The Specker-Blatter Theorem Revisited

E. Fischer* and J.A. Makowsky**

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

Abstract. In this paper we study the generating function of classes of graphs and hypergraphs modulo a fixed natural number m. 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 $MSOL$ 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.
In this paper we show how the Specker-Blatter Theorem depends on the choice of constants and relations allowed in the definition of $\mathcal{C}$. Among the main results we have the following:

- For 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, a linear recurrence relation holds, irrespective of the arity of the relations involved.
- In all the results $MSOL$ can be replaced by $CMSOL$, Monadic Second Order Logic with (modular) Counting. This covers many new cases, for which such a recurrence relation was not known before.

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 due to E. Specker on counting labeled graphs (and more generally, labeled binary relational structures), first

* Partially supported by the VPR fund – Dent Charitable Trust – non-military research fund of the Technion-Israeli Institute of Technology.

** Partially supported by a Grant of the Fund for Promotion of Research of the Technion-Israeli Institute of Technology.

T. Warnow and B. Zhu (Eds.): COCOON 2003, LNCS 2697, pp. 90–101, 2003.

announced by C. Blatter and E. Specker 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 $R_1, \ldots, R_k$, 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. In the long survey version of the paper, [FM] we shall also give numerous examples, mostly taken from [HP73,Ges84,Wil90], 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 8 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 for every $\mathfrak{A}$ we have

$\mathfrak{A} \in \mathcal{C}$ if and only if $\mathfrak{A} \models \phi$. 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 $1 < 2 < \cdots < n$.

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

(i) *$\mathcal{C} = ORD$, the class of all linear orders, satisfies $f_{ORD}(n) = n!$. ORD is $FOL(R)$-definable.*

(ii) *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 $\{1, \ldots, n\}$.*

(iii) *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.*

(iv) *The class REG_r of simple regular graphs where every vertex has degree r is FOL-definable (for any fixed r). Counting the number of labeled regular graphs is treated completely in [HP73, Chapter 7]. For cubic graphs, the function is explicitly given in [HP73, page 175] as $f_{\mathcal{R}_3}(2n+1) = 0$ and*

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

which is ultimately 0 for every modulus m.

(v) *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. For $CONN$ [HP73, page 7] gives the following recurrence:*

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

(vi) *Counting labeled connected graphs is treated in [HP73, Chapter 1] and in [Wil90, Chapter 3]. But our Theorem 5 will give directly, that for every m this functionis ultimately 0 modulo m.*

[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 $K_{n,n}$ with both parts of equal size is definable in $MSOL_2$ and $f_{CBIPEQ}(2n) = \frac{1}{2}\binom{2n}{n}$.

(vii) *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].*

(viii) *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.

(ix) *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 was announced in [BS81], and proven in [BS84,Spe88]:

Theorem 1 (Specker and Blatter, 1981). *Let $\mathcal{C}$ be definable in Monadic Second Order Logic with unary and binary relation symbols only. For every $m \in \mathbb{N}$, there are $d_m, a_j^{(m)} \in \mathbb{N}$ such that the function $f_{\mathcal{C}}$ satisfies the linear recurrence relation $f_{\mathcal{C}}(n) \equiv \sum_{j=1}^{d_m} a_j^{(m)} f_{\mathcal{C}}(n-j) \pmod m$, and hence is ultimately periodic modulo m.*

The case of ternary relation symbols, and more generally of arity $k \geq 3$, was left open in [BS84,Spe88] and appears 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. [Fis03].

Theorem 2 (Fischer, 2002). *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 (1) above.

1.4 Improvements and Variations

The purpose of this paper is to explore variations and extensions of the Specker-Blatter Theorem.

First, we note that in the case of unary relations symbols, $MSOL_{ord}$ and $CMSOL_{ord}$ have the same expressive power, [Cou90], and $MSOL_{ord}$-sentences define exactly the regular languages. Schützenberger's Theorem characterizes 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]. Hence, the Specker-Blatter Theorem has an important precursor in formal language theory, reformulated for our purposes as:

Theorem 3 (Schützenberger).
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 4. *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 sketched in Section 3.2.

Theorem 4 covers cases not covered by the Specker-Blatter Theorem (Theorem (1). Although $EVENDEG$ is not $MSOL$-definable, it is $CMSOL$-definable, and its function satisfies $f_{EVENDEG}(n+1) = f_{GRAPHS}(n)$. However, it seems not very obvious that the function f_{EULER} satisfies modular recurrence relations.

Finally, we study the case of relations of bounded degree.

Definition 1.

(i) 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.*

(ii) *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.*

(iii) *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 $Gaif(\mathfrak{A})$.*

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

Theorem 5. *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 given in Section 4.

2 Variations and Counterexamples

2.1 Why Modular Recurrence?

Theorem 1 provides linear recurrence relations modulo m for every $m \in \mathbb{N}$. Theorem 3 provides a uniform linear recurrence relation over $\mathbb{Z}$.

For the following FOL-definable $\mathcal{C}$, with one binary relation symbol, $f_{\mathcal{C}}(n)$ 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 $f_{\mathcal{C}}(n) = n!$.

This follows from the well known fact, cf. [LN83], that every function $f : \mathbb{Z} \to \mathbb{Z}$, which satisfies a linear recurrence relation $f(n+1) = \sum_{i=0}^{k} a_i f(n-i)$ over $\mathbb{Z}$, grows at most exponentially, i.e. there is a constant $c \in \mathbb{Z}$ such that $f(n) \leq 2^{cn}$.

