CONTEMPORARY MATHEMATICS

558

Model Theoretic Methods in Finite Combinatorics

AMS-ASL Joint Special Session
January 5–8, 2009
Washington, DC

Martin Grohe
Johann A. Makowsky
Editors

American Mathematical Society

Model Theoretic Methods in Finite Combinatorics

This page intentionally left blank

CONTEMPORARY MATHEMATICS

558

Model Theoretic Methods in Finite Combinatorics

AMS-ASL Joint Special Session
January 5–8, 2009
Washington, DC

Martin Grohe
Johann A. Makowsky
Editors

American Mathematical Society
Providence, Rhode Island

2010 *Mathematics Subject Classification*. Primary 03–02, 03–06, 05–02, 05–06, 68–02, 68–06.

Library of Congress Cataloging-in-Publication Data

Model theoretic methods in finite combinatorics : AMS-ASL special session, January 5–8, 2009 Washington, DC / Martin Grohe, Johann A. Makowsky, editors.
p. cm. — (Contemporary mathematics ; v. 558)
Includes bibliographical references.
ISBN 978-0-8218-4943-9 (alk. paper)
1. Finite model theory—Congresses. 2. Combinatorial probabilities—Congresses. I. Grohe, M. (Martin). II. Makowsky, Johann A., 1948–

QA9.7.M583 2011
519.2—dc23

2011031229

---

10 9 8 7 6 5 4 3 2 1 16 15 14 13 12 11

Contents

This page intentionally left blank

Preface

From the very beginnings of *Model Theory*, applications to Algebra, Geometry and Number Theory accompanied its development and served as a source of inspiration. In contrast to this, applications to combinatorics emerged more recently and more slowly. Between 1975 and 1990 we saw the discovery of 0-1 laws for finite models of sentences of Predicate Logic by R. Fagin[1] and Y.V. Glebskii, D.I. Kogan, M.I. Liogonkii and V.A. Talanov[2], applications of Model Theory to generating functions by K.J. Compton[3] and to counting functions by C. Blatter and E. Specker[4]; and the emergence of algorithmic meta-theorems initiated by B. Courcelle[5]. The best known of these are the 0-1 laws which were, and still are, widely studied. The least known are the applications of Model Theory to combinatorial functions.

Methodologically, in the early applications of Model Theory to Algebra and Number Theory, elimination of quantifiers and the Compactness Theorem play crucial rôles, and the logic is always First Order Logic. In the applications to Combinatorics the logic often is Monadic Second Order Logic, and the tools are refinements of Ehrenfeucht-Fraïssé Games and of the Feferman-Vaught Theorem.

In recent years other very promising interactions between Model Theory and Combinatorics have been developed in areas such as extremal combinatorics and graph limits, graph polynomials, homomorphism functions and related counting functions, and discrete algorithms, touching the boundaries of computer science and statistical physics.

This volume highlights some of the main results, techniques, and research directions of the area. Topics covered in this volume include recent developments on 0-1 laws and their variations, counting functions defined by homomorphisms and graph polynomials and their relation to logic, recurrences and spectra, the logical complexity of graphs, algorithmic meta theorems based on logic, universal and homogeneous structures, and logical aspects of Ramsey theory. Most of the articles are expository and contain comprehensive surveys as well as new results on particular aspects of how to use Model Theoretic Methods in Combinatorics. They make

---

[1] R. Fagin, *Probabilities on finite models*, J. Symb. Log. **41** (1976), no. 1, 50–58.

[2] Y.V. Glebskii, D.I. Kogan, M.I. Liogonkii, and V.A. Talanov, *Range and degree of realizability of formulas in the restricted predicate calculus*, Cybernetics **5** (1969), 142–154.

[3] K.J. Compton, *Applications of logic to finite combinatorics*, Ph.D. thesis, University of Wisconsin, 1980.

[4] 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 (1981), 41–44, and *Recurrence relations for the number of labeled structures on a finite set.* In Logic and Machines: Decision Problems and Complexity (E. Börger, G. Hasenjaeger, and D. Rödding, eds.), Lecture Notes in Computer Science, vol. 171, Springer, 1984, pp. 43–61.

[5] B. Courcelle, *Graph rewriting: An algebraic and logic approach*, Handbook of Theoretical Computer Science (J. van Leeuwen, ed.), vol. 2, Elsevier Science Publishers, 1990, pp. 193–242.

the volume suitable to serve both as an introduction to current research in the field and as a reference text.

In spite of the title of this book, not all the articles deal with finite combinatorics in a strict sense. Two articles deal with countable homogeneous and countable universal structures, one article deals with partition properties of group actions, and two articles explore various aspects of Ramsey's Theorem. Still, they all share a model theoretic view and involve notions from finite Combinatorics.

Two applications of Model Theory to Combinatorics are, unfortunately, not covered in this volume, as they were published, or promised for publication elsewhere, before this volume came into being. G. Elek and B. Szegedy[6] used ultraproducts and Loeb measures to prove generalizations of Szemerédi's Regularity Lemma to hypergraphs. The Regularity Lemma plays a crucial rôle in the study of very large finite graphs. A. Razborov[7] used Model Theory to develop *flag algebras*, which allow one to derive results in extremal graph theory in a uniform way.

The volume is the outcome of the special session on

Model Theoretic Methods in Finite Combinatorics

that was held at the AMS-ASL Meeting of January 2009 in Washington D.C.

Most speakers of the special session, and a few other prominent researchers in the area, were invited to contribute to the volume. In the name of all the authors we would like to thank the referees for their careful reading of the contributions and their many suggestions which were incorporated into the final articles. Special thanks go to Christine Thivierge at the American Mathematical Society for patiently shepherding us through the publication process.

Martin Grohe
Johann Makowsky

[6] G. Elek and B. Szegedy, *Limits of hypergraphs, removal and regularity lemmas. A non-standard approach*, Available at: arXiv: 0705.2179, 2007.

[7] A. Razborov, *Flag algebras*, Journal of Symbolic Logic **72.4** (2007), 1239–1282.

Contemporary Mathematics
Volume **558**, 2011

Application of Logic to Combinatorial Sequences and Their Recurrence Relations

Eldar Fischer, Tomer Kotek, and Johann A. Makowsky

2010 *Mathematics Subject Classification.* 03-02, 03C98, 05-02, 05A15, 11B50 .

Partially supported by an ISF grant number 1101/06 and an ERC-2007-StG grant number 202405.

Partially supported by the Fein Foundation and the Graduate School of the Technion - Israel Institute of Technology.

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

Part 1. Introduction and Synopsis

1. Sequences of integers and their combinatorial interpretations

We discuss sequences $a(n)$ of natural numbers or integers which arise in combinatorics. In some cases such sequences satisfy linear recurrence relations with constant or polynomial coefficients. In this paper we discuss sufficient structural conditions on $a(n)$ which imply the existence of various linear recurrence relations.

The traditional approach for studying such sequences consists of interpreting $a(n)$ as the coefficients of a generating function $g(x) = \sum_n a(n)x^n$, and of using *analytic methods*, to derive properties of $a(n)$, cf. [**38**]. Another general framework of *species of structures* for combinatorial interpretations of counting functions is described in [**9, 10**]. Lack of space does not allow us to use this formalism in this paper. There is also a substantial theory of how to algorithmically verify and prove identities among the terms of $a(n)$, see [**64**].

We are interested in the case where the sequence $a(n)$ admits a *combinatorial* or a *logical* interpretation, i.e., $a(n)$ counts the number of relations or functions over the set $[n] = \{1, \ldots, n\}$ which have a certain property, possibly definable in some logical formalism (with or without its natural order). To make this precise, we assume the reader is familiar with the basics of Logic and Finite Model Theory, cf. [**30, 56**] and also [**46**]. We shall mostly deal with the logics **SOL**, Second Order Logic, and **MSOL**, Monadic Second Order Logic. Occasionally, we formulate statements in the language of automata theory and regular languages and use freely the Büchi-Elgot-Trakhtenbrot Theorem, which states that a language is regular iff it is definable in **MSOL** when we view its words of length n as ordered structures over a set of n elements equipped with unary predicates, cf. [**30**]. More details on the logical tools used will be given wherever needed.

We define a general notion of *combinatorial interpretations* for finite ordered relational structures.

Definition 1.1 (Combinatorial interpretation). *A* combinatorial interpretation $\mathcal{K}$ of $a(n)$ *is given by*

(i) a class $\mathcal{K}$ of finite structures over a vocabulary
$\tau = \{R_1, \ldots R_r\} = \{\overline{R}\}$ *or* $\tau_{ord} = \{<_{nat}, \overline{R}\}$
where the universe of a structure of size n in $\mathcal{K}$ is $[n] = \{1, \ldots, n\}$, and the relation symbol $<_{nat}$ is interpreted as the natural order on $[n]$.

(ii) The counting function $d_{\mathcal{K}}(n)$,

$$d_{\mathcal{K}}(n) = \left|\{\overline{R} \text{ on } [n] : \langle [n], <_{nat}, \overline{R}\rangle \in \mathcal{K}\}\right|$$

which counts the number of relations[1]*, is such that $d_{\mathcal{K}}(n) = a(n)$.*

(iii) A combinatorial interpretation $\mathcal{K}$ is a pure combinatorial interpretation *of $a(n)$ if $\mathcal{K}$ is closed under τ-isomorphisms. In particular, if $\mathcal{K}$ does not depend on the natural order $<_{nat}$ on $[n]$, but only on τ.*

Our counting functions count *labeled* structures. In the example of counting linear orders we have $n!$ linear orders on $[n]$ in the labeled case, whereas only one linear order in the unlabeled case. In this article we do not deal with the unlabeled case.

[1]In enumerative combinatorics there are various terminologies. If the counting function is monotone, it is called speed in [**5**].

The way we defined combinatorial interpretations does not require *uniformity*. In the ordered case $\mathcal{K}$ could be patched together arbitrarily, in the pure case we only have to require that it is closed under isomorphisms. Uniformity can be formulated in terms of some device (a Turing machine, a logical formula) or closure conditions (closed under substructures, products, disjoint unions). In this article we are mostly concerned with classes defined by logical formulas. Intuitively speaking, a combinatorial interpretation $\mathcal{K}$ of $a(n)$ is a *logical interpretation of* $a(n)$ if $\mathcal{K}$ is definable by a formula in some logic formalism, say full Second Order Logic.

DEFINITION 1.2 (Logical interpretation and Specker sequences).

(i) A combinatorial interpretation $\mathcal{K}$ *of* $a(n)$ *is an* **SOL**-interpretation (**MSOL**-interpretation) of $a(n)$*, if* $\mathcal{K}$ *is definable in* $\mathbf{SOL}(\tau_{ord})$ *(*$\mathbf{MSOL}(\tau_{ord})$*).*

(ii) Pure **SOL**-interpretations (**MSOL**-interpretation) of $a(n)$ *are defined analogously.*

(iii) We call a sequence $a(n)$ *which has a logical interpretation in some fragment* $\mathcal{L}$ *of* **SOL** *an* $\mathcal{L}$-Specker sequence*, or just a* Specker sequence *if the fragment is* $\mathbf{SOL}$[2]*.*

REMARK 1.1. *In the sequel of this article we will encounter sequences* $a(n)$ *which are the normalized difference* $\frac{a_1(n)-a_2(n)}{f(n)}$ *of two sequences* $a_1(n), a_2(n)$*, which both have a logical interpretation, and* $f(n)$ *is a normalizing function depending on* $a(n)$*. As this is not a logical interpretation of* $a(n)$ *we speak here of a* logical characterization *of* $a(n)$*.*

The following two propositions are straightforward.

PROPOSITION 1.1.

(i) If $a(n)$ *has a combinatorial interpretation then for all* $n \in \mathbb{N}$ *we have* $a(n) \geq 0$*.*

(ii) If $a(n)$ *has a combinatorial interpretation then for all* $n \in \mathbb{N}$ *we have* $a(n) \leq 2^{n^{d(\tau)}}$*, where* $d(\tau)$ *is a constant depending on the vocabulary* τ*.*

(iii) There are uncountably many sequences $a(n)$ *which have a combinatorial interpretation.*

PROOF. We only prove (iii). Let $\mathcal{K}_0$ be the class of structures with one unary predicate, where the structures have the form $\langle [n], \emptyset \rangle$. Let $\mathcal{K}_1$ be the class of structures with one unary predicate, where the structures have the form $\langle [n], P \rangle$, with no restriction on P. We have $d_{\mathcal{K}_0}(n) = 1$ and $d_{\mathcal{K}_1}(n) = 2^n$. Now let $A \subseteq \mathbb{N}$ and define

$$\mathcal{K}_A = \{\langle [n], \emptyset \rangle : n \in A\} \cup \{\langle [n], P \rangle : n \notin A\}$$

Clearly, for each $A \subseteq \mathbb{N}$, the class $\mathcal{K}_A$ gives a pure combinatorial interpretation for the sequence

$$d_{\mathcal{K}_A}(n) = \begin{cases} 1 & n \in A \\ 2^n & n \notin A \end{cases}$$

□

PROPOSITION 1.2.

