OMITTING TYPES, BOUNDED WIDTH AND THE ABILITY TO COUNT

BENOIT LAROSE, MATT VALERIOTE, AND LÁSZLÓ ZÁDORI

Abstract. We say that a finite algebra $\mathbb{A} = \langle A; F\rangle$ has the *ability to count* if there are subalgebras C of $\mathbb{A}^3$ and Z of $\mathbb{A}$ such that the structure $\langle A; C, Z\rangle$ has the ability to count in the sense of Feder and Vardi. We show that for a core relational structure $\mathbf{A}$ the following conditions are equivalent: (i) the variety generated by the algebra $\mathbb{A}$ associated to $\mathbf{A}$ contains an algebra with the ability to count; (ii) $\mathbb{A}^2$ has the ability to count; (iii) the variety generated by $\mathbb{A}$ admits the unary or affine type. As a consequence, for CSP's of finite signature, the bounded width conjectures stated in Feder-Vardi [9], Larose-Zádori [16] and Bulatov [4] are identical.

1. Introduction

Constraint Satisfaction Problems (CSP's) provide a convenient framework in which to express several standard combinatorial problems such as graph colouring, graph reachability, satisfiability and so on. Roughly speaking, a CSP consists of a set of constraints placed on a collection of variables, and one must decide if values can be assigned to the variables so as to satisfy all constraints. In general this class of problems is NP-complete but by restricting the types of constraints one may obtain tractable subclasses. It is also convenient to consider CSP's as homomorphism problems: given a fixed relational structure $\mathbf{A}$, $CSP(\mathbf{A})$ denotes the decision problem consisting of all structures $\mathbf{A}'$ that admit a homomorphism to $\mathbf{A}$. The question is to determine the complexity of $CSP(\mathbf{A})$ in terms of the structure $\mathbf{A}$.

In particular, a central problem in the study of these non-uniform constraint satisfaction problems is the *Dichotomy Conjecture*, that asserts that for every structure $\mathbf{A}$, $CSP(\mathbf{A})$ is either polynomial-time solvable or NP-complete [8]. Problems of so-called *bounded width* form a large class of tractable CSP's. These problems, which are equivalently described in terms of the query language Datalog or as those problems having bounded treewidth duality, are solvable by local consistency methods. An important problem in the area is to determine whether the property of having bounded width is actually decidable (see [9, 5].) In their seminal paper [9], Feder

Date: September 10, 2008.

2000 *Mathematics Subject Classification.* Primary 68Q25; Secondary 08B05.

Key words and phrases. constraint satisfaction problems, bounded width, variety, Hobby-McKenzie types, monotone circuits.

The first author's research is supported by grants from NSERC, FQRNT and the Centre de Recherches Mathématiques.

The second author's research is supported by a grant from NSERC.

The third author's research is supported by OTKA grants no. T48809 and K60148.

Part of this research was conducted while the first author was visiting the Bolyai Institute in Szeged. The support of NSERC, OTKA and the Bolyai Institute is gratefully acknowledged.

and Vardi introduced the notion of "ability to count" for CSP's. This simple combinatorial property is sufficient to guarantee that a CSP does not have bounded width. Two proofs of this fact are found in [9], one of them relying on Razborov's monotone circuit lower bound for matching [18]. A related result can be found in Atserias, Bulatov and Dawar [2]: it is proved there that the problem of existence of solutions for systems of equations over an Abelian group cannot be expressed in certain counting logics, which properly extend the language Datalog.

In [16], the universal algebraic aspects of CSP's with bounded width are investigated. To the problem $CSP(\mathbf{A})$ one associates naturally an algebra on the universe A whose basic operations are those that are compatible with the basic relations of $\mathbf{A}$. It is proved in [16] that if $CSP(\mathbf{A})$ has bounded width, then the associated algebra must satisfy certain special identities; it is conjectured that in fact this condition, that the variety generated by the algebra associated to the CSP omits the unary and affine types, is equivalent to having bounded width (see Conjecture 1 below). A similar conjecture is described in [4]: a colour scheme is defined on CSP's: certain pairs of elements are called edges, some of which are red, yellow or blue. A variant of the notion of bounded width called *relational width* (which for CSP's of finite signature is equivalent to bounded width, see [5]) is conjectured to be equivalent to the absence of blue edges. This conjecture, for CSP's of finite signature, is equivalent to Conjecture 1 [6]. In the paper mentioned earlier Feder and Vardi outline a conjecture stating that the ability to count essentially captures those CSP's that do not have bounded width. In the present paper we will make this connection more precise, and show that it actually matches the algebraic conjectures mentioned above.

We now give a detailed outline of the contents of the paper. In section 2 we present some required basic concepts and definitions; in particular, we introduce the notion of an algebra with the ability to count (Definition 2.4). In section 3, we prove a special case of our main result, namely that a finite idempotent algebra has the ability to count precisely if the variety it generates admits either the unary or affine type (Theorem 3.1). In section 4, we extend the result to the general case: we introduce the notion of finite core algebra (Definition 4.1), and show that the variety generated by a finite, core algebra contains an algebra with the ability to count if and only if it admits the unary or affine type (Theorem 4.7). As a byproduct of our investigations we also show that there is a polynomial-time algorithm to determine whether the variety generated by a finite core algebra omits the types in an order ideal of types (Corollary 4.5). In particular, this shows that having the ability to count is a decidable property for idempotent algebras. In section 5 we present an interesting connection to the circuit complexity of CSP's (Proposition 5.1). Finally, in section 6, we discuss the implications and limitations of our results.

2. Preliminaries

In this section we introduce the notation and concepts we require for our results (see also [5].)

2.1. **CSP's and the ability to count.** Let $\tau = \{R_1, \dots, R_m\}$ be a *vocabulary*, i.e., a finite set of *relational symbols*. Each relational symbol R_i has a positive integer r_i associated to it called its *arity*. A *τ-structure* is a relational structure $\mathbf{A} = \langle A; R_1(\mathbf{A}), \dots, R_m(\mathbf{A})\rangle$ where $R_i(\mathbf{A}) \subseteq A^{r_i}$ for each $1 \leq i \leq m$. Throughout the paper we use the same boldface and capital letters to denote a structure and its universe, respectively. A *homomorphism* from a τ-structure $\mathbf{A}'$ to a τ-structure $\mathbf{A}$ is a mapping $h : A' \to A$ such that for every r-ary $R \in \tau$ and every $(a_1, \ldots, a_r) \in R(\mathbf{A}')$, we have $(h(a_1), \ldots, h(a_r)) \in R(\mathbf{A})$.

Given a τ-structure $\mathbf{A}$, we let $CSP(\mathbf{A})$ denote the class of all τ-structures $\mathbf{A}'$ that admit a homomorphism to $\mathbf{A}$. A structure $\mathbf{A}$ is a *core* if every homomorphism $h : \mathbf{A} \to \mathbf{A}$ is onto. It is easy to see that for every structure $\mathbf{A}$ there exists a core structure $\mathbf{A}'$, unique up to isomorphism, such that $CSP(\mathbf{A}) = CSP(\mathbf{A}')$. We denote the class of all τ-structures that do not admit a homomorphism to $\mathbf{A}$ by $\neg CSP(\mathbf{A})$.

Datalog was originally introduced as a database query language. We view it here simply as a means to define classes of τ-structures. Let τ be a signature. A Datalog program over the signature τ consists of a finite set of *rules* of the form $h \leftarrow b_1 \wedge \ldots \wedge b_k$ where each of the b_i and h are atomic formulas of the form $R(x_{j_1}, \ldots, x_{j_r})$ where R is a relational symbol from the signature $\tau'' = \tau \cup \tau'$ where τ' is disjoint from τ. The left side of the rule is called the *head* of the rule, and the righthand side is the *body*. Symbols from τ' are called *intensional database predicates* (IDBs) while the symbols in τ, which can occur only in the body of a rule, are called *extensional database predicates* (EDBs). Roughly speaking, a Datalog program receives a τ-structure as input, and computes recursively the contents of the IDB's. We are interested here in Datalog programs equipped with a special 0-ary IDB which signals, when it becomes non-empty, that the input τ-structure is *accepted* (precise definitions of the semantics of Datalog can be found in [12, 7], see also [9]). It is immediate from the definition that the class of structures accepted by a Datalog program is homomorphism closed, i.e., if there is a homomorphism $\mathbf{A} \to \mathbf{B}$ and $\mathbf{A}$ is accepted then so is $\mathbf{B}$.

Definition 2.1. *Let $\mathbf{A}$ be a τ-structure. We say that $CSP(\mathbf{A})$ has* bounded width *if $\neg CSP(\mathbf{A})$ is definable in Datalog, i.e., if there exists a Datalog program that accepts precisely those structures that do not admit a homomorphism to $\mathbf{A}$.*

Various equivalent formulations of bounded width may be found in [5]. If $CSP(\mathbf{A})$ has bounded width then it is tractable; for instance well-known decision problems such as 2-colouring, HORN 3-SAT, directed and undirected reachability all have bounded width. However, other standard problems in PTIME such as solving linear equations over finite fields are known not to have bounded width. In [9], Feder and Vardi investigate the problem of determining which CSP's have bounded width and introduce the concept of the "ability to count" as an attempt to capture those problems that do not have bounded width. It is still open whether the property of bounded width is actually decidable.

