The Specker-Blatter theorem revisited: generating functions for definable classes of structures

The Specker-Blatter Theorem Revisited: Generating Functions for Definable Classes of Structures

E. Fischer and J.A. Makowsky

Faculty of Computer Science
Technion - Israel Institute of Technology
Haifa, Israel

Subject Classification:
Combinatorics related algorithms and complexity, Generating functions,
Automata, logic and computability, Graph theory,

Abstract.
In this paper we study the generating function of classes of graphs and hypergraphs. For a class of labeled graphs $\mathcal{C}$ we denote by $f_{\mathcal{C}}(n)$ the number of structures of size n. For $\mathcal{C}$ definable in Monadic Second Order Logic with unary and binary relation symbols only, E. Specker and C. Blatter showed in 1981 that for every $m \in \mathbb{N}$, $f_{\mathcal{C}}(n)$ satisfies a linear recurrence relation $f_{\mathcal{C}}(n) = \sum_{j=1}^{d_m} a_j^{(m)} f_{\mathcal{C}}(n-j)$, over $\mathbb{Z}_m$, and hence is ultimately periodic for each m. To show this they introduced what we call the Specker-index of $\mathcal{C}$ and first showed the theorem to hold for any $\mathcal{C}$ of finite Specker-index, and then showed that every $\mathcal{C}$ definable in Monadic Second Order Logic is indeed of finite Specker-index. E. Fischer showed in 2002 that the Specker-Blatter Theorem does not hold for quaternary relations.
In this paper we show how the Specker-Blatter Theorem is related to Schützenberger's Theorem and the Myhill-Nerode criterion for the characterization of regular languages, and discuss in detail how the behavior of this generating function depends on the choice of constant and relation symbols allowed in the definition of $\mathcal{C}$. Among the main results we have the following:

- We consider n-ary relations of degree at most d, where each element a is related to at most d other elements by any of the relations. We show that the Specker-Blatter Theorem holds for those, irrespective of the arity of the relations involved.
- Every $\mathcal{C}$ definable in Monadic Second Order Logic with (modular) Counting (CMSOL) is of finite Specker-index. This covers many new cases, for which such a recurrence relation was not known before.
- There are continuum many $\mathcal{C}$ of finite Specker-index. Hence, contrary to the Myhill-Nerode characterization of regular languages, the recognizable classes of graphs cannot be characterized by the finiteness of the Specker-index.

1 Introduction and main results

Counting objects of a specified kind belongs to the oldest activities in mathematics. In particular, counting the number of (labeled or unlabeled) graphs satisfying a given property is a classic undertaking in combinatorial theory. The first deep results for counting unlabeled graphs are due to J.H. Redfield (1927) and to G. Polya (1937), but were only popularized after 1960. F. Harary, E.M. Palmer and R.C. Read unified these early results, as witnessed in the still enjoyable [HP73].

It is unfortunate that a remarkable theorem by E. Specker and C. Blatter on counting labeled graphs (and more generally, labeled binary relational structures), first announced in 1981, cf. [BS81,BS84,Spe88], has not found the attention it deserves, both for the beauty of the result and the ingenuity in its proof.

E. Specker and C. Blatter look at the function $f_{\mathcal{C}}(n)$ which counts the number of labeled relational structures of size n with k relations $R_1, \dots, R_k$, which belong to a class $\mathcal{C}$. We shall call this function the *density function for $\mathcal{C}$*. It is required that $\mathcal{C}$ be definable in Monadic Second Order Logic and that the relations are all unary or binary relations. The theorem says that under these hypotheses the function $f_{\mathcal{C}}(n)$ satisfies a linear recurrence relation modulo m for every $m \in \mathbb{Z}$. Special cases of this theorem have been studied extensively, cf. [HP73,Ges84,Wil90] and the references therein. However, the possibility of using a formal logical classification as a means to collect many special cases seems to have mostly escaped notice in this case.

In the present paper, we shall discuss both the Specker-Blatter theorem, and its variations and limits of generalizabilty. We first set up our framework of logic. For the reader not familiar with logic, we recommend consulting [EF95]. We also give numerous examples in Appendix E, which in turn provide combinatorial corollaries to the Specker-Blatter Theorem. Proving directly the linear recurrence relations over every modulus m for all the given examples would have been a nearly impossible undertaking. We should also note that counting structures up to isomorphism is a very different task, cf. [HP73]. From Proposition 11 below one can easily deduce that the Specker-Blatter Theorem does not hold in this setting.

1.1 Counting labeled structures

Let $\bar{R} = \{R_1, \ldots, R_\ell\}$ be a set of relation symbols where each R_i is of arity $\rho(i)$. Let $\mathcal{C}$ be a class of relational $\bar{R}$-structures. For an $\bar{R}$-structure $\mathfrak{A}$ with universe A we denote the interpretation of R_i by $R_i(A)$. We denote by $f_{\mathcal{C}}(n)$ the number of structures in $\mathcal{C}$ over the labeled set $A_n = \{1, \ldots, n\}$, i.e.,

$$f_{\mathcal{C}}(n) =\mid \{(R_1(A_n), \ldots, R_\ell(A_n)) : \langle A_n, R_1(A_n), \ldots, R_\ell(A_n)\rangle \in \mathcal{C}\} \mid .$$

The notion of $\bar{R}$-isomorphism is the expected one: Two structures $\mathfrak{A}, \mathfrak{B}$ are isomorphic, if there is a bijection between their respective universes which preserves relations in both directions.

> **Proviso:** When we speak of a class of structures $\mathcal{C}$, we always assume that $\mathcal{C}$ is closed under $\bar{R}$-isomorphisms. However, we count two isomorphic but differently labeled structures as two different members of $\mathcal{C}$.

1.2 Logical formalisms

First Order Logic $FOL(\bar{R})$, Monadic Second Order Logic $MSOL(\bar{R})$, and Counting Monadic Second Order Logic $CMSOL(\bar{R})$ are defined as usual, cf. [EF95]. A class of $\bar{R}$-structures $\mathcal{C}$ is is called *$FOL(\bar{R})$-definable* if there exists an $FOL(\bar{R})$ formula ϕ with no free (non-quantified) variables such that $\mathfrak{A} \in \mathcal{C}$ if and only if $\mathfrak{A} \models \phi$ for every $\mathfrak{A}$. Definability for $MSOL(\bar{R})$ and $CMSOL(\bar{R})$ is defined analoguously.

We shall also look at two variations[1] of $CMSOL(\bar{R})$, and analogously for FOL and $MSOL$. The first variation is denoted by $CMSOL_{lab}(\bar{R})$, where the set of relation symbols is extended by an infinite set of constant symbols $c_i, i \in \mathbb{N}$. In a labeled structure over $\{1, \ldots, n\}$ the constant $c_i, i \leq n$ is interpreted as i. If $\phi \in MSOL_{lab}(\bar{R})$ and c_k is the constant occurring in ϕ with largest index, then the universe of a model of ϕ has to contain the set $\{1, \ldots, k\}$.

The second variation is denoted by $CMSOL_{ord}(\bar{R})$, where the set of relation symbols is augmented by a binary relation symbol $R_<$ which is interpreted on $\{1, \ldots, n\}$ as the natural order $1 < 2 < \cdots < n$.

[1] In [Cou90] another version, $MSOL_2$ is considered, where one allows also quantification over sets of edges. The Specker-Blatter Theorem does not hold in this case, as the class $CBIPEQ$ of complete bipartite graphs $K_{n,n}$ is definable in $MSOL_2$ and $f_{CBIPEQ}(2n) = \frac{1}{2}\binom{2n}{n}$.

Example 1 *Let $\bar{R}$ consist of one binary relation symbol R.*

1. *$\mathcal{C} = ORD$, the class of all linear orders, satisfies $f_{ORD}(n) = n!$. ORD is $FOL(R)$-definable.*
2. *In FOL_{lab} we can look at the above property and additionally require by a formula ϕ_k that the elements $1, \ldots, k \in [n]$ indeed occupy the first k positions of the order defined by R, preserving their natural order. It is easily seen that $f_{ORD \wedge \phi_k}(n) = (n-k)!$. In FOL_{ord} we can express even more stringent compatibilities of the order with the natural order of $\{1, \ldots, n\}$.*
3. *For $\mathcal{C} = GRAPHS$, the class of simple graphs (without loops or multiple edges), $f_{GRAPHS}(n) = 2^{\binom{n}{2}}$. $GRAPHS$ is $FOL(R)$-definable.*
4. *The class REG_r of simple regular graphs where every vertex has degree r is FOL-definable (for any fixed r). Details are given in Appendix E.2.*
5. *The class $CONN$ of all connected graphs is not $FOL(R)$-definable, but it is $MSOL(R)$-definable using a universal quantifier over set variables. Counting labeled connected graphs is treated in [HP73, Chapter 1] and in [Wil90, Chapter 3]. Details are given in Appendix E.1.*
6. *Let $\mathcal{C} = BIPEQ$ be the class of simple bipartite graphs with m elements on each side (hence $n = 2m$). $BIPEQ$ is not $CMSOL(R)$-definable. However, the class BIP of bipartite graphs with unspecified number of vertices on each side is $MSOL$-definable. Again this is treated in [HP73, Chapter 1]. Details are given in Appendix E.5.*
7. *Let $\mathcal{C} = EVENDEG$ be the class of simple graphs where each vertex has an even degree. $EVENDEG$ is not $MSOL$-definable, but it is $CMSOL$-definable. $f_{EVENDEG}(n) = 2^{\binom{n-1}{2}}$, cf. [HP73, page 11].*

   Let $\mathcal{C} = EULER$ be the class of simple connected graphs in $EVENDEG$. $EULER$ is not $MSOL$-definable, but it is $CMSOL$-definable. In [HP73, page 7] a recurrence formula for the number of labeled eulerian graphs is given. Details are given in Appendix E.7.
8. *Let $\mathcal{C} = EQCLIQUE$ be the class of simple graphs which consist of two disjoint cliques of the same size. Then we have $f_{EQCLIQUE}(2n) = \frac{1}{2}\binom{2n}{n}$ and $f_{EQCLIQUE}(2n+1) = 0$. $EQCLIQUE$ is not even $CMSOL(R)$-definable, but it is definable in Second Order Logic SOL, when we allow quantification also over binary relations.*

   We can modify $\mathcal{C} = EQCLIQUE$ by adding another binary relation symbol R_1 and expressing in $FOL(R_1)$ that R_1 is a bijection between the two cliques. We denote the resulting class of structures by $\mathcal{C} = EQCLIQUE_1$. $f_{EQCLIQUE_1}(2n) = n!\frac{1}{2}\binom{2n}{n}$ and $f_{EQCLIQUE_1}(2n+1) = 0$.

