices are used to study definability properties of graph parameters and graph polynomials.

Outline of the Talk. In this talk we summarize the results from [FLS07] and [GKM08] and discuss further applications and open problems. In Section

2 introduce connection matrices of τ-invariants and their rank. We illustrate their uses in the case of graph parameters. This paraphrases the main results of [FLS07], and explains the Freedman-Lovász-Schrijver Theorem which gives a characterization of graph parameters arising from counting weighted homomorphisms. In Section 3 we show how one can use a Feferman-Vaught-style Theorem from [Mak04] for τ-invariants definable in Monadic Second Order Logic *MSOL* to show that the rank of many connection matrices has to be finite. The exact statement of this is the Finite Rank Theorem (Theorem 9). In Section 4 we give applications of the Finite Rank Theorem, mostly taken from [GKM08]. We conclude with a list of open problems for further investigations.

2 Properties of Connection Matrices of τ-invariants

The Rank of Connection Matrices. Besides multiplicative τ-invariants we consider also τ-invariants f with the following properties:

(i) f is $\Box$-*additive* if $f(G_1 \Box G_2) = f(G_1) + f(G_2)$.

(ii) f is $\Box$-*maximizing*, respectively $\Box$-*minimizing* if there exist infinite sequences of graphs $(G_i)_{i\in\mathbb{N}}, (H_i)_{i\in\mathbb{N}}$ such that for all $(i,j) \in \mathbb{N}^2$ we have

$$f(G_i \Box H_j) = \max f(G_i), f(H_j), \text{respectively} f(G_i \Box H_j) = \min f(G_i), f(H_j).$$

Furthermore, for all $i \in \mathbb{N}$ the sequence $f(i)_j = f(G_i \Box H_j)$ is strictly monotone increasing. If the two sequences consist of a ll τ-structures, we say f is *strictly* $\Box$-*maximizing*, respectively *strictly* $\Box$-*minimizing*.

(iii) A τ-invariant f is *weakly* $(\Box, \gamma)$-*multiplicative*, if there exists a finite set of graph parameters $f_i : i \leq \gamma$ with $i, \gamma \in \mathbb{N}$ with $f = f_0$, and a matrix $N^k \in \mathbb{R}^{\gamma\times\gamma}$, such that $f_0(\mathcal{A}_1 \Box \mathcal{A}_2) = \sum_{i,j} f_i(\mathcal{A}_1) N^k_{ij} f_j(\mathcal{A}_2)$.
In other words, $f(\mathcal{A}_1 \Box \mathcal{A}_2)$ is given by a quadratic form defined by $N^k_{i,j}$ of rank at most γ.

Typical examples[1] of $\sqcup$-additive parameters are the cardinality of the vertex set, the cardinality of the edge-set, $k(G)$, $b(G)$, number of connected components and number of blocks (doubly connected components), respectively. An example of a $\bowtie$-additive graph parameter is the diameter of a graph. Among the $\sqcup$-maximizing graph parameters we have: the chromatic number $\chi(G)$, the edge chromatic number $\chi_e(G)$, and the total coloring number $\chi_t(G)$, the size of a maximal clique $\omega(G)$, the size of the maximal degree $\Delta(G)$, the tree-width $tw(G)$, and the clique-width $cw(G)$. Note that, for example, $\chi(G)$, $\omega(G)$ and $\Delta(G)$, are $\bowtie$-additive.

Proposition 2 *Let f be a τ-invariant.*

(i) If f is $\Box$-multiplicative, $r(f, \Box) = 1$.

[1] Almost all graph parameters discussed are taken from [Die96]. One exception is the clique-width, which was introduced in [CO00], and, in connection to graph polynomials, in [Mak04].

(ii) *If f is $\Box$-additive, $r(f, \Box) = 2$, unless $M(f, \Box)$ is the zero matrix.*
(iii) *If f is $\Box$-maximizing or $\Box$-minimizing, $r(f, \Box)$ is infinite.*
(iv) *Let f be a graph parameter which is weakly $(\Box, \gamma)$-multiplicative. Then $r(f, \Box) \leq \gamma$.*

Proof. (i) was already stated in Proposition 1. (ii), (iii) and (iv) are easy to verify.

Counting Weighted Homomorphisms of Graphs . A k-graph is a graph $G = (V(G), E(G))$ with k distinct vertices labeled with $0, 1, \ldots, k-1$. We denote by $\mathcal{K}_k$ the class of finite k-graphs. $\mathcal{G}_0 = \mathcal{G}$ the set of all finite graphs without labels.