(i) There are only countably many Specker sequences.

[2] E. Specker was to the best of our knowledge the first to introduce **MSOL**-definability as a tool in analyzing combinatorial interpretations of sequences of non-negative integers.

(ii) *Every Specker sequence $a(n)$ is computable, and in fact it is in the counting class $\sharp \cdot$ **PH**, [**49**], where **PH** is the polynomial hierarchy and the input n is measured in unary presentation. Hence $a(n)$ is computable in exponential time, and using polynomial space.*

(iii) *The set of Specker sequences is closed under the point-wise operations of addition and multiplication. The same holds for **MSOL**-Specker sequences.*

We shall discuss examples of sequences $a(n)$ which have a combinatorial interpretation in great detail in Part 2. Among them we find the counting functions for trees, graphs, planar graphs, binomial coefficients, factorials, and many more. The reader may want to consult the *The On-Line Encyclopedia of Integer Sequences (OEIS)* [**1**] which contains close to 200 000 integer sequences studied in the literature. The non-monotonic sequence $a(n)$ beginning with

$$\begin{aligned} &1,1,1,2,5,4,7,6,7,9,11,10,13,13,13,14,\\ &17,16,19,18,19,21,23,22,25,25,25,26,29,\\ &28,31,30,31,33,35,34,37,37,37,38,41,\\ &23,22,25,25,25,26,29,28,31,30,31,33,\\ &35,34,37,37,37,38,41,40,43,42,43,45,47,\\ &46,49,49,49,50,53,52,55,54,55,57,59,58,\dots \end{aligned} \tag{A085801}$$

appears there as Sequence A085801 which has an ordered combinatorial interpretation as the *maximum number of nonattacking queens on an $(n \times n)$ toroidal board.* A more theoretical source is the erudite monograph [**32**].

To the best of our knowledge no systematic study of sequences $a(n)$ which have a combinatorial interpretation has been undertaken so far. Notable exceptions are the cases where $\mathcal{K}$ is a graph property which is hereditary (closed under induced subgraphs) or monotone (closed under subgraphs), cf. [**72, 5, 6, 4**].

PROBLEM 1.

(i) *Characterize the sequences of integers which have (pure) combinatorial interpretations under various restrictions imposed on $\mathcal{K}$. What are the additional restrictions on the counting function besides those listed in Proposition 1.1?*

(ii) *Characterize the $\mathcal{L}$-Specker sequences for various sublogics of **SOL**.*

In this paper we investigate sufficient conditions for Specker sequences to satisfy linear recurrence relations. We also study the converse question: For a given class of linear recurrence relations REC, can we find a family of logical interpretations $\mathcal{I}$ such that every sequence in REC has a logical interpretation in $\mathcal{I}$?

2. Linear recurrences

We are in particular interested in linear recurrence relations which may hold over $\mathbb{Z}$ or $\mathbb{Z}_m$.

DEFINITION 2.1 (Recurrence relations). *Given a sequence $a(n)$ of integers we say $a(n)$ is*

(i) C-finite *or* rational *if there is a fixed* $q \in \mathbb{N} \backslash \{0\}$ *for which* $a(n)$ *satisfies for all* $n > q$

$$a(n+q) = \sum_{i=0}^{q-1} p_i a(n+i)$$

where each $p_i \in \mathbb{Z}$.

(ii) P-recursive *or* holonomic *if there is a fixed* $q \in \mathbb{N} \backslash \{0\}$ *for which* $a(n)$ *satisfies for all* $n > q$

$$p_q(n) \cdot a(n+q) = \sum_{i=0}^{q-1} p_i(n) a(n+i)$$

where each p_i *is a polynomial in* $\mathbb{Z}[X]$ *and* $p_q(n) \neq 0$ *for any* n. *We call it* Simply-P-recursive or SP-recursive, *if additionally* $p_q(n) = 1$ *for every* $n \in \mathbb{Z}$.

(iii) hypergeometric *if* $a(n)$ *satisfies for all* $n > 2$

$$p_1(n) \cdot a(n+1) = p_0(n) a(n)$$

where each p_i *is a polynomial in* $\mathbb{Z}[X]$ *and* $p_1(n) \neq 0$ *for any* n. *In other words,* $a(n)$ *is P-recursive with* $q = 1$.

(iv) MC-finite *(modularly C-finite), if for every* $m \in \mathbb{N} \backslash \{0\}, m > 0$ *there is* $q(m) \in \mathbb{N} \backslash \{0\}$ *for which* $a(n)$ *satisfies for all* $n > q(m)$

$$a(n + q(m)) = \sum_{i=0}^{q(m)-1} p_i(m) a(n+i) \mod m$$

where $q(m)$ *and* $p_i(m)$ *depend only on* m, *and* $p_i(m) \in \mathbb{Z}$. *Equivalently,* $a(n)$ *is MC-finite, if for all* $m \in \mathbb{N}$ *the sequence* $a(n) \pmod m$ *is ultimately periodic.*

(v) trivially MC-finite, *if for each* $m \in \mathbb{N}$ *and large enough* n, $f(n) \equiv 0 \pmod m$.

The terminology C-finite and holonomic are due to [**83**]. P-recursive is due to [**76**]. P-recursive sequences were already studied in [**12, 13**].

The following are well known, see [**38, 32**].

Lemma 2.1.

(i) *Let* $a(n)$ *be C-finite. Then there is a constant* $c \in \mathbb{Z}$ *such that* $a(n) \leq 2^{cn}$.

(ii) *Furthermore, for every holonomic sequence* $a(n)$ *there is a constant* $\gamma \in \mathbb{N}$ *such that* $\mid a(n) \mid \leq n!^{\gamma}$ *for all* $n \geq 2$.

(iii) *The sets of C-finite, MC-finite, SP-recursive and P-recursive sequences are closed under addition, subtraction and point-wise multiplication.*

In general, the bound on the growth rate of holonomic sequences is best possible, since $a(n) = n!^m$ is easily seen to be holonomic for any integer m, [**41**].

Proposition 2.2. *Let* $a(n)$ *be a function* $a : \mathbb{N} \to \mathbb{Z}$.

(i) *If* $a(n)$ *is C-finite then* $a(n)$ *is SP-recursive.*

(ii) *If* $a(n)$ *is SP-recursive then* $a(n)$ *is P-recursive.*

(iii) *If* $a(n)$ *is SP-recursive then* $a(n)$ *is MC-finite.*

(iv) *If* $a(n)$ *is hypergeometric then* $a(n)$ *is P-recursive.*

Moreover, the converses of (i), (ii), (iii) and (iv) do not hold, and no implication holds between MC-finite and P-recursive.

PROOF. The implications follow from the definitions.
$n!$ is SP-recursive, but not C-finite, as, by Lemma 2.1 it grows too fast.
$\frac{1}{2}\binom{2n}{n}$ is P-recursive but not MC-finite, see the discussion in Example 7.4.
The Bell numbers $B(n)$ are MC-finite, but not P-recursive, hence not SP-recursive; see the examples in Example 7.5.
The derangement numbers $D(n)$ in Example 7.3 are SP-recursive but not hypergeometric, cf. [**64**]. To see that no implication holds between MC-finite and P-recursive, note that n^n is MC-finite, but not P-recursive. Furthermore, $\frac{1}{2}\binom{2n}{n}$ is P-recursive, but not MC-finite; see the Examples 7.4 and 7.2. □

REMARKS 2.1. *There are sequences with pure combinatorial interpretations which are neither MC-finite nor P-recursive. As an example, take the sequence* $\frac{1}{2}\binom{2n}{n} + 2n^{2n}$. *It counts the number of binary relations* $R \subseteq [2n]^2$ *such that* R *is either the edge relation of a graph which consists of two cliques of equal size, or* R *is a function* $R : [2n] \to [2n]$.

PROPOSITION 2.3.

(i) *There are only countably many P-recursive sequences* $a(n)$.
(ii) *There are continuum many MC-finite sequences.*
Furthermore, let $a(n)$ *be an MC-finite sequence such that* $a(n) = 0 \pmod{m}$ *for all* m *and* $n \geq q(m)$, *and* $a(n)$ *is nonzero for all large enough* n. *For* $A \subseteq \mathbb{N}$, *let*

$$a_A(n) = \begin{cases} a(n) & n \in A \\ 2 \cdot a(n) & n \notin A \end{cases}$$

Then $a_A(n) = 0 \pmod{m}$ *for all* m *and* $n \geq q(m)$.

3. Logical formalisms

3.1. Fragments of SOL. Let $\overline{R}$ be a finite set of relation symbols. First Order Logic (**FOL**) over $\overline{R}$ has the atomic formulas "$R_i(x_1, \ldots, x_{\rho(i)})$" and "$x_1 = x_2$" where $x_1, x_2, \ldots$ are any individual variables. The set of formulas **FOL**$(\overline{R})$ denotes all formulas composed from the atomic ones using boolean connectives, and quantifications of individual variables "$\exists x$" and "$\forall x$".

Second Order Logic formulas **SOL**$(\overline{R})$ are obtained by allowing also relation variables $V_{i,\rho(i)}$, where $\rho(i)$ is the arity of $V_{i,\rho(i)}$, and atomic formulas of the type "$V_{i,\rho(i)}(x_1, \ldots x_{\rho(i)})$"

Monadic Second Order Logic formulas **MSOL**$(\overline{R})$ are obtained by allowing as relation variables only set variables (unary predicates) U_i, atomic formulas of the type "$U_i(x_j)$" (also expressible as "$x_j \in U_i$"), and quantifications of the form $\exists U$ and $\forall U$.

Counting Monadic Second Order Logic formulas **CMSOL**$(\overline{R})$ are obtained by allowing additional quantifiers for individual variables; for every $m, n \in \mathbb{N}$ we allow for the quantification "$C_{m,n}x$" – if $\phi(x)$ is a **CMSOL**$(\overline{R})$-formula then so is $C_{m,n}x\phi(x)$.

The satisfaction relation between an $\overline{R}$-structure $\mathfrak{A}$ and an **SOL**-formula ϕ is defined as usual (for example, if there exists $A' \subset A$ such that $\mathfrak{A}$ satisfies $\phi(A')$, then $\mathfrak{A}$ satisfies $\exists U \phi(U)$), and is denoted by $\mathfrak{A} \models \phi$. For **CMSOL**, we define $\mathfrak{A} \models C_{m,n} x \phi(x)$ to hold if the number of elements $a \in A$ for which $\mathfrak{A} \models \phi(a)$ is equivalent to n modulo m.

A class of $\overline{R}$-structures $\mathcal{C}$ is is called **FOL**($\overline{R}$)-*definable* if there exists an **FOL**($\overline{R}$) formula ϕ with no free (non-quantified) variables such that $\mathfrak{A} \in \mathcal{C}$ if and only if $\mathfrak{A} \models \phi$ for every $\mathfrak{A}$. We similarly define **MSOL**($\overline{R}$)-definable classes and **CMSOL**($\overline{R}$)-definable classes.

We use $\mathfrak{A}, \mathfrak{B}, \ldots$ for relation structures, but G, H for graphs, in particular K_n for complete graphs of n vertices, $K_{n,m}$ for complete bipartite graphs on $n+m$ vertices, C_n for cycles on n vertices, etc. Classes of structures and graphs are denoted by $\mathcal{C}, \mathcal{G}, \mathcal{K}$ and the like.

3.2. Proving non-definability. For the purpose of this article we just list a few useful facts and examples of classes of graphs definable and/or not definable in **CMSOL** and its sublogics.

First we note that for words, i.e., ordered structures with unary predicates only (besides the order relation), **MSOL** and **CMSOL** have the same expressive power. For graphs $G = (V, E)$, represented as structures with one binary edge relation E and a universe V of vertices, this is not the case.

Typical examples of classes of graphs not definable in **CMSOL** can therefore be obtained using reductions to non-regular languages. For example the language $a^n b^n$ over the alphabet $\{a, b\}$ is well known not to be regular, and therefore neither **MSOL**-definable nor **CMSOL**-definable. The class of complete bipartite graphs $K_{m,n}$ with a linear ordering is first order bi-reducible to the language $a^m b^n$. $K_{m,n}$ has a Hamiltonian cycle iff $m = n$, and therefore Hamiltonian graphs, even with a linear order, are not **CMSOL**-definable. The class EQ_2CLIQUE of graphs consisting of two equal-sized cliques is also not **CMSOL**-definable, even in the presence of a linear order, because it consists of the complement graphs of $K_{n,n}$. Using the same method, one can construct other examples.