The ability to count is a natural combinatorial generalisation of the following elementary property of an Abelian group: suppose that we are given a system of linear equations on a finite Abelian group, such that each equation is of the form (i) $x + y + z = \alpha$ or (ii) $x = 0$, where α is some non-zero element of the group. If we can find two sets L and R of equations of the system such that each variable appears in exactly one equation from L and one equation from R, and furthermore the set L contains one more equation of form (i) than R, then the system has no solution. Indeed, it suffices to sum all equations in L and R respectively, yielding

$$\sum_{x \in X} x = k\alpha \text{ and } \sum_{x \in X} x = (k-1)\alpha$$

for some positive integer k; subtracting the second equation from the first yields $0 = \alpha$, contrary to our choice of α; hence the system cannot have a solution.

The following definition is sketched on p. 85 of [9]:

Definition 2.2. *Let $\mathbf{A} = \langle A; C, Z\rangle$ be a structure where C is a 3-ary relation and Z is unary. We say that $\mathbf{A}$* has the ability to count *if the following two conditions hold:*

(1) *A contains elements 0 and 1 such that*
 (a) $0 \in Z$*,*
 (b) $(0,0,1)$*,* $(0,1,0)$ *and* $(1,0,0) \in C$*;*

(2) *Suppose $\mathbf{A}' = \langle A'; C', Z'\rangle$ is a structure similar to $\mathbf{A}$, such that there exist subsets L and R of $C' \cup Z'$ and the following conditions are satisfied:*
 (a) *every a' in A' appears in exactly one tuple from L and exactly one tuple from R;*
 (b) *there is precisely one more element in $L \cap C'$ than in $R \cap C'$.*

 Then there is no *homomorphism from $\mathbf{A}'$ to $\mathbf{A}$.*

By extension we say that a structure $\mathbf{B}$ has the ability to count *if among its basic relations there are relations C and Z such that $\langle B; C, Z\rangle$ has the ability to count.*

For example, consider the structure $\mathbf{A}' = \langle A'; C', Z'\rangle$ where $A' = \{x, y, z\}$, $C' = \{(x, y, z)\}$ and $Z' = \{x, y, z\}$. We may partition $C' \cup Z'$ as $L = \{(x, y, z)\}$ and $R = \{x, y, z\}$ (of course, the elements of R are viewed as 1-ary tuples.) Hence if the structure $\mathbf{A} = \langle A; C, Z\rangle$ has the ability to count then there is no homomorphism from $\mathbf{A}'$ to $\mathbf{A}$. In particular, we conclude that no triple of elements from Z can be in C, and that $1 \notin Z$ (note that this implies that $0 \neq 1$.) Notice also that it is not clear at first glance that the ability to count is a decidable property of a structure since there is no a priori bound on the size of the input structures.

Feder and Vardi show that if a structure $\mathbf{A}$ has the ability to count, then the problem $\neg CSP(\mathbf{A})$ does not have bounded width [9]. They proceed to conjecture that in essence, all CSP's without bounded width should occur in this way. One of our goals in this paper is to make this link precise, and to show that this conjecture is equivalent to the bounded width conjecture (Conjecture 1 below.) For this we shall require some algebraic machinery that we develop next.

2.2. **CSP's and algebras.** An n-ary operation on the set A is a map $f : A^n \to A$. If θ is a k-ary relation on A we say that f is a *polymorphism* of θ (or *preserves* θ, or that θ is *invariant* under f) if, given any $k \times n$ matrix M whose columns are in θ, applying f to the rows of M yields a tuple in θ. By extension, if Γ is a set of relations on A then f is a polymorphism of Γ if it is a polymorphism of every relation in Γ, and we denote by $Pol(\Gamma)$ the set of all these operations. To every τ-structure $\mathbf{A}$ is naturally associated an (non-indexed) algebra $\mathbb{A}_{\mathbf{A}} = \langle A; Pol(\Gamma)\rangle$ where A is the universe of $\mathbf{A}$ and Γ consists of all the basic relations of $\mathbf{A}$. If we fix some ordering of the fundamental operations of $\mathbb{A}_{\mathbf{A}}$ and view this algebra as an indexed algebra, we can consider standard algebraic constructions such as homomorphic images, subalgebras and products. A *variety* is a class of similar algebras closed under homomorphic images, subalgebras and products; the variety *generated* by the algebra $\mathbb{A}$, denoted by $\mathcal{V}(\mathbb{A})$, is the smallest variety containing $\mathbb{A}$; if $\mathbb{A}$ is finite, then the finite members of $\mathcal{V}(\mathbb{A})$ are precisely those algebras obtained by taking homomorphic images of subalgebras of finite powers of $\mathbb{A}$.

An n-ary operation f on a set A is *idempotent* if it satisfies $f(x,...,x) = x$ for all $x \in A$, i.e., if it preserves all one-element subsets of A. An algebra $\mathbb{A}$ is idempotent if all its basic operations are idempotent, and a variety of algebras is idempotent if all of its members are. For an arbitrary algebra $\mathbb{A}$, its *full idempotent reduct*, denoted by $\mathbb{A}^\diamond$, is the algebra whose fundamental operations are all idempotent term operations of $\mathbb{A}$. In particular for a structure $\mathbf{A}$, the full idempotent reduct of the algebra $\mathbb{A}_\mathbf{A}$ is the algebra on the same universe whose fundamental operations are the idempotent operations in $Pol(\Gamma)$ where as above Γ is the set of basic relations of $\mathbf{A}$; put differently, the full idempotent reduct of $\mathbb{A}_\mathbf{A}$ is the algebra $\mathbb{A}_{\mathbf{A}'}$ where $\mathbf{A}'$ is the structure obtained from $\mathbf{A}$ by adding all the one-element unary relations. If the structure $\mathbf{A}$ is a core, then it is known that the problems $CSP(\mathbf{A})$ and $CSP(\mathbf{A}')$ are equivalent under first-order reductions[1], and furthermore, if one of these CSP's has bounded width then so does the other (see [15], Theorem 2.1.)

An operation f on A is *affine* if there exists an Abelian group structure $\mathcal{G} = \langle A, +, -, 0\rangle$ on A such that f commutes with the operation $x - y + z$, i.e., f can be written in the form

$$f(x_1, \ldots, x_n) = a + \sum_i a_i x_i$$

for some $a_i \in End(\mathcal{G})$; if f is idempotent then $a = 0$ and $\sum a_i = 1$. An algebra is *affine* if there is an abelian group structure on its base set such that (i) $m(x, y, z) = x - y + z$ is a term operation of the algebra and (ii) every term operation of the algebra is an affine operation. Equivalently, an idempotent algebra is affine if and only if it is the full idempotent reduct of a module.

Let $\mathbb{A}$ be an algebra. The *term* (*polynomial*) operations of $\mathbb{A}$ are all operations on A that can be constructed from projections and the fundamental operations of $\mathbb{A}$ (and constant operations) using composition. Two algebras are said to be *term-equivalent* (*polynomially equivalent*) if they have the same universe and the same term (polynomial) operations. Notice that for any structure $\mathbf{A}$ the term operations of the algebra $\mathbb{A}_\mathbf{A}$ are precisely its fundamental operations.

Tame congruence theory, developed by Hobby and McKenzie [11], is a powerful tool for the analysis of finite algebras. Every finite algebra has a *typeset*, which describes the local behaviour of the algebra, which consists of one or more of the following 5 *types*: (**1**) the unary type, (**2**) the affine type, (**3**) the Boolean type, (**4**) the lattice type and (**5**) the semilattice type. The typeset of a variety is the union of the typesets of its finite members. We say that a variety *admits type* **i** if its typeset contains the type **i**; otherwise the variety is said to *omit type* **i**. There is a very tight connection between the kind of equations that are satisfied by the algebras in a variety and the types that are admitted (omitted) by a variety.

The next result is a special case of Lemma 3.1 in [22]; we shall need it in the proof of our main result, and we state it here also to give the reader at least some partial insight into the nature of algebras whose variety admits the unary or affine types. We'll require a few preliminary definitions and results.

A *divisor* of an algebra $\mathbb{A}$ is a homomorphic image of a subalgebra of $\mathbb{A}$. An algebra is *strictly simple* if it has no divisors other than itself or one-element algebras. A strictly simple algebra has a unique type associated to it. A. Szendrei has characterised all idempotent strictly simple algebras ([19] Theorem 6.1), in particular,

[1] To be precise we also have to assume that the basic relations of $\mathbf{A}$ are irredundant, but this is only a minor technicality.

up to term equivalence, the only idempotent strictly simple algebra of unary type is the 2-element set, i.e., the 2-element algebra with no basic operations $\langle\{0,1\};\emptyset\rangle$. The strictly simple idempotent algebras of affine type are affine algebras. We'll need the following result in section 3:

Proposition 2.3 ([22]). *Let $\mathbb{A}$ be a finite, idempotent algebra. Then $\mathcal{V}(\mathbb{A})$ admits the unary type or the affine type if and only if $\mathbb{A}$ has a divisor which is either affine or term equivalent to the 2-element set.*

We may now state the bounded width conjecture:

Conjecture 1 ([16]). *Let $\mathbf{A}$ be a core structure. Then $CSP(\mathbf{A})$ has bounded width if and only if $\mathcal{V}(\mathbb{A}_{\mathbf{A}})$ omits the unary and affine types.*

For our purposes, it will be convenient to extend the notion of ability to count to algebras.