Given two k-graphs G_1, G_2 we define the k-sum $G_1 \sqcup_k G_2$ as the disjoint union of G_1 and G_2 where we identify correspondingly labeled vertices. In [FLS07] the connection matrices $M(f, \sqcup_k)$ on $\mathcal{G}_k$ are used to characterize those graph parameters f which can be represented as counting functions of weighted homomorphisms. The setup is as follows:

Let $H = (V(H), E(H)) \in \mathcal{G}$ be a fixed graph, possibly with loops. Let $\alpha : V(H) \to \mathbb{R}^+$ and $\beta : E(H) \to \mathbb{R}$ be weight functions of vertices and edges respectively, and let $h : G \to H$ be a homomorphism. We define weights of h by

$$\alpha_h = \prod_{u \in V(G)} \alpha(h(u)) \quad \text{and} \quad \beta_h = \prod_{u,v \in E(G)} \beta(h(u), h(v))$$

Finally, we sum over all homomorphisms

$$Z_{H,\alpha,\beta}(G) = \sum_{h:G \to H} \alpha_h \cdot \beta_h.$$

$Z_{H,\alpha,\beta}(G)$ is often called a *partion function* or a *vertex coloring model.*

Observation 1 *Partition functions are multiplicative.*

Example 1 *The following are simple partition functions:*

(i) *For $H = K_m$, a clique with m vertices,*

$$Z_{K_m,1,1}(G) = \chi(G, m)$$

which counts the number of proper m-colorings.
(ii) *For $H = L_1$, an isolated loop, $\alpha = \lambda$, $\beta = \mu$,*

$$Z_{L_1,\lambda,\mu}(G) = \lambda^{|V(G)|} \cdot \mu^{|E(G)|}$$

(iii) *For $H = L_m$ consisting of m isolated loops, $\alpha = \lambda$, $\beta = \mu$,*

$$Z_{L_m,\lambda,\mu}(G) = m^{k(G)} \cdot \lambda^{|V(G)|} \cdot \mu^{|E(G)|}$$

(iv) *For* $H = K_1 \bowtie L_1$ *with vertices* v, ℓ *respectively, and* $\alpha(v) = X, \alpha(\ell) = 1$, $\beta = 1$ *we get*

$$Z_{K_1 \bowtie L_1, \alpha, \beta}(G) = \sum_i ind_i(G) \cdot X^i$$

where $ind_i(G)$ *is the number of independents sets of size* i *in* G*.*

In [FLS07] it is proved that the connection matrices $M(f, \sqcup_k)$ for $f = Z_{H,\alpha,\beta}(G)$ have the following properties:

Proposition 3 (M. Freedman, L. Lovász and A. Schrijver, 2007)

(i) *For every weighted graph* (H, α, β) *we have*

$$r(Z_{H,\alpha,\beta}(G), \sqcup_k) \leq |V(H)|^k$$

(ii) *If* (H, α, β) *has no automorphisms and no twins, then*

$$r(Z_{H,\alpha,\beta}(G), k) = |V(H)|^k$$

Autmorphisms here are weight preserving. Two vertices $u, v \in V(H)$ of (H, α, β) are *twins* if for every $w \in V(H)$ we have that $\beta(u, w) = \beta(v, w)$. Being twins does not depend on α.

Proposition 4 (M. Freedman, L. Lovász and A. Schrijver, 2007) *For every weighted graph* (H, α, β) *the matrix* $M(Z_{H,\alpha,\beta}(G), \sqcup_k)$ *is positive semi-definite.*

Example 2 *(i)* *Let* $pm(G)$ *denote the number of perfect matchings of* G*.* $pm(G)$ *is multiplicative and* $r(pm, k) = 2^k$*, but* $M(pm, 1)$ *is not positive definite.*

(ii) *For* $\chi(-, \lambda)$*,* $\lambda \in \mathbb{Z}$ *we have:* $M(\chi(-, \lambda), k)$ *is positive-semi-definite, and* $r(\chi(-, \lambda), k)$ *is finite, but exponentially bounded only for* $\lambda \in \mathbb{Z}^+$*, otherwise it grows superexponentially.*

The Freedman-Lovász-Schreijver Theorem. We say that a a numeric graph invariant is *hom-presentable* if there is a weighted graph (H, α, β) such that for every G $f(G) = Z_{H,\alpha,\beta}(G)$. We have seen in Example 1 that $2^{|V(G)|}, 2^{|E(G)|}, 2^{k(G)}$ are hom-presentable, but by Proposition 2 and 3, $|V(G)|, |E(G)|, k(G)$ are not hom-presentable, as their connection matrices have infinite rank. $\chi(-, \lambda)$ is hom-representable for every $\lambda \in \mathbb{Z}^+$, but the *choice of* (H, α, β) *depends on* λ.