   A further modification is $\mathcal{C} = EQCLIQUE_2$, which is $FOL_{ord}(R, R_1)$-definable. We require additionally that the bijection R_1 is such that the first elements (in the order $R_<$) of the cliques are matched, and if $(v_1, v_2) \in R_1$ then the $R_<$- successors $(suc(v_1), suc(v_2)) \in R_1$. This makes the matching unique (if it exists), and we have $f_{EQCLIQUE}(n) = f_{EQCLIQUE_2}(n)$. Similarly, we can look at $EQ_mCLIQUE$, $EQ_mCLIQUE_1$ and $EQ_mCLIQUE_2$ respectively, where we require m equal size cliques instead of two. Here we also have $f_{EQ_mCLIQUE}(n) = f_{EQ_mCLIQUE_2}(n)$.

The non-definability statements are all relatively easy, using Ehrenfeucht-Fraïssé Games, cf. [EF95].

1.3 The Specker-Blatter Theorem

The following remarkable theorem due to E. Specker and C. Blatter was announced in [BS81], and proven in [BS84,Spe88]:

Theorem 2. *For any $\mathcal{C}$ definable in Monadic Second Order Logic with unary and binary relation symbols only, the function $f_{\mathcal{C}}$ satisfies a linear recurrence relation $f_{\mathcal{C}}(n) \equiv \sum_{j=1}^{d_m} a_j^{(m)} f_{\mathcal{C}}(n-j) \pmod{m}$, for every $m \in \mathbb{N}$, and hence is ultimately periodic for each m.*

The case of ternary relation symbols, and more generally of arity $k \geq 3$, was left open in [BS84,Spe88]. The question as to whether Theorem 2 holds for these appears, together with other questions concerning this theorem, in the list of open problems in Finite Model Theory, [Mak00, Problem 3.5]. Counterexamples for quaternary relations were first found by E. Fischer, cf. [Fis02].

Theorem 3. *For every prime p there exists a class of structures $\mathcal{C}_p$ which is definable in first order logic by a formula ϕ_{Im_p}, with one binary relation symbol E and one quaternary relation symbol R, such that $f_{\mathcal{C}_p}$ is not ultimately periodic modulo p.*

From this theorem the existence of such classes are easily deduced also for every non-prime number m (just take p to be a prime divisor of m). The proof of the theorem is based on the class $EQ_pCLIQUE$ from Example 8 above, and, for completeness, is outlined in Appendix B.

1.4 Improvements and variations

The purpose of this paper is to explore variations and extensions of the Specker-Blatter Theorem and its relationship to Schützenberger's characterization of regular languages.

First, we study the case of unary relations symbols. We shall see in Section 3 that for unary relations Theorems 2 and 5 can be strengthened using Schützenberger's approach to regular languages.

Theorem 4. *For any $\mathcal{C}$ definable in Counting Monadic Second Order Logic with an order, $CMSOL_{ord}(\bar{R})$, where $\bar{R}$ contains only unary relations, the function $f_{\mathcal{C}}$ satisfies a linear recurrence relation $f_{\mathcal{C}}(n) = \sum_{j=1}^{d} a_j f_{\mathcal{C}}(n-j)$ over the integers $\mathbb{Z}$, and in particular satisfies the same relation for every modulus m.*

Next we extend the Specker-Blatter Theorem to allow $CMSOL$, rather then $MSOL$.

Theorem 5. *For any $\mathcal{C}$ definable in Counting Monadic Second Order Logic ($CMSOL$) with unary and binary relation symbols only, the function $f_{\mathcal{C}}$ satisfies a linear recurrence relation $f_{\mathcal{C}}(n) \equiv \sum_{j=1}^{d_m} a_j^{(m)} f_{\mathcal{C}}(n-j) \pmod{m}$, for every $m \in \mathbb{N}$.*

The proof is given in Appendix D.

Theorem 5 covers cases not covered by the Specker-Blatter Theorem (Theorem 2). Although $EVEN$ is not $MSOL$-definable, it is $CMSOL$-definable, and its function satisfies $f_{EVEN}(n+1) = f_{GRAPHS}(n)$. However, it seems not very obvious that the function f_{EULER} satisfies modular recurrence relations. Many more examples are discussed in Appendix E, especially in E.7.

Finally, we study the case of relations of bounded degree. For any element $a \in A$, we define the *degree* of a to be the number of elements $b \neq a$ for which there exists a relation $R \in \bar{R}$ and some $\bar{a} \in R(A)$ such that both a and b appear in $\bar{a}$ (possibly with other members of A as well). We say that R is of *bounded degree d* if every $a \in A$ has degree at most d. We say that an $\bar{R}$-structure is *connected*, if for any $A' \subsetneq A$ there is a relation $R(A)$ with $R \in \bar{R}$ and some $\bar{a} \in R(A)$ containing both an element from A' and an element from $A - A'$. We say that a function $f(n)$ satisfies a *trivial modular recurrence* if there exist functions $g(n), h(n)$ with $g(n)$ tending to infinity such that $f(n) = g(n)! \cdot h(n)$; this is equivalent to saying that for every m there exists N_m such that if $n > N_m$ then $g(n) \equiv 0 \pmod{m}$. Clearly, $f_{EQCLIQUE_1}(n)$ satisfies a trivial modular recurrence.

For bounded degree models we prove the following.

Theorem 6. *For any $\mathcal{C}$ definable in Counting Monadic Second Order Logic $CMSOL$, with all relations in all members of $\mathcal{C}$ being of bounded degree d, the function $f_{\mathcal{C}}$ satisfies a linear recurrence relation $f_{\mathcal{C}}(n) \equiv \sum_{j=1}^{d_m} a_j^{(m)} f_{\mathcal{C}}(n-j) \pmod{m}$, for every $m \in \mathbb{N}$. Furthermore, if all the models in $\mathcal{C}$ are connected, then $f_{\mathcal{C}} = 0 \pmod{m}$ for $m \in \mathbb{N}$ large enough.*

The proof is sketched in Section 4 and completed in Appendix C.

2 Variations and counterexamples

2.1 Why modular recurrence?

Theorem 2 provides linear recurrence relations modulo m for every $m \in \mathbb{N}$. Theorem 4 provides a uniform linear recurrence relation over $\mathbb{Z}$. We show that even for $\phi \in FOL(R)$ with one binary relation symbol only, A uniform linear recurrence over $\mathbb{Z}$ does not hold. We begin with the following well known lemma, cf. [LN83].

Lemma 7 *Let $f : \mathbb{Z} \to \mathbb{Z}$ be a function which satisfies a linear recurrence relation $f(n+1) = \sum_{i=0}^{k} a_i f(n-i)$ over $\mathbb{Z}$. Then there is a constant $c \in \mathbb{Z}$ such that $f(n) \leq 2^{cn}$.*

Hence, for the following $\mathcal{C}$, $f_{\mathcal{C}}(n)$ does not satisfy a linear recurrence over $\mathbb{Z}$: The class of all binary relations over any finite set, for which $f_{\mathcal{C}}(n) = 2^{n^2}$, and the class of all linear orders over any finite set, for which $f_{\mathcal{C}}(n) = n!$.

2.2 Trivial recurrence relations

We say that a function $f(n)$ satisfies a *trivial modular recurrence* if there are functions $g(n), h(n)$ with $g(n)$ tending to infinity such that $f(n) = g(n)! \cdot h(n)$. We call this a trivial recurrence, because it is equivalent to the statement that for every $m \in \mathbb{N}$ and large enough n, $f(n) \equiv 0 \pmod{m}$. The most obvious example is the number of labeled linear orderings, given by $f_{ord}(n) = n!$ and $g(n) = f(n)$. Clearly, also $f_{EQCLIQUE_1}(n)$ satisfies a trivial modular recurrence. For the class of all graphs the recurrences are non-trivial. More generally, for a set of relation symbols $\bar{R}$ with k_j many j-ary relation symbols, the set of all labeled structures on n elements is given by $f_{\bar{R}}(n) = 2^{\sum_j k_j n^j}$ which is only divisible by 2. It follows immediately that

Observation 8 *If $\mathcal{C}$ is a class of $\bar{R}$-structures, and $\bar{\mathcal{C}}$ its complement, then at least one of $f_{\mathcal{C}}(n)$ or $f_{\bar{\mathcal{C}}}(n)$ does not satisfy the trivial modular recurrence relations.*

2.3 Existential second order logic is too strong

In the following, we let p be a prime number, and state some lemmas and definitions; in particular, we show that $EQ_pCLIQUE$ is a graph property for which the number of models is not periodic modulo p. It is not $CMSOL$ definable, but in Appendix B we construct first order properties that are related to it.

We denote by $b_p(n) = f_{EQ_pCLIQUE}(n) = f_{EQ_pCLIQUE_2}(n)$ the number of graphs with $[n]$ as a set of vertices which are disjoint unions of exactly p same-size cliques, that is, $b_p(n) = f_{EQC_p}(n)$. As an example for $p = 2$, note that $b_2(2k+1) = 0$ and $b_2(2k) = \frac{1}{2}\binom{k}{2}$ for every k.

Proposition 9 *For every n which is not a power of p, we have $b_p(n) \equiv 0 \pmod{p}$, and for every n which is a power of p we have $b_p(n) \equiv 1 \pmod{p}$. In particular, $b_p(n)$ is not ultimately periodic modulo p.*

The proof is given in Appendix A.

The example $EQCLIQUE$ is definable in Second Order Logic with existential quantification over one binary relation. But $b_p(n) = f_{EQ_pCLIQUE}(n) = f_{EQ_pCLIQUE_2}(n)$ is not periodic modulo p. Hence we obtain, using Proposition 9, the following.

Proposition 10 *$EQ_pCLIQUE$ is definable in Existential Second Order Logic but $f_{EQ_pCLIQUE}$ is not periodic modulo p, and hence does not satisfy a linear recurrence relation modulo p.*

2.4 Using the labels

Labeled structures have additional structure which can not be exploited in defining classes of models in $CMSOL(\bar{R})$. The additional structure consists of the labels. We can import them into our language as additional constants (with fixed interpretation) as in $CMSOL_{lab}(\bar{R})$ or, assuming the labels are linearly ordered, as a linear order with a fixed interpretation, as in $CMSOL_{ord}(\bar{R})$. Theorem 4 states that, when we restrict $\bar{R}$ to unary predicates, adding the linear order still gives us even a uniform recurrence relation. There are $\phi \in FOL_{ord}(R)$ with binary relation symbols only, such that even the non-uniform linear recurrences over $\mathbb{Z}_p$ do not hold. Here we use $EQ_pCLIQUE_2$ from Example 8, with Proposition 9.

Proposition 11 *$EQ_pCLIQUE_2$ is FOL_{ord}-definable, using the order. However $f_{EQ_pCLIQUE}$ is not ultimately periodic modulo p. Therefore $f_{EQ_pCLIQUE_2}$ does not satisfy a linear recurrence relation modulo p.*

In fact, it is not too hard to formulate in FOL_{ord} a property with one binary relation symbol that has the same density function as $EQ_pCLIQUE$.