Definition 2.4. *Let $\mathbb{A}$ be a finite algebra with universe A. We say that $\mathbb{A}$ has* the ability to count with C and Z *if C is a subuniverse of $\mathbb{A}^3$ and Z is a subuniverse of $\mathbb{A}$ such that the structure $\langle A; C, Z\rangle$ has the ability to count. We'll say $\mathbb{A}$ has* the ability to count *if there exist C and Z such that the above holds.*

It is easy to verify the following:

Proposition 2.5. *Let $\mathbb{A}$ be a finite algebra. If some divisor of $\mathbb{A}$ has the ability to count then so does $\mathbb{A}$.*

Proof. Suppose that $\mathbb{B}$ is a subalgebra of $\mathbb{A}$ with the ability to count, with relations C and Z. Then C and Z are subuniverses of $\mathbb{A}^3$ and $\mathbb{A}$ respectively, and the structure $\langle A; C, Z\rangle$ has the ability to count: this follows easily from the fact that if a structure $\mathbf{C}$ admits a homomorphism into $\langle A; C, Z\rangle$, then it admits one whose image is contained in B. Indeed, if an element of $\mathbf{C}$ appears in a tuple then it must be mapped to B, and otherwise it can be mapped to any element of B. Hence $\mathbb{A}$ has the ability to count. If π is a homomorphism of $\mathbb{A}$ onto $\mathbb{D}$ where $\mathbb{D}$ has the ability to count with relations C and Z, then π is a structure homomorphism from $\langle A; C', Z'\rangle$ to $\langle D; C, Z\rangle$, where $C' = \pi^{-1}(C)$ and $Z' = \pi^{-1}(Z)$ are subuniverses of $\mathbb{A}^3$ and $\mathbb{A}$ respectively. It follows immediately that $\mathbb{A}$ has the ability to count. □

3. Idempotent algebras and the ability to count

The main result of this section is the following:

Theorem 3.1. *Let $\mathbb{A}$ be a finite, idempotent algebra. Then the following are equivalent:*

1. *$\mathcal{V}(\mathbb{A})$ contains an algebra with the ability to count;*
2. *$\mathbb{A}$ has the ability to count;*
3. *$\mathcal{V}(\mathbb{A})$ admits the unary or affine type.*

Before we prove the theorem, we require two auxiliary results, the first of which is immediate.

Lemma 3.2. *If $\mathbb{A}$ has the ability to count with C and Z, and C', Z' are subuniverses of $\mathbb{A}^3$ and $\mathbb{A}$ respectively such that*

$$\{(0,0,1),(0,1,0),(1,0,0)\} \subseteq C' \subseteq C \text{ and } \{0\} \subseteq Z' \subseteq Z$$

then $\mathbb{A}$ has the ability to count with C' and Z'.

Lemma 3.3. *Let $\mathcal{V}$ be an idempotent variety that contains an algebra with the ability to count. Then $\mathcal{V}$ contains an algebra $\mathbb{A}$ with the ability to count with C and Z satisfying the following conditions:*

1. *$\mathbb{A}$ is generated by $\{0,1\}$;*
2. *C is symmetric, i.e., if $(x_1,x_2,x_3) \in C$ then $(x_{\pi(1)},x_{\pi(2)},x_{\pi(3)}) \in C$ for any permutation π;*
3. *$Z = \{0\}$;*
4. *any tuple of C is uniquely determined by any two of its entries;*
5. *for every $x \in A$ there exists a unique x' such that $(x,x',0) \in C$.*

Proof. (1) Let $\mathbb{A} \in \mathcal{V}$ be an algebra with the ability to count with some C and Z, chosen so that its universe A has minimal size. Let $\mathbb{A}'$ be the subalgebra of $\mathbb{A}$ generated by 0 and 1 and let A' denote its universe. Let $C' = C \cap A'^3$ and let $Z' = Z \cap A'$. It is easy to verify that the algebra $\mathbb{A}'$ has the ability to count with C' and Z'. By minimality we conclude that $\mathbb{A} = \mathbb{A}'$.

(2) Let C' be the subuniverse of $\mathbb{A}^3$ generated by the tuples $(1,0,0)$, $(0,1,0)$ and $(0,0,1)$. It is clear that $C' \subseteq C$ is symmetric and the result follows from the lemma.

(3) The claim is immediate from the last lemma since $\mathbb{A}$ is idempotent and hence $Z' = \{0\}$ is a subuniverse of $\mathbb{A}$.

(4) Define the following relations:

$$\alpha_0 = \{(x,y) : \exists u,v\ (x,u,v),(y,u,v) \in C\}$$

and let α be the transitive closure of α_0. Clearly α_0 is symmetric. It is also reflexive: indeed, let A' be the projection of C on its first coordinate. Then A' is a subuniverse of $\mathbb{A}$ containing 0 and 1, so by (1) $A' = A$ and hence α_0 is reflexive. It follows that α is a congruence of the algebra. Let $\pi : \mathbb{A} \to \mathbb{A}/\alpha$ be the canonical homomorphism. Let $0' = \pi(0)$ and let $1' = \pi(1)$. Finally let $C' = \pi(C)$ and $Z' = \{0'\}$. Clearly C' and Z' are subuniverses of $(\mathbb{A}/\alpha)^3$ and $\mathbb{A}/\alpha$ respectively. We prove that $\langle A/\alpha; C', Z'\rangle$ has the ability to count.

Clearly $(0',0',1')$, $(0',1',0')$, and $(1',0',0')$, are in C'. Let $\langle X; C'', Z''\rangle$ be a structure of the same type with sets L and R as specified in the definition of the ability to count; we must show that it does not admit a homomorphism to $\langle A/\alpha; C', Z'\rangle$.

Create a new structure of the same type as follows. Since A is finite there exists a positive integer k such that, for every $(a,b) \in \alpha$ there exists $c_0,\dots,c_{k+1} \in A$ such that $a = c_0$, $b = c_{k+1}$ and $(c_{i-1},c_i) \in \alpha_0$ for all $1 \le i \le k+1$.

The universe of our new structure is $Y = X \times \{0,\dots,k+1,0',\dots,k',0'',\dots,k''\}$ where $0,\dots,k,k+1,0',\dots,k',0'',\dots,k''$ are all distinct; for ease of notation we'll denote the elements (x,i), (x,i') and (x,i'') by x_i, x'_i and x''_i respectively. We define two sets L' and R' as follows: L' is the union of the following sets:

$$\{(x_0,y_0,z_0) : (x,y,z) \in L\},$$
$$\{x_0 : x \in L\},$$
$$\{(x_{i+1},x'_i,x''_i) : x \in X,\ 0 \le i \le k\};$$

the set R' is the union of the following sets:

$$\{(x_{k+1},y_{k+1},z_{k+1}) : (x,y,z) \in R\},$$

$$\{x_{k+1} : x \in R\},$$

$$\{(x_i, x'_i, x''_i) : x \in X,\, 0 \leq i \leq k\}.$$

The relation C_0 is defined as $Y^3 \cap (L' \cup R')$ and let $Z_0 = Y \cap (L' \cup R')$. Notice that by hypothesis on L and R, the sets L' and R' satisfy the following conditions: every element of Y appears in exactly one tuple of each set, and one of the sets contains one more triple than the other.

We claim that if there exists a homomorphism $\phi : \langle X; C'', Z''\rangle \to \langle A/\alpha; C', Z'\rangle$ then there is a homomorphism $f : \langle Y; C_0, Z_0\rangle \to \langle A; C, Z\rangle$. Indeed, for every triple $(a_1, a_2, a_3) \in C''$, fix a triple $(a'_1, a'_2, a'_3) \in C$ such that $a'_i \in \phi(a_i)$ for all $1 \leq i \leq 3$. Now define f as follows: for any $x \in X$, one of two cases holds: (i) there exists a unique triple in L with x appearing in it, say $(a_1, a_2, a_3) \in L$ with $x = a_j$; then define $f(x_0) = a'_j$; otherwise (ii) $x \in L$, and then let $f(x_0) = 0$. Similarly, either (i) there exists a unique triple in R with x appearing in it, say $(a_1, a_2, a_3) \in R$ with $x = a_j$; then define $f(x_{k+1}) = a'_j$; otherwise (ii) $x \in R$, and then let $f(x_{k+1}) = 0$. Since $f(x_0)$ and $f(x_{k+1})$ belong to the same α block, it means we can find elements $z_0, \dots, z_{k+1} \in A$ such that $f(x_0) = z_0$, $f(x_{k+1}) = z_{k+1}$ and $(z_i, z_{i+1}) \in \alpha_0$ for all $0 \leq i \leq k$. This in turn means there exist elements $u_i, v_i \in A$, $0 \leq i \leq k$, such that (z_i, u_i, v_i) and (z_{i+1}, u_i, v_i) belong to C for all $0 \leq i \leq k$. Now define f in the obvious way: let $f(x'_i) = u_i$ and $f(x''_i) = v_i$ for all $0 \leq i \leq k$ and let $f(x_i) = z_i$ for all $1 \leq i \leq k$. It is clear that f is a well-defined homomorphism, contradicting the fact that $\langle A; C, Z\rangle$ has the ability to count. It follows that ϕ cannot exist and this concludes the proof.

(5) Uniqueness follows from (4) so all we need to show is existence. Define $A' = \{x : \exists y\, (x, y, 0) \in C\}$. Then A' contains 0 and 1, and by (1) we conclude that $A' = A$.