Theorem 5 (M. Freedman, L. Lovász and A. Schrijver, 2007) *Let* f *be a real-valued graph parameter.* f *is hom-presentable iff for every* $k \in \mathbb{N}$

(i) $M(f, k)$ *is positive semi-definite, and*
(ii) $r(f, k) \leq q^k$ *for some* $q \in \mathbb{N}^+$*.*

There are various generalizations of Theorem 5. B. Szegedy [Sze07] considers *edge coloring models* and connection matrices $S(f, k)$ based on identification of k unfinished edges. A. Schrijver [Sch08] unifies the proofs of [FLS07] and [Sze07] using further variations of connection matrices defined also for hyper-graphs and directed graphs.

3 Enter Logic

Monadic Second Order Logic. A vocabulary is a finite set of relation and constant symbols. We define the logic *MSOL* for τ-structures inductively. We have first order variables $x_i : i \in \mathbb{N}$ which range over elements of A, the universe of a τ-structure, and (monadic) second order variables $U_i : i \in \mathbb{N}$, which range over subsets of A. Terms $t, t', \ldots$ are either first order variables or constant symbols from τ. Atomic formulas are of the form $t = t'$, $R(\bar{t})$, where R is a relation symbol of τ $U_i(t)$ and have the natural interpretation. Formulas are built inductively using the connectives $\vee, \wedge, \rightarrow, \leftrightarrow, \neg$, and the quantifiers $\forall x_i$, $\exists x_i$, $\forall U_i$, $\exists U_i$ with their natural interpretation. The quantifier rank of an *MSOL*-formula ϕ is defined as usual and denoted by $qr(\phi)$ and for the rank we do not distinguish between first order and second variables.

***MSOL*-definable τ-Polynomials in Normal Form.** A *MSOL*-definable polynomial in indeterminates $X_1, \ldots, X_\ell$ in *normal form* has the form

$$\sum_{U_1:\Phi_1(U_1)} \sum_{U_2:\Phi_2(U_2)} \cdots \sum_{U_{\ell_1}:\Phi_{\ell_1}(U_{\ell_1})} \left(\prod_{\bar{x}_1:\phi_1(\bar{x}_1)} X_1 \prod_{\bar{x}_2:\phi_2(\bar{x}_2)} X_2 \ldots \prod_{\bar{x}_\ell:\phi_\ell(\bar{x}_\ell)} X_\ell \right)$$

where all the formulas Φ_i and ϕ_i are *MSOL*-formulas with the iteration variables (for summation and products) indicated. There may be additional parameters in the formulas. However, Φ_i may not contain the variables U_j for $j > i$, and ϕ_i may not contain $\bar{x}_j$ for $j > i$. Both Φ_i and ϕ_i are referred to as iteration formulas.

Looking at the partition function

$$Z_{H,\alpha,\beta}(G) = \sum_{h:G\rightarrow H} \alpha_h \cdot \beta_h. \tag{1}$$

we can rewrite it as follows: Let $G = (V(G), E(G))$, $H = (V(H), G(H))$ and $V(H) = \{v_0, \ldots, v_{n-1}\}$. We introduce, for each $v_i : i \leq n-1$ a set variable U_i. Let $\phi_{hom(H)}(U_0, \ldots, U_{n-1})$ be the formula $U_0, \ldots, U_{n-1}$ is a partition of $V(G)$ and that for all $x, y \in V(G)$, if $(x, y) \in E(G)$ then there is a $(v_i, v_j) \in E(H)$ such that $x \in U_i$ and $y \in U_j$. The formula $\phi_{hom(H)}(U_0, \ldots, U_{n-1})$ is a first order formula over the relation symbols for $E(G)$ and $U_0, \ldots, U_{n-1}$. It can also be viewed as a formula in Monadic Second Order Logic *MSOL* over the vocabulary consisting only of the binary relation symbol for $E(G)$.