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 $f_{EQCLIQUE_1}(n)$ and $f_{REG_3}(n)$ satisfy trivial modular recurrences. 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 6. *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

The example $EQ_pCLIQUE$, cf. Example 1 (1) is definable in Second Order Logic with existential quantification over one binary relation, but it is not $CMSOL$ definable.

Let p be a prime number $b_p(n) = f_{EQ_pCLIQUE}(n) = f_{EQ_pCLIQUE_2}(n)$ 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 7. *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 [Fis03].

Therefore, $f_{EQ_pCLIQUE}$ 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 3 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 1, with Proposition 7.

Proposition 8. *$EQ_pCLIQUE_2$ 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 4 also holds for $CMSOL_{lab}$. This is proven using standard reduction techniques, and the proof is omitted.

Proposition 9. *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].

3 *DU*-Index and Specker Index

Specker's proof of Theorem 1 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.

3.1 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].

Definition 2. *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. The disjoint union of two structures $\mathfrak{A}$ and $\mathfrak{B}$ is denoted by $\mathfrak{A} \sqcup \mathfrak{B}$.

Definition 3. *Let $\mathcal{C}$ be a class of, possibly pointed, $\bar{R}$-structures. We define two equivalence relations between $\bar{R}$-structures:*

- *We say that $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are $Su(\mathcal{C})$-*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 $Subst((\mathfrak{S}, s), \mathfrak{A}_2) \in \mathcal{C}$.*
- *Similarly, We say that $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are $DU(\mathcal{C})$-*equivalent*, denoted $\mathfrak{A}_1 \sim_{DU(\mathcal{C})} \mathfrak{A}_2$, if for every structure $\mathfrak{B}$ we have that $\mathfrak{A}_1 \sqcup \mathfrak{B} \in \mathcal{C}$ iff $\mathfrak{A}_2 \sqcup \mathfrak{B} \in \mathcal{C}$.*
- *The* Specker index *(resp. DU-index) of $\mathcal{C}$ is the number of equivalence classes of $\sim_{Su(\mathcal{C})}$ (resp. of $\sim_{DU(\mathcal{C})}$).*

Specker's proof in [Spe88] of Theorem 1 has a purely combinatorial part:

Lemma 10 (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 $m \in \mathbb{N}$.*

In Section 4 we shall prove an analogue of Specker's lemma (Theorem 14) for $\mathcal{C}$ of finite D-index with structures of bounded degree, which generalizes a similar statement du to I. Gessel, [Ges84].

3.2 Classes of Finite Specker Index

Clearly, if $\mathcal{C}$ has finite Specker index, then it has finite DU-index. Furthermore, every class of connected graphs has DU-index 2. The class $EQ_2CLIQUE$ has an infinite Specker index.

Let the class $CONN - EQ_2CLIQUE$ be the class of all graphs obtained from members of $EQ_2CLIQUE$ by connecting any two vertices from different cliques. We note that $CONN - EQ_2CLIQUE$ contains structures of arbitrary large degree. The class $CONN - EQ_2CLIQUE$ has DU-index 2, but infinite Specker index. It is an easy exercise to show the same for the class of graphs which contain a hamiltonian cycle. None of these classes with an infinite Specker index are $CMSOL$-definable. 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 11. *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 [FM]. It uses a form of the Feferman-Vaught Theorem for $CMSOL$ due to Courcelle, [Cou90], see also [Mak01]. Specker[2] noted that there is a continuum of classes (of graphs, of $\bar{R}$-structures) of finite Specker index which are not $CMSOL$-definable.

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

Definition 4. *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 12. *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 $\mathcal{K}^0_j$ the equivalence classes and for each $j \leq \alpha$

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

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

Corollary 13.

(i) If $\mathcal{C}_0, \mathcal{C}_1$ are classes of finite $\mathcal{F}$-index, then so are all their boolean combinations.

[2] Personal communication

(ii) 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.

Proof. For (i) take the coarsest common refinement of the $\mathcal{F}(\mathcal{C}_0)$-equivalence and the $\mathcal{F}(\mathcal{C}_1)$-equivalence relations.
(ii) is left to the reader. □

4 Structures of Bounded Degree

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. In this section we prove Theorem 5 in the following form:

Theorem 14. *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, $f_{\mathcal{C}}^{(d)}(n)$ 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 15. *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 14 we define orbits for permutation groups.

Definition 5. *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 16. *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 17. *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 17 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 16 allows us to prove the following.

Lemma 18. *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 16 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 15 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 (of Theorem 14:). We use Lemma 18: 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 18 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 16. 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 $f_{\mathcal{C}}^{(d)}(n)$ ultimately vanishes modulo m. □

Acknowledgment. We are grateful to E. Specker, for his encouragement and interest in our work, and for his various suggestions and clarifications.

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.

[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 théorie 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.

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

[Fis03] E. Fischer. The Specker-Blatter theorem does not hold for quaternary relations. *Journal of Combinatorial Theory, Series A*, 2003. in press.

[FM] E. Fischer and J.A. Makowsky. The Specker-Blatter theorem revisited. in preparation.

[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.

[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`.

[Mak01] J.A. Makowsky. Algorithmic uses of the Feferman-Vaught theorem. Lecture delivered at the Tarski Centenary Conference, Warsaw, May 2001, paper submitted to APAL in January 2003, special issue of the conference, 2001.

[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.