□

Proof of Theorem 3.1. (3) ⇒ (2): suppose that $\mathcal{V}(\mathbb{A})$ admits the unary or affine type. Then by Proposition 2.3 there exist a divisor $\mathbb{D}$ of $\mathbb{A}$ which is an affine algebra or a two-element set. We define relations C and Z on the universe D of $\mathbb{D}$ such that the structure $\langle D; C, Z\rangle$ has the ability to count. Indeed, in both cases there exists an Abelian group structure on D and a non-zero element α (any will do) such that $C = \{(x, y, z) : x + y + z = \alpha\}$ is a subuniverse of $\mathbb{D}^3$ and $Z = \{0\}$ is a subuniverse of $\mathbb{D}$. This is trivial in the case where $\mathbb{D}$ is a set since for any n every n-ary relation is a subuniverse of $\mathbb{D}^n$, and in the affine case, it is a simple exercise to verify that idempotent operations that commute with $m(x, y, z) = x - y + z$ preserve C and Z as defined above. It follows that $\mathbb{D}$ has the ability to count, and hence so does $\mathbb{A}$ by Proposition 2.5.

(2) ⇒ (1): trivial.

(1) ⇒ (3): Suppose that $\mathbb{A}$ is an algebra in $\mathcal{V}$ with the ability to count with C and Z. We may assume that C and Z satisfy all of the conditions (1)-(4) of Lemma 3.3. As we noted earlier, we have that $0 \neq 1$; let α be a congruence of $\mathbb{A}$ which is maximal with the property that $(0, 1) \notin \alpha$, and let β be a cover of α. By Theorem 2.8 (4) of [11], there exists a polynomial retraction r of A onto an (α, β)-minimal set U such that $(r(0), r(1)) \in \beta \setminus \alpha$.

Suppose for a contradiction that the $(\alpha|_U, \beta|_U)$-minimal algebra $\mathbb{A}|_U$ is of type **3**, **4** or **5**. By Lemmas 4.15 and 4.17 of [11], in each case there exists a pseudo-meet operation u of $\mathbb{A}$; in particular, u is a binary polynomial operation of $\mathbb{A}$ that acts on

$\{r(0), r(1)\}$ as a semilattice operation. Since $u(r(x), r(y))$ is a binary polynomial of $\mathbb{A}$, there exist an $(n+2)$-ary term t of $\mathbb{A}$ and a fixed n-tuple a over A such that $u(r(x), r(y)) = t(x, y, a)$ for all $x, y \in A$. Recall from Lemma 3.3 (4) that for every $x \in A$ there exists a unique $x' \in A$ such that $(x, x', 0) \in C$; if $a = (a_1, \ldots, a_n)$, let $a' = (a'_1, \ldots, a'_n)$.

We assume that $r(1)$ is the absorbing element of u on the set $\{r(0), r(1)\}$; the case when $r(1)$ is the neutral element is similar.

By using the five conditions in the previous lemma, the properties of u and the fact that t preserves C and Z, we deduce that $t(1, 0, a') = r(1)'$ by considering the following:

$$\begin{array}{rcl} t(0,1,a) & = & r(1) \\ t(0,0,0) & = & 0 \\ t(1,0,a') & = & r(1)' \end{array}$$

There exists a unique element w of A such that $(r(0), r(1)', w) \in C$ and $t(0, 1, 0) = w$ since

$$\begin{array}{rcl} t(0,0,a) & = & r(0) \\ t(1,0,a') & = & r(1)' \\ t(0,1,0) & = & w. \end{array}$$

We can also deduce that $t(0, 0, a') = r(1)'$ by considering the following:

$$\begin{array}{rcl} t(1,1,a) & = & r(1) \\ t(0,0,0) & = & 0 \\ t(0,0,a') & = & r(1)'; \end{array}$$

It follows that $t(0, 1, 0) = 0$ since

$$\begin{array}{rcl} t(0,0,a') & = & r(1)' \\ t(1,0,a) & = & r(1) \\ t(0,1,0) & = & 0; \end{array}$$

Hence $w = 0$, which means that $r(0) = r(1)$, a contradiction.

□

4. The general case

If the structure $\mathbf{A}$ is a core, then the associated algebra $\mathbb{A}_{\mathbf{A}}$ has the property that every one of its unary term operations is a permutation of A. This observation motivates the following definition.

Definition 4.1. *An algebra $\mathbb{A}$ is a* core algebra *if all unary term operations of $\mathbb{A}$ are permutations of A.*

Notice that the notion of core algebra is a varietal property, i.e., if $\mathbb{A}$ is a core algebra then so is every finite algebra in the variety generated by $\mathbb{A}$. Clearly, any idempotent algebra is a core algebra since the identity map is the only unary term operation of such algebras. Recall that for an arbitrary algebra $\mathbb{A}$, we denote its full idempotent reduct by $\mathbb{A}^{\diamond}$. We gather some elementary features of finite core algebras in the following proposition.

Proposition 4.2. *Let $\mathbb{A}$ be a finite core algebra. Let G be the set of unary term operations of $\mathbb{A}$ and let $\mathcal{C}$ the set of all idempotent term operations of $\mathbb{A}$. Then*

(1) *G forms a group under composition,*
(2) *for all term operations $t(x_1, \dots, x_n)$ of $\mathbb{A}$, there is some $g \in G$ and $t^\diamond(x_1 \dots, x_n)$ of $\mathcal{C}$ such that $t(x_1, \dots, x_n) = gt^\diamond(x_1, \dots, x_n)$ for all $x_i \in A$,*
(3) *if $t(x_1, \dots, x_n)$ is an idempotent term operation of $\mathbb{A}$ and $g_i \in G$, $1 \le i \le n$, then there is some $g \in G$ and $t^\diamond \in \mathcal{C}$ with*
$$t(g_1(x_1), g_2(x_2), \dots, g_n(x_n)) = gt^\diamond(x_1, \dots, x_n).$$
If $g_i = g_1$ for all i then we can take $g = g_1$.
(4) *If $\mathbb{B}$ is a subalgebra of $\mathbb{A}^\diamond$ then $\mathrm{Sg}^{\mathbb{A}}(B) = \bigcup_{g \in G} g(B)$.*
(5) *If $a \in A$, then $\mathrm{Sg}^{\mathbb{A}}(\{a\}) = \{g(a) : g \in G\}$.*
(6) *If $\mathbb{B}$ is a subalgebra of $\mathbb{A}^\diamond$ such that $\mathrm{Sg}^{\mathbb{A}}(B) = A$ and δ is a congruence of $\mathbb{B}$ then $\delta' = \mathrm{Cg}^{\mathbb{A}}(\delta)$ is equal to the transitive closure of $\bigcup_{g \in G} g(\delta)$.*

In this section we will extend our results for idempotent algebras to core algebras. In order to do so, we first extend Proposition 3.1 of [22] (see also Proposition 2.3 above). It is observed in [11] that the set of tame congruence theoretic types is naturally ordered according to the ordering in Figure 1. In Chapter 9 of [11] it is shown that if T is any order ideal of types then the class of locally finite varieties that omit the types in T can be characterized in a number of ways, and in particular by an idempotent Maltsev condition. Thus omitting an order ideal of types is a feature of the full idempotent reduct of a finite algebra.

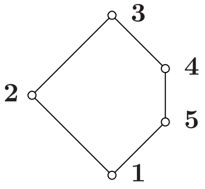

Figure 1. The Pentagon of Types

Lemma 4.3. *Let $\mathbb{A}$ be a finite core algebra and $\mathbb{A}^\diamond$ its full idempotent reduct. If $\mathbf{j} \in \mathrm{typ}\{\mathsf{S}(\mathbb{A}^\diamond)\}$ then $\mathbf{i}$ is admitted by some 2-generated subalgebra of $\mathbb{A}$ for some type $\mathbf{i} \le \mathbf{j}$.*

Proof. We will argue by contradiction. Choose a finite core algebra $\mathbb{A}$ of minimal cardinality such that for some type $\mathbf{j}$, $\mathbf{j} \in \mathrm{typ}\{\mathsf{S}(\mathbb{A}^\diamond)\}$ but $\mathbf{i}$ is omitted by all 2-generated subalgebras of $\mathbb{A}$ for all types $\mathbf{i} \le \mathbf{j}$. Arguing directly, or using Proposition 3.1 of [22], it follows that there is 2-generated subalgebra $\mathbb{B}$ of $\mathbb{A}^\diamond$ and some congruence δ of $\mathbb{B}$ such that $\mathbb{B}/\delta$ is a strictly simple algebra of type $\mathbf{j}'$ for some type $\mathbf{j}' \le \mathbf{j}$. We choose $\mathbb{B}$ with this property with $|B|$ as small as possible. Note that we may assume that $\mathbf{j}' = \mathbf{j}$ since our assumptions about $\mathbb{A}$ will still hold.

Let $\{0, 1\}$ be a generating set for $\mathbb{B}$ and let δ' be the congruence of $\mathbb{A}$ generated by δ. Since any pair of elements from B that are not δ-related generate $\mathbb{B}$, we may select 0 and 1 so that they belong to some $(\delta, 1_B)$-trace. By the minimality of $|A|$ it follows that $\{0, 1\}$ is a generating set for the algebra $\mathbb{A}$ as well.