Now the expression (1) can be, using $\bar{U} = (U_0, U_1, \ldots U_{n-1})$, written as

$$Z_{H,\alpha,\beta}(G) = \sum_{\bar{U}:\phi_{hom(H)}} \left(\left(\prod_{i=0}^{n-1} \prod_{x\in U_i} \alpha(x)) \right) \left(\prod_{(j,k)\in E(H)} \prod_{(y\in U_j \wedge z\in U_k)} \beta(y,z)) \right) \right) \tag{2}$$

If we consider all the $\alpha(v_i)$ and $\beta(v_j, v_k)$ as indeterminates, the left hand side of the expression (2) is a typical instance of a *MSOL*-definable graph polynomial introduced in [Mak04]. For fixed values of $\alpha(v_i)$ and $\beta(v_j, v_k)$ this gives an *MSOL*-definable graph parameter, and, more generally, if we replace graphs by relational structures, of *MSOL*-definable τ-invariants. Hence we have shown:

Proposition 6 *For every α, β the graph parameter $Z_{H,\alpha,\beta}(G)$ is an MSOL-definable τ_1-invariant with $\tau_1 = \{E\}$.*

Using Finite Rank to Compute Partition Functions. Let $TW(k)$ and $CW(k)$ denote the class of graphs of tree-width and clique-width at most k, respectively. It was shown in [CO00] that $TW(k) \subseteq CW(2^{k+1} + 1)$. Using the main results of [CMR01,Mak04] combined with [Oum05] we get from Proposition 6 the following complexity result.

Proposition 7 *On the class $CW(k)$ of graphs of clique-width at most k the graph invariants $Z_{H,\alpha,\beta}(G)$ can be computed in polynomial time, and are fixed parameter tractable, i.e., the exponent of the polynomial is independent of k, but the upper bounds for the constants are simply exponential in the case of $TW(k)$, but at least doubly exponential in k in the case of $CW(k)$.*

For graphs in $TW(k)$ this was already observed in [Lov07]. To get the better bound on the constants in the case of $TW(k)$, we can use Proposition 3 in the dynamic programming algorithm underlying the proofs in [CMR01,Mak04].

***MSOL*-compatible Operations on τ-structures.** Two τ-structures $\mathcal{A}, \mathcal{B}$, are said to be k-equivalent for *MSOL*, if they satisfy the same *MSOL*-sentences of quantifier rank k. We denote this equivalence relation by $\mathcal{A} \equiv_k \mathcal{B}$.

A binary operation $\Box$ on τ-structures is called *MSOL-k-compatible* if for $k \in \mathbb{N}$ we have that $\mathcal{A} \equiv_{m+k} \mathcal{A}'$ and $\mathcal{B} \equiv_{m+k} \mathcal{B}'$ implies that

$$\mathcal{A} \Box \mathcal{B} \equiv_m \mathcal{A}' \Box \mathcal{B}'.$$

The operation $\Box$ is called *MSOL-compatible* if there is some $k \in \mathbb{N}$ such that $\Box$ is *MSOL*-k-compatible.

In [Mak04] the case of $k = 0$ is called *MSOL-smooth*. The disjoint union of τ-structures is *MSOL*-smooth. So are the operations $\sqcup_k$ on k-graphs. The cartesian product $\times$ is not *MSOL*-compatible. However, the notion of *MSOL*-compatible operation is sensitive to the choice of the representation of,say, graphs as τ-structures. If we represent graphs $G = (V(G), E(G))$ as τ_1-structures with $\tau_1 = \{E\}$, which have universe $V(G)$ and a binary relation $E(G)$, the join operation $G_1 \bowtie G_2$ is *MSOL*-smooth. This is so, because it can be obtained from the disjoint union by the application of a quantifierfree transduction. If, however, we represent graphs as a two-sorted $\{R\}$ τ_2-structures, with $\tau_2 = \{P_V, P_E, R\}$, with sorts $P_V = V(G)$ and $P_E = E(G)$, and a binary incidence relation $R(G) \subset V(G) \times E(G)$, then $G_1 \bowtie G_2$ contains the cartesian product $V(G_1) \times V(G_2)$ in $E(G_1 \bowtie G_2)$ and behaves more like a cartesian product, which is not even *MSOL*-compatible. It is important to note that the operations $\sqcup_k$ are *MSOL*-smooth for graphs as τ_1-structures and as τ_2-structures.

The following theorem is proven in [Mak04, Theorem 6.4]:

Theorem 8 *Let f be a graph parameter which is the evaluation $f(G) = p(G, \bar{x}_0)$ of an MSOL-definable τ-polynomial $p(G, \bar{X})$. Furthermore, let $\Box$ be a binary operation on τ-structures which is MSOL-k-compatible. Then f is weakly $(\Box, \gamma)$)-multiplicative for some $\gamma \in \mathbb{N}$ which depend on τ, the polynomial p, k, but not on $\bar{x}_0$.*

The Finite Rank Theorem. As in [GKM08], we get immediately, using Proposition 2 and Theorem 8 the following Theorem.