A typical example of a graph class which is not **MSOL**-definable, but **CMSOL**-definable, is the class of Eulerian graphs. A clique K_n is Eulerian iff n is odd. But the cliques of odd size are not **MSOL**-definable. An undirected graph G has an Eulerian cycle iff G is connected and every vertex has even degree. So graphs with Eulerian cycles are **CMSOL**-definable.

In contrast to this, the class of graphs where every induced cycle has an even size is **MSOL**-definable. To see this, one notes that an induced cycle is even iff it is bipartite.

4. Finiteness conditions

In order to prove modular linear recurrences (MC-finiteness) for a sequence $a(n)$ with an **MSOL**-interpretation $\mathcal{K}$ one proves first that $\mathcal{K}$ satisfies a certain finiteness condition derived using Ehrenfeucht-Fraïssé Games for **MSOL**. But often a weaker finiteness condition suffices to get the desired recurrence relation. We now discuss these finiteness conditions. They are all of the following form.

Let $\mathcal{K}$ be a class of τ-structures and let $\otimes$ be a binary operation on all τ-structures. We associate with $\mathcal{K}$ and $\otimes$ an equivalence relation $\sim_{\mathcal{K}}^{\otimes}$ on τ-structures: $\mathfrak{A}_1 \sim_{\mathcal{K}}^{\otimes} \mathfrak{A}_2$ iff for all τ-structures $\mathfrak{B}$ we have that $\mathfrak{A}_1 \otimes \mathfrak{B} \in \mathcal{K}$ iff $\mathfrak{A}_2 \otimes \mathfrak{B} \in \mathcal{K}$. Instead of $\mathfrak{A}_1 \sim_{\mathcal{K}}^{\otimes} \mathfrak{A}_2$ we also say that $\mathfrak{A}_1$ is $(\otimes, \mathcal{K})$-equivalent to $\mathfrak{A}_2$.

Our finiteness conditions require that the number of $\otimes(\mathcal{K})$-equivalence classes is finite. We call the number of $\otimes(\mathcal{K})$-equivalence classes the $\otimes$-*index* of $\mathcal{K}$ and denote it by $\otimes(\mathcal{K})$.

There are three operations $\otimes$ which are of interest:

(i) The *disjoint union* of τ-structures, denoted by $\sqcup$. In this case we speak of $(\sqcup, \mathcal{K})$-equivalence, and of the DU-index of $\mathcal{K}$, denote by $DU(\mathcal{K})$.
(ii) The *ordered sum* of τ_{ord}-structures, denoted by $\sqcup_{ord}$. In this case we speak of $(\sqcup_{ord}, \mathcal{K})$-equivalence, and of the OS-index of $\mathcal{K}$, denote by $OS(\mathcal{K})$.
(iii) The *substitution* of τ-structures into a pointed τ_a-structure, where τ_a has one distinguished constant symbol, and where the result is an unpointed τ-structure. In this case we speak of $(subst, \mathcal{K})$-equivalence, and of the *Specker*-index of $\mathcal{K}$, denote by $SP(\mathcal{K})$.
(iv) Similarly, substitution is defined also for ordered structures, where all the elements of the substituted structure fall in their order between all the elements smaller than a and all the elements bigger than a, and the result is unpointed.

In the case of $\mathcal{K}$ being a set of words, the ordered sum corresponds to the concatenation of words and the OS-index is finite iff $\mathcal{K}$ is regular. This is the classical Myhill-Nerode characterization of regular languages. If we combine this with the Büchi-Elgot-Trakhtenbrot characterization of regular languages, we get that $\mathcal{K}$ has finite OS-index iff $\mathcal{K}$ is definable in **MSOL**.

In the case of $\mathcal{K}$ being a class of finite directed graphs, we say that $\mathcal{K}$ is a *Compton-Gessel class*[3], if $\mathcal{K}$ is closed under disjoint unions and components.

We shall prove in Section 13 that

THEOREM 4.1. *Let $\mathcal{K}$ be a class of τ-structures.*

(i) $DU(\mathcal{K}) \leq SP(\mathcal{K})$;
(ii) $OS(\mathcal{K}) \leq OSP(\mathcal{K})$;
(iii) If $\mathcal{K}$ is a Compton-Gessel class, then $DU(\mathcal{K}) \leq 2$.

Furthermore, we shall see in Section 14:

THEOREM 4.2. *If $\mathcal{K}$ is* **CMSOL***-definable then $SP(\mathcal{K})$ is finite.*

The theorems we discuss in this chapter all show that a sequence $a(n)$ satisfies some linear recurrence relation, provided $a(n)$ has a combinatorial interpretation with a finite $\otimes$-index for a suitable choice of $\otimes$.

To show that the $\otimes$-index is very small, say 2 or 3, a direct argument may suffice. However, to establish that the $\otimes$-index is finite, without computing an explict bound, it is often more convenient to use Theorem 4.2.

The exact relationship between finite indices and logical definability will be discussed in Section 14.

---

[3]The usefulness of the property of being closed under disjoint unions and components in the study of generating functions was pointed in [**22, 23**] and, independently, for modular recurrencies in [**43**]. K. Compton also gives a characterization of the sentences of First Order Logic which define Compton-Gessel classes. In Compton's papers Compton-Gessel classes are called *admissible.* and in [**18**] *adequate.*

5. Logical interpretations of integer sequences

5.1. Logical interpretations of C-finite sequences. As our first example of the use of logical interpretations we reinterpret a classical theorem about regular languages. What we obtain is a characterization of C-finite sequences of integers as differences of two counting functions of regular languages.

Let $\mathcal{K}$ be a class of ordered structures with a fixed finite number of unary predicates. Such structures are conveniently identified with words over a fixed alphabet, and a class of such structures is called a language and is denoted by $L = \mathcal{K}$.

Let us recall the following characterization of languages which have finite OS-index, [**51, 30**].

THEOREM 5.1. *Let L be a language. The following are equivalent:*

(i) L has finite OS-index.
(ii) L is a regular language.
(iii) L is **MSOL**-*definable.*

PROOF. The equivalence of (i) and (ii) is the classical Myhill-Nerode Theorem, whereas the equivalence of (ii) and (iii) was proven by Büchi, Trakhtenbrot and Elgot independently. □

THEOREM 5.2 (N. Chomsky and M. Schützenberger, [**20**]). *Let $d_L(n)$ be a counting function of a regular language L. Then $d_L(n)$ is C-finite.*

The converse is not true. However, we proved recently the following:

THEOREM 5.3 ([**52**]). *Let $a(n)$ be a function $a : \mathbb{N} \to \mathbb{Z}$ which is C-finite. Then there are two regular languages L_1, L_2 with counting functions $d_1(n), d_2(n)$ such that $a(n) = d_1(n) - d_2(n)$.*

The proof will be given in Section 5.1.

REMARK 5.1. *We could replace the difference of two sequences in the expression $a(n) = d_1(n) - d_2(n)$ by $a(n) = d_3(n) - c^n$ where $d_3(n)$ also comes from a regular language, and $c \in \mathbb{N}$ is suitably chosen.*

Using the characterization of regular languages of Theorem 5.1, Theorem 5.2 and Theorem 5.3 can be combined to characterize the C-finite sequences of integers.

THEOREM 5.4. *Let $a(n)$ be a function $a : \mathbb{N} \to \mathbb{Z}$. $a(n)$ is C-finite iff there are two* **MSOL**-*Specker sequences $d_1(n), d_2(n)$, where the sequences $d_1(n), d_2(n)$ have* **MSOL**-*interpretations over a fixed finite vocabulary which contains $<_{nat}$ and otherwise only unary relation symbols, such that $a(n) = d_1(n) - d_2(n)$.*

5.2. Logical interpretations of P-recursive (holonomic) sequences. An infinite set of holonomic sequences can be obtained from counting restricted lattice walks. A step in a lattice walk is a pair $a = (x, y) \in \mathbb{Z}^2$. For a set $\mathcal{Y}$ of steps, a lattice walk is a word $w \in \mathcal{Y}^\star$. $\mathcal{Y}$ is *symmetric* if for all $(i, j) \in \mathcal{Y}$ also $(i, -j) \in \mathcal{Y}$. For a lattice walk $w = (x_1, y_1)(x_2, y_2) \ldots (x_m, y_m)$ we define $X_i(w) = \sum_{j \leq i} x_j$ and $Y_i(w) = \sum_{j \leq i} y_j$. A lattice walk over the *quarter plane* is a lattice walk w such that for all $i \leq \ell(w)$ we have that $X_i(w), Y_i(w) \geq 0$. A lattice walk over the quarter plane stays below the diagonal, if additionally we have for all $i \leq \ell(w)$ that

$Y_i(w) \leq X_i(w)$. An *n-lattice path* is a self-avoiding lattice walk starting at a point $(0, y_0)$ and ending at a point (n, Y_m). Usually one counts lattice walks by their length m and lattice paths by prescribing their origin and their end points on an $(n \times n)$-grid. Note that an n-path may have length different from n.

The next two theorems show that there are different ways of counting lattice walks and paths (or combinations thereof) which yield holonomic sequences. In analogy to Theorems 5.3 and 5.4, we are looking for a way to represent all holonomic sequences in a uniform way.

Let $\mathcal{Y} \subset \mathbb{Z} \times \mathbb{Z} - \{(0,0)\}$.

Denote by $A_{\mathcal{Y},q}(m)$ ($A_{\mathcal{Y},d}(m)$) the number of lattice walks of length m, i.e., words of length m, in the quarter plane with steps in $\mathcal{Y}$ (which stay below the diagonal).

Denote by $a_{\mathcal{Y},q}(n)$ ($a_{\mathcal{Y},d}(n)$) the number of n-lattice paths in the quarter plane with steps in $\mathcal{Y}$ (which stay below the diagonal). For example, if $\mathcal{Y} = \{1\} \times \mathbb{Z}$, then $a_{\mathcal{Y},q}(n)$ is infinite, but $a_{\mathcal{Y},d}(n) = n!$.

THEOREM 5.5 (Bousquet-Mélou, [**17**]). *Let $\mathcal{Y} \subset \{-1,0,1\} \times \mathbb{Z} - \{(0,0)\}$. If $\mathcal{Y}$ is finite and symmetric, then $A_{\mathcal{Y},q}(m)$ is holonomic.*

The symmetry assumption on $\mathcal{Y}$ is essential. In [**61**] it is shown that there are asymmetric finite $\mathcal{Y} \subseteq \{1,-1,0\}^2 - \{(0,0)\}$ such that $A_{\mathcal{Y},q}(m)$ is not holonomic.

To get many examples, we make the set of allowed steps dependent on its position. Let $w \in \mathcal{Y}^n$, $a \in \Sigma$ and $u \in \Sigma^n$. We say that w follows u at a if the following holds: If $u[i] = a$ then $w[i] \in \mathcal{Y}^n$, else $w[i] = (1,0)$.

$a_{\mathcal{Y},d,L,a}(n)$ counts the number of pairs (w,u) such that $u \in L$ and $\ell(w) = \ell(u) = n$, and w is an n-path below the diagonal which follows u at a. Similarly, for $\overline{a} = (a_1, \ldots, a_k)$ the function $a_{\mathcal{Y},d,L,\overline{a}}(n)$ counts the number of tuples $(w_1, \ldots w_k, u)$ such that $u \in L$ and for $j \leq k$, we have that $\ell(w_j) = \ell(u) = n$ and w_j is a path below the diagonal which follows u at a_j.

DEFINITION 5.1. *Let $\mathcal{Y} = \{1\} \times \{-1,0,1\}$. A sequence $d(n)$ of integers is an LP-sequence if there is a regular language $L \subset \Sigma^\star$ and elements $a_1, \ldots, a_k \in \Sigma$ such that $d(n) = a_{\mathcal{Y},d,L,\overline{a}}(n)$.*

THEOREM 5.6 ([**53**]). *Let $d(n)$ be an LP-sequence of integers. Then $d(n)$ is holonomic, and even SP-recursive.*

LP-sequences count combinations of lattice paths with a fixed set $\mathcal{Y}$ of steps, and which all follow along words of length n of a suitable regular language L. A further degree of freedom is given by the choice of letters $\overline{a}$.