We observe that in $\mathbb{B}$ the congruence generated by $\{(0,1)\}$ is 1_B since $\mathbb{B}$ is an idempotent algebra and $\{0,1\}$ is a generating set. Thus the congruence generated by $\{(0,1)\}$ in $\mathbb{A}$ contains δ, and hence δ'. Our proof breaks into two cases depending on whether or not $(0,1) \in \delta'$.

Case 1: $(0,1) \notin \delta'$**:**

Choose some congruence α of $\mathbb{A}$ maximal with $\delta' \leq \alpha$ and $(0,1) \notin \alpha$. We argue that $\alpha = 0_A$. Considered as a congruence of $\mathbb{A}^\diamond$ we have that the restriction of α to B is equal to δ since $\alpha|_B \supseteq \delta$ and $\mathbb{B}/\delta$ is a simple algebra. From this we conclude that the algebra $\mathbb{A}^\diamond/\alpha$ has a 2-generated subalgebra that admits type $\mathbf{j}$, since the subalgebra of this quotient generated by $\{0/\alpha, 1/\alpha\}$ is isomorphic to $\mathbb{B}/\delta$.

Unless $\alpha = 0_A$ it follows, by the minimality of $|A|$, that some 2-generated subalgebra of the core algebra $\mathbb{A}/\alpha$ admits some type $\mathbf{i} \leq \mathbf{j}$ since its full idempotent reduct has a 2-generated subalgebra that admits type $\mathbf{j}$. But then $\mathbb{A}$ also has this property, contrary to our choice of $\mathbb{A}$. Thus $\alpha = 0_A$ and hence $0_A \prec \beta = \mathrm{Cg}^{\mathbb{A}}((0,1))$. Furthermore, since $\delta \leq \alpha$, it follows that $\mathbb{B}$ is a strictly simple idempotent algebra of type $\mathbf{j}$.

We will show that the type of $(0_A, \beta) = \mathbf{j}$, and so conclude that no such algebra $\mathbb{A}$ can exist. Let U be a $(0_A,\beta)$-minimal set and let $p(x)$ be a unary polynomial of $\mathbb{A}$ with range U and such that $p(0) \neq p(1)$. Since $\mathbb{A}$ is a core algebra then we can select U and p so that p is also a polynomial of $\mathbb{A}^\diamond$ and so can be written in the form $p(x) = t(x,0,1)$ for some term $t \in \mathcal{C}$ (we use here that $\{0,1\}$ generates $\mathbb{A}$). Thus the elements $0' = p(0)$ and $1' = p(1)$ belong to B and also to some $(0_A,\beta)$-trace. Since $\{0,1\}$ and $\{0',1'\}$ are polynomially isomorphic in $\mathbb{B}$ (and $\mathbb{A}$) then we may assume that in fact $0 = 0'$ and $1 = 1'$.

To show that $\mathrm{typ}(0_A,\beta) = \mathbf{j}$, we make use of the characterization of strictly simple idempotent algebras found in [19]. If $\mathbf{j} = \mathbf{1}$ then $B = \{0,1\}$ and $\mathbb{B}$ is essentially unary. If $\mathrm{typ}(0_A,\beta) \neq \mathbf{1}$ then there is a binary polynomial of $\mathbb{A}$ whose restriction to B is not essentially unary. Indeed, we consider two cases: if $\mathrm{typ}(0_A,\beta)$ is not affine, then the entire trace is $\{0,1\}$, and the result is immediate; if the type is affine, the whole trace supports a Mal'tsev polynomial $p(x,y,z)$: if we let $g(x,y) = p(x,0,1)$, then $g(0,0) = 0$ and $g(1,0) = g(0,1) = 1$ and so g is not essentially unary on $\{0,1\}$. Given a binary polynomial $h(x,y)$ of $\mathbb{A}$ that is not essentially unary on $\{0,1\}$, then we can write $h(x,y) = gt(x,y,0,1)$ for some unary term g of $\mathbb{A}$ and idempotent term t (since $\{0,1\}$ generates $\mathbb{A}$). Since h is not essentially unary on $\{0,1\}$, then neither is $t(x,y,0,1)$ and hence neither is the 4-ary term $t(x,y,z,w)$. This leads to a term operation of $\mathbb{B}$ that is not essentially unary and so in this case we conclude that $\mathrm{typ}(0_A,\beta) = \mathbf{1}$.

If $\mathbf{j} = \mathbf{2}$ then $\mathbb{B}$ is an affine algebra. Suppose that $\mathrm{typ}(0,\beta) \neq \mathbf{2}$; this type cannot be unary because otherwise β is strongly abelian, but since $\mathbb{B}$ is affine it has a polynomial (and hence so does $\mathbb{A}$) that fails the strong term condition. In all other cases β is non-abelian and in particular there is a binary polynomial of $\mathbb{A}$ that acts as meet on the trace $\{0,1\}$. So there is a polynomial $h(x,y)$ with $h(0,1) = h(0,0)$ but $h(1,1) \neq h(1,0)$. Arguing as above, it follows that there is a term of $\mathbb{B}$ that, when restricted to $\{0,1\}$ fails the term condition, implying that $\mathbb{B}$ can't be abelian, and so can't be affine.

If $\mathbf{j} = \mathbf{4}$ then $B = \{0,1\}$ and we may assume that $\mathbb{B}$ has $x \wedge y$ as a term operation, since all of its term operations are lattice operations and $\mathbb{B}$ is not essentially unary. The type of $(0_A,\beta)$ is at least that of the type of $\mathbb{B}$ and so is either $\mathbf{3}$ or $\mathbf{4}$. If it is of type **3** then there is a unary polynomial $e(x)$ of $\mathbb{A}$ with $e(0) = 1$ and $e(1) = 0$. This polynomial can be written in the form $hs(x, 0, 1)$ for some $h \in G$ and $s \in \mathcal{C}$. Since the restriction of s to B is a lattice term it follows that $s(0, 0, 1) = 0$ and $s(1, 0, 1) = 1$ and so $h(0) = 1$ and $h(1) = 0$. Since $\mathbb{A}$ is a core algebra, no such unary term can exist, because the unary term $h(x) \wedge x$ is not a permutation of $\mathbb{A}$. Thus $\mathrm{typ}(0_A, \beta) = \mathbf{4}$.

If $\mathbf{j} = \mathbf{5}$ then $\mathbb{B}$ is term equivalent to a 2 element (meet) semilattice. As in the previous paragraph, we can rule out that $\mathrm{typ}(0_A, \beta) = \mathbf{3}$ since there can be no polynomial of $\mathbb{A}$ that maps 0 to 1 and 1 to 0. If $\mathrm{typ}(0_A, \beta) = \mathbf{4}$ then $\mathbb{A}$ will have a binary polynomial whose restriction to $\{0, 1\}$ defines the lattice join operation. From this it follows that the clone of $\mathbb{B}$ cannot be that of a semilattice, contrary to $\mathbf{j} = \mathbf{5}$. Finally, if $\mathbf{j} = \mathbf{3}$ then $\mathrm{typ}(0_A, \beta) = \mathbf{3}$ since this type is at least as rich as the type of $\mathbb{B}$.

Case 2: $(0, 1) \in \delta'$**:**

In this case, it follows that $\delta' = \mathrm{Cg}^{\mathbb{A}}((0, 1))$. Let α be a congruence of $\mathbb{A}$ with $\alpha \prec \delta'$. Since $(0, 1) \notin \alpha$ then $\alpha|_B \leq \delta$ and so, as in Case 1, we conclude that $\alpha = 0_A$ (or else $\mathbb{A}/\alpha$ provides a smaller counter example to the lemma). As in Proposition 4.2 we let G denote the set of unary term operations of $\mathbb{A}$ and let $\mathcal{C}$ the set of all idempotent term operations of $\mathbb{A}$.

Claim 1. *The action of G on A is transitive and so for some $g \in G$, $g(0) = 1$. Furthermore, $\mathbb{A}$ has no proper subuniverses.*

Using part 6 of Proposition 4.2 it follows that since $(0, 1) \in \delta' \setminus \delta$ there is a chain of overlapping sets of the form $h(D)$, where $h \in G$ and D is a δ-class, that connects 0 to 1. From this it follows that there is some c, $d \in B$ and $h \in G$ with $(0, c) \in \delta$, $(0, d) \notin \delta$ and $h(c) = d$. As remarked earlier, any pair of elements from B that are not δ-related generate both $\mathbb{B}$ and $\mathbb{A}$. In particular, $\{c, d\}$ is such a generating set. But since $d = h(c)$, we have that in fact $\{c\}$ generates $\mathbb{A}$. From part 5 of Proposition 4.2 it follows that the action of G on A is transitive and so there must be some $g \in G$ with $g(0) = 1$. It also follows that $\mathbb{A}$ can have no proper subuniverse.

Claim 2. *$\delta' = 1_A$ and so $\mathbb{A}$ is a strictly simple core algebra.*

Let C be the δ'-class that contains B. Since δ' is also a congruence of the idempotent algebra $\mathbb{A}^\diamond$, it follows that C is closed under the set of idempotent term operations of $\mathbb{A}$. Let H be the set of unary term operations of $\mathbb{A}$ under which C is closed. Clearly H is closed under composition and so is a subgroup of G, the group of all unary term operations of $\mathbb{A}$. Let $\mathbb{C}$ be the algebra with universe C and whose basic operations are the restrictions of the operations in H and $\mathcal{C}$ to C. It is not hard to see that $\mathbb{C}$ is a core algebra, since it is a subalgebra of a reduct of a core algebra.