Theorem 9 (Finite Rank Theorem) *Let $p(G, \bar{X})$ be an MSOL-definable τ-polynomial with values in $\mathbb{R}[\bar{X}]$ with m indeterminates, and let $\Box$ be a binary operation on τ-structures which is MSOL-k-compatible. There is $\gamma_{\tau,\Box}(p) \in \mathbb{N}$ depending on τ, the polynomial p, and k only, such that for all $\bar{x}_0 \in \mathbb{R}^m$, we have $r(p(G, \bar{x}_0), \Box) \leq \gamma_{\tau,\Box}(p)$.*

The upper bound on the rank obtained in Theorem 9 again is very large. In the case of partition functions this bound is computed precisely in Proposition 3.

4 Applications of the Finite Rank Theorem

4.1 Non-definability in *MSOL*

Counting hamiltonian circuits. We shall look at the graph parameter $hc(G)$ which count the number of hamiltonian circuits of a graph G, and the graph property HAM, which consists of all graphs which do have a hamiltonian circuit. If we represent graphs $G = (V(G), E(G))$ as τ_1-structures with $\tau_1 = \{E\}$, which have universe $V(G)$ and a binary relation $E(G)$, it is well known, cf. [dR84], that HAM is not *MSOL*-definable. If, however, we represent graphs as a two-sorted $\{R\}$ τ_2-structures, with $\tau_2 = \{P_V, P_E, R\}$, with sorts $P_V = V(G)$ and $P_E = E(G)$, and a binary incidence relation $R(G) \subset V(G) \times E(G)$, HAM is *MSOL*-definable.

Let E_m be the graph with m vertices and no edges. It is easy to see that $E_m \bowtie E_n$a contains exactly one hamiltonian circuit if and only if $m = n$. Therefore, $M(hc, \bowtie)$ and $M(HAM, \bowtie)$ both contain the infinite unit matrix as a submatrix, and $r(HAM, \bowtie)$ is infinite over $\mathbb{Q}$, whereas $r(HAM, \bowtie)$ is infinite over $\mathbb{Z}_2$. We conclude that, HAM is not an *MSOL*-definable property of τ_1-structures, and that hc is not an evaluation of an τ_1-polynomial.

The subtle point is, that the join of two graphs is *MSOL*-smooth only for graphs as τ_1-structures. In the presentation as τ_2-structures, the sort $E(G_1 \bowtie G_2)$ grows quadratically in the size of $V(G_1)$ and $V(G_2)$, and is not even *MSOL*-compatible.

Graph colorings with no large monochromatic components.
The same happens with the chromatic polynomial, and its relatives, the polynomials $mcc_t(G, k)$ for $t \in \mathbb{N} - \{0\}$. Following [LMST07], we denote by $mcc_t(G, k)$

the number of functions $f : V(G) \to [k]$ such that for each $i \leq k$, the set $f^{-1}(i)$ induces a graph which consist of connected components of size at most t. Clearly, we have $\chi(G, k) = mcc_1(G, k)$. It follows from results in [KMZ08] that for each $t \in \mathbb{N}$ the counting function $mcc_t(G, k)$ is a polynomial in k.

Proposition 10 (T. Kotek) *For each $t \in \mathbb{N} - \{0\}$ the rank $r(mcc_t(G, k), \bowtie)$ tends to infinity with k.*

Corollary 11 *The polynomial $mcc_t(G, k)$ is not a τ_1-polynomial.*

But for connected graphs, we have $\chi(G, k) = T(G; 1 - k, 0)$, where $T(G, X, Y)$ is the Tutte polynomial, which is *MSOL*-definable over the vocabulary $\tau_3 = \tau_2 \cup \{<_E\}$, where $<_E$ is a linear ordering of $E(G)$.

4.2 Evaluations of Well Known Graph Polynomials

A particular graph polynomial is considered interesting if it encodes many useful graph parameters. Let $G = (V(G), E(G))$ be a graph. The characteristic polynomial $P(G, X)$ of a graph is defined as the characteristic polynomial (in the sense of linear algebra) of the adjacency matrix A_G of G. The coefficients of $P(G, X)$ are defined by

$$det(X \cdot \mathbf{1} - A_G) = \sum_{i=0}^{n} c_i(G) \cdot X^i.$$

It is well known that $n = |V(G)|$, $-c_2(G) = |E(G)|$, and $-c_3(G)$ equals twice the number of triangles of G. The second largest zero $\lambda_2(G)$ of $P(G; X)$ gives a lower bound to the conductivity of G, cf. [GR01].