Conversely, every SP-recursive sequence can be obtained from sequences with *LP*-interpretations:

THEOREM 5.7 ([**53**]). *A sequence $d(n)$ of integers is SP-recursive iff there are two LP-sequences $d_1(n), d_2(n)$ such that $d(n) = d_1(n) - d_2(n)$.*

We consider Theorem 5.7 a “logical characterization” of SP-recursive sequences in the same sense as Theorems 5.3 and 5.4 are a “logical characterization” of C-finite sequences. The general case of characterizing P-recursive (holonomic) sequences will be discussed in Section 12. In Subsection 12.4 we discuss a different approach for characterizing holonomic sequences which uses position specific weights on words from [**54**].

5.3. Logical interpretations and modular recurrences. Modular recurrence relations for sequences with combinatorial interpretations are studied widely, cf. [**37, 43**]. A logical approach to this topic was pioneered in [**14, 15**] and further pursued in [**74, 75**]. In this section we only outline what we discuss in greater detail in Part 4.

C. Blatter and E. Specker have shown:

THEOREM 5.8 (C. Blatter and E. Specker, [**14**]). *Let $a(n)$ be a Specker sequence which has a* pure *combinatorial interpretation $\mathcal{K}$ over a finite vocabulary which contains only relation symbols of arity at most* 2. *If $\mathcal{K}$ has finite Specker index then $a(n)$ is MC-finite.*

Using Theorem 4.2 we get:

COROLLARY 5.1. *Let $a(n)$ be a Specker sequence which has a* pure *combinatorial interpretation $\mathcal{K}$ over a finite vocabulary which contains only relation symbols of arity at most* 2. *If $\mathcal{K}$ is* **MSOL**-*definable then $a(n)$ is MC-finite.*

REMARKS 5.1.

(i) *Corollary 5.1 is not true for* **MSOL**-*interpretations with an order, i.e. which are not pure, cf.* [**36**].
(ii) *E. Fischer,* [**35**]*, showed that Corollary 5.1 is also not true if one allows relation symbols of arity ≥ 4; see also* [**75**].
(iii) *In the light of Proposition 1.2(i) and Proposition 2.3(ii) there cannot be a converse of Corollary 5.1; the set of MC-finite sequences of integers cannot be characterized by a set of Specker sequences.*

In 1984 I. Gessel proved the following related result:

THEOREM 5.9 (I. Gessel, [**43**]). *Let $\mathcal{K}$ be a class of (possibly) directed graphs of bounded degree d which is a Compton-Gessel class, i.e., closed under disjoint unions and components. Then $d_{\mathcal{K}}(n)$ is MC-finite.*

REMARK 5.2. *Theorem 5.9 does not use logic. However, let $\mathcal{K}$ be a class of connected finite directed graphs, and let $\overline{\mathcal{K}}$ be the closure of $\mathcal{K}$ under disjoint unions. It is easy to see that $\mathcal{K}$ is* **MSOL**-*definable iff $\overline{\mathcal{K}}$ is* **MSOL**-*definable. Therefore, naturally arising Compton-Gessel classes are likely to be definable in* **SOL** *or even* **MSOL**.

The notion of degree can be extended to arbitrary relational structures $\mathcal{A}$ via the Gaifman graph[4] of $\mathcal{A}$, cf. [**30**].

DEFINITION 5.1.

(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 $\overline{a} \in R_i^A$ containing both u and v.*
(ii) *We define the* Compton-Gaifman graph $\mathrm{Gaif}(\mathfrak{A})$ *of a structure $\mathfrak{A}$ as the graph with the vertex set A and the neighbor relation defined above.*

[4]The terminology "Gaifman graph" is by now standard in Finite Model Theory. H. Gaifman used this definition in [**39**]. However, K. Compton used the same definition already in [**22, 23**]. The terminology "Compton-Gaifman graph" would be more appropriate.

(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 Compton-Gaifman graph* $\mathrm{Gaif}(\mathfrak{A})$.

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

OBSERVATION 5.2. *Every relational structure $\mathfrak{A}$ has a unique (up to isomorphism) decomposition into maximal connected substructures.*

Inspired by Theorem 5.8 and Theorem 5.9, E. Fischer and J.A. Makowsky showed:

THEOREM 5.10 (E. Fischer and J.A. Makowsky, [**36**]). *Let $a(n)$ be a Specker sequence which has a* pure *combinatorial-interpretation $\mathcal{K}$ over any finite relational vocabulary (without restrictions on the arity of the relation symbols), but which is of bounded degree.*

(i) If $\mathcal{K}$ has finite DU-index then $a(n)$ is MC-finite.

(ii) If additionally all structures in $\mathcal{K}$ are connected, then $a(n)$ is trivially MC-finite.

The proof will be given in Section 15.

Part 2. Guiding Examples

We now discuss various combinatorial functions with respect to their logical interpretations and the nature of their recurrence relations.

6. The classical recurrence relations

6.1. C-finite with positive coefficients. The Fibonacci sequence $f(n)$ is defined by $f(n+2) = f(n+1) + f(n)$ with $f(1) = 1$ and $f(2) = 2$. It is therefore C-finite.

To illustrate Theorem 5.3, let L_{Fib} be given by the regular expression $(a \vee ab)^\star$ with counting function $d_{Fib}(n)$. It is easy to see that $d_{Fib}(n) = f(n)$. Similarly, if $g(n+2) = 2g(n+1) + 3g(n)$ and L_g is given by

$$(a_1 \vee a_2 \vee b_1^2 \vee b_2^2 \vee b_3^2)^\star$$

with counting $d_g(n)$, then $g(n) = d_g(n)$.

It is not difficult to generalize this to any C-finite sequence with positive coefficients. For the general case, see [**7, 52**].

6.2. Growth arguments. Let $\mathcal{K}$ be a class of graphs and $d_{\mathcal{K}}(n)$ its counting function. We denote by $\overline{\mathcal{K}}$ the complement of $\mathcal{K}$. If $\mathcal{K}$ is the class of all graphs, then $d_{\mathcal{K}}(n)$ is not holonomic, since $2^{\binom{n}{2}}$ grows faster than $(n!)^c$. It follows that for any graph property $\mathcal{K}$ either $d_{\mathcal{K}}(n)$ or $d_{\overline{\mathcal{K}}}(n)$ is not holonomic, because the sum of two holonomic sequences is again holonomic.

Let $\mu_{\mathcal{K}}(n) = \frac{d_{\mathcal{K}}(n)}{2^{\binom{n}{2}}}$ denote that fraction of graphs of size n which are in $\mathcal{K}$. From the above we immediately get:

PROPOSITION 6.1. *Let $\mathcal{K}$ be any graph property, i.e., a class of finite simple graphs closed under isomorphisms.*

(i) If $\lim \mu_{\mathcal{K}}(n) = \alpha$ exists and $\alpha \neq 0$, then $d_{\mathcal{K}}(n)$ is not holonomic.

(ii) If $\lim \mu_{\mathcal{K}}(n) = \alpha$ exists and $\alpha < 1$ then $d_{\overline{\mathcal{K}}}(n)$ is not holonomic.

7. Functions, permutations and partitions

7.1. Factorials. The factorial function $n!$ is SP-recursive and hypergeometric by $(n+1)! = (n+1) \cdot n!$. This shows that it is also trivially MC-finite. $n!$ is not C-finite because it grows too fast.

$n!$ has several combinatorial interpretations: It counts the number of functions $f : [n] \to [n]$ which are bijective, which is pure and **MSOL**-definable, and it also counts the number of functions $f : [n] \to [n]$ such that $f(i) <_{nat} i+1$, which is not pure, but also **MSOL**-definable.

7.2. The function n^n. The function n^n is not P-recursive, [**41**]. It is MC-finite, which is an easy consequence of the Little Fermat Theorem, i.e., $n^{p-1} = 1 \pmod p$ provided p does not divide n. The function n^n counts the number of structures on universe $[n]$ over the vocabulary $\tau_F = \langle F \rangle$, where F is an unary function symbol. In other words, n^n is the number of functions $f : [n] \to [n]$. This gives a pure **MSOL**-definable interpretation.

7.3. Derangement numbers. The derangement numbers $D(n)$ are usually defined by their pure combinatorial definition as the set of functions $f : [n] \to [n]$ such that f is bijective and for all $i \in [n]$ we have $f(i) \neq i$. This is **MSOL**-definable. Their explicit definition is given by

$$D(n) = n! \sum_{i=0}^{n} \frac{(-1)^i}{i!} = \left\lfloor \frac{n!}{e} + \frac{1}{2} \right\rfloor$$

They are SP-recursive by $D(n) = (n-1)(D(n-1) + D(n-2))$ with $D(0) = 1$ and $D(1) = 0$, hence MC-finite, but not C-finite, by the growth argument from Subsection 6.2. In [**64**, Example 8.6.1.] they are shown not to be hypergeometric.

7.4. Central binomial coefficient. The function $\binom{2n}{n}$, the central binomial coefficient, is P-recursive and hypergeometric since

$$(n+1)^2 \cdot \binom{2(n+1)}{n+1} = 2 \cdot \binom{2(n+1)}{2} \cdot \binom{2n}{n}$$

$\binom{2n}{n}$ has many combinatorial interpretations. It counts the number of ordered partitions of $[2n]$ into two equal sized sets. If the partitions are not ordered, the counting function is $\frac{1}{2}\binom{2n}{n}$. $\binom{2n}{n}$ also counts the number of functions $f : [n+1] \to [n+1]$ such that $f(i+1) \geq f(i)$ and $f(n+1) = n+1$. This is not pure, but it is **MSOL**-definable. $\frac{1}{2}\binom{2n}{n}$ also counts the number of graphs with vertex set $[2n]$ which consists of the disjoint union of two equal sized cliques. We denote this class of graphs by $\mathrm{EQ}_2\mathrm{CLIQUE}$. Both of the above combinatorial interpretations are pure, but not **MSOL**-definable in the language of graphs; [**74**].

Similarly, the class $\mathrm{EQ}_p\mathrm{CLIQUE}$ denotes the class of graphs with vertex set $[pn]$ which consist of p disjoint cliques of equal size. We denote by $b_p(n)$ the number of graphs with $[n]$ as a set of vertices which are in $\mathrm{EQ}_p\mathrm{CLIQUE}$. Clearly,

$$b_p(n) = \begin{cases} \frac{1}{p!}\binom{pn}{n} \cdot \binom{(p-1)n}{n} \cdot \ldots \binom{n}{n} & \text{if } p \text{ divides } n \\ 0 & \text{otherwise} \end{cases}$$

Congruence relations of binomial coefficients and related functions have received a lot of attention in the literature, starting with Lucas's famous result for $b_2(n)$, [**57**]. For $b_p(n)$ modulo p, a prime, we have:

LEMMA 7.1. *For every $k > 1$, $b_p(pk) \equiv b_p(k) \pmod{p}$.*

The proof uses the method of combinatorial proof of Fermat's congruence theorem by J. Petersen from 1872, given in [**43**, page 157].

PROPOSITION 7.2. *For every n which is not a power of the prime 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.*

PROOF. By induction on n, where the basis is $n = p$ (for which $b_p(n) = 1$) and every n which is not divisible by p (for which $b_p(n) = 0$); the induction step follows from Lemma 7.1. □

From the above one can see that $b_2(n) = \frac{1}{2}\binom{2n}{n}$, and more generally $b_p(n)$, are P-recursive, but are not MC-finite. Therefore, for p a prime, $b_p(n)$ is is neither SP-recursive nor C-finite.

REMARK 7.1. *Let $a(n)$ be a sequence of integers. The* zero-set[5] $Z(a(n))$ *of* $a(n)$ *is defined by*

$$Z(a(n)) = \{n : a(n) = 0\}.$$

A subset S of the positive integers is said to be ultimately periodic *if its characteristic function $\chi_S(n)$ is ultimately periodic. The Skolem-Mahler-Lech Theorem shows that the zero-set of a C-finite seuqence is ultimately periodic. For a simple proof, see* [**47, 32**]. *L. Rubel* [**71**] *asked if it is the case that the zero-set of a P-recursive sequence is ultimately periodic. We do not have an answer to this question, but the example of $a(n) = b_2(n)$ shows that if reduce a P-recursive sequence modulo m, then the zero-set of the resulting sequence need not be ultimately periodic.*

7.5. Bell numbers. The Bell numbers B_n count the number of partitions of an n-element set. They also count the number of equivalence relations on an n-element set, which gives a pure **MSOL**-interpretation, and Theorem 5.8 applies, hence they are MC-finite. Theorem 5.8 only proves the existence of modular recurrence relations, and no explicit modular recurrence relations appear in the literature which cover all $m \in \mathbb{N}$. For prime moduli p, however, they satisfy the simple Touchard congruence $B_{p+n} = B_n + B_{n+1} \pmod{p}$, [**42**]. They satisfy the recurrence relation

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k,$$

but in [**41**] it is shown that they are not holonomic. For more properties of Bell numbers, cf. [**68**], and for congruences, cf. [**42**].