Let $p(x_1, \ldots, x_n)$ be a polynomial of $\mathbb{A}$ under which C is closed. We can write $p(x_1, \ldots, x_n)$ as $gt(x_1, \ldots, x_n, 0, 1)$ for some $g \in G$ and $t \in \mathcal{C}$, since $\{0, 1\}$ generates $\mathbb{A}$. Since C is closed under p then $p(0, 0, \ldots, 0) \in C$ and so $gt(0, \ldots, 0, 0, 1) \in C$. As $t(0, \ldots, 0, 0, 1) \in B$ it follows that g maps C onto C and so $g \in H$. Thus the restriction to C of any polynomial of $\mathbb{A}$ under which C is closed, is actually a polynomial of $\mathbb{C}$. In other words, $\mathbb{A}|_C$ is polynomially equivalent to $\mathbb{C}$. It also follows that since $0_A \prec \delta'$ then $\mathbb{C}$ is a simple algebra.

From Claim 1 we conclude that $\mathbb{C}$ has no proper subuniverses, since the action of G on A is transitive. Thus $\mathbb{C}$ is a strictly simple core algebra. Using Lemma 6.16 of [11] and our observation from the previous paragraph it follows that the type of the simple algebra $\mathbb{C}$ is equal to $\operatorname{typ}(0_A, \delta')$. We note that since $\mathbb{B}$ is a subalgebra of $\mathbb{C}^\diamond$ then, by the minimality of $|A|$ it follows that if C is a proper subset of A, then for some type $\mathbf{i} \leq \mathbf{j}$, some subalgebra of $\mathbb{C}$ admits $\mathbf{i}$. Since $\mathbb{C}$ is strictly simple, then $\mathbb{C}$ must have type $\mathbf{i}$. But then $\mathbf{i} = \operatorname{typ}(0_A, \delta')$, contrary to our assumptions on $\mathbb{A}$. Thus $C = A$ and so $\delta' = 1_A$.

To conclude the proof of this lemma we make use of the characterization of finite strictly simple surjective algebras having no trivial subuniverses given by Theorem 6.3 in [19]. It is shown that such an algebra is either quasi-primal, affine, or a matrix power of a G-set (and so is strongly abelian). From this it is almost immediate that the type of $\mathbb{A}$ is equal to $\operatorname{typ}(\delta, 1_B) = \mathbf{j}$ since being quasi-primal, affine, or strongly abelian is preserved under taking full idempotent reducts and subalgebras. □

Proposition 4.4. *Let $\mathbb{A}$ be a finite core algebra and T an order ideal of types. Then $\mathcal{V}(\mathbb{A})$ omits the types in T if and only if every 2-generated subalgebra of $\mathbb{A}$ omits the types in T.*

Proof. As noted earlier, omitting (or failing to omit) the types in T is a feature of the idempotent term operations of an algebra and so, with $\mathbb{A}^\diamond$ the full idempotent reduct of $\mathbb{A}$, we conclude that $\mathcal{V}(\mathbb{A}^\diamond)$ admits some type in T if $\mathcal{V}(\mathbb{A})$ does. In this case, we have from Proposition 3.1 of [22] that there is some (2-generated) subalgebra of $\mathbb{A}^\diamond$ that admits some type $\mathbf{j} \in T$. So by the previous lemma we conclude that some 2-generated subalgebra of $\mathbb{A}$ admits some type $\mathbf{i} \leq \mathbf{j} \in T$, as required. □

Corollary 4.5. *There is a polynomial time algorithm to decide whether the variety generated by a given finite core algebra omits the types in a given order ideal of types.*

Proof. By the previous proposition, testing whether the variety generated by a finite core algebra $\mathbb{A}^\diamond$ omits the types in some order ideal T amounts to computing the typesets of all 2-generated subalgebras of $\mathbb{A}$. According to [3], this calculation can be performed in time bounded by a polynomial in the size of $\mathbb{A}$. □

Theorem 4.6. *Let $\mathbb{A}$ be a finite core algebra. The variety generated by $\mathbb{A}$ admits the unary or affine type if and only if $\mathbb{A}^2$ has the ability to count.*

Proof. Suppose that $\mathbb{A}^2$ has the ability to count. Then so does $(\mathbb{A}^\diamond)^2$. By Theorem 3.1, it follows that $\mathcal{V}(\mathbb{A}^\diamond)$ admits the unary or affine type, and hence $\mathcal{V}(\mathbb{A})$ admits the unary or affine type.

For the converse we proceed by contradiction. Choose a finite core algebra $\mathbb{A}$ of smallest size such that $\mathcal{V}(\mathbb{A})$ admits the unary or affine type but $\mathbb{A}^2$ does not have the ability to count. Let G be the set of unary term operations of $\mathbb{A}$. By Propositions 2.5 and 4.4 we may assume that $\mathbb{A}$ is subdirectly irreducible with least non-zero congruence μ such that $\operatorname{typ}(0_A, \mu)$ is $\mathbf{1}$ or $\mathbf{2}$. Furthermore, $\mathbb{A}$ is generated by any set $\{u, v\}$ with $(u, v) \in \mu \setminus 0_A$. We aim to show that $\mu = 1_A$, and hence that $\mathbb{A}$ is a strictly simple algebra.

Let D be a non-trivial μ-class and suppose that μ has k classes. Then by Proposition 4.2 (4), since D is a subuniverse of $\mathbb{A}^\diamond$ and $\mathrm{Sg}^{\mathbb{A}}(D) = A$, there are $g_i \in G$, $1 \leq i \leq k$ such that each μ-class is of the form $g_i(D)$. We may assume that g_1 is the identity map. Note that since the g in G are permutations and D is a congruence class, it follows that for any $g \in G$, either $g(D) = D$ or $g(D)$ is a μ-class that is disjoint from D. Thus G acts transitively on the μ-classes by the natural action. Let N be the stabilizer subgroup of D in G. Then for every $g \in G$ there is a unique $i \leq k$ such that $g = g_i \circ n$ for some $n \in N$.

Let $\mathbb{D}$ be the algebra obtained from the subalgebra of $\mathbb{A}^\diamond$ with universe D by adding, for each $n \in N$, the restriction $n|_D$ as a basic operation. We claim that $\mathbb{D}$ is an abelian core algebra. To show that $\mathbb{D}$ is a core algebra it suffices to note that for all idempotent term operations $t(x_1, \ldots, x_k)$ of $\mathbb{A}$ and $n_i(x) \in N$ for $1 \leq i \leq k$, the map $g(x) = t(n_1(x), \ldots, n_k(x))$ maps D into D and so belongs to N. Since μ is an abelian congruence, then it readily follows that $\mathbb{D}$ is abelian. If $\mu < 1_A$ then D is a proper subset of A and so, by our choice of $\mathbb{A}$, $\mathbb{D}^2$ has the ability to count.

We next show that since $\mathbb{D}^2$ has the ability to count, then so does $\mathbb{A}^2$, and from this conclude that $D = A$ and hence that $\mathbb{A}$ is a strictly simple core algebra. Let $\langle D^2; C, Z\rangle$ be a structure that has the ability to count with C and Z subuniverses of $(\mathbb{D}^2)^3$ and $\mathbb{D}^2$ respectively. Let C' be the subuniverse of $(\mathbb{A}^2)^3$ generated by C and Z' the subuniverse of $\mathbb{A}^2$ generated by Z.

We claim that there is a morphism from $\langle A^2; C', Z'\rangle$ onto $\langle D^2; C, Z\rangle$ and so the algebra $\mathbb{A}^2$ will have the ability to count. Define the function f from A^2 to D^2 as follows: given an element $(a_1, a_2) \in A^2$, there will be unique $1 \leq i, j \leq k$ with $a_1 \in g_i(D)$ and $a_2 \in g_j(D)$. Define $f(a_1, a_2)$ to be $(g_i^{-1}(a_1), g_j^{-1}(a_2))$. To see that f preserves Z', let $(a_1, a_2) \in Z'$. Since Z' is the subuniverse of $\mathbb{A}^2$ generated by Z then there is some term $t(x_1, \ldots, x_m)$ of $\mathbb{A}$ and pairs $(c_i, d_i) \in Z$, for $1 \leq i \leq m$ such that $(a_1, a_2) = t((c_1, d_1), \ldots, (c_m, d_m))$.

By Proposition 4.2 (2) there is some idempotent term operation $t^\diamond$ of $\mathbb{A}$ and some $g \in G$ with $t(\bar{x}) = gt^\diamond(\bar{x})$ for all $\bar{x}$ from A. As noted earlier, there is a unique i with $g = g_i n$ for some $n \in N$, and so

$$t(\bar{x}) = g_i(nt^\diamond(\bar{x}))$$

for all $\bar{x}$ from A. It follows that (a_1, a_2) is equal to $(g_i(c), g_i(d))$, where $(c, d) = nt^\diamond((c_1, d_1), \ldots, (c_m, d_m))$ and so $f(a_1, a_2) = (c, d)$. The restriction of the operation $nt^\diamond$ to D is a term operation of $\mathbb{D}$ and so the subuniverse Z of $\mathbb{D}^2$ is closed under this operation, applied component-wise. Thus the pair $(c, d) \in Z$ and so f preserves Z'. Similarly one can show that f maps C' into C, and so f is a homomorphism from $\langle A^2; C', Z'\rangle$ to $\langle D^2; C, Z\rangle$. From this we conclude that $\mathbb{A}^2$ has the ability to count, contrary to our assumption.