The Tutte polynomial of G is defined as

$$T(G; X, Y) = \sum_{F \subseteq E(G)} (X - 1)^{r\langle E\rangle - r\langle F\rangle}(Y - 1)^{n\langle F\rangle} \tag{3}$$

where $k\langle F\rangle$ is the number of connected components of the spanning subgraph defined by F, $r\langle F\rangle = |V| - k\langle F\rangle$ is its rank and $n\langle F\rangle = |F| - |V| + k\langle F\rangle$ is its nullity.

The Tutte polynomial $T(G; X, Y)$ has remarkable evaluations which count certain configurations of the graph G, cf. [Wel93].

(i) $T(G; 1, 1)$ is the number of spanning trees of G,
(ii) $T(G; 1, 2)$ is the number of connected spanning subgraphs of G,
(iii) $T(G; 2, 1)$ is the number of spanning forest of G,
(iv) $T(G; 2, 2) = 2^{|E|}$ is the number of spanning subgraphs of G,
(v) For connected graphs, $T(G; 1 - k, 0)$ is the number of proper k-vertex colorings of G,
(vi) For connected graphs, $T(G; 2, 0)$ is the number of acyclic orientations of G,
(vii) $T(G; 0, -2)$ is the number of Eulerian orientations of G.

All these are also graph parameters which take values in $\mathbb{N}$. More sophisticated evaluations of the Tutte polynomial can be found in [Goo06,Goo08].

For now it suffices to know that the Tutte polynomial, the matching polynomial, the characteristic polynomial, all discussed in [GR01,Mak07], and the interlace polynomial, defined in [ABS04,Cou], and virtually all the prominent graph polynomials in the literature, are *MSOL*-definable τ_3-polynomials, independently of the order of the order $<_E$. Furthermore, the operations $\sqcup_k$ are all *MSOL*-smooth on τ_3-structures for order-invariant sentences.

The following is a consequence of the Finite Rank Theorem (Theorem 9):

Theorem 12 ([GKM08]) *Let f be a τ-invariant and $\Box$ be an MSOL-compatible operation on τ-structures. If $r(f,\Box)$ is infinite, then f is not an evaluation of an MSOL-definable τ-polynomial.*

In [GKM08] many examples are given for graph parameters. This includes all $\sqcup$-maximizing (minimizing) graph parameters, such as the clique number $\omega(G)$, the chromatic number $\chi(G)$, but also the average degree of a graph.

4.3 More Graph Polynomials which are not *MSOL*-definable

Harmonious and complete colorings. Complete colorings, also called achromatic colorings, were introduced in [HHP67]. Harmonious colorings were introduced in [HK83]. For surveys, cf. [Edw97,HM97].

Definition 1 *(i) A proper vertex coloring is* harmonious*, if each pair of colors appears at most once along an edge. We denote by $\chi_{harm}(G)$ the least k such that G has a harmonious proper k-coloring.*

(ii) A proper vertex coloring is complete*, if each pair of colors appears at least once along an edge. We denote by $\chi_{comp}(G)$ the largest k such that G has a complete proper k-coloring.*

(iii) Let $\chi_{harm}(G;k)$ and $\chi_{comp}(G;k)$ denote the number of harmonious, respectively complete proper k-colorings of G.

Proposition 13 *(i) $\chi_{harm}(G;k)$ is a polynomial in k.*

(ii) $\chi_{comp}(G;k)$ is not a polynomial in k.

Proof. (i) follows from [MZ06], but it is not difficult to prove it directly.
(ii) $\chi_{comp}(G;k) = 0$ for large enough k. □

Theorem 14 *$\chi_{harm}(G)$ and $\chi_{comp}(G)$ are graph parameters which are not evaluations of invariantly CMSOL-definable graph polynomials.*

Proof. $\chi_{comp}(G)$ is maximizing, so we can apply Proposition 2.

For $\chi_{harm}(G)$ we observe that, for stars S_n, a set of n edges which meet all in one single vertex, we have

$$\chi_{harm}(S_n \sqcup S_m) = \max\{\chi_{harm}(S_m), \chi_{harm}(S_n)\} + 1.$$

Now the argument proceeds like in the case a maximizing graph parameter.

Theorem 15 $\chi_{harm}(G;k)$ *is not an invariantly CMSOL-definable graph polynomial.*

Proof. Let L_i denote the graph which consists of i vertex disjoint edges. We look at $M(\chi_{harm}(G,k),0)$ restricted to the graphs $L_i, i \in \mathbb{N}$, which we denote by $M_L(k)$ and its rank by $r_L(k)$. We note that $\chi_{harm}(L_i \sqcup L_j) = 0$ iff $i+j > \binom{k}{2}$. Therefore, $r_L(k) = \binom{k}{2}$ which is not bounded, contradicting Theorem 9.