7.6. Stirling numbers of the first kind. The Stirling numbers of the first kind $\left[{n \atop k}\right]$ count arrangements of $[n]$ into k non-empty cycles (where a single element and a pair of elements are considered cycles). In other words, $\left[{n \atop k}\right]$ counts permutations with k cycles. For fixed k this is a pure **MSOL**-interpretation. They satisfy the following recurrence relation

$$\left[{n \atop k}\right] = (n-1)\left[{n-1 \atop k}\right] + \left[{n-1 \atop k-1}\right] .$$

[5]The complement of the zero-set of $a(n)$ is called the *spectrum* of $a(n)$, written $\mathrm{Spec}(a(n))$ in [**8**]. Since $Z(a(n))$ is ultimately periodic iff $\mathrm{Spec}(a(n))$ is ultimately periodic, the study of $Z(a(n))$ is closely linked to that of $\mathrm{Spec}(a(n))$. The study of spectra of **FOL**-Specker sequences was initiated by H. Scholz and is known as the *Spectrum Problem* of **FOL**. For a recent survey see [**29**].

Using the growth argument from Subsection 6.2 we can see that the Stirling numbers of the first kind are not C-finite, because $\left[{n \atop 1}\right]$ grows like the factorial $(n-1)!$. Using the Cayley-Hamilton theorem, one can deduce from the above equation that for fixed k, the sequence $\left[{n \atop k}\right]$ is SP-recursive, and therefore also MC-finite.

7.7. Stirling numbers of the second kind. The Stirling numbers of the second kind $\left\{{n \atop k}\right\}$ count the number of partitions of $[n]$ into k non-empty parts. For fixed k this is a pure **MSOL**-interpretation. $\left\{{n \atop k}\right\}$ is given explicitly by

$$\left\{{n \atop k}\right\} = \frac{1}{k!}\sum_{i=0}^{k}(-1)^{k-i}\binom{k}{i}i^n .$$

They satisfy the following recurrence relation

$$\left\{{n \atop k}\right\} = k\left\{{n-1 \atop k}\right\} + \left\{{n-1 \atop k-1}\right\} .$$

Writing this for all $k' \leq k$, we can extract from this recurrence, using the Cayley-Hamilton Theorem, that for fixed k the sequence $\left\{{n \atop k}\right\}$ is C-finite. We shall show (Proposition 11.1) in Section 5.1 that $\left\{{n \atop k}\right\}$ is C-finite without using the above recurrence relation. Instead we will exhibit a regular language the counting function of which is given by the Stirling numbers of the second kind.

Note that the Bell numbers can be expressed in terms of the Stirling numbers of the second kind:

$$B_n = \sum_{i=0}^{n}\left\{{n \atop i}\right\} .$$

7.8. Catalan numbers. The Catalan numbers $C(n)$ are defined by $C(n) = \frac{1}{n+1}\cdot\binom{2n}{n}$. They satisfy the recurrence relation

$$C(n+1) = \frac{2(2n+1)}{n+2}\cdot C(n) \tag{1}$$

In general, concerning the recurrence relations $C(n)$ behaves like $\frac{1}{2}\binom{2n}{n}$. In [**77**] there is an abundance of combinatorial interpretations which are not pure. Many of these are based on functions $f : [n] \to [n]$ which represent lattice paths subject to various conditions. One of these is the set of weakly monotonic functions $f : [n] \to [n]$ such that $f(1) = 1$, $f(n) = n$ and $f(i) \leq i$.

8. Trees and forests

8.1. Trees. Trees are (undirected) connected acyclic graphs. They are not **FOL**-definable but are **MSOL**-definable, and have finite Specker index, hence finite DU-index. Acyclicity is expressed by saying that there is no subset of size at least three such that the induced graph on it is 2-regular and connected. Denote by T_n the number of of labeled trees on n vertices. From the **MSOL**-definability it follows that the sequence T_n is MC-finite.

Labeled trees were among the first objects to be counted explicitly, cf.[**48**, Theorem 1.7.2].

THEOREM 8.1 (A. Cayley 1889). *$T_n = n^{n-2}$.*

Here the modular linear recurrences can be given explicitly: For $m = 2$ we have

$$T_1 = T_2 = 1, T_3 = 3, T_4 = 16, T_5 = 125, \dots$$

and $T_n = n \pmod 2$ for $n \geq 3$. The function n^{n-2} is not P-recursive, [**41**]. The same holds for n^{n-1} which counts the number of rooted trees.

For the number of trees of outdegree bounded by k we get the following corollary of Theorem 5.10:

COROLLARY 8.2. *The number of labeled trees of outdegree at most k is, for every $m \in \mathbb{N}$, ultimately constant* $\pmod m$*.*

In [**48**, Chapter 3] there is a wealth of results on counting various labeled trees and tree-like structures. It is worth noting that the notion of k-tree, and more generally the property of a graph of having tree-width at most k are **MSOL**-definable, cf. [**26**].

8.2. Forests. Forests are disjoint unions of trees, or equivalently, they are acyclic graphs. Therefore they are **MSOL**-definable, and have finite Specker index. In fact, they form a Compton-Gessel class.

It is well known, cf. [**81**], that the number of rooted forests on n vertices is $RF_n = (n+1)^{n-1}$. Again this is not holonomic but MC-finite.

The number of forests F_n (of non-rooted trees) is more complicated. L. Takács, [**80**] showed that

$$F_n = \frac{n!}{n+1} \sum_{j=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^j \frac{(2j+1)(n+1)^{n-2j}}{2^j \cdot j! \cdot (n-2j)!} \tag{2}$$

A simpler proof is given in [**19**].

From Theorem 5.8 we know that F_n is MC-finite, which is not obvious from the formula, as it is not obvious that the sum has integer value. Whether F_n is holonomic or not seems to be open.

9. Graph properties

In this section we list examples of graph properties $\mathcal{K}$. By definition $d_{\mathcal{K}}(n)$ has a pure combinatorial interpretation. We discuss the definability of $\mathcal{K}$ and the behaviour of $d_{\mathcal{K}}(n)$ in terms of recurrence relations. Our main sources for definability are [**30, 26**], and for the behaviour of $d_{\mathcal{K}}(n)$, [**48, 81**].

9.1. Connected graphs. The class $\mathcal{K} = \mathrm{CONN}$ is not $\mathbf{FOL}(R)$-definable, but it is $\mathbf{MSOL}(R)$-definable using a universal quantifier over set variables. We just say that every subset of vertices which is closed under the edge relation has to be the set of all vertices. For a detailed discussion of the exact status of definability, cf. [**2**].

Counting labeled connected graphs is treated in [**48**, Chapters 1 and 7] and in [**81**, Chapter 3]. For CONN [**48**, page 7] gives the following recurrence:

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

It is well known, see [**56**, Page 236], that $\lim_{n\to\infty} \mu_{\mathrm{CONN}}(n) = 1$. To see that $d_{\mathrm{CONN}}(n)$ is not holonomic we use Proposition 6.1.

9.2. Regular graphs. The class $\mathrm{REG_r}$ of simple regular graphs where every vertex has degree r is **FOL**-definable (for fixed r). The formula says that every vertex has exactly r different neighbors. The formula grows with r. The class REG of regular graphs without specifying the degree is not **FOL**-definable, and actually not even **CMSOL**-definable. To see this we note that a complete bipartite graph $K_{m,n}$ is regular iff $m = n$, but equi-cardinality of definable relations is not **CMSOL**-definable. The class $\mathrm{CREG_r}$ of connected r-regular graphs is **MSOL**-definable (for fixed r).

Counting the number of labeled regular graphs is treated completely in [**48**, Chapter 7], where an explicit formula is given, essentially due to J.H. Redfield [**67**] and rediscovered by R.C. Read [**65, 66**]. However, the formula is very complicated.

For cubic graphs, the function is explicitly given in [**48**, page 175] as $d_{\mathcal{R}_3}(2n+1) = 0$ and

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

Both REG and REG_r are Compton-Gessel classes, i.e., closed under taking compononts and disjoint unions. Furthermore, REG_r is of bounded degree. Applying Theorem 5.9, we see that $d_{\mathrm{REG}_r}(n)$ is MC-finite with a simple two term recurrence relation. Applying Theorem 5.10 we get that $d_{\mathrm{CREG}_r}(n)$ is trivially MC-finite.

I. Gessel [**45**] showed that for fixed $r \in \mathbb{N}$ the sequence $d_{\mathrm{REG}_r}(n)$ is holonomic. The problem of counting r-regular graphs with a specified set of forbidden subgraphs is one whose holonomicity remains open. N.C. Wormald [**82**] showed that the counting sequence for 3-regular graphs without triangles is holonomic.

Let h_n be the number of claw-free cubic graphs on $2n$ labeled nodes. Recently, E. Palmer, R. Read and R. Robinson, [**63**], have shown that "the enumeration of labeled claw-free cubic graphs can be added to the handful of known counting problems for regular graphs with restrictions which have been proved P-recursive". Actually, they showed that h_n is SP-recursive, giving explicitly a linear homogeneous recurrence of order 12 in which the coefficients are polynomials of degree up to 23 and all have integer coefficients. Therefore, h_n is also MC-finite. However, MC-finiteness for all the above examples follows also, without giving the recurrence explicitly, since the classes of triangle-free, or claw-free cubic graphs are **FOL**-definable.

9.3. Bipartite graphs. The class BIPART of bipartite graphs is **MSOL**-definable, and so is the class of connected bipartite graphs. We say that there is a partition of the vertex set into two independent sets (and add the statement for connectedness). Let $\beta(n)$ be the number of labeled bipartite graphs. Therefore $\beta(n)$ is MC-finite. Furthermore, the counting function for r-regular connected bipartite graphs is trivially MC-finite. BIPART is also a Compton-Gessel class. Therefore, the counting function for r-regular bipartite graphs is MC-finite with a simple two term recurrence relation.

Note that the class BIPART is not **FOL**-definable, since a regular graph of degree two is bipartite iff all its components are cycles of even size. But large enough cycles of even or odd degree cannot be distinguished by an **FOL**-formula.

In [**48**, Page 17] an explicit formula is given:

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

This also shows that $\beta(n)$ is not holonomic, because it grows too fast.

9.4. Graphs of even degree and Eulerian graphs. Let $\mathcal{C} = \mathrm{EVENDEG}$ be the class of simple graphs where each vertex has even degree. EVENDEG is not **MSOL**-definable but is **CMSOL**-definable. By Theorem 4.2 it has finite Specker index, hence finite DU-index.

$d_{\mathrm{EVENDEG}}(n) = 2^{\binom{n-1}{2}}$, cf. [**48**, page 11].

This is an example where the same function has two combinatorial interpretations: $d_{\mathrm{EVENDEG}}(n+1) = d_{\mathrm{GRAPHS}}(n)$ where the former is not **MSOL**-definable, but the latter is even **FOL**-definable.

Let $\mathcal{C} = \mathrm{EULER}$ be the class of simple connected graphs in EVENDEG. EULER is not **MSOL**-definable, but is **CMSOL**-definable. In [**48**, page 7] the following recurrence for $d_{\mathrm{EULER}}(n)$ is given:

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

The number of Eulerian graphs of degree at most r is also **CMSOL**-definable. To find an explicit formula of its counting function seems very hard. However, our Theorem 5.10 shows that the number of such graphs is trivially MC-finite.

9.5. Planar graphs and grid graphs. Planar graphs are **MSOL**-definable. To see this one can use the Kuratowski-Wagner Theorem characterizing planar graphs with forbidden (topological) minors, cf. [**28**].

A special class of planar graphs is GRIDS, the class of rectangular grids which look like rectangular checker boards, with the north-south and east-west neighborhood relation. *Partial rectangular grids* PGRIDS are subgraphs of rectangular grids. It is easy to see that both GRIDS and PGRIDS have finite Specker index, but GRIDS are **MSOL**-definable while PGRIDS are not **CMSOL**-definable, cf. [**24, 26, 73, 69**].