Thus the algebra $\mathbb{A}$ is a finite strictly simple abelian core algebra. If $\mathbb{A}$ happens to be idempotent, then by Theorem 3.1 we conclude that $\mathbb{A}$, and hence $\mathbb{A}^2$ has the ability to count. On the other hand, if $\mathbb{A}$ is not idempotent, then, depending on whether the type of $\mathbb{A}$ is unary or affine, one can use Corollary 3.10 and Lemma 4.1 of [20] or Theorem 12.4 of [10] to show that $\mathbb{A}^2$ has an idempotent abelian homomorphic image, and hence that $\mathbb{A}^2$ has the ability to count. This is contrary to our assumptions on $\mathbb{A}$ and so we conclude that no such algebra can exist. □

Theorem 4.7. *Let $\mathbb{A}$ be a finite core algebra. The following are equivalent:*

(1) $\mathcal{V}(\mathbb{A})$ *contains an algebra with the ability to count;*
(2) $\mathbb{A}^2$ *has the ability to count;*
(3) $\mathcal{V}(\mathbb{A})$ *admits the unary or affine type;*
(4) $\mathsf{S}(\mathbb{A})$ *admits the unary or affine type.*

5. Monotone Circuits and Typesets

We have seen how the ability to count, a combinatorial condition on CSP's introduced by Feder and Vardi, is related to the typeset of the variety of the associated algebra. Boolean circuit size is another facet of the ability to count that Feder and Vardi exploited to show that certain CSP's do not have bounded width. Every CSP can be viewed naturally as a Boolean query and hence as a family of Boolean functions (see below for a precise description.) Obviously if a structure $\mathbf{B}$ belongs to $\neg CSP(\mathbf{A})$ then so does every structure obtained from $\mathbf{B}$ by adding tuples to its basic relations, i.e., the query $\neg CSP(\mathbf{A})$ is *monotone.* If $\{C_n\}$ is a family of monotone Boolean circuits computing the family of functions representing our query, one may ask about the size $S(n)$ of the smallest such circuits. Afrati, Cosmadakis and Yannakakis ([1], Theorem 3.1) prove that if $\neg CSP(\mathbf{A})$ is expressible in Datalog, then it can be computed by polynomial size monotone circuits. On the other hand, Feder and Vardi show that if $CSP(\mathbf{A})$ has the ability to count then $\neg CSP(\mathbf{A})$ cannot be computed by polynomial size monotone circuits ([9], Theorem 30). The next result extends the Feder Vardi result to all core CSP's whose algebra generates a variety admitting the unary or affine type.

Before stating and proving the result, we require a few definitions.

We first describe how we view relational structures as words on $\{0, 1\}$ (see [17] p. 88 and [1] for details.) Let $\tau = (R_1, \ldots, R_s)$ be a signature. We consider structures with universe of size n to be over the fixed universe $\{0, 1, \ldots, n-1\}$. Encode a k-ary relation θ over this set as a word of length n^k as follows: the j-th bit of this word is 1 if and only if the j-th tuple in the lexicographic ordering of the n^k k-tuples belongs to θ. Then a structure $\mathbf{B}$ is encoded as the word 0^n1 (to indicate the size of its universe) followed by the concatenation of the encodings of its basic relations $R_1(\mathbf{B})$, $R_2(\mathbf{B})$, etc.

For the formal definition of Boolean circuits we refer the reader to Chapter 6 of [17].

Next we describe the reductions that we use (for details see [15]). Let σ and $\tau = (R_1, \ldots, R_s)$ be two signatures. A *k-ary first-order interpretation with p parameters of τ in σ* is an $(s+1)$-tuple $\mathcal{I} = (\phi_U, \phi_{R_1}, \ldots, \phi_{R_s})$ of first-order formulas over the vocabulary σ, where $\phi_U = \phi_U(x, y)$ has $k + p$ free variables $x = (x^1, \ldots, x^k)$ and $y = (y^1, \ldots, y^p)$ and $\phi_{R_i} = \phi_{R_i}(x_1, \ldots, x_r, y)$ has $kr + p$ free variables where r is the arity of R_i and each $x_j = (x_j^1, \ldots, x_j^k)$ and $y = (y^1, \ldots, y^p)$.

Let $\mathbf{G}$ be a σ-structure. A tuple $c = (c_1, \ldots, c_p)$ of elements of $\mathbf{G}$ is said to be *proper* if $c_i \neq c_j$ when $i \neq j$. Let $c = (c_1, \ldots, c_p)$ be proper. The *interpretation of* $\mathbf{G}$ *through* $\mathcal{I}$ *with parameters* c, denoted by $\mathcal{I}(\mathbf{G}, c)$, is the τ-structure whose universe is

$$\{a \in G^k : \phi_U(a, c)\}$$

and whose interpretation for R_i is

$$\{(a_1, \ldots, a_r) \in (G^k)^r : \phi_U(a_1, c) \wedge \cdots \wedge \phi_U(a_r, c) \wedge \phi_{R_i}(a_1, \ldots, a_r, c)\}.$$

Let σ and τ be finite relational vocabularies, let $\mathcal{C}$ be a class of σ-structures and let $\mathcal{D}$ be a class of τ-structures closed under isomorphisms. We say that a first-order interpretation $\mathcal{I}$ with p parameters of τ in σ is a *first-order reduction of* $\mathcal{C}$ *to* $\mathcal{D}$ if for every σ-structure $\mathbf{G}$ with at least p points the following two equivalences hold:

(A) $\mathbf{G} \in \mathcal{C} \Leftrightarrow \mathcal{I}(\mathbf{G}, c) \in \mathcal{D}$ for every proper c,

(B) $\mathbf{G} \in \mathcal{C} \Leftrightarrow \mathcal{I}(\mathbf{G}, c) \in \mathcal{D}$ for some proper c.

A first-order reduction is *positive* if it satisfies the following conditions: the formula ϕ_U is quantifier-free and for every θ in τ, ϕ_θ is built from atomic formulas and equalities using only the existential quantifier, disjunction and conjunction.

We can now prove the main result of this section:

Proposition 5.1. *Let* $\mathbf{A}$ *be a core structure such that* $\mathcal{V}(\mathbb{A}_{\mathbf{A}})$ *admits the unary or affine type. Then* $\neg CSP(\mathbf{A})$ *cannot be computed by polynomial size monotone circuits.*

Proof. Let $\mathbb{A}^\diamond$ denote the full idempotent reduct of $\mathbb{A}_{\mathbf{A}}$. Since $\mathcal{V}(\mathbb{A}_{\mathbf{A}})$ admits the unary or affine type, so does $\mathcal{V}(\mathbb{A}^\diamond)$. By Theorem 3.1 the algebra $\mathbb{A}^\diamond$ has the ability to count; hence there exist a subuniverse C of $(\mathbb{A}^\diamond)^3$ and a subuniverse Z of $\mathbb{A}^\diamond$ such that the structure $\mathbf{B} = \langle A; C, Z \rangle$ has the ability to count. By Theorem 30 of [9] $\neg CSP(\mathbf{B})$ cannot be computed by polynomial size monotone circuits. We require the following three observations:

(1) according to the proof of Theorem 2.1 of [15], there exists a sequence of structures $\mathbf{B} = \mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_m = \mathbf{A}$ such that for each $i = 1, \ldots, m-1$, there is a positive first-order reduction with parameters of $\neg CSP(\mathbf{G}_i)$ to $\neg CSP(\mathbf{G}_{i+1})$.

(2) it is easy to see, inspecting Lemmas 2.2 to 2.11 of [15], that in each case we may assume that the formula ϕ_U is simply set to $TRUE$, i.e. that the universe of $\mathcal{I}(\mathbf{G}, c)$ consists of G^k (indeed, this must hold since we are dealing with homomorphism closed classes.)

(3) the problem $CSP(\mathbf{B})$ is Mod$_p$ L-hard by Theorem 4.1 of [15], and in particular L-hard, hence $CSP(\mathbf{B})$ is not first-order definable (see the discussion preceding Theorem 3.3 in [15]). In particular, for each i there exists at least one structure of the same signature as $\mathbf{G}_i$ which does not belong to $CSP(\mathbf{G}_i)$, and obviously one may take this to be the one-element "loop" structure of the given signature, i.e. it has universe $\{0\}$ and each k-ary basic relation is equal to $\{(0, \ldots, 0)\}$.

So to prove the result it now suffices to show that each of the positive first-order reductions with parameters in (1) above can be computed by a family of monotone circuits of polynomial size. Let $\mathcal{I}$ denote our interpretation with p parameters. Let c be some fixed proper p-tuple on $\{0, \ldots, n-1\}$. First assume that our input structures have universe of size $n \geq p$. We construct a monotone circuit (i.e. involving only OR and AND gates) which, given as input the encoding of a structure $\mathbf{G}$, will output the encoding of the structure $\mathcal{I}(\mathbf{G}, c)$. Since the universe of $\mathcal{I}(\mathbf{G}, c)$ is G^k, the prefix which encodes the size of the universe is easy to output, it is simply the sequence 0^{n^k} by observation (2). Next we must output the encodings of the basic relations: this in fact is a straightforward encoding of the formulas describing the output relations of our interpretation, see for instance Theorem 6.4 of [17]; in particular, in is not hard to see that the size of the circuit is polynomial in n. Since our reduction is positive, we use no negations and our circuit is monotone. The fact that the output of the circuit belongs to $\neg CSP(\mathbf{G}_{i+1})$ precisely when the input is in $\neg CSP(\mathbf{G}_i)$ follows from the defining property (A) of reductions with parameters that guarantees *any* proper tuple will work.

