

it has less than 2 vertices. Finally, one adds a 6th, nullary operation, of type \mathbf{o} : the constant 0, standing for the empty graph with no vertices. \square

A Equivalences of logical formulas

In this appendix, we discuss some equivalences and transformations of logical formulas which can be used to give upper bounds for the index of congruences considered in this paper, and to complete the proof of the effectiveness of certain notions (e.g. quantifier-free definition schemes).

More specifically, we make precise in what sense we can state, as we do in the body of the paper, that the set of first-order (resp. monadic second-order) formulas over finite sets of relations, constants and free variables, and with a bounded quantification depth, can be considered as finite. Moreover, explicit upper bounds on the size of these finite sets are derived, which can be used to justify the termination of some of our algorithms, and in evaluating their complexity. That these upper bounds have unbounded levels of exponentiation is not unexpected.

A.1 Boolean formulas

Let p_1, \dots, p_n be Boolean variables and let B_n be the set of Boolean formulas written with these variables. It is well known that B_n is finite up to logical equivalence. For further reference, we record the following more precise statement.

Proposition A.1 *There exists a subset B_n^{red} of B_n , of cardinality 2^{2^n} such that every formula in B_n can be effectively reduced to an equivalent formula in B_n^{red} .*

Proof. We let B_n^{red} be the set of Boolean formulas in disjunctive normal form, where in each disjunct, variables occur at most once and in increasing order, no two disjuncts are equal, and disjuncts are ordered lexicographically. These constraints guarantee the announced cardinality of B_n^{red} ; the rest of the proof is classical. \square

Of course, the formula in B_n^{red} equivalent to a given formula, is not always the shortest possible.

A.2 First-order formulas, semantic equivalence

Let us consider finite sets R and C , of relational symbols and of constants (nullary relations, source labels) as in Section 3.1. Recall that, if X is a finite set, $FO(R, C, X)$ denotes the set of first-order formulas in the language of (R, C) -structures, with free variables in X . For unproved results in this section, we refer the reader to [4].

Several notions of semantic equivalence of formulas can be defined. If $\varphi, \psi \in FO(R, C, X)$, say that $\varphi \equiv \psi$ if for every (R, C) -structure S and for every assignment of values in S for the elements of X , φ and ψ are both true or both false. Say also that $\varphi \equiv_\omega \psi$ if the same holds for every finite or countable (R, C) -structure S , and $\varphi \equiv_f \psi$ if S is restricted to being finite.

The equivalences \equiv and \equiv_ω coincide by the Löwenheim-Skolem theorem. Indeed this theorem states that if a closed formula has an infinite model, then it has one of each infinite cardinality: it suffices to apply it to the formula $\exists \vec{x} \neg(\varphi(\vec{x}) \Leftrightarrow \psi(\vec{x}))$. We note that this equivalence cannot be extended to monadic second-order formulas: there exists an MS formula with a unique model, isomorphic to the set of integers \mathbb{N} with its order.

Each of these three equivalences is known to be undecidable.

The equivalence \equiv (or \equiv_ω since we consider only first-order formulas) is semi-decidable: by Gödel's completeness theorem, $\varphi \equiv \psi$ if and only if the formula $\forall \vec{x} (\varphi(\vec{x}) \Leftrightarrow \psi(\vec{x}))$ has a proof, which is a recursively enumerable property.

Trakhtenbrot proved that one cannot decide whether a first-order formula is true in every finite structure, thus proving that \equiv_f is not decidable. However, the negation of \equiv_f is semi-decidable: if $\varphi \not\equiv_f \psi$, a counter-example can be produced by exploring systematically all finite (R, C) -structures. This is a proof also that \equiv and \equiv_f do not coincide.

A.3 First-order formulas, a syntactic equivalence

We now describe a syntactic equivalence \approx on formulas, which refines the semantic equivalences \equiv and \equiv_f : that is, if $\varphi \approx \psi$, then $\varphi \equiv \psi$ and $\varphi \equiv_f \psi$.

If $b \in B_n$, and if $\varphi_1, \dots, \varphi_n \in FO(R, C, X)$, we denote by $b(\varphi_1, \dots, \varphi_n)$ the formula in $FO(R, C, X)$ obtained by replacing each occurrence of p_i in b by φ_i . It is clear that if b and b' are equivalent Boolean formulas, then $b(\varphi_1, \dots, \varphi_n) \equiv b'(\varphi_1, \dots, \varphi_n)$.

A *Boolean transformation step* consists in replacing in a first-order formula, a sub-formula of the form $b(\varphi_1, \dots, \varphi_n)$ by the equivalent formula $b'(\varphi_1, \dots, \varphi_n)$, where $b, b' \in B_n$ are equivalent. Then we let $\varphi \approx \psi$ if φ can be transformed into ψ by a sequence of Boolean transformation steps and of renaming of bound variables.