On the other hand, using the labels as constants does not change the situation, Theorem 5 also holds for $CMSOL_{lab}$. This is proven using standard reduction techniques, and the proof is omitted.

Proposition 12 *For $\phi \in CMSOL_{lab}(\bar{R})$ (resp. $MSOL_{lab}(\bar{R})$, $FOL_{lab}(\bar{R})$), where the arities of the relation symbols in $\bar{R}$ are bounded by r and there are k labels used in ϕ, there exists $\psi \in MSOL(\bar{S})$ (resp. $MSOL(\bar{S})$, $FOL(\bar{S})$) for suitable $\bar{S}$ with the arities of $\bar{S}$ bounded by r such that $f_\phi(n) = f_\psi(n-k)$*

We finally note that in the presence of a fixed order, the modular counting quantifiers are definable in $MSOL_{ord}$. They are, however, not definable in FOL_{ord}. This was already observed in [Cou90].

Proposition 13 *For every $\phi \in CMSOL_{ord}(\bar{R})$ there is an equivalent $\psi \in MSOL_{ord}(\bar{R})$.*

3 Generating functions for formal languages

The Specker-Blatter Theorem has an important precursor in formal language theory: Schützenberger's Theorem characterizing regular languages in terms of the properties of the power series of their generating function. The property in question is $\mathbb{N}$-rationality, which implies rationality. For details the reader should consult [BR84] and for constructive versions [BDFR01].

3.1 Generating functions

We put Theorem 2 into a more general context and study the (ordinary) generating function $F_{\mathcal{C}}^m(X) = \sum_{n=0}^{\infty} f_{\mathcal{C}}^m(n) X^n$ Using [LN83, Theorem 8.40 in chapter 8], Theorem 5 (and hence Theorem 2) can now be rephrased as

Theorem 14 *Let $\mathcal{C}$ be definable in $CMSOL(\bar{R})$, where $\bar{R}$ consists of unary and binary relation symbols only. For every $m \in \mathbb{N}$, $F_{\mathcal{C}}^m(X) = \sum_{n=0}^{\infty} f_{\mathcal{C}}^m(n) X^n$ satisfies a linear recurrence relation over $\mathbb{Z}_m$, $f_{\mathcal{C}}^m(n+k) = \sum_{i=0}^{k-1} a_i^m f_{\mathcal{C}}^m(n-i)$, and hence it is rational with $F_{\mathcal{C}}^m(X) = \frac{G(X)}{H(X)}$, where $H(X) = 1 - \sum_{i=1}^{k} a_{k-i}^m X^i$ and $G(X) = \sum_{j=0}^{k-1} \left(f_{\mathcal{C}}^m(j) - \sum_{i=0}^{j-1} a_{k+i-j}^m f_{\mathcal{C}}^m(i) \right) X^j$.*

3.2 Regular languages

If we restrict $\bar{R}$ to consist only of unary relation symbols $\bar{R} = \bar{U} = U_1, U_2, \dots, U_k$, but allow a fixed linear order on the universe, then the corresponding structures can be viewed as words over an alphabet with 2^k letters. We assume that the reader is familiar with the basics of formal language theory, as given in [HU80,BR84].

From Proposition 13 we know that every $CMSOL_{ord}(\bar{U})$ formula is equivalent to an $MSOL_{ord}(\bar{U})$ formula. Combining this with the $MSOL$-characerization of regular languages we get

Theorem 15. *A language $\mathcal{C}$ is $CMSOL_{ord}$-definable if and only if it is regular.*

M.P. Schützenberger introduced generating functions into the study of formal languages, cf. [?]. In the light of Corollary 15, his theorem is equivalent to the following:

Theorem 16 *Let $\mathcal{C}$ be definable in $CMSOL_{ord}(\bar{U})$, where $\bar{U}$ consists of unary relation symbols only. Then $F_{\mathcal{C}}(X) = \sum_{n=0}^{\infty} f_{\mathcal{C}}(n)X^n$ is rational, and hence satisfies over $\mathbb{Z}$ a linear recurrence relation $f_{\mathcal{C}}(n+k) = \sum_{k=0}^{k-1} a_i^m f_{\mathcal{C}}(n-i)$.*

Theorem 4 is now proved.

4 Myhill-Nerode and Specker index

Specker's proof of Theorem 2 is based on the analysis of an equivalence relation induced by a class of structures $\mathcal{C}$. It is reminiscent of the Myhill-Nerode congruence relation for words, cf. [HU80], but generalized to graph grammars, and to general structures. Note however, that the Myhill-Nerode congruence is, strictly speaking, not a special case of the Specker equivalence. What one gets is the syntactic congruence relation for formal languages.

4.1 The Myhill-Nerode Theorem

Let $\mathcal{C}$ be a set of words over a fixed alphabet. We say that two words v, w over the same alphabet are $MN(\mathcal{C})$-equivalent iff for every word u the concatenations vu, wu satisfy $vu \in \mathcal{C}$ iff $wu \in \mathcal{C}$. This equivalence relation was introduced by J. Myhill and A. Nerode, cf. [HU80]. The *Myhill-Nerode index of $\mathcal{C}$* is the number of equivalence classes of $MN(\mathcal{C})$-equivalence.

Theorem 17 (Myhill and Nerode) *A language $\mathcal{C}$ is regular iff $\mathcal{C}$ has a finite Myhill-Nerode index.*

4.2 Substitution of structures

A pointed $\bar{R}$-structure is a pair $(\mathfrak{A}, a)$, with $\mathfrak{A}$ an $\bar{R}$-structure and a an element of the universe A of $\mathfrak{A}$. In $(\mathfrak{A}, a)$, we speak of the structure $\mathfrak{A}$ and the *context* a.

The terminology is borrowed from the terminology used in dealing with tree automata, cf. [GS97].

Given two pointed structures $(\mathfrak{A}, a)$ and $(\mathfrak{B}, b)$ we form a new pointed structure $(\mathfrak{C}, c) = Subst((\mathfrak{A}, a), (\mathfrak{B}, b))$ defined as follows:

- The universe of $\mathfrak{C}$ is $A \cup B - \{a\}$.
- The context c is given by b, i.e., $c = b$.

– For $R \in \bar{R}$ of arity r, R^C is defined by $R^C = (R^A \cap (A - \{a\})^r) \cup R^B \cup I$ where for every relation in R^A which contains a, I contains all possibilities for replacing these occurrences of a with a member of B.

We similarly define $Subst((\mathfrak{A}, a), \mathfrak{B})$ for a structure $\mathfrak{B}$ that is not pointed, in which case the resulting structure $\mathfrak{C}$ is also not pointed.

Let $\mathcal{C}$ be a class of, possibly pointed, $\bar{R}$-structures. We define an equivalence relation between $\bar{R}$-structures:

– We say that $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are equivalent, denoted $\mathfrak{A}_1 \sim_{Su(\mathcal{C})} \mathfrak{A}_2$, If for every pointed structure $(\mathfrak{S}, s)$ we have that $Subst((\mathfrak{S}, s), \mathfrak{A}_1) \in \mathcal{C}$ if and only if $Subst((\mathfrak{S}, s), \mathfrak{A}_2) \in \mathcal{C}$.
– The *Specker index* of $\mathcal{C}$ is the number of equivalence classes of $\sim_{Su(\mathcal{C})}$.

Specker's proof in [Spe88] of Theorem 2 has a purely combinatorial part:

Lemma 18 (Specker's Lemma) *Let $\mathcal{C}$ be a class of $\bar{R}$-structures of finite Specker index with all the relation symbols in $\bar{R}$ at most binary. Then $f_{\mathcal{C}}(n)$ satisfies modular linear recurrence relations for every $m \in \mathbb{N}$.*

4.3 Classes of finite Specker index

Proposition 19 *The class $EQ_2CLIQUE$ has an infinite Specker index.*

Proof. We show that for all $i, j \in \mathbb{N}, 1 \leq i \leq j$, the pairs of cliques $\langle C_i, C_j \rangle$ are inequivalent with respect to $\sim_{Su(EQ_2CLIQUE)}$. The key observation is that substituting a clique in a clique gives again a clique. Hence we can make $C_{i+j} \sqcup C_j$ into $C_{i+j} \sqcup C_{i+j}$ substituting a C_{j-i}. □

It is an easy exercise to show the same for the class of graphs which contain a hamiltonian cycle. Again, these graphs are not $CMSOL$-definable. So far, all the classes of infinite Specker index were not definable in $CMSOL$. This is no accident. Specker noted that all $MSOL$-definable classes of $\bar{R}$-structures (with all relations at most binary) have a finite Specker index. We shall see that this can be extended to $CMSOL$.

Theorem 20 *If $\mathcal{C}$ is a class of $\bar{R}$-structures (with no restrictions on the arity) which is $CMSOL$-definable, then $\mathcal{C}$ has a finite Specker index.*

The proof is given in Appendix D. It uses a form of the Feferman-Vaught Theorem for $CMSOL$ due to Courcelle, [Cou90].

Without logic, the underlying principle for establishing a finite Specker index of a class $\mathcal{C}$ is the following:

Definition 21 *Let $\mathcal{C}$ be a class of graphs and $\mathcal{F}$ be a binary operation on $\bar{R}$-structures which is isomorphism invariant. We say that $\mathfrak{A}_0$ and $\mathfrak{A}_1$ are $\mathcal{F}(\mathcal{C})$-equivalent if for every $\mathfrak{B}$, $\mathcal{F}(\mathfrak{A}_0, \mathfrak{B}) \in \mathcal{C}$ iff $\mathcal{F}(\mathfrak{A}_1, \mathfrak{B}) \in \mathcal{C}$.*
$\mathcal{C}$ has a finite $\mathcal{F}$-index if the number of $\mathcal{F}(\mathcal{C})$-equivalence classes is finite.