Next we must deal with structures with universe of size $n < p$. Intuitively, our circuit outputs the one-element loop structure if $\mathbf{G}$ is in $\neg CSP(\mathbf{G}_i)$ and outputs the one-element structure with empty relations otherwise. More precisely: the prefix of our output is 01; if $\mathbf{G}$ is a structure with universe $\{0,\ldots,n-1\}$ in $\neg CSP(\mathbf{G}_i)$, let $\phi_{\mathbf{G}}$ be an AND gate with inputs all the tuples in all the relations of $\mathbf{G}$; then every output bit is an OR of all the $\phi_{\mathbf{G}}$ with $\mathbf{G}$ in $\neg CSP(\mathbf{G}_i)$ and $|G| < p$. By observation (3) our circuit outputs a structure in $CSP(\mathbf{G}_{i+1})$ if and only if the input is in $CSP(\mathbf{G}_i)$.

□

6. Concluding Remarks

We have shown that for a core structure $\mathbb{A}$, the variety $\mathcal{V}(\mathbb{A}_{\mathbf{A}})$ admits the unary type or the affine type precisely if it contains an algebra with the ability to count; in fact if this is the case then $(\mathbb{A}_{\mathbf{A}})^2$ has the ability to count, and furthermore if $\mathbb{A}_{\mathbf{A}}$ is idempotent then the algebra $\mathbb{A}_{\mathbf{A}}$ itself has the ability to count. The next example shows that we cannot expect this to happen if the algebra is not idempotent:

Example. Consider the 2-element structure $\mathbf{A} = \langle\{0,1\};\theta\rangle$ where θ consists of all triples (x,y,z) such that x,y,z are not all equal. It is not hard to see that the algebra $\mathbb{A}_{\mathbf{A}}$ is a core algebra that generates a variety admitting the unary type; however it does not have the ability to count, in fact, it has no one-element subalgebras.

The following example shows that we cannot expect our results to extend to non-core algebras either:

Example. Consider the structure $\mathbf{A}$ on $\{0,1,2,3,4,5\}$, with a single binary relation θ which is the symmetric, irreflexive 6-cycle, i.e., $\theta = \{(i,j) : |i-j| = 1\}$ where the difference is taken modulo 6. The algebra $\mathbb{A}_{\mathbf{A}}$ generates a variety which admits the unary type (the 6-cycle admits no Taylor operation, see [13], Theorem 4.4.) However, the core $\mathbf{A}'$ of this graph is the 2-element edge, so $CSP(\mathbf{A}) = CSP(\mathbf{A}')$ has bounded width (the edge admits a majority operation, see e.g. [5]). Let $\mathbb{C} \in \mathcal{V}(\mathbb{A}_{\mathbf{A}})$ and let $\mathbf{C}$ be a relational structure whose basic relations are subuniverses of powers of $\mathbb{C}$; by Theorem 2.1 of [15], $CSP(\mathbf{C})$ reduces to $CSP(\mathbf{A})$ via Datalog-preserving reductions, and hence has bounded width. In particular, $CSP(\mathbf{C})$ cannot have the ability to count, and consequently no algebra in $\mathcal{V}(\mathbb{A}_{\mathbf{A}})$ has the ability to count.

As a consequence of Corollary 4.5 and Theorem 4.7, the problem of determining whether an idempotent algebra has the ability to count is decidable. However, the analogous question remains open for core algebras, general algebras and also for relational structures.

Let $\mathbf{A}$ be a core structure, and consider the following conditions:

(1) $CSP(\mathbf{A})$ has bounded width;
(2) $\neg CSP(\mathbf{A})$ is computable by polynomial-size monotone circuits;
(3) $\mathcal{V}(\mathbb{A}_{\mathbf{A}})$ omits the unary and affine types;
(4) $\mathcal{V}(\mathbb{A}_{\mathbf{A}})$ contains no algebra with the ability to count;
(5) $(\mathbb{A}_{\mathbf{A}})^2$ does not have the ability to count.

Our results, combined with Theorem 3.1 of [1], show that the following implications hold: (1) implies (2), and (2) implies the equivalent conditions (3), (4) and (5). Furthermore, the bounded width conjectures of [16] and [8] are equivalent, and predict that the 5 conditions should all be equivalent.

Acknowledgements. We are grateful to Andrei Bulatov and Tomás Feder for useful discussions. Special thanks to Albert Atserias for his pertinent comments and suggestions.

References

[1] F. Afrati, S. S. Cosmadakis, and M. Yannakakis. On Datalog vs. polynomial time. *Proc. 10th ACM SIGACT-SIGMOD-SIGART Symp. on Principles of Database Systems* (1991), 13–25.

[2] A. Atserias, A. Bulatov, A. Dawar. Affine systems of equations and counting infinitary logic. In *ICALP'07*, volume 4596 of *LNCS*, pages 558–570, 2007.

[3] J. D. Berman, E. W. Kiss, P. Pröhle, Á. Szendrei. The set of types of a finitely generated variety, *Discrete Math.*, **112** no. 1-3, (1993), 1–20.

[4] A. Bulatov. A graph of a relational structure and constraint satisfaction problems. In *LICS'04*, pages 448–457, 2004.

[5] A. Bulatov, A.Krokhin, B. Larose. Dualities for Constraint Satisfaction Problems. 32 pages, *LNCS*, to appear.

[6] A. Bulatov, M. Valeriote. Recent results on the algebraic approach to the CSP. 26 pages, *LNCS*, to appear.

[7] V. Dalmau. Linear Datalog and bounded path duality for relational structures. *Logical Methods in Computer Science*, 1(1), 2005. (electronic).

[8] T. Feder, M. Y. Vardi. Monotone monadic SNP and constraint satisfaction. in *Proceedings of the 25rd Annual ACM Symposium on Theory of Computing (STOC)*, San Diego, California, (1993), 612–622.

[9] T. Feder, M.Y. Vardi. The computational structure of monotone monadic SNP and constraint satisfaction: A study through Datalog and group theory. *SIAM Journal on Computing*, 28:57–104, 1998.

[10] R. Freese, R. McKenzie. *Commutator Theory for Congruence Modular Varieties*, Cambridge University Press, London Mathematical Society Lecture Note Series, volume 125, 1987.

[11] D. Hobby, R. McKenzie. *The Structure of Finite Algebras.* volume 76 of Contemporary Mathematics. American Mathematical Society, 1988.

[12] Ph.G. Kolaitis, M.Y. Vardi. On the expressive power of Datalog: tools and a case study. *Journal of Computer and System Sciences*, 51:110–134, 1995.

[13] B. Larose, L. Haddad. Colourings of hypergraphs, permutation groups ans CSP's. *Logic Colloquium '04*, 93-108, Lect. Notes Log., **27**, *Assoc. Symbol. Logic*, La Jolla, CA, 2008.

[14] B. Larose, C. Loten, C. Tardif. A characterisation of first-order constraint satisfaction problems. *Logical Methods in Computer Science*(Special issue: selected contributions of LICS '06) 3 (4), 2007.

[15] B. Larose, P. Tesson. Universal algebra and hardness results for constraint satisfaction problems. *Theoret. Comput. Sci.* (Special issue for selected papers from ICALP'07), 32 pages, to appear.

[16] B. Larose, L. Zádori. Bounded width problems and algebras. *Algebra Universalis*, 56(3-4):439–466, 2007.

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

[18] A. A. Razborov. Lower bounds on monotone complexity of the logical permanent. *Mathematical Notes of the Academy of Sciences of the USSR*, **37** (6), 485-493, 1985.

[19] Á. Szendrei. A survey on strictly simple algebras and minimal varieties, pages 209239. Research and Exposition in Mathematics. Heldermann Verlag, 1992.

[20] Á. Szendrei. Simple surjective algebras having no proper subalgebras, *J. Austral. Math. Soc. Ser. A*, **48** no. 3, (1990), 434–454.

[21] Á. Szendrei. *Clones in Universal Algebra*, volume 99 of *Seminaires de Mathematiques Superieures*. University of Montreal, 1986.

[22] M. Valeriote. A subalgebra intersection property for congruence distributive varieties. Canadian J. of Math., 18 pages, to appear.

Department of Mathematics and Statistics, Concordia University, 1455 de Maisonneuve West, Montréal, Qc, Canada, H3G 1M8
E-mail address: larose@mathstat.concordia.ca
URL: http://cicma.mathstat.concordia.ca/faculty/larose/

Bolyai Intézet, Aradi vértanúk tere 1, H-6720, Szeged, Hungary
E-mail address: zadori@math.u-szeged.hu
URL: http://www.math.u-szeged.hu/~zadori/index.htm

Department of Mathematics and Statistics, McMaster University, 1280 Main Street West Hamilton, Ontario, Canada L8S 4K1
E-mail address: matt@math.mcmaster.ca
URL: http://www.math.mcmaster.ca/~matt/