9.6. Forbidden subgraphs and forbidden minors. To get many classes of graphs which are **MSOL**-definable it is useful to note the following:

PROPOSITION 9.1. *Let $\mathcal{H}$ be a fixed finite set of finite graphs. Then the following are* **MSOL***-definable:*

(i) The class of graphs which have no subgraph isomorphic to some $H \in \mathcal{H}$.
(ii) The class of graphs without an induced subgraph isomorphic to some $H \in \mathcal{H}$.
(iii) The class of graphs without a minor isomorphic to some $H \in \mathcal{H}$.
(iv) The class of graphs without a topological minor isomorphic to some $H \in \mathcal{H}$.

PROOF. (i) and (ii) are easily expressed on **FOL**.

We sketch a proof of (iii) in the case where $\mathcal{H}$ consists of a single graph. Assume $H = (V(H), E(H))$ with $V(H) = \{1, \ldots, , n\}$. Now G has H as a minor if and only if the following holds:

there are disjoint connected subsets $U_i \subseteq V(G)$, $i = 1, \ldots, n$, such that there is an edge $e \in E(G) \cap (U_i \times U_j)$ if $(i, j) \in E(H)$.

Clearly this can be expressed in **MSOL**.

To prove (iv), we note that H is a topological minor of G if and only if for each edge e of $E(H)$, there is a set U_e of vertices inducing a connected subgraph such that

(i) if e_1, e_2, e_3 are distinct edges of $E(H)$, then $U_{e_1} \cap U_{e_2} \cap U_{e_3} = \emptyset$,
(ii) if e_1, e_2 are distinct edges of $E(H)$, then $U_{e_1} \cap U_{e_2} \neq \emptyset$ iff e_1 and e_2 have a common vertex, and
(iii) if $U_{e_1} \cap U_{e_2} \neq \emptyset$ then it has exactly one element.

Again these conditions can be expressed in **MSOL**.

□

It follows that every minor closed class of graphs is **MSOL**-definable. To see this, one uses the Graph Minor Theorem of P. Robertson and N. Seymour, which states that every minor closed class of graphs can be represented as a class of graphs with a finite set of forbidden minors, see [**28**].

The power of **MSOL** in graph theory has been studied extensively by B. Courcelle and J. Engelfriet [**27**].

9.7. Perfect graphs. A graph is perfect if for every induced subgraph (including the graph itself) the chromatic number equals the clique number. On the face of it, this does not seem **MSOL**- or **CMSOL**-definable. However, the following was conjectured by Berge [**16**, Chapter V.5] and proved by M. Chudnovsky, N. Robertson, P. Seymour, R. Thomas in [**21**].

THEOREM 9.1 (Strong Perfect Graph Theorem). *A graph G is perfect iff neither G nor its complement graph contains an odd induced cycle of size at least* 5.

This gives us an **MSOL**-definition of perfect graphs. Furthermore, the Specker index for perfect graphs can be computed directly and is much smaller than one would get using the **MSOL**-definition.

PROPOSITION 9.2. *If G and H are graphs, and a is a vertex of G, then the graph obtained from the substitution of H at a into G, $Subst(G, a, H)$, is perfect iff both G and H are perfect.*

PROOF. One direction follows from the definition of substitution 13.3 given precisely in Section 13, the other direction is by now classic, cf. [**16**, Chapter V.5, Theorem 19]. □

Using Corollary 13.4 we get:

COROLLARY 9.3. *The Specker index of perfect graphs is* 2.

9.8. More CMSOL-definable classes. We can now combine previous properties and see that the following are **CMSOL** definable classes of graphs. We have not found any explicit discussion of their counting functions in the literature, but the Specker-Blatter Theorem (or one of our generalizations) applies to all of these cases. The following list can be extended ad libitum.

- Planar Eulerian graphs and Eulerian graphs of bounded degree d.
- Graphs where every clique has odd cardinality.

- Graphs where every minimal cycle has even cardinality.
- Planar regular graphs of odd (even) degree.
- Planar graphs with a finite set of forbidden induced subgraphs.

10. Latin squares

In the Specker-Blatter Theorem, 5.8, the pure combinatorial interpretations are required to use relations of arity at most 2. The theorem is known to be false for arity 4, as was shown by E. Fischer, [**35**], and its status is open for arity 3.

Latin rectangles are matrices of size $(k \times n)$ with entries from $[n]$ such that in each row and column each number appears at most once. Latin squares are of the form $(n \times n)$. We denote by $L(k, n)$ the number of Latin rectangles of size $(k \times n)$ A Latin rectangle is reduced if the the first row is $(1, 2, \ldots, n)$ and the first column is $(1, 2, \ldots, k)$. We denote by $R(k, n)$ the number of reduced Latin rectangles of size $(k \times n)$. It is well known, [**59**], that

$$L(k,n) = \frac{n!(n-1)!}{(n-k)!} \cdot R(k,n) \tag{3}$$

The sequences $L(n, n)$ and $n \cdot R(n, n)$ have pure **MSOL**-interpretations with one ternary predicate. To see this we note that $L(n, n)$ counts the number of relations $M \subseteq [n]^3$ such that

- for every $i, j \in [n]$ there is exactly one $k \in [n]$ with $(i, j, k) \in M$, and
- for every $i, k \in [n]$ there is exactly one $j \in [n]$ with $(i, j, k) \in M$, and
- for every $k, j \in [n]$ there is exactly one $i \in [n]$ with $(i, j, k) \in M$.

Similarly, $n \cdot R(n, n)$ counts the number of relations $M \subseteq [n]^3$ which additionally satisfy

- there is $i \in [n]$ such that for all $j \in [n]$ we have $(i, j, j) \in M$, and $(j, i, j) \in M$.

For fixed k the sequences $L(k, n)$ and $n \cdot R(k, n)$ have pure **MSOL**-interpretations with k binary predicates $P_i(j, k)$ and the corresponding properties.

From Equation (3) it follows that $L(k, n)$ and $L(n, n)$ are both trivially MC-finite. This is also true for $R(k, n)$ or $R(n, n)$, by a theorem due to E.B. McKay and I.M. Wanless [**60**], cf. also [**78**, Theorem 4.1.], as they proved:

THEOREM 10.1. *Let $m = \lfloor n/2 \rfloor$. For all $n \in N$, $R(n, n)$ is divisible by $m!$.*

D. S. Stones and I.M. Wanless, [**79**] also showed, cf. [**78**, Theorem 4.4.]:

THEOREM 10.2. *If $k \leq n$ then $R(k, n+t) = ((-1)^{k-1}(k-1)!)^t \cdot R(k, n) \pmod t$ for all $t \geq 1$.*

In some cases, this shows that $R(k, n)$ is indivisible by some t for all $n > k$, when k is fixed and $t > k$. Nevertheless, Theorem 5.8 shows that, for fixed k, the sequence $R(k, n)$ is MC-finite.

On the other hand, $L(n, n)$ is not holonomic, as by [**31**] $L(k, n)$ grows asymptotically like

$$L(k,n) \approx (n!)^k \cdot exp\left(\frac{-k(k-1)}{2}\right). \tag{4}$$

Using Equation (3), it follows that $R(n, n)$ is also not holonomic. For fixed k, I. Gessel has shown, [**44**], that $R(k, n)$ and $L(k, n)$ are holonomic, without giving the recurrence explicitly. From Equation (4) one can also see that they are not C-finite.

Part 3. C-Finite and Holonomic Sequences

11. C-Finite sequences

We now return to the characterization of C-finite sequences as stated in Section 5. In Subsection 11.1 we prove the missing direction of Theorem 5.3 as follows:

THEOREM 11.1. *Let $a(n)$ be a C-finite sequence of integers. There exist a regular language L and a constant $c \in \mathbb{N}$ such that $a(n) = d_L(n) - c^n$, where $d_L(n)$ is the number of words of length n in L.*

The proof is based on the proof presented in [**52**], but uses the framework of the theory of regular languages instead of Monadic Second Order Logic. The other direction of Theorem 5.3 follows directly from the Chomsky-Schützenberger theorem, Theorem 5.2, and the closure of C-finite integer sequences under difference.

In Subsection 11.2 we show that the Stirling numbers of the second kind $\left\{{n \atop k}\right\}$ are C-finite for fixed k.

11.1. Proof of Theorem 11.1.

Let $a(n)$ satisfy the following recurrence

$$a(n+q) = \sum_{i=0}^{q-1} p_i a(n+i)$$

for $n \geq q$ with $p_i \in \mathbb{Z}$. Unwrapping the recurrence we get an expression for $a(n)$ as a sum over all monomials of the form

$$p_{r_1} p_{r_2} \cdots p_{r_k} \cdot a(n - r_1 - \cdots - r_k)$$

where $r_1, \ldots, r_k \in [q]$ and $n - r_1 - \ldots - r_k \in [q]$. We would like to write $a(n)$ as the sum over words of a regular language. Let $\Sigma_1 = [q]$, $\Sigma_2 = \{\widetilde{1}, \ldots, \widetilde{q}\}$ and $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \{b\}$. Let $L_{rec(q)}$ be the regular language over Σ given by the regular expression

$$(\widetilde{1} + b\widetilde{2} + bb\widetilde{3} + \ldots + b^{q-1}\widetilde{q}) \cdot (1 + b2 + bb3 + \ldots + b^{q-1}q)^\star .$$

The following function f is a one-to-one map between:

(i) tuples $\overline{r} = (r_1, \ldots, r_k)$ satisfying $n - r_1 - \ldots - r_k \in [q]$ and $r_i \in [q]$ for $i = 1, \ldots, k$.
(ii) words w in $L_{rec(q)}$.

f is given by

$$f(\overline{r}) = b^{t_{\overline{r}}-1}\widetilde{t_{\overline{r}}} \cdot (b^{r_1-1} r_1 \cdots b^{r_k-1} r_k) ,$$

where $t_{\overline{r}} = n - r_1 - \cdots - r_k$. It is not difficult to see that $f(\overline{r})$ belongs to $L_{rec(q)}$ by construction and that f is indeed a bijection between (i) and (ii). The sequence $a(n)$ can now be written explicitly as

$$a(n) = \sum_{w \in L_{rec(q)}} p_1^{|\{j : w[j]=1\}|} \cdots p_q^{|\{j : w[j]=p_q\}|} \cdot a(1)^{|\{j : w[j]=\widetilde{1}\}|} \cdots a(q)^{|\{j : w[j]=\widetilde{q}\}|} \tag{5}$$

Let $\Sigma_{(1)} = \Sigma - \{1\} \cup \{1_{(1)}, \ldots, 1_{(|p_1|)}\}$ be the alphabet obtained by replacing the letter 1 by $|p_1|$ many new letters. Let $h_1 : \Sigma_{(1)} \to \Sigma^\star$ be given by $h_1(1_{(i)}) = 1$ for all i and $h_1(j) = j$ otherwise. The function h_1 is a homomorphism from $\Sigma_{(1)}$

to $\Sigma^\star$. By the closure of regular languages under inverse homomorphisms, cf. [**51**, Page 61], the language $h^{-1}(L_{rec(q)})$ is regular, where

$$h^{-1}(L_{rec(q)}) = \{x \in \Sigma^\star : h(x) \in L_{rec(q)}\}\,.$$

It holds that we can replace p_1 in Equation (5) by 1 if we replace $L_{rec(q)}$ by $L_{(1)}$:

$$\begin{aligned} a(n) \quad = \quad & \sum_{w \in L_{(1)}} (sign(p_1))^{|\{j:h(w[j])=1\}|} \cdot (p_2)^{|\{j:w[j]=2\}|} \cdots p_q^{|\{j:w[j]=p_q\}|} \cdot \\ & a(1)^{|\{j:w[j]=\widetilde{1}\}|} \cdots a(q)^{|\{j:w[j]=\widetilde{q}\}|} \end{aligned}$$

where $sign(p_1) = 1$ if $p_1 > 0$ and otherwise $sign(p_1) = -1$. Continuing similarly we may replace each $p_1, \ldots, p_q$ and $a(1), \ldots, a(q)$ with 1 or -1 and obtain a regular language L' over

$$\Sigma' = \{j_{(i)} : j \in [q], i \in [|p_j|]\} \cup \{\widetilde{j_{(i)}} : j \in [q], i \in [|a(j)|]\} \cup \{b\}\,,$$

a homomorphism $\Sigma' \to \Sigma^\star$, $h(j_{(i)}) = j$, and two sets $S_1 \subseteq \Sigma_1$ and $S_2 \subseteq \Sigma_2$ for which

$$a(n) = \sum_{w \in L'} (-1)^{|\{j:h'(w[j]) \in S_1 \cup S_2\}|}$$

For any set of letters $D \subseteq \Sigma'$, the languages $L_{even(D)}$ and $L_{odd(D)}$ over Σ', which consist of words with an even (respectively odd) number of occurrences of the letters of D, are regular. Hence by the closure of regular languages under intersection and union, $a(n)$ can be written in the form $a(n) = d_{L_A}(n) - d_{L_B}(n)$, where $d_{L_A}(n)$ and $d_{L_B}(n)$ are the counting functions of regular languages L_A and L_B.