Remark 1 *It is shown in [EM95], that computing* $\chi_{harm}(G)$ *is* **NP***-complete already for trees. This, together with the fact, proven in [Mak05], that evaluations of invariantly CMSOL-definable graph polynomials are polynomial time for graphs of tree-width at most* k*, shows that* $\chi_{harm}(G;X)$ *is not invariantly CMSOL-definable, unless* $\mathbf{P} = \mathbf{NP}$*. Our proof above eliminates the complexity theoretic hypothesis* $\mathbf{P} = \mathbf{NP}$*.*

Convex colorings. A vertex coloring of a graph $G = (V,E)$ with k colors ($k \in \mathbb{N}$) is a function $f : V \to [k]$. f is *convex* if for every $i \in [k]$ the colorclass $f^{-1}(i)$ induces a connected subgraph. For a partial function $f_0 : V \to [k]$ we say that f_0 is convex if there a is a total function f extending f_0 which is convex. In this case we also say that f is a *convex extension* of f_0. Convex extensions of partial colorings of trees have been introduced in the context of phylogenetic trees by S. Moran and S. Snir [MS07].

The existence problem of convex colorings for an arbitrary graph G is easily solved by trying to color every connected component by one color, and only depends on the number of colors available and the number of connected components of G. It follows from [MZ06,KMZ08] that the number of convex colorings of a graph G is a polynomial in k, which we denote by $conv(G,k)$. For $k = 1$ we have $conv(G,1) = 1$, if G is connected, and $conv(G,1) = 0$ otherwise. It has been shown by S. Noble and A. Goodall[2] that computing $conv(G,2)$ is $\sharp\mathbf{P}$-hard. It follows, using a similar argument as in [Lin86], that computing $conv(G,k)$ is $\sharp\mathbf{P}$-hard for every $k \in \mathbb{N} - \{0,1\}$. On the graphs E_n convex colorings have to color every vertex with a different color. It follows again that $r(conv(G,k),\sqcup)$ tends to infinity with k, and we get

Proposition 16 *The graph polynomial* $conv(G,k)$ *is not MSOL-definable.*

5 Open Problems

We have discussed various aspects of connection matrices of graph parameters introduced in [FLS07], and have generalized them for τ-invariants and various binary operations between τ-structures. We have shown that the rank of connection matrices is finite for *MSOL*-definable τ-invariants and *MSOL*-compatible binary operations between τ-structures. We used this to show that various graph parameters and graph polynomials are not *MSOL*-definable.

[2] Personal communication

In the case of partition functions knowing the exact rank $r(f, \sqcup_k)$ allows us to compute f on graphs of tree-width at most k in polynomial time with improved constants on the running time. Can this be generalized?

This leads us the the following questions about τ-invariants in general, although we formulate them for graphs..

Open Problem 1 *Assume $M(f, \Box)$ has rank r and an $(r \times r)$-submatrix M_r of maximal rank is given. Under what conditions on $\Box$ can we compute all the entries of $M(f, \Box)$ from M_r and the computability of $\Box$? What is the complexity of computing the entry $f(G_i \Box G_j)$ of $M(f, \Box)$?*

Open Problem 2 *Under what conditions on the graph parameter f and on $\Box$ can we compute the rank $r(f, \Box)$ precisely, or at least give reasonable lower and upper bounds?*

Open Problem 3 *Let f be a graph parameter on k-graphs and let $r(f, \sqcup_j)$ be finite for every $j \leq k$. Is it true that f can be computed in polynomial time on graphs of tree-width at most k.*

Open Problem 4 *In case the Open Problem 3 has a positive answer, is there an analogue for clique-width?*

I am pretty convinced that the answer are positive. In order to attack the Open Problems above it may be useful look at connection matrices restricted to a graph property Φ and an operation $\Box$ such that

(i) $\Box$ preserves Φ, i.e., if $G_1 \in \Phi$ and $G_2 \in \Phi$ then also $G_1 \Box G_2 \in \Phi$, and
(ii) the size of $G_1 \Box G_2$ bigger than the size of G_1 and G_2, for example

$$|V(G_1 \Box G_2)| \geq |V(G_1)| + |V(G_2)|$$

(iii) For every graph $G \in \Phi$ we can effectively find non-trivial G_1 and G_2 such that $G = G_1 \Box G_2$.

Examples for Φ and $\Box$ satisfying these conditions are trees with root a with an additional distinguished node b and $\sqcup_1$ identifying a from one tree with b from the other. Another example are the cliques with the join operation $\bowtie$, or graphs with no edges and the disjoint union $\sqcup$.