Proposition 22 *A class of $\bar{R}$-structures $\mathcal{C}$ has a finite $\mathcal{F}$-index iff there are $\alpha \in \mathbb{N}$ and classes of $\bar{R}$-structures $\mathcal{K}^i_j$ $(0 \leq j \leq \alpha, 0 \leq i \leq 1)$ such that $\mathcal{F}(\mathfrak{A}_0, \mathfrak{A}_1) \in \mathcal{C}$ iff there exists j such that $\mathfrak{A}_0 \in \mathcal{K}^0_j$ and $\mathfrak{A}_1 \in \mathcal{K}^1_j$.*

Proof. If $\mathcal{C}$ is of finite $\mathcal{F}$-index α then we can choose for $\mathcal{K}_j^0$ the equivalence classes and for each $j \leq \alpha$

$$\mathcal{K}_j^1 = \{\mathfrak{A} \in Str(\bar{R}) : \mathcal{F}(\mathfrak{A}', \mathfrak{A}) \in \mathcal{C} \text{ for } \mathfrak{A}' \in \mathcal{K}_j^0\}$$

Conversely, if the $\mathcal{K}_j^0$ are all disjoint, the pairs $(\mathfrak{A}, \mathfrak{A}')$ with $\mathfrak{A} \in \mathcal{K}_j^0, \mathfrak{A}' \in \mathcal{K}_j^0$ are all in the same equivalence class. But without loss of generality, but possibly increasing α, we can assume that the the $\mathcal{K}_j^0$ are all disjoint. □

Corollary 23 *If $\mathcal{C}_0, \mathcal{C}_1$ are classes of finite $\mathcal{F}$-index, so are all their boolean combinations.*

Proof. Take the coarsest common refinement of the $\mathcal{F}(\mathcal{C}_0)$-equivalence and the $\mathcal{F}(\mathcal{C}_1)$-equivalence relations. □

We also have

Corollary 24 *If $\mathcal{C}$ is a class of $\bar{R}$-structures such that $\mathcal{F}(\mathfrak{A}, \mathfrak{B}) \in \mathcal{C}$ iff both $\mathfrak{A}, \mathfrak{B} \in \mathcal{C}$ then the $\mathcal{F}(\mathcal{C})$-index of $\mathcal{C}$ is at most 2.*

4.4 A continuum of classes of finite index

As there are only countably many regular languages over a fixed alphabet, the Myhill-Nerode theorem implies that there are only countably many languages with finite MN-index. in contrast to this, for general relational structures there are plenty of classes of graphs which are of finite Specker index.

Definition 25 *Let C_n denote the cycle of size n, i.e. a regular connected graph of degree 2 with n vertices. Let $A \subseteq \mathbb{N}$ be any set of natural numbers and $Cycle(A) = \{C_n : n \in A\}$.*

Proposition 26 (Specker) *$Cycle(A)$ has Specker index at most 5.*

Proof. All binary structures with three or more vertices fall into two classes, the class of graphs G for which $Subst((\mathfrak{A}, a), G) \in \mathcal{C}$ if and only if $\mathfrak{A}$ has a single element a (this equals the class $Cycle(A)$), and the class of graphs G for which $Subst((\mathfrak{A}, a), G) \in \mathcal{C}$ never occurs (which contains all binary structures which are not graphs, and all graphs with at least three elements which are not in $Cycle(A)$). Binary structures with less than three vertices which are not graphs also fall into the second class above, while the three possible graphs with less then three vertices may form classes by themselves (depending on A). □

Corollary 27 (Specker) *There is a continuum of classes (of graphs, of $\bar{R}$-structures) of finite Specker index which are not $CMSOL$-definable.*

Proof. Clearly, there is a continuum of classes of the type $Cycle(A)$, and hence a continuum of classes that are not definable in $CMSOL$ (or even in second order logic, SOL).
Now $f_{Cycle(A)}(n) = 0$ if $n \notin A$ and $f_{Cycle(A)}(n) = (n-1)!$ otherwise. Hence it satisfies trivial recurrences. Using Observation 8 we know that the complement $\overline{Cycle}(A)$ does satisfy a non-trivial recurrence relation. □

This shows that, in contrast to the Myhill-Nerode Theorem, no characterization of the classes of finite Specker index in terms of their definability in $CMSOL$, or any other logic with countably many formulas, is possible. However, it makes sense to ask whether among the classes of graphs definable in, say, Second Order Logic, the classes of finite Specker index can be characterized.

But $CMSOL$ will not suffice, as one can easily find an A such that $Cycle(A)$ is not $CMSOL$-definable, but definable in Second Order Logic. A could be chosen as, e.g., the set of primes, or the set of squares. Candidates for such characterization could be classes of graphs generated by some graph grammars, possibly different from the usual HR-grammars (Hyperedge replacement grammars) and VR-grammars (Vertex replacement grammars, which can be characterized in terms of $MSOL$-transductions, cf. [Cou94].

Acknowledgments We are grateful to E. Specker, for his encouragement and interest in our work, and for his various suggestions and clarifications, which were incorporated into this paper.

References

[BDFR01] E. Barcucci, A. Del Lungo, A Forsini, and S Rinaldi. A technology for reverse-engineering a combinatorial problem from a rational generating function. *Advances in Applied Mathematics*, 26:129–153, 2001.

[Bol99] B. Bollobás. *Modern Graph Theory.* Springer, 1999.

[BR84] J. Berstel and C. Reutenauer. *Rational Series and their languages*, volume 12 of *EATCS Monographs on Theoretical Computer Science.* Springer, 1984.

[BS81] C. Blatter and E. Specker. Le nombre de structures finies d'une th'eorie à charactère fin. *Sciences Mathématiques, Fonds Nationale de la recherche Scientifique, Bruxelles*, pages 41–44, 1981.

[BS84] C. Blatter and E. Specker. Recurrence relations for the number of labeled structures on a finite set. In E. Börger, G. Hasenjaeger, and D. Rödding, editors, *In Logic and Machines: Decision Problems and Complexity*, volume 171 of *Lecture Notes in Computer Science*, pages 43–61. Springer, 1984.

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

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

[Cou97] B. Courcelle. The expression of graph properties and graph transformations in monadic second-order logic. In G. Rozenberg, editor, *Handbook of graph grammars and computing by graph transformations, Vol. 1 : Foundations*, chapter 5, pages 313–400. World Scientific, 1997.

[Die90] R. Diestel. *Graph Decompositions, A Study in Infinite Graph Theory.* Clarendon Press, Oxford, 1990.

[EF95] H.D. Ebbinghaus and J. Flum. *Finite Model Theory.* Perspectives in Mathematical Logic. Springer, 1995.

[Fin47] N.J. Fine. Binomial coefficients modulo a prime. *American Mathematical Monthly*, 54:589–592, 1947.

[Fis02] E. Fischer. The Specker-Blatter theorem does not hold for quaternary relations. *Journal of Combinatorial Theory, Series B*, x:xx–yy, 2002. submitted.

[Ges84] I. Gessel. Combinatorial proofs of congruences. In D.M. Jackson and S.A. Vanstone, editors, *Enumeration and design*, pages 157–197. Academic Press, 1984.

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

[HP73] F. Harary and E. Palmer. *Graphical Enumeration.* Academic Press, 1973.

[HU80] J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages and Computation.* Addison-Wesley Series in Computer Science. Addison-Wesley, 1980.

[LN83] R. Lidl and H. Niederreiter. *Finite Fields*, volume 20 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, 1983.

[Luc78] E. Lucas. Théorie des fonctions numériques simplement périodiques. *Amer. J. Math.*, 1:184–240, 289–321, 1878.

[Mak00] J.A. Makowsky. Specker's problem. In E. Grädel and C. Hirsch, editors, *Problems in Finite Model Theory.* THE FMT Homepage, 2000. Last version: June 2000, `http://www-mgi.informatik.rwth-aachen.de/FMT/problems.ps`.

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

[Spe88] E. Specker. Application of logic and combinatorics to enumeration problems. In E. Börger, editor, *Trends in Theoretical Computer Science*, pages 141–169. Computer Science Press, 1988. Reprinted in: Ernst Specker, Selecta, Birkhäuser 1990, pp. 324-350.

[Wil90] H.S. Wilf. *generatingfunctionology.* Academic Press, 1990.

A Counting modulo p

In the following, we let p be a prime number, and state some lemmas and definitions; in particular, we show that $EQ_pCLIQUE$ is a graph property for which the number of models is not periodic modulo p. It is not $CMSOL$ definable, but later we construct first order properties that are related to it.

To help us count modulo p, we make extensive use of the following simple lemma. Similar methods have been extensively used before, at least as early as in the 1872 combinatorial proof of Fermat's congruence theorem by J. Petersen, given in [Ges84, page 157].

Lemma 28 *Suppose that $\mathcal{F}$ is a family of structures over $[n] = \{0, \ldots, n-1\}$ which is preserved under permutations of $[n]$ (e.g. a family defined by a first order expression over some language). Let $\sigma : [n] \to [n]$ be a permutation such that $\sigma \neq \mathrm{Id}$ but $\sigma^p = \mathrm{Id}$. We let σ operate on $\mathcal{F}$ in the natural manner.*

Let $\mathcal{F}' \subset \mathcal{F}$ be a family of structures such that σ preserves membership in $\mathcal{F}'$, and which contains all structures that are invariant with respect to σ. Then $|\mathcal{F}'| \equiv |\mathcal{F}| \pmod{p}$.

Proof. By the above definitions, σ defines a permutation over $\mathcal{F}$, which preserves $\mathcal{F}'$. Decomposing this permutation to disjoint orbits, it is not hard to see that every member of $\mathcal{F}$ which is not invariant under σ is in an orbit of size p (using the information that p is prime); in particular $\mathcal{F} - \mathcal{F}'$ is a disjoint union of such orbits, and so its size is divisible by p.

We denote by $b_p(n) = f_{EQ_pCLIQUE}(n) = f_{EQ_pCLIQUE_2}(n)$ the number of graphs with $[n]$ as a set of vertices which are disjoint unions of exactly p same-size cliques, that is, $b_p(n) = f_{\mathrm{EQC}_p}(n)$. As an example for $p = 2$, note that $b_2(2k+1) = 0$ and $b_2(2k) = \frac{1}{2}\binom{k}{2}$ for every k.

Congruence classes of binomial coefficients and related functions have received a lot of attention in the literature, starting with Lucas's famous result [Luc78]. For an accessible proof, cf, [Fin47].

Theorem 29 (Lucas) *Let p be prime and*

$$r = r_m p^m + \ldots + r_1 p + r_0 (0 \leq r_i < p)$$
$$k = k_m p^m + \ldots + k_1 p + k_0 (0 \leq k_i < p)$$

then

$$\binom{r}{k} \equiv \prod_{i=0}^{m} \binom{r_i}{k_i} \pmod{p}$$

We now investigate the congruences of $b_p(n)$ modulo p.

Lemma 30 *For every $k > 1$, $b_p(pk) \equiv b_p(k) \pmod{p}$.*

Proof. We define $\sigma : [pk] \to [pk]$ by $\sigma(pi + j) = \sigma(pi + j + 1)$ for $0 \leq i < k$ and $0 \leq j < p - 1$, and $\sigma(pi + p - 1) = \sigma(pi)$ for $0 \leq i < k$ (so σ is composed of k disjoint orbits of size p).

We now use Lemma 28. We first note that all graphs for which any clique contains more than one member, but not all members, of $\{pi, \ldots, pi+p-1\}$ for some i, are not invariant with respect to σ. We also note that all graphs for which some clique contains all members of $\{pi, \ldots, pi+p-1\}$, but only one member of $\{pj, \ldots, pj+p-1\}$ for some other j, are not invariant with respect to σ.