Finally, let L_B^c be the language obtained by replacing all the letters in the complement of L_B, $\Sigma' - L_B$, by new letters which do not appear in Σ'. Then $d_{L_B}(n) = |\Sigma'|^n - d_{L_B^c}(n)$. Since the alphabets of L_A and L_B^c are disjoint, $d_{L_B^c \cup L_A}(n) = d_{L_B^c}(n) + d_{L_A}(n)$. Therefore, $d_{L_B^c \cup L_A}(n) - |\Sigma'|^n = d_{L_A}(n) - d_{L_B}(n) = a(n)$. By the closure of regular languages under complement and union, $L_B^c \cup L_A$ is a regular language, implying the required result.

□

11.2. The Stirling Numbers of the Second Kind. The Stirling numbers of the second kind were discussed in Subsection 7.7. Recall the Stirling numbers of the second kind $\left\{ {n \atop k} \right\}$ count partitions of $[n]$ into k non-empty parts.

Proposition 11.1. *Let k be fixed. There exists a regular lanuage L_k such that $\left\{ {n \atop k} \right\} = d_{L_k}(n)$.*

Proof. Let $\Sigma = [k]$ and let L_k be the language which consists of all words over Σ in which every letter of Σ occurs at least once, and furthermore, a letter $i \in [k]$ may only occur in a word w of L_k if $i-1$ occurs before it in w. The language L_k is given by the regular expression

$$1 \cdot 1^\star \cdot 2 \cdot \{1,2\}^\star \cdot 3 \cdot \{1,2,3\}^\star \cdots k \cdot [k]^\star\,,$$

where $\{1, 2, \ldots, i\}^\star$ denotes the set of words with letters $\{1, 2, \ldots, i\}$.

Let P be the set of all partitions of $[n]$ with exactly k non-empty parts. Let $f : L_k \to P$ be the function given by $f(w) = p_w$, where p_w is the following partition:

$$p_w = \{\{i : w[i] = 1\}, \ldots, \{i : w[i] = k\}\}\,.$$

By the definition of L_k, p_w is indeed a partition of $[n]$ which consists of k non-empty parts. Moreover, f is a bijection, implying $\{{n \atop k}\} = d_{L_k}(n)$. □

12. Holonomic sequences

In this section we discuss an interpretation of SP-recursive and P-recursive integer sequences. We give two logical interpretations and characterizations, one with lattice paths, and one with positionally weighted words. The lattice path approach is more suitable for SP-recursive sequences, whereas in the case of P-recursive sequences which are not SP-recursive, weights are more appropriate. We omit the proofs and emphasize the concepts. Missing proofs may be found in [**54, 53**].

12.1. SP-recursive Sequences and Lattice Paths. Various P-recursive sequences have interpretations as counting lattice paths. A prominent example is given by the Catalan numbers $C(n)$, cf. Section 7.8. The Catalan numbers $C(n)$ count the number of paths in an $n \times n$ grid from the lower left corner to the upper right corner which have steps $\rightarrow$ and $\uparrow$ and which never go above the diagonal line, see Figure 1. The central binomial coefficient, discussed in Section 7.4, counts paths

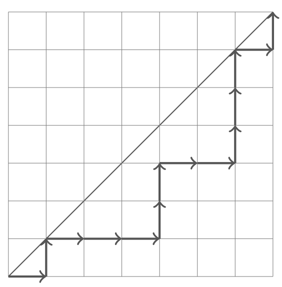


FIGURE 1. A legal path counted by $C(n)$

similar to those counted by $C(n)$, with the exception that the paths are allowed to go above the diagonal. Many other P-recursive sequences which can be interpreted as counting lattice paths can be found in [**77**]. Among them we find the Motzkin numbers and the Schröder numbers.

Possibly the simplest SP-recursive sequence is the factorial $n!$, satisfying the recurrence $(n+1)! = (n+1) \cdot n!$. The factorial was discussed in Section 7.1. $n!$ can be interpreted as counting lattice paths in an $(n+1) \times (n+1)$ grid which:

(Req. a) start from the lower left corner and end at the upper right corner,
(Req. b) consist of steps from $\rightarrow, \uparrow, \downarrow$,
(Req. c) do not cross the diagonal line, and
(Req. d) are self-avoiding.

For an example of a path satisfying these requirements see Figure 2. Notice that removing the requirement that the lattice paths are self-avoiding will mean that there are always infinitely many such paths. Furthermore, notice that this requirement is not needed in the case of the Catalan numbers since the paths are monotonic.

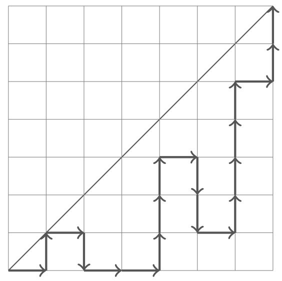


FIGURE 2. A legal path counted by $n!$

12.2. Combining Lattice Paths with Regular Languages. The lattice paths of interest to us are lattice paths in an $(n+1)\times(n+1)$ grid which satisfy requirements (Req. a), (Req. b), (Req. c) and (Req. d) and an additional requirement (Req. e). We define these lattice paths now.

DEFINITION 12.1. *Let w be a word of length n over an alphabet Σ and let $\sigma \in \Sigma$. A (w,σ)-*path *is a lattice path in an $(n+1)\times(n+1)$ grid which:*

(Req. a) starts from the lower left corner and ends at the upper right corner,
(Req. b) consists of steps from $\rightarrow, \uparrow, \downarrow$,
(Req. c) does not cross the diagonal line,
(Req. d) is self-avoiding, and
(Req. e) for any $j \in [n]$, if the j-th letter of w is not σ, then any step starting at (i,j) for any $i \in [n]$ must be a right step $\rightarrow$.

We think of the word w as labeling the columns of the grid. In columns labeled with a letter which is not σ, the only step allowed is $\rightarrow$. A legal $(ccbcbbb, b)$-path is shown in Figure 3.

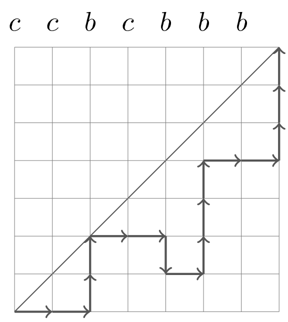


FIGURE 3. A legal $(ccbcbbb, b)$-path

DEFINITION 12.2. *Let L be a regular language over Σ and let $\overline{\sigma} = (\sigma_1, \dots, \sigma_r)$ be a tuple of Σ letters. We define the function $m_{L,\overline{\sigma}} : \mathbb{N} \rightarrow \mathbb{N}$ as follows: $m_{L,\overline{\sigma}}(n)$ is the number of tuples $(w, p_1, \dots, p_r)$ where $w \in L$ and each p_i is a (w,σ_i)-path.*

We say a sequence $a(n)$ has an LP-interpretation *if there exists a regular language L and a tuple of letters $\overline{\sigma}$ such that $a(n) = m_{L,\overline{\sigma}}(n)$.*

THEOREM 12.1 ([**53**]). *Let $a(n)$ be a sequence of integers.*

(i) $a(n)$ is SP-recursive iff $a(n)$ is the difference of two sequences $d_1(n)$ and $d_2(n)$ which have LP-interpretations,

$$a(n) = d_1(n) - d_2(n) .$$

(ii) $a(n)$ is P-recursive with leading polynomial $p_q(x)$ iff there exist $d_1(n)$ and $d_2(n)$ which have LP-interpretations such that

$$a(n) = \frac{d_1(n) - d_2(n)}{\prod_{s=1}^{n} p_q(s)} .$$

12.3. Permutations and Lattice Paths. We now discuss two examples of interpreting SP-recursive sequences which arise from counting two types of permutations as LP-interpretations. In both cases we already know that they are SP-recursive. The point here is to exhibit explicitly how they can be seen as LP-sequences.

12.3.1. *Counting permutations with a fixed number of cycles.* The Stirling numbers of the first kind $\left[{n \atop k}\right]$ were discussed in Section 7.6. They count the number of permutations of $[n]$ with exactly k cycles and are SP-recursive. They satisfy

$$\left[{n+1 \atop k}\right] = n \cdot \left[{n \atop k}\right] + \left[{n \atop k-1}\right] .$$

$\left[{n+1 \atop k}\right]$ can be interpreted naturally as counting (w, σ)-paths.

Let L_k be the set of words over alphabet $\{0, 1\}$ in which 1 occurs exactly $k - 1$ times. This is a regular language, given by the regular expression $(0^\star 1)^{k-1} 0^\star$. We will see that

$$m_{L_k,0}(n) = \left[{n+1 \atop k}\right] . \tag{6}$$

By definition, $m_{L_k,0}(n)$ counts (w, σ)-paths, where $w \in L_k$. Let $u \in L_k$ and let $A_u = \{j + 1 \mid u[j] = 1\}$. The number of (u, σ)-paths equals

$$\prod_{j : j+1 \notin A_u} j . \tag{7}$$

On the other hand, we want to count the permutations of $[n + 1]$ such that $i \in [n + 1]$ is the minimal element in its cycle iff $i \in A_u \cup \{1\}$. Let $i \in [n + 1]$. Assume we have a permutation π_i of $[i]$ such that the set of elements which are minimal in their cycle in π_i is $(A_u \cup \{1\}) \cap [i]$. We want to count the number of ways of adding the element $i + 1$ to π_i and getting a permutation π_{i+1} of $[i + 1]$ for which the set of elements which are minimal in their cycle is $(A_u \cup \{1\}) \cap [i + 1]$.

If $i + 1 \in A_u$ then $i + 1$ must be the minimal element in its cycle. This means that $i + 1$ must form a new cycle of its own and so there is exactly one permutation π_{i+1} which extends π_i in this way. Otherwise, if $i + 1 \notin A_u$ then $i + 1$ must not be the minimal element in its cycle in π_{i+1}. Hence, it must be added to one of the existing cycles. There are i ways to do so, which correspond to choosing the element $j \in [i]$ which $i + 1$ will follow in π_{i+1}.

We get that the number of permutations of $[n + 1]$ for which the set of elements which are minimal in their cycle is $A_u \cup \{1\}$ is given in Equation (7) and is equal to the number of (u, σ)-paths. Summing over words $u \in L_k$ or equivalently, over sets $A = A_u \cup \{1\}$ of size k, we get Equation (6).

12.3.2. *Permutations without fixed points.* The derangement numbers $D(n)$ were defined in Section 7.3. $D(n)$ counts permutations of $[n]$ with no fixed-point. They satisfy the SP-recurrence

$$D(n+1) = n \cdot D(n) + n \cdot D(n-1) \tag{8}$$

with initial conditions $D(0) = 1$ and $D(1) = 0$. Let $D'(n)$ be the sequence such that $D(n+1) = D'(n)$, i.e. $D'(n)$ is the number of permutations of $[n+1]$ without fixed-points. Then

$$D'(n) = n \cdot D'(n-1) + n \cdot D'(n-2) \tag{9}$$

with initial conditions $D'(1) = D(0) = 1$ and $D'(2) = D(1) = 0$.

It can be shown by induction that

$$D'(n) = \sum_{w \in \{a,b,c,d\}^n \cap L_{der}} \prod_{w[i] \in \{a,b\}} i \tag{10}$$

where the summation is over words of length n in L_{der}. The language L_{der} consists of all words w such that $w = d$ or

(i) $w[i] = c$ iff $w[i+1] = b$, and
(ii) $w[1] \cdot w[2] \cdot w[3] = dcb$.

We can think of Equation (10) as the sum over all paths in the recurrence tree of equation (9) from the root to a leaf $D'(1)$ (notice a path in the recurrence tree that ends in $D'(2) = 0$ has value 0). Such a path can be described by $1 = t_1 \leq \ldots \leq t_r = n$ such that for each i, $1 < i \leq r$, the difference of subsequent elements $t_i - t_{i-1}$ is either 1 or 2. The elements t_i in $[n]$ for which a recurrence step of the form $i \cdot D'(i-1)$ was chosen (i.e., those for which $t_i - t_{i-1} = 1$) are assigned the letter a, whereas b is assigned to those elements t_i of $[n]$ which correspond to a choice of the form $i \cdot D'(i-2)$ (i.e., those t_i for which $t_i - t_{i-1} = 2$). We assign c to all the elements $i \in [n] - \{t_1, \ldots, t_r\}$, which are skipped by a recursive choice $j \cdot D'(j-2)$, where $j = i+1$. The letter d is assigned to the leafs $D'(1)$. Condition (i) requires that $i-1$ is skipped iff $i \cdot D'(i-2)$ is chosen for i. Condition (ii) requires that the path in the recurrence tree does not end in $D'(2) = 0$, but rather skips from $D'(3)$ to $D'(1)$. Notice this is a regular language, given by the regular expression

$$dcb \cdot \left(a^\star \left(cb\right)^\star\right)^\star + d$$

Equation (10) can be interpreted as counting the number of tuples (w, p_0, p_1) where $w \in L_{der}$ is of length $|w| = n$, p_0 is a (w,a)-path and p_1 is a (w,b)-path.