Acknowledgments

I wish to thank my co-authors B. Godlin and T. Kotek for their contributions in [GKM08], to T. Kotek who allowed me to include her Proposition 10 from his thesis work, and to L. Lovász and A. Schrijver for their interest and encouragement. Finally I would like to thank R. Ramanujam and S. Sarukkai for inviting me to the 3rd Indian Conference on Logic and its Applications (ICLA) and their insistence that I write up the outline of my invited lecture.

References

[ABS04] R. Arratia, B. Bollobás, and G.B. Sorkin. The interlace polynomial of a graph. *Journal of Combinatorial Theory, Series B*, 92:199–233, 2004.

[CMR01] 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.

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

[Cou] B. Courcelle. A multivariate interlace polynomial. Preprint, December 2006.

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

[dR84] M. de Rougemont. Uniform definability on finite structures with successor. *SIGACT'84*, pages 409–417, 1984.

[Edw97] 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.

[EM95] K. Edwards and C. McDiarmid. The complexity of harmonious colouring for trees. *Discrete Appl. Math.*, 57(2-3):133–144, 1995.

[FLS07] M. Freedman, László Lovász, and A. Schrijver. Reflection positivity, rank connectivity, and homomorphisms of graphs. *Journal of AMS*, 20:37–51, 2007.

[GKM08] B. Godlin, T. Kotek, and J.A. Makowsky. Evaluation of graph polynomials. In *WG'08*, volume 5xxx of *Lecture Notes in Computer Science*, pages xx–yy, 2008.

[Goo06] A. J. Goodall. Some new evaluations of the Tutte polynomial. *Journal of Combinatorial Theory, Series B*, 96:207–224, 2006.

[Goo08] A. J. Goodall. Parity, eulerian subgraphs and the Tutte polynomial. *Journal of Combinatorial Theory, Series B*, 98.3:599–628, 2008.

[GR01] C. Godsil and G. Royle. *Algebraic Graph Theory*. Graduate Texts in Mathematics. Springer, 2001.

[HHP67] F. Harary, S. Hedetniemi, and G. Prins. An interpolation theorem for graphical homomorphisms. *Portugal. Math.*, 26:453–462, 1967.

[HK83] J.E. Hopcroft and M.S. Krishnamoorthy. On the harmonious coloring of graphs. *SIAM J. Algebraic Discrete Methods*, 4:306–311, 1983.

[HM97] F. Hughes and G. MacGillivray. The achromatic number of graphs: a survey and some new results. *Bull. Inst. Combin. Appl.*, 19:27–56, 1997.

[KMZ08] T. Kotek, J.A. Makowsky, and B. Zilber. On counting generalized colorings. In *CSL'08*, volume 5213 of *Lecture Notes in Computer Science*, pages xx–yy, 2008.

[Lin86] M. Linial. Hard enumeration problems in geometry and combinatorics. *SIAM Journal of Algebraic and Discrete Methods*, 7:331–335, 1986.

[LMST07] N. Linial, J. Matousek, O. Sheffet, and G. Tardos. Graph coloring with no large monochromatic components. arXiv:math/0703362, 2007.

[Lov07] L. Lovász. Connection matrics. In G. Grimmet and C. McDiarmid, editors, *Combinatorics, Complexity and Chance, A Tribute to Dominic Welsh*, pages 179–190. Oxford University Press, 2007.

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

[Mak05] J.A. Makowsky. Colored Tutte polynomials and Kauffman brackets on graphs of bounded tree width. *Disc. Appl. Math.*, 145(2):276–290, 2005.

[Mak07] J.A. Makowsky. From a zoo to a zoology: Towards a general theory of graph polynomials. *Theory of Computing Systems*, online first:http://dx.doi.org/10.1017/s00224–007–9022–9, July 2007.

[MS07] S. Moran and S. Snir. Efficient approximation of convex recolorings. *Journal of Computer and System Sciences*, 73.7:1078–1089, 2007.

[MZ06] 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.

[Oum05] S. Oum. Approximating rank-width and clique-width quickly. In *Graph Theoretic Concepts in Computer Science, WG 2005*, volume 3787 of *Lecture Notes in Computer Science*, pages 49–58, 2005.

[Sch08] A. Schrijver. Polynomial and tensor invariants and combinatorial parameters. Preprint: Amsterdam, 2008.

[Sze07] B. Szegedy. Edge coloring models and reflection positivity. Available at: arXiv: math.CO/0505035, 2007.

[Wel93] D.J.A. Welsh. *Complexity: Knots, Colourings and Counting*, volume 186 of *London Mathematical Society Lecture Notes Series*. Cambridge University Press, 1993.