We let $\mathcal{F}'$ be the family of all other graphs which are disjoint union of p same-size cliques. It is not hard to see that $\mathcal{F}'$ contains two types of graphs: those for which every $\{pi, \ldots, pi+p-1\}$ is contained in one of the cliques, whose number is $b_p(k)$, and those for which every $\{pi, \ldots, pi+p-1\}$ contains exactly one member from each clique, whose number $(p!)^{k-1}$ is divisible by p if $k > 1$.

To prove Proposition 9 we proceed by induction on n, where the basis is $n = p$ (for which $b_p(n) = 1$) and every n which is not divisible by p (for which $b_p(n) = 0$); the induction step follows from Lemma 30.

B An *FOL* sentence without modular recurrence

In this section we give a proof of Theorem 3, and follow closely [Fis02]. The basic idea is to find for every prime number p an FOL-definable class $\mathcal{C}$ such that $f_{\mathcal{C}}(n) \equiv f_{\mathrm{EQ_pCLIQUE}}(n) \pmod p$. It is easy to find an FOL formula that states that a graph is a disjoint union of p cliques; we now need to show how to equate the cliques in a way that will preserve the number of models modulo p.

B.1 Comparing sizes in a modulo-preserving manner

We recall that the naive way to ensure (with a first order property) that the sizes of p sets $A_0, \ldots, A_{p-1}$ are all equal (by adding a binary relation and stating that it is a perfect matching between each pair of these sets), leads to a property with a trivial linear recurrence relation. We thus have to find another method. We start with one that does not ensure that the sets are equal, and later show how to iterate it in a manner that indeed provides a good substitute for the notion of a perfect matching.

Definition 31 *A* preserving p-matching *between $A_0, \ldots, A_{p-1}$ is a set consisting of $\frac{1}{p}\sum_{i=0}^{p-1}|A_i|$ vertex disjoint p-cliques on $\bigcup_{i=0}^{p-1} A_i$, such that every clique is either fully contained in one of the A_i's, or contains exactly one vertex from each A_i.*

Note that for $p = 2$, every perfect matching on $A_0 \cup A_1$ (in the usual graph-theoretic sense) is a preserving 2-matching. The enumeration of preserving p-matchings modulo p is given by the following.

Lemma 32 *If $|A_0| \equiv, \ldots, \equiv |A_{p-1}| \pmod p$ then the number of preserving p-matchings is 1 modulo p. Otherwise, there are no preserving p-matchings at all.*

Proof. The proof of the second part (where the $|A_i|$ are not all equivalent modulo p) is simple. The proof of the first part is done by induction on $\sum_{i=0}^{p-1}|A_i|$.

The base case is where all $|A_i|$ are equal to some $k < p$. It is clear that in this case a preserving matching consists of k cliques such that each of them contains exactly one vertex from each A_i. Denoting $A_i = \{v_{i,0}, \ldots, v_{i,k-1}\}$, define σ by $\sigma(v_{i,j}) = v_{i+1,j}$ for every $0 \leq j \leq k-1$ and $0 \leq i < p-1$, and $\sigma(v_{p-1,j}) = v_{0,j}$ for every $0 \leq j \leq k-1$. Since $k < p$, for every clique with vertices $\{v_{0,j_0}, \ldots, v_{p-1,j_{p-1}}\}$ there exist $i \neq i'$ such that $j_i = j_{i'}$; from this it is not hard to show that the matching is not invariant with respect to σ unless for every such clique, $j_i = j_{i''}$ for every i''. Thus there exists only one preserving p-matching which is invariant with respect to σ, and using Lemma 28 the base case is proven.

For the induction step, let i_0 be such that $|A_{i_0}| \geq p$, and let $v_0, \ldots, v_{p-1}$ be p vertices in A_{i_0}. In this case we define σ by $\sigma(v_j) = v_{j+1}$ for $0 \leq j < p-1$, $\sigma(v_{p-1}) = v_0$, and $\sigma(u) = u$ for every $u \notin \{v_0, \ldots, v_p - 1\}$. It is clear that the only invariant preserving p-matchings are those for which $\{v_0, \ldots, v_{p-1}\}$ is one of the p-cliques, and using Lemma 28 the induction step follows.

To fully equate the sizes of the sets $A_0, \ldots, A_{p-1}$, we use the following notion of a matching between the sets.

Definition 33 *Given disjoint sets $A_0, \dots, A_{p-1}$, an* iterative p-matching *between these sets is a sequence of graphs $\{\mathcal{M}_i\}_{i \geq 0} = \mathcal{M}_0, \mathcal{M}_1, \dots$ where each has its own vertex set, satisfying the following.*

- *If $A_i = \emptyset$ for every i then $\mathcal{M}_0 = \emptyset$.*
- *Otherwise, $\mathcal{M}_0$ is a preserving p-matching between $A_0, \dots, A_{p-1}$.*
- *Defining by A'_i the set of p-cliques of $\mathcal{M}_0$ inside A_i for every i, $\mathcal{M}_1, \mathcal{M}_2, \dots$ is an iterative p-matching between $A'_0, \dots, A'_{p-1}$.*

The above sequences may look infinite, but one can see that if $A_0, \dots, A_{p-1}$ are all finite, then the number of non-empty elements in an iterative p-matching is also finite. We shall also use the following alternative definition of iterative matchings.

Definition 34 *Given disjoint sets $A_0, \dots, A_{p-1}$, a* graphic iterative p-matching *between these sets is a sequence of graphs $\{M_i\}_{i \geq 0} = M_0, M_1, \dots$ which all have $\bigcup_{i=0}^{p-1} A_i$ as a vertex set, satisfying the following.*

- *Each M_i consists of isolated vertices and vertex disjoint copies of the complete p-partite graph with p color classes of size p^i.*
- *Each of the p-partite graphs in M_i is either fully contained in one of the A_1's, or is such that each of its color classes is fully contained in a different A_i.*
- *For $i > 1$, each color class of a p-partite graph in M_i consists of all vertices of one of the p-partite graphs in M_{i-1} which are fully contained in one of $A_0, \dots, A_{p-1}$; moreover, for each of the p-partite graphs of M_{i-1} with the above property there exists a complete p-partite graph in M_i containing its vertices in this manner.*

It easily follows that M_0 in a graphic iterative matching is a preserving p-matching between $A_0, \dots, A_{p-1}$, like $\mathcal{M}_0$ in an iterative matching. It is not very hard to see that the correspondence defined below is in fact a one to one and onto correspondence between all possible iterative matchings and all possible graphic iterative matchings between $A_0, \dots, A_{p-1}$.

Definition 35 *Given a graphic iterative matching $\{M_i\}_{i \geq 0}$ we construct the corresponding iterative matching $\{\mathcal{M}_i\}_{i \geq 0}$ as follows.*

- *$\mathcal{M}_0$ is M_0.*
- *For every i we let A'_i be the set of p-cliques of M_0 that are fully contained in A_i. We then construct $M'_1, M'_2, \dots$ by defining M'_j to have an edge between $u \in \bigcup_{i=0}^{p-1} A'_i$ and $v \in \bigcup_{i=0}^{p-1} A'_i$ if and only if M_j has an edge between the corresponding cliques. It is not hard to see that $M'_1, M'_2, \dots$ is a graphic iterative p-matching between $A'_0, \dots, A'_{p-1}$; we then define $\mathcal{M}_1, \mathcal{M}_2, \dots$ as the iterative matching corresponding to $M_1, M_2, \dots$ inductively.*

Henceforth, we use the term "iterative matchings" for both point of views. We now show how iterative matchings are useful for equating sets in the modulo p setting.

Lemma 36 *If $|A_i|$ are all equal, then the number of iterative p-matchings between $A_0, \dots, A_{p-1}$ is 1 modulo p. Otherwise, there are no such matchings.*

Proof. The proof is by induction on $\sum_{i=1}^{p-1} |A_i|$. The case where this sum is zero is clear (in this case $A_i = \emptyset$ for every i and indeed there exists exactly one possible iterative p-matching), as well as all cases where the $|A_i|$ are not all equivalent modulo p (in which there is no possibility for constructing even the first preserving p-matching $\mathcal{M}_0$).

In any other case the number of ways to construct $\mathcal{M}_0$ is 1 modulo p by Lemma 32. For each such construction, if we construct the appropriate $A'_0, \ldots, A'_{p-1}$ as per the definition above, it is easy to see that $\sum_{i=1}^{p-1} |A'_i| < \sum_{i=1}^{p-1} |A_i|$, as well as that $|A'_i|$ are all equal if and only if $|A_i|$ are all equal. The latter occurs since if we denote by r the number of cliques in $\mathcal{M}_0$ not fully contained in any of the A_i, we get $|A'_i| = \frac{|A_i|-r}{p}$ for every i.

If $|A_i|$ are all equal, then by the induction hypothesis for each choice of $\mathcal{M}_0$ the number of choices for $\mathcal{M}_1, \mathcal{M}_2, \ldots$ is 1 modulo p, and thus their sum over all choices of $\mathcal{M}_0$ is 1 modulo p. If $|A_i|$ are not all equal, then by the induction hypothesis there exists no good choice of $\mathcal{M}_1, \mathcal{M}_2, \ldots$ for any choice of $\mathcal{M}_0$, completing the proof.

We end this section with a simple lemma which is not directly related to counting, but is used in the following.

Lemma 37 *For every iterative matching between $A_0, \ldots, A_{p-1}$ (by Lemma 36 we need only consider sets with equal sizes), every vertex in $\bigcup_{i=0}^{p-1} A_i$ is eventually matched (a vertex in A_i is considered eventually matched if it has a neighbor outside of A_i in some M_k, when we consider the graphic version $\{M_i\}_{i\geq 0}$ of the iterative matching).*

Proof. In this case it is better to look at $\{\mathcal{M}_i\}_{i\geq 0}$ which corresponds to $\{M_i\}_{i\geq 0}$, and note that a vertex $v \in A_i$ is eventually matched if and only if it is either contained in a clique of $\mathcal{M}_0$ which is not internal to A_i, or contained in a clique of $\mathcal{M}_0$ which is internal to A_i but which is eventually matched by $\mathcal{M}_1, \mathcal{M}_2, \ldots$; the proof is then completed by an easy induction on $|A_0|$.

B.2 Constructing the first order property

We now construct a first order property that in essence counts $b_p(n)$ times the number of possible iterative matchings between the p sets of size $\frac{n}{p}$; by Lemma 36 this is equivalent modulo p to $b_p(n)$.

We look at structures $\langle [n], E, R\rangle$ where E is a binary relation and R is a quaternary (arity four) relation. The property will state that E is a union of p vertex-disjoint cliques and that R is a representation (we prove later that it is unique) of an iterative p-matching between the cliques in E. Instead of defining the property all at once we define it as the conjunction of several properties defined below. All the properties are first order, and whenever proving this part is clear we omit all further mention thereof. In the presentation we also define and use some relations that can be expressed using first order expressions over E and R.