It is clear that if $\varphi \approx \psi$, then $\varphi \equiv \psi$. We want to show that each first-order formula is effectively equivalent to an \approx -equivalent formula of the same quantifier height, and to give an upper bound on the number of \approx -equivalence classes of formulas of a given height.

A.3.1 Quantifier-free formulas

Let $QF(R, C, X)$ be the set of quantifier-free formulas in $FO(R, C, X)$. Such formulas are Boolean combinations of atomic formulas. Let $Atom(R, C, X)$ be the set of these atomic formulas. Note that each atomic formula is either of the form $x = y$, where x and y are in $X \cup C$, or $r(x_1, \dots, x_{\rho(r)})$ where r is a $\rho(r)$ -ary

relation in R and the x_i are in $X \cup C$. Letting $n = \mathbf{card}(X)$ and $c = \mathbf{card}(C)$, it is easily verified that

$$\mathbf{card}(Atom(R, C, X)) = (n + c)^2 + \sum_{r \in R} (n + c)^{\rho(r)}.$$

We let $f(R, c, n)$ be this function. Note that if we allow for the (effective) syntactic simplifications of identifying the formulas of the form $x = x$ with the constant **true**, and of identifying the formulas $x = y$ and $y = x$, we can lower the value of $f(R, c, n)$ to $1 + \frac{1}{2}(n + c)(n + c - 1) + \sum_{r \in R} (n + c)^{\rho(r)}$.

We then have the following.

Proposition A.2 *There exists a subset $QF^{red}(R, C, X)$ of $QF(R, C, X)$, of cardinality $2^{2^{f(R, c, n)}}$, such that every formula in $QF(R, C, X)$ can be effectively transformed to an \approx -equivalent formula in $QF^{red}(R, C, X)$.*

Proof. By definition of quantifier-free formulas, $QF(R, C, X)$ is the set of all formulas of the form $b(\varphi_1, \dots, \varphi_n)$, where b is a Boolean formula and the φ_i are atomic formulas. Now let $QF^{red}(R, C, X)$ be the set of all formulas of the form $b(\varphi_1, \dots, \varphi_n)$, where $b \in B_n^{red}$ and the φ_i are pairwise distinct atomic formulas. The proof of the precise statement is now immediate, using Proposition A.1. \square

Example A.3 Let us consider graphs with sources, so that R consists of a single, binary edge relation. Then $f(R, c, 0) = 2c^2$ and $\mathbf{card}(QF^{red}(R, C, \emptyset)) = 2^{2^{2c^2}} = q(c)$. Thus the type equivalence ζ (see Section 3.3 and Lemma 3.7) has at most $2^{q(c)}$ classes in $\mathcal{GS}(C)$. \square

Remark A.4 Again, we are not claiming that the set $QF^{red}(R, C, X)$ is as small as possible. On quantifier-free formulas, the equivalence \equiv is decidable, because $\varphi \equiv \psi$ is false if and only if the closed formula $\exists \vec{x}(\varphi(\vec{x}) \not\equiv \psi(\vec{x}))$ is satisfiable, and the satisfiability problem for existential formulas in prenex normal form is decidable (see [4]). Thus one can modify Proposition A.2 by letting $QF^{red}(R, C, X)$ be the set of lexicographically minimal formulas in each \equiv -class: the same statement of Proposition A.2 would then hold with \equiv instead of \approx . In particular, the transformation would still be effective, although very inefficient. It is not clear whether the cardinality of the new set of reduced quantifier-free formulas would be significantly smaller. \square

A.3.2 Quantifier depth of first-order formulas

Recall that the quantifier depth of a first-order formula is the maximal number of nested quantifiers. If we let $FO_k(R, C, X)$ be the set of formulas in $FO(R, C, X)$ of quantifier depth at most k , a formal definition is as follows:

$FO_0(R, C, X) = QF(R, C, X)$ and, for each $k \geq 0$, $FO_{k+1}(R, C, X)$ is the set of Boolean combinations of formulas in

$$\begin{aligned} \widehat{FO}_k(R, C, X) &= FO_k(R, C, X) \\ &\cup \{\exists y \varphi \mid \varphi \in FO_k(R, C, X \cup \{y\})\} \\ &\cup \{\forall y \varphi \mid \varphi \in FO_k(R, C, X \cup \{y\})\}. \end{aligned}$$

Using the same recursion, let us define sets of “reduced” formulas of every quantifier depth. First we fix an enumeration of the countable set of variables. Next, we let $FO_0^{red}(R, C, X) = QF^{red}(R, C, X)$. For each $k \geq 0$, we then let $FO_{k+1}^{red}(R, C, X)$ be the set of formulas of the form $b(\varphi_1, \dots, \varphi_n)$ where $b \in B_n^{red}$ and the φ_i are in

$$\begin{aligned} \widehat{FO}_k^{red}(R, C, X) &= FO_k^{red}(R, C, X) \\ &\cup \{\exists y \varphi \mid \varphi \in FO_k^{red}(R, C, X \cup \{y\}), y \text{ minimal not in } X\} \\ &\cup \{\forall y \varphi \mid \varphi \in FO_k^{red}(R, C, X \cup \{y\}), y \text{ minimal not in } X\}. \end{aligned}$$