Definition 38 *Property* $\mathrm{Cl}_p(E)$ *states that E is a non-directed simple graph which is the disjoint union of exactly p cliques.*

In the sequel we denote by $A_0, \ldots, A_{p-1}$ the p cliques. We note however that the labeling of these cliques is arbitrary, and make sure that all the logical constructions below are invariant with respect to permuting the labels $A_0, \ldots, A_{p-1}$; note that in particular the definition of a preserving p-matching is such a construction.

Definition 39 *Property* $\mathrm{Edg}_p(R)$ *states that if (e_1, e_2, o_1, o_2) is in R then $e_1 \neq e_2$, and also (e_2, e_1, o_1, o_2) and (e_1, e_2, o_2, o_1) and (e_2, e_1, o_2, o_1) are in R. We say in this case that* the edge (e_1, e_2) has (o_1, o_2) as an origin. *We say that* (e_1, e_2) has an origin *if there exist (o_1, o_2) for which $(e_1, e_2, o_1, o_2) \in R$. Note that there is the possibility that $o_1 = o_2$.*

In the sequel we shall usually refer by the term ‘edge’ to an (e_1, e_2) that has an origin according to R, and only refer indirectly (e.g. by the definition of $A_0, \ldots, A_{p-1}$) to the graph E.

Definition 40 *If (e_1, e_2) which has an origin satisfies $(e_1, e_2) \notin E$ (that is, it is an edge between A_i and A_j for some $i \neq j$) then we say that (e_1, e_2) is a* bridge. *Otherwise we say that (e_1, e_2) is* internal *to the clique that contains e_1 and e_2 (which is one of $A_0, \ldots, A_{p-1}$).*

We use the definition of bridge and internal edges to define the property of R representing an iterative p-matching $\{M_i\}_{i \geq 0}$, while distinguishing which edge belongs to which M_i will result from the above definition of an origin. First we deal with M_0.

Definition 41 *Property* $\mathrm{Base}_p(E, R)$ *states the following.*

- *If (e_1, e_2) has some (o, o) as an origin, then for every (o_1, o_2) it has (o_1, o_2) as an origin if and only if $o_1 = o_2$.*
- *For every o, the set of edges having (o, o) as an origin is a preserving p-matching between $A_0, \ldots, A_{p-1}$.*

The reason for requiring that an edge has either no origin of the type (o, o) or has all of them is to ensure that there is only one way to represent M_0 using R. We shall now require a representation of M_{i+1} given that we already have the representations of $M_0 \ldots, M_i$.

The following definition makes use of the notion of connected components, which is not first order definable. However, whenever this is mentioned, it can be replaced with the first order notion of all vertices having distance no more than two from a given vertex, since we prove later that for any (o_1, o_2) the set of edges having it as an origin forms a disjoint union of isolated vertices and complete p-partite graphs, so in particular all the connected components have diameter at most 2. We shall also prove that each such component is either internal to one of $A_0, \ldots, A_{p-1}$, or brings together a component of M_i from every A_i. This will be proven by induction; the basis $o_1 = o_2$ is relatively easy using the property $\mathrm{Base}_p(E, R)$.

Definition 42 *Property* $\mathrm{Next}_p(E, R)$ *states the following.*

- *If (e_1, e_2) has (o_1, o_2) with $o_1 \neq o_2$ as an origin, then for every (o'_1, o'_2) it has (o'_1, o'_2) as an origin if and only if (o_1, o_2) and (o'_1, o'_2) have the same origin (i.e. if there exists (r_1, r_2) such that $(o_1, o_2, r_1, r_2) \in R$ and $(o'_1, o'_2, r_1, r_2) \in R$).*
- *For every $o_1 \neq o_2$, we look at the set of connected components of the set of edges having the same origin as (o_1, o_2), apart from those which are isolated vertices and those which are not internal to one of $A_0, \ldots, A_{p-1}$; denote them by $C_1, \ldots, C_l$. We also denote by G the graph resulting from the set of edges having (o_1, o_2) as an origin.*
  - *G consists of isolated vertices and vertex disjoint copies of complete p-partite graphs, each of which has p members of $C_1, \ldots, C_l$ as its color classes.*
  - *Each of the complete p-partite graphs in G is either fully contained in one of $A_0, \ldots, A_{p-1}$, or is such that each of its color classes is fully contained in a different A_i.*
  - *Each of $C_1, \ldots, C_l$ intersects one of the complete p-partite graphs of G.*

To justify the use of the notion of complete p-partite graphs in the definition of a first order property, note that the following property of a vertex v_0 is first order, and that it is equivalent to the property that the connected component containing v_0 is a complete p-partite graph: "There exists $v_1, \ldots, v_{p-1}$ such that $\{v_0, \ldots, v_{p-1}\}$ is a clique, that every vertex with distance 3 or less from v_0 is adjacent to exactly $p - 1$ of the vertices $\{v_0, \ldots, v_{p-1}\}$, and that every two such vertices are adjacent to each other if and only if they are not adjacent to the same $p - 1$ members of $\{v_0, \ldots, v_{p-1}\}$".

To finalize the definition of our first order property, we make sure that vertex pairs incident with bridge edges are 'out of the game', to avoid multiplicities in counting that may result from assigning them arbitrary origins.

Definition 43 *Property* $\mathrm{Clear}_p(E, R)$ *states that for every* (o_1, o_2), *no edges that are incident with a bridge edge having* (o_1, o_2) *as an origin may have any origin, except possibly the edges which are internal to the connected components of the graph of edges having* (o_1, o_2) *as an origin.*

We now state and prove the concrete form of Theorem 5.

Theorem 44 *Let* $\mathrm{Im}_p(E, R) = \mathrm{Cl}_p(E) \wedge \mathrm{Edg}_p(R) \wedge \mathrm{Base}_p(E, R) \wedge \mathrm{Next}_p(E, R) \wedge \mathrm{Clear}_p(E, R)$. *Denote by* $f_{\mathrm{Im}_p}(n)$ *the number of structures* $\langle [n], E, R \rangle$ *satisfying* Im_p. *Then* $f_{\mathrm{Im}_p}(n) \equiv b_p(n) \pmod{p}$, *and so it is not ultimately periodic modulo p.*

To prove it we consider an E which satisfies $\mathrm{Cl}_p(E)$, and define a way to encode an iterative matching between the cliques $A_0, \dots, A_{p-1}$ of E, as a relation R for which Im_p is satisfied. Then we prove that such encodings are the only instances which satisfy Im_p for any given E.

Definition 45 *Suppose that* $\{M_i\}_{i \geq 0}$ *is an iterative matching (we use the graphic definition) between the cliques of E. We define an R which is* the encoding of $\{M_i\}_{i \geq 0}$ *as follows.*

- *Every edge of* M_0 *is an edge according to R that has every* (o, o) *and no other pair as an origin.*
- *For* $i > 1$, *we let every edge of* M_i *have every edge of* M_{i-1} *and no other pair as an origin.*
- *No other combinations of edges with origins exist apart from those constructed above.*

It is not extremely hard to prove the following.

Claim. An encoding of an iterative matching satisfies Im_p. Moreover, for any two distinct iterative matchings, the corresponding encodings are also distinct. □

Suppose now that we are given a structure $\langle [n], E, R \rangle$ that satisfies Im_p. To prove that it is an encoding of some iterative matching we first define inductively the graphs $\{M_i\}_{i \geq 0}$ and then prove that they form the matching which $\langle [n], E, R \rangle$ encodes.

Definition 46 *Given a structure* $\langle [n], E, R \rangle$ *satisfying* Im_p *we define a sequence* $\{M_i\}_{i \geq 0} = M_0, M_1, \dots$ *of graphs on* $[n]$ *inductively as follows.*

- M_0 *consists of all the edges having any* (o, o) *as an origin.*
- M_i *for* $i > 0$ *consists of all the edges having any edge from* M_{i-1} *as an origin.*

Lemma 47 *The following holds for the above defined graphs.*

- *Every edge in* M_0 *has every* (o, o) *and no other pair as an origin, and every edge in* M_i *has every edge in* M_{i-1} *and no other pair as an origin.*
- *There is no edge in* $M_i \cap M_j$ *for any* $i < j$.
- M_0 *is a preserving matching between the p cliques of E.*
- $\{M_i\}_{i \geq 0}$ *is an iterative matching between the p cliques of E (in particular, the connected components of each* M_i *are isolated vertices and complete p-partite graphs).*
- *There are no other edges with origins (according to R) apart from those in* $\bigcup_{i \geq 0} M_i$.

Proof. The first two items follow by induction from $\langle [n], E, R\rangle$ satisfying the first item of Base_p and the first item of Next_p. The third item follows from the second item of Base_p. The fourth item follows by induction from the above together with the second item in Next_p (with all its subitems). Finally, the fifth item follows from $\langle [n], E, R\rangle$ satisfying Clear_p, when used in conjunction with Lemma 37.

Lemma 47 directly provides the final component required for the proof of Theorem 44.

Consequence 48 *For every $\langle [n], E, R\rangle$ satisfying Im_p, the relation R is an encoding of an iterative matching between the p cliques of E.* □

Proof (Proof of Theorem 44:). Claim B.2 and Consequence 48 imply that the number of structures $\langle [n], E, R\rangle$ equals $b_p(n)$ times the number of possible iterative matchings between p sets of size $\frac{n}{p}$, and by Lemma 36 the latter number is 1 modulo p.

Finally, we note that it is possible to formulate a property similar to Im_p that uses only a single quaternary relation R, by using "$R(u, u, v, v)$" to represent "$E(u, v)$" and changing the formulation of the property accordingly.

C Structures of bounded degree

Definition 49 *1. Given a structure $\mathfrak{A} = \langle A, R_1^A, \ldots, R_k^A\rangle$, $u \in A$ is called a* neighbor *of $v \in A$ if there exists a relation R_i^A and some $\bar{a} \in R_i^A$ containing both u and v.*

2. We define the Gaifman graph *$Gaif(\mathfrak{A})$ of a structure $\mathfrak{A}$ as the graph with the vertex set A and the neighbor relation defined above.*

3. The degree *of a vertex $v \in A$ in $\mathfrak{A}$ is the number of its neighbors. The degree of $\mathfrak{A}$ is defined as the maximum over the degrees of its vertices. It is the degree of its Gaifman graph $Gaif(\mathfrak{A})$.*

4. A structure $\mathfrak{A}$ is connected *if its Gaifman graph $Gaif(\mathfrak{A})$ is connected.*

Definition 50 *For an* MSOL *class $\mathcal{C}$, denote by $f_{\mathcal{C}}^{(d)}(n)$ the number of structures over $[n]$ that are in $\mathcal{C}$ and whose degree is at most d.*

The DU-index of a class of structures is the $\mathcal{F}$-index for the case that $\mathcal{F}$ is the disjoint union of two structures.