Proposition A.5 *For each $k \geq 0$, the set $FO_k^{red}(R, C, X)$ is finite. Moreover, every formula in $FO_k(R, C, X)$ can be effectively transformed to an \approx -equivalent formula in $FO_k^{red}(R, C, X)$.*

Proof. Let $n = \text{card}(X)$ and $c = \text{card}(C)$, let $g(k, R, c, n)$ be the cardinality of $FO_k^{red}(R, C, X)$, and let $h(k, R, c, n)$ be the cardinality of $\widehat{FO}_k^{red}(R, C, X)$. It is elementary to verify that these functions can be bounded as follows:

$$\begin{aligned} g(0, R, c, n) &\leq 2^{f(R, c, n)} \text{ and for } k > 0 \\ g(k, R, c, n) &\leq 2^{2^{h(k, R, c, n)}} \\ h(k, R, c, n) &\leq 3g(k-1, R, c, n+1). \end{aligned}$$

The rest of the proof is immediate, from the recursive definitions. \square

Remark A.6 Since there is a procedure to transform each first-order formula into an \approx -equivalent formula in “reduced form”, we can consider a new equivalence relation on first-order formulas: to yield the same reduced formula. This equivalence is decidable and it refines \approx (and hence \equiv). \square

Remark A.7 In the Proposition A.5, we can still consider replacing each formula by the lexicographically least equivalent formula, but this method is not effective, since the equivalence of first-order formulas is not decidable. \square

A.4 Monadic second-order formulas

A very similar analysis can be conducted for monadic second-order formulas of bounded quantifier depth. One difference is that the Löwenheim-Skolem theorem does not hold for these formulas, so the semantic equivalence of formulas

based on coincidence on all finite or countable models does not imply coincidence on all models. Moreover, since there is no complete proof systems for such formulas, the equivalences \equiv and \equiv_ω are not semi-decidable.

For the rest, one can follow the same techniques as above, to prove the following result. We denote by $MS_k(R, C, W)$ the set of monadic second-order formulas of quantification depth k in the language of (R, C) -structures, with their first- and second-order free variables in W .

Proposition A.8 *For every finite R, C, W, k , one can construct a finite subset $MS_k^{red}(R, C, W)$ of $MS_k(R, C, W)$ such that, for every formula in $MS_k(R, C, W)$, one can construct effectively an \equiv -equivalent formula in $MS_k^{red}(R, C, W)$.*

References

- [1] K. ABRAHAMSON, M. FELLOWS, Finite automata, bounded tree-width and well-quasiordering. In: N. ROBERTSON, P. SEYMOUR (eds) *Graph Structure Theory*. Contemporary Mathematics **147** (1993), 539-564.
- [2] S. ARNBORG, B. COURCELLE, A. PROSKUROWSKI, D. SEESE, An algebraic theory of graph reduction. *J. ACM* **40** (1993), 1134-1164.
- [3] B. BOLLOBAS, *Extremal graph theory*. Academic Press (1978).
- [4] E. BÖRGER, E. GRÄDEL, YU. GUREVICH, *The classical decision problems*. Springer, 1997.
- [5] D. G. CORNEIL, M. HABIB, J.-M. LANLIGNEL, B. REED, U. ROTICS, Polynomial time recognition of clique-width ≤ 3 graphs, Extended abstract. In: G.H. GONNET, D. PANARIO, A. VIOLA (eds.) *LATIN'2000*. LNCS **1776** (2000), 126-134.
- [6] B. COURCELLE, The monadic second-order logic of graphs I: recognizable sets of finite graphs. *Information and Computation* **85** (1990), 12-75.
- [7] B. COURCELLE, The monadic second-order logic of graphs VII: Graphs as relational structures. *Theoretical Computer Science* **101** (1992) 3-33.
- [8] B. COURCELLE, Recognizable sets of graphs: equivalent definitions and closure properties. *Mathematical Structures in Computer Science* **4** (1994), 1-32.
- [9] B. COURCELLE, The monadic second-order logic of graphs VI: on several representations of graphs by relational structures. *Discrete Applied Mathematics* **54** (1994) 117-149. Erratum **63** (1995) 199-200.
- [10] B. COURCELLE, Basic notions of universal algebra for language theory and graph grammars. *Theoretical Computer Science* **163** (1996), 1-54.
- [11] B. COURCELLE, The monadic second-order logic of graphs X: Linear orders. *Theoretical Computer Science* **160** (1996), 87-143.
- [12] B. COURCELLE, Structural properties of context-free sets of graphs generated by vertex-replacement. *Information and Computation* **116** (1995), 275-293.