Theorem 51 *If $\mathcal{C}$ is a class of $\bar{R}$-structures which has a finite DU-index, then $f_{\mathcal{C}}^{(d)}(n)$ is ultimately periodic modulo m, hence, trivially, $f_{\mathcal{C}}^{(d)}(n)$ satisfies for every $m \in \mathbb{N}$ a linear recurrence relation modulo m.*

Furthermore, if all structures of $\mathcal{C}$ are connected, then this modular linear recurrence is trivial.

Lemma 52 *If $\mathfrak{A} \sim_{Du(\mathcal{C})} \mathfrak{B}$, then for every $\mathfrak{C}$ we have*

$$\mathfrak{C} \sqcup \mathfrak{A} \sim_{Du(\mathcal{C})} \mathfrak{C} \sqcup \mathfrak{B}.$$

Proof. Easy, using the associativity of the disjoint union.

To prove Theorem 51 we define orbits for permutation groups rather than for single permutations.

Definition 53 *Given a permutation group G that acts on A (and in the natural manner acts on models over the universe A), the* orbit *in G of a model $\mathfrak{A}$ with the universe A is the set* $\mathrm{Orb}_G(\mathfrak{A}) = \{\sigma(\mathfrak{A}) : \sigma \in G\}$.

For $A' \subset A$ we denote by $S_{A'}$ the group of all permutations for which $\sigma(u) = u$ for every $u \notin A'$. The following lemma is useful for showing linear congruences modulo m.

Lemma 54 *Given $\mathfrak{A}$, if a vertex $v \in A - A'$ has exactly d neighbors in A', then $|\mathrm{Orb}_{S_{A'}}(\mathfrak{A})|$ is divisible by $\binom{|A'|}{d}$.*

Proof. Let N be the set of all neighbors of v which are in A', and let $G \subset S_{A'}$ be the subgroup $\{\sigma_1\sigma_2 : \sigma_1 \in S_N \wedge \sigma_2 \in S_{A'-N}\}$; in other words, G is the subgroup of the permutations in $S_{A'}$ that in addition send all members of N to members of N. It is not hard to see that $|\mathrm{Orb}_{S_{A'}}(\mathfrak{A})| = \binom{|A'|}{|N|}|\mathrm{Orb}_G(\mathfrak{A})|$.

The following simple observation is used to enable us to require in advance that all structure in $\mathcal{C}$ have a degree bounded by d.

Observation 55 *We denote by $\mathcal{C}_d$ the class of all members of $\mathcal{C}$ that in addition have bounded degree d. If $\mathcal{C}$ has a finite DU-index then so does $\mathcal{C}_d$.* □

In the following we fix m and d. Instead of $\mathcal{C}$ we look at $\mathcal{C}_d$, which by Observation 55 also has a finite DU-index. We now note that there is only one equivalence class containing any structures whose maximum degree is larger than d, which is the class $\mathcal{N}_{\mathcal{C}}^{(d)} = \{\mathfrak{A} : \forall_{\mathfrak{B}}(\mathfrak{B} \sqcup \mathfrak{A}) \not\models \mathcal{C}_d)\}$ In order to show that $f_{\mathcal{C}}^{(d)}(n)$ is ultimately periodic modulo m, we show a linear recurrence relation modulo m on the vector function $(f_{\mathcal{E}}(n))_{\mathcal{E}}$ where $\mathcal{E}$ ranges over all other equivalence classes with respect to $\mathcal{C}_d$.

Let $C = md!$. We note that for every $t \in \mathbb{N}$ and $0 < d' \leq d$, m divides $\binom{tC}{d'}$. This with Lemma 54 allows us to prove the following.

Lemma 56 *Let $\mathcal{D} \neq \mathcal{N}_\phi$ be an equivalence class for ϕ, that includes the requirement of the maximum degree not being larger than d. Then*

$$f_{\mathcal{D}}(n) \equiv \sum_{\mathcal{E}} a_{\mathcal{D},\mathcal{E},m,(n \bmod C)} f_{\mathcal{E}}(C\lfloor \frac{n-1}{C} \rfloor) \pmod{m},$$

for some fixed appropriate $a_{\mathcal{D},\mathcal{E},m,(n \bmod C)}$.

Proof. Let $t = \lfloor \frac{n-1}{C} \rfloor$. We look at the set of structures in $\mathcal{D}$ with the universe $[n]$, and look at their orbits with respect to $S_{[tC]}$. If a model $\mathfrak{A}$ has a vertex $v \in [n] - [tC]$ with neighbors in $[tC]$, let us denote the number of its neighbors by d'. Clearly $0 < d' \leq d$, and by Lemma 54 the size of $\mathrm{Orb}_{S_{[tC]}}(\mathfrak{A})$ is divisible by $\binom{tC}{d'}$, and therefore it is divisible by m. Therefore, $f_{\mathcal{D}}(n)$ is equivalent modulo m to the number of structures in $\mathcal{D}$ with the universe $[n]$ that in addition have no vertices in $[n] - [tC]$ with neighbors in $[tC]$.

We now note that any such structure can be uniquely written as $\mathfrak{B} \sqcup \mathfrak{C}$ where $\mathfrak{B}$ is any structure with the universe $[n - tC]$, and $\mathfrak{C}$ is any structure over the universe $[tC]$. We also note using Lemma 52 that the question as to whether $\mathfrak{A}$ is in $\mathcal{D}$ depends only on the equivalence class of $\mathfrak{C}$ and on $\mathfrak{B}$ (whose universe size is bounded by the constant C). By summing over all possible $\mathfrak{B}$ we get the required linear recurrence relation (cases where $\mathfrak{C} \in \mathcal{N}_{\mathcal{C}}^{(d)}$ do not enter this sum because that would necessarily imply $\mathfrak{A} \in \mathcal{N}_{\mathcal{C}}^{(d)} \neq \mathcal{D}$).

Proof (Proof of Theorem 51:). We use Lemma 56: Since there is only a finite number of possible values modulo m to the finite dimensional vector $(f_{\mathcal{E}}(n))_{\mathcal{E}}$, the linear recurrence relation in Lemma 56 implies ultimate periodicity for n's which are multiples of C. From this the ultimate periodicity for other values of n follows, since the value of $(f_{\mathcal{E}}(n))_{\mathcal{E}}$ for an n which is not a multiple of C is linearly related modulo m to the value at the nearest multiple of C.

Finally, if all structures are connected we use Lemma 54. Given $\mathfrak{A}$, connectedness implies that there exists a vertex $v \in A'$ that has neighbors in $A - A'$. Denoting the number of such neighbors by d_v, we note that $|\mathrm{Orb}_{S'_A}(\mathfrak{A})|$ is divisible by $\binom{|A'|}{d_v}$, and since $1 \leq d_v \leq d$ (using $|A'| = tC$) it is also divisible by m. This makes the total number of models divisible by m (remember that the set of all models with $A = [n]$ is a disjoint union of such orbits), so $f_C^{(d)}(n)$ ultimately vanishes modulo m.

D Specker index and *CMSOL*

Although Theorem 2 is stated for classes of structures definable in some logic, logic is only used to verify the hypothesis of Specker's Lemma, 18. In this Appendix we develop the machinery which serves this purpose. The crucial property needed to prove Theorem 20 is a reduction property which says that both for the disjoint union $\mathfrak{A} \sqcup \mathfrak{B}$ and for the substitution $Subst((\mathfrak{A}, a), \mathfrak{B})$ the truth value of a sentence $\phi \in CMSOL(\bar{R})$ depends only on the truth values of the sentences of the same quantifier rank in the structures $\mathfrak{A}$ and $\mathfrak{B}$, respectively $\langle \mathfrak{A}, a \rangle$ and $\mathfrak{B}$. For the case of $MSOL$ this follows either from the Feferman-Vaught Theorem for disjoint unions together with some reduction techniques, or using Ehrenfeucht-Fraïssé games. The latter is used in [Spe88]. We shall use the former, as it is easier to adapt for $CMSOL$.

D.1 Quantifier rank

We define the quantifier rank $qr(\phi)$ of a formula ϕ of $CMSOL(\bar{R})$ inductively as usual: For quantifier free formulas ϕ we have $qr(\phi) = 0$. For boolean operations we take the maximum of the quantifier ranks. Finally, $qr(\exists U \phi) = qr(\exists x \phi) = qr(C_{p,q} x \phi) = qr(\phi) + 1$. We denote by $CMSOL^q(\bar{R}, \bar{x}, \bar{U})$ the set of $CMSOL(\bar{R})$-formulas with free variables $\bar{x}$ and $\bar{U}$ which are of quantifier rank at most q. When there are no free variables we write $CMSOL^q(\bar{R})$.

We write $\mathfrak{A} \sim^q_{CMSOL} \mathfrak{B}$ for two $\bar{R}$-structures $\mathfrak{A}$ and $\mathfrak{B}$ if they satisfy the same $CMSOL^q(\bar{R})$-sentences.

The following is folklore, cf. [EF95].

Proposition 57 *There are, up to logical equivalence, only finitely many formulas in $CMSOL^q(\bar{R}, \bar{x}, \bar{U})$. In particular, the equivalence relation $\mathfrak{A} \sim^q_{CMSOL} \mathfrak{B}$ is of finite index.*

D.2 A Feferman-Vaught Theorem for *CMSOL*

We are now interested in how the truth of a sentence in $CMSOL$ in the disjoint union of two structures $\mathfrak{A} \sqcup \mathfrak{B}$ depends on the truth of other properties expressible in $CMSOL$ which hold in $\mathfrak{A}$ and $\mathfrak{B}$ separately.

Theorem 58 (Courcelle)

1. *For every formula $\phi \in CMSOL^q(\tau)$ one can compute in* polynomial time *a sequence of formulas*

$$\langle \psi_1^A, \ldots, \psi_m^A, \psi_1^B, \ldots, \psi_m^B \rangle \in CMSOL^q(\tau)^{2m}$$

and a boolean function $B_\phi : \{0,1\}^{2m} \rightarrow \{0,1\}$ *such that*

$$\mathfrak{A} \sqcup \mathfrak{B} \models \phi$$

if and only if

$$B_\phi(b_1^A, \dots b_m^A, b_1^B, \dots b_m^B) = 1$$

where $b_j^A = 1$ *iff* $\mathfrak{A} \models \psi_j^A$ *and* $b_j^B = 1$ *iff* $\mathfrak{B} \models \psi_j^B$.

A detailed proof is found in [Cou90, Lemma 4.5, page 46ff].

D.3 Quantifier free transductions and *CMSOL*

FOL-reductions are widely used in descriptive complexity theory, cf. [EF95]. They are also called transductions, cf. [Cou94]. Quantifier free $\bar{R}$-transductions are $FOL(\bar{R})$-reductions with the defining formulas quantifier free. They are called scalar, when the defining formula for the universe has one free variable only.

Lemma 59 *Let* Φ^* *be a quantifier free scalar* $\bar{R}$*-transduction. Assume* $\mathfrak{A}_1, \mathfrak{A}_2$ *are* $\bar{R}$*-structures and* $\mathfrak{A}_1 \sim^q_{CMSOL} \mathfrak{A}_2$. *Then* $\Phi^*(\mathfrak{A}_1) \sim^q_{CMSOL} \Phi^*(\mathfrak{A}_2)$.

Lemma 60 $Subst((\mathfrak{A}, a), (\mathfrak{B}, b))$ *can be obtained from the disjoint union of* $(\mathfrak{A}, a)$ *and* $(\mathfrak{B}, b))$ *by a quantifier free transduction.*

Proof (Sketch of proof:). The universe of the structure is $C = (A \sqcup B) - \{a\}$. For each relation symbol $R \in \bar{R}$ we put

$$R^C = R^A|_{A-\{a\}} \cup R^B \cup \{(a', b) : (a', a) \in R^A, b \in B\}$$

This is clearly expressible as a quantifier free transduction from the disjoint union.

Proposition 61 *Assume* $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{B}_1, \mathfrak{B}_2$ *are* $\bar{R}$*-structures and with context* a_1, a_2, b_1, b_2, *respectively, and*

$$(\mathfrak{A}_1, a_1) \sim^q_{CMSOL} (\mathfrak{A}_2, a_2) \text{ and } (\mathfrak{B}_1, b_1) \sim^q_{CMSOL} (\mathfrak{B}_2, b_2).$$

Then $Subst((\mathfrak{A}_1, a_1), (\mathfrak{B}_1, b_1)) \sim^q_{CMSOL} Subst((\mathfrak{A}_2, a_2), (\mathfrak{B}_2, b_2))$.

Proof. Use Theorem 58, Lemma 59 and Lemma 60.

D.4 Finite index theorem for *CMSOL*

Now we can state and prove the Finite Index Theorem;

Theorem 62 *Let* $\mathcal{C}$ *be defined by an* $CMSOL(\bar{R})$*-sentence* ϕ *of quantifier rank* q*. Then* $\mathcal{C}$ *has a finite Specker index, which is bounded by the number of inequivalent* $CMSOL^q(\bar{R})$*-sentences. This number is finite by Proposition 57.*

Proof. We have to show that the equivalence relation $\mathfrak{A} \sim^q_{CMSOL} \mathfrak{B}$ is a refinement of $\mathfrak{A} \sim_{\mathcal{C}} \mathfrak{B}$. But this follows from Proposition 61.

E Detailed discussion of examples

In this appendix we list examples with their definability properties and give their density functions or generating functions, as far as we could find them in the literature. The non-definability results are fairly standard in logic, cf. [EF95,Cou97], and we state them bona fide. For the density and generating function, our main sources are [HP73,Wil90]. Additional related results may be found in I. Gessel's seminal work [Ges84], where many explicit formulas are given for various graph classes of bounded degree.

E.1 Connected graphs

The class CONN is not FOL(R)-definable, but it is MSOL(R)-definable using a universal quantifier over set variables. We just say that every subset of vertices which is closed under the edge relation has to be the set of all vertices.

Counting labeled connected graphs is treated in [HP73, Chapters 1 and 7] and in [Wil90, Chapter 3]. For CONN [HP73, page 7] gives the following recurrence:

$$f_{\mathrm{CONN}}(n) = 2^{\binom{n}{2}} - \frac{1}{n}\sum_{k=1}^{n-1} k\binom{n}{k} 2^{\binom{n-k}{2}} f_{\mathrm{CONN}}(k).$$

E.2 Regular graphs

The class $\mathrm{REG_r}$ of simple regular graphs where every vertex has degree r is FOL-definable (for fixed r). The formulas says that every vertex has exactly r different neighbors. The formula grows with r. Regularity without specifying the degree is not FOL-definable, actually not even $CMSOL$-definable.

Counting the number of labeled regular graphs is treated completely in [HP73, Chapter 7]. However, the formula is very complicated.

For cubic graphs, the function is explicitly given in [HP73, page 175] as $f_{\mathcal{R}_3}(2n+1) = 0$ and

$$f_{\mathcal{R}_3}(2n) = \frac{(2n)!}{6^n}\sum_{j,k}\frac{(-1)^j(6k-2j)!6^j}{(3k-j)!(2k-j)!(n-k)!}48^k\sum_i \frac{(-1)^i j!}{(j-2i)!i!}$$

E.3 Trees and acyclic digraphs

Trees are (undirected) connected acyclic graphs. They are not FOL-definable but $MSOL$-definable. Acyclicity is expressed by saying there is no subset of size at least three such that the induced graph on it is 2-regular and connected. Labeled trees were among the first objects to be counted, cf.[HP73, Theorem 1.7.2].

Theorem 63 (A. Cayley 1889) *The number of labeled trees on n vertices is $T_n = n^{n-2}$.*

Here the modular linear recurrences can be given explicitly: We have $T_1 = T_2 = 1$, $T_3 = 3$, $T_4 = 16$, $T_5 = 125$, ... and $T_n = n \pmod 2$ for $n \geq 3$.

For the number of trees of outdegree bounded by k we get the following corollary of Theorem 6:

Corollary 64 *The number of labeled trees of outdegree at most k is, for every $m \in \mathbb{Z}$, ultimately constant* $\pmod{m}$.

In [HP73, Chapter 3] there is a wealth of results on counting various labeled trees and tree-like structures. It is worth noting that the notion of k-tree, and more generally the property of a graph of having tree-width at most k are $MSOL$-definable, cf. [Cou97]

E.4 Directed acyclic graphs

If we look at trees as directed graphs where there is exactly one node with indegree 0 and all others have indegree 1, the orientation is unique, hence counting those gives the same function.

Directed acyclic graphs (DAG's) my have vertices with arbitrary indegree and do not have to be connected. DAG's are again $MSOL$-definable, but not FOL-definable. Let $a_{n,m}$ be the number of labeled acyclic digraphs with exactly m vertices of indegree 0. The [HP73, Theorem 1.6.4] give

$$a_{n,m} = \sum_{k=1}^{n-m} (2^m - 1)^k 2^{m(n-m-k)} \binom{n}{m} a_{n-m,k}.$$

E.5 Bipartite graphs

Bipartite graphs are $MSOL$-definable, and so are connected bipartite graphs. We say that there is partition of the vertex set into two independent sets (and add the statement for connectedness). Let β_n be the number of labeled bipartite graphs. In [Wil90, Page 79ff]. we find that the exponential generating function associated with β_n satisfies the following identity:

$$\left(\sum_{n=0} \beta_n \frac{x_n}{n!}\right)^2 = \sum_{n=0} \left(\sum_{k} \binom{n}{k} 2^{k(n-k)}\right) \frac{x_n}{n!}$$

From [HP73, Page 17] we also get

$$\beta_n = \frac{1}{2} \sum_{k=1}^{n-1} \binom{n}{k} 2^{k(n-k)}.$$

E.6 k-colored graphs

k-colored graphs are $MSOL$-definable with a formula depending on k. We say that there is a partition of the vertices into k independent sets. Let γ_n^k denote the number of k-colored labeled graphs with n vertices. The case of bipartite graphs is a special case: $\beta_n = \gamma_n^2$. The formula for $beta_n$ is generalized in [HP73, Page 17]:

$$\gamma_n^k = \frac{1}{k} \sum_{k=1}^{n-1} \binom{n}{k} 2^{k(n-k)} \gamma_n^{k-1}$$

E.7 Even and eulerian graphs

Let $\mathcal{C} = EVENDEG$ the class of simple graphs where each vertex has even degree. $EVENDEG$ is not $MSOL$-definable but $CMSOL$-definable.

$f_{EVENDEG}(n) = 2^{\binom{n-1}{2}}$, cf. [HP73, page 11].

Let $\mathcal{C} = EULER$ the class of simple connected graphs in $EVENDEG$. $EULER$ is not $MSOL$-definable, but $CMSOL$-definable. In [HP73, page 7] 5stacs the following recurrence for $f_{\mathrm{EULER}}(n)$:

$$f_{\mathrm{EULER}}(n) = 2^{\binom{n-1}{2}} - \frac{1}{n}\sum_{k=1}^{n-1} k\binom{n}{k} 2^{\binom{n-k-1}{2}} f_{\mathrm{EULER}}(k).$$

is given. The number of labeled r-regular eulerian graphs is also $CMSOL$-definable. To find an explicit formula of its density function seems very hard. However, our Theorem 6 gives

Corollary 65 *The number of labeled r-regular eulerian graphs is, for every $m \in \mathbb{Z}$, ultimately constant* $\pmod{m}$.

E.8 Planar graphs

Planar graphs are $MSOL$-definable. To see this one can use Kuratowski's Theorem characterizing planar graphs with topological minors, cf. [Die90]. We have not found any formula countin the number of labeled planar graphs in the literature. But the Specker-Blatter Theorem and its variations can be applied.

A special kind of planar graphs are the *rectangular grids $GRIDS$*, which look like rectangular checker boards, with the north-south and east-west neighborhood relation. *Partial rectangular grids $PGRIDS$* are subgraphs of rectangular grids. It is easy to see that both $GRIDS$ and $PGRIDS$ have finite Specker index, but $GRIDS$ are $MSOL$-definable while $PGRIDS$ are not $CMSOL$-definable, cf. [Cou97,Rot98].

E.9 Perfect graphs

A graph is perfect of the for every induced subgraph (including the graph itself) the chromatics number equals the clique number. On the face of it, this does not seem $MSOL$- or $CMSOL$-definable. However, it was conjectured by Berge[2], [Bol99, Chapter V.5]

Conjecture 1 (Strong perfect graph conjecture). A graph $\mathcal{G}$ is perfect iff neither $\mathcal{G}$ nor its complement graph contains contains an odd cycle of size at least 5.

If the conjecture is true, this gives as a $MSOL$-definition of perfect graphs. However, the Specker index for perfect graphs is much smaller than one would get using the $MSOL$-definition.

Proposition 66 *Let $\mathcal{G}$ and $\mathcal{H}$ be graphs, and a is a vertex of $\mathcal{G}$.*
Then $Subst(\mathcal{G}, a, \mathcal{H})$ is perfect iff both $\mathcal{G}$ and $\mathcal{H}$ are perfect.

Proof. One direction follows from the definition, the other direction is by now classic, cf. [Bol99, Chapter V.5, Theorem 19].

Using Proposition 22 we get

Corollary 67 *The Specker index of perfect graphs is* 2.

[2] It was recently announced as proven by M. Cudnovski and R. Seymour