12.4. P-recursive sequences and positional weights. In this subsection we present a logical interpretation for P-recursive sequences which uses positional weights. The interpretation sheds light on P-recurrences as a generalization of hypergeometric recurrences.

Recall a sequence of integers $a(n)$ is hypergeometric if for all $n > 1$,

$$a(n) = \frac{p_0(n)}{p_1(n)} a(n-1) \tag{11}$$

where $p_1(x), p_0(x) \in \mathbb{Z}[x]$ are polynomials and $p_1(n)$ does not vanish for any n. A prominent example of a hypergeometric sequence is the binomial coefficient $\binom{n}{k}$ fixed k and $n > k$, given by the recurrence

$$\binom{n}{k} = \frac{n}{n-k}\binom{n-1}{k}.$$

In Sections 7.8 and 7.4 we saw that the Catalan numbers $C(n)$ and the central binomial coefficient satisfy hypergeometric recurrences.

Given $a(n)$ which satisfies Equation (11), one can write $a(n)$ explicitly as

$$a(n) = a(1) \cdot \prod_{j=2}^{n} \frac{p_0(j)}{p_1(j)}\,. \tag{12}$$

We may rewrite Equation (12) as follows:

$$a(n) = \sum_{w \in L_{ba^\star} \cap \{a,b\}^n} \prod_{j:w[j]=b} a(1) \prod_{j:w[j]=a} \frac{p_0(j)}{p_1(j)}\,,$$

where $L_{ba^\star}$ is the language specified by the regular expression $ba^\star$, where $w[j]$ is the j-th letter of w, and where the products range over elements $j \in [n]$ such that $w[j] = b$ or $w[j] = a$ respectively. We will show that any P-recursive sequence can be interpreted in a similar way. Let Σ be an alphabet. For $s \in \Sigma$, let $\alpha_s : \mathbb{N} \to \mathbb{Z}$ be a function. We define the *weight* $\alpha(w)$ *of a word* $w \in \Sigma^\star$ by

$$\alpha(w) = \prod_{j=1}^{|w|} \alpha_{w[j]}(j)$$

where $|w|$ is the length of w. For a language $L \subseteq \Sigma^\star$ we define its *positionally weighted density* by

$$d_{L,\alpha}(n) = \sum_{w \in L \cap \Sigma^n} \alpha(w)$$

where the summation is over all words of L of length n.

DEFINITION 12.3. *A sequence $a(n)$ of integers has a* PW-interpretation *if there exists a regular language $L \subseteq \Sigma^\star$, and for each $s \in \Sigma$ a rational function $\alpha_s \in \mathbb{Q}(x)$ such that $a(n) = d_{L,\alpha}(n)$.*

THEOREM 12.2. *Let $a(n)$ be a sequence of integers. Then $a(n)$ is P-recursive iff $a(n)$ has a PW-interpretation.*

Theorem 12.2 can be modified to characterize SP-recursive sequences. In this case one simply needs to restrict the rational functions $\alpha_1, \ldots, \alpha_k$ to be polynomials over the integers.

12.5. Two explicit examples. We now give two examples. The Apéry numbers show how to get a PW-interpretation from its P-recurrence. The derangement numbers illustrate how to use an explicit summation formula for a sequence to get a P-recurrence.

12.5.1. *A Non-trivial Example of a Holonomic Sequence.* The Apéry numbers appear in Apéry's proof that $\zeta(3)$ is irrational and are known to be P-recursive, cf. [**3, 76**]. They satisfy the P-recurrence

$$n^3 b_n = (34n^3 - 51n^2 + 27n - 5) b_{n-1} - (n-1)^3 b_{n-2}\,.$$

The purpose of this subsection is to show how the polynomials of the P-recurrence of $a(n)$ are used to compute the weights for the PW-interpretation of $a(n)$.

Using a similar argument to the one used in Subsection 12.3.2 it holds that:

$$b_n = \sum_{w\in L_{rec(2)},|w|=n}$$

$$\left(\prod_{j:w[j]=1} \frac{(34j^3-51j^2+27j-5)}{j^3} \cdot \prod_{j:w[j]=2} \frac{-(j-1)^3}{j^3} \prod_{j:w[j]=\widetilde{1}} b(1) \prod_{j:w[j]=\widetilde{2}} b(2) \right),$$

where the language $L_{rec(2)}$ is from Subsection 11.1 with $q = 2$.

Let

$$\alpha_1(x) = \frac{(34x^3-51x^2+27x-5)}{x^3}, \quad \alpha_2(x) = \frac{-(x-1)^3}{x^3},$$

$\alpha_{\widetilde{1}}(x) = b_1$, $\alpha_{\widetilde{2}}(x) = b_2$, and $\alpha_b(x) = 1$. b_n has the following PW-interpretation:

$$b_n = \sum_{w\in L_{rec(2)},|w|=n} \left(\prod_{j=1}^{|w|} \alpha_{w[j]}(j) \right) = d_{L_{rec(2)},\alpha}(n)\,.$$

12.5.2. *Derangement numbers again.* The derangement numbers are given explicitly by the formula

$$D(n) = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}\,.$$

Using this formula one can easily see that $D(n)$ has a PW-interpretation,

$$D(n) = \sum_{w\in L_{0^\star 1^\star},|w|=n} \left(\prod_{j:w[j]=0} -1 \prod_{j:w[j]=1} j \right),$$

where $L_{0^\star 1^\star}$ is the language obtained from the regular expression $0^\star 1^\star$.

Part 4. Modular Recurrence Relations

13. DU-index and Specker index

In this section we give the detailed definitions of the finiteness conditions mentioned in Section 4. Specker's proof in [**74**] of Theorem 5.8 is based on the analysis of an equivalence relation $\sim_{\mathcal{C}}$ induced by a class of structures $\mathcal{C}$. However, we first look at a simpler case of disjoint unions of structures.

13.1. DU-index of a class of structures. We denote by $\mathfrak{A} \sqcup \mathfrak{B}$ the disjoint union of two $\overline{R}$-structures $\mathfrak{A}$ and $\mathfrak{B}$.

Definition 13.1. *Let $\mathcal{C}$ be a class of $\overline{R}$-structures.*

(i) We say that $\mathfrak{A}_1$ is $DU(\mathcal{C})$-equivalent to $\mathfrak{A}_2$ and write $\mathfrak{A}_1 \sim_{DU(\mathcal{C})} \mathfrak{A}_2$, if for every $\overline{R}$-structure $\mathfrak{B}$, $\mathfrak{A}_1 \sqcup \mathfrak{B} \in \mathcal{C}$ if and only if $\mathfrak{A}_2 \sqcup \mathfrak{B} \in \mathcal{C}$.

(ii) The DU-index of $\mathcal{C}$ is the number of $DU(\mathcal{C})$-equivalence classes.

Definition 13.2. *A class of structures $\mathcal{C}$ is a* Compton-Gessel class *if for every $\mathfrak{A}$ and $\mathfrak{B}$, $\mathfrak{A} \sqcup \mathfrak{B} \in \mathcal{C}$ iff both $\mathfrak{A} \in \mathcal{C}$ and $\mathfrak{B} \in \mathcal{C}$.*

I. Gessel in [**43**, Theorem 4.2] looks at Compton-Gessel classes of directed graphs which in addition have a bounded degree. He proves the following congruence theorem:

THEOREM 13.1 (I. Gessel 1984). *If $\mathcal{C}$ is a Compton-Gessel class of directed graphs of degree at most d, then*

$$d_{\mathcal{C}}(m+n) \equiv d_{\mathcal{C}}(m) \cdot d_{\mathcal{C}}(n) \pmod{\frac{m}{\ell}}$$

where ℓ is the least common multiple of all divisors of m not greater than d.

In particular, $d_{\mathcal{C}}(n)$ satisfies for every $m \in \mathbb{N}$ the linear recurrence relation $d_{\mathcal{C}}(n) \equiv a^{(m)} d_{\mathcal{C}}(n - d!m) \pmod{m}$ where $a^{(m)} = d_{\mathcal{C}}(d!m)$.

A less informative version of this theorem was stated in the introduction as Theorem 5.9.

In its proof, the following simple observation was implicitly used:

OBSERVATION 13.2. *If $\mathcal{C}$ is a Compton-Gessel class of $\overline{R}$-structures then $\mathcal{C}$ has DU-index at most 2.*

PROOF. We observe that all members of $\mathcal{C}$ are in one equivalence class, as well as that all other $\overline{R}$-structures are in one equivalence class (which is usually different than that of the members of $\mathcal{C}$). □

REMARK 13.1. *The converse is not true, because if $\mathcal{C}$ is a class of connected graphs then $\mathcal{C}$ has DU-index 1, but is not Gessel.*

This may seem unnatural, as we would expect to have at least two classes, the connected graphs and the non-connected graphs. If we allowed the empty structure to be a structure, this would indeed be the case. But in logic and model theory empty structures are traditionally avoided, because $\forall x \phi(x) \to \exists x \phi(x)$ is considered a tautology. This is in contrast to graph theory, where the empty graph is allowed.

Theorem 5.10 of the introduction can be viewed as a strong variation of Gessel's Theorem for arbitrary $\overline{R}$-structures of bounded degree. Note that the formulation of Gessel's Theorem as Theorem 13.1 contains much more information on the recurrence relation than Theorem 5.10.

13.2. Substitution of structures. A pointed $\overline{R}$-structure is a pair $(\mathfrak{A}, a)$, with $\mathfrak{A}$ an $\overline{R}$-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. [**70, 40**].

DEFINITION 13.3. *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:*

- *(i) The universe of $\mathfrak{C}$ is $A \cup B - \{a\}$.*
- *(ii) The context c is given by b, i.e., $c = b$.*
- *(iii) For $R \in \overline{R}$ 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 (identical or differing) members of B.

- *(iv) 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.*

DEFINITION 13.4. *Let $\mathcal{C}$ be a class of $\overline{R}$-structures.*

(i) We define an equivalence relation between $\overline{R}$-structures; we say that $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are equivalent, denoted $\mathfrak{B}_1 \sim_{Su(\mathcal{C})} \mathfrak{B}_2$, if for every pointed structure $(\mathfrak{A}, a)$ we have that $Subst((\mathfrak{A}, a), \mathfrak{B}_1) \in \mathcal{C}$ iff $Subst((\mathfrak{A}, a), \mathfrak{B}_2) \in \mathcal{C}$.
(ii) The Specker index *of $\mathcal{C}$ is the number of equivalence classes of $\sim_{Su(\mathcal{C})}$.*

The Specker index is related to the DU-index by the following.

Proposition 13.3. *Let $\mathcal{C}$ be a class of $\overline{R}$-structures and $\mathfrak{A}_1$ and $\mathfrak{A}_2$ be two $\overline{R}$-structures.*
(i) If $\mathfrak{A}_1 \sim_{Su(\mathcal{C})} \mathfrak{A}_2$, then $\mathfrak{A}_1 \sim_{DU(\mathcal{C})} \mathfrak{A}_2$.
(ii) The Specker index of $\mathcal{C}$ is at least as big as the DU-index of $\mathcal{C}$. In particular, if the Specker index of $\mathcal{C}$ is finite, then so is its DU-index.

We also have in analogy to Observation 13.2,

Observation 13.4. *If $\mathcal{C}$ is a class of pointed $\overline{R}$-structures such that both $(\mathfrak{A}, a), (\mathfrak{B}, b) \in \mathcal{C}$ iff $Subst((\mathfrak{A}, a), (\mathfrak{B}, b)) \in \mathcal{C}$ then the Specker-index of $\mathcal{C}$ is 2.*

Specker's proof of Theorem 5.8 consists of a purely combinatorial part:

Lemma 13.5 (Specker's Lemma). *Let $\mathcal{C}$ be a class of $\overline{R}$-structures of finite Specker index with all the relation symbols in $\overline{R}$ of arity at most 2. Then $d_{\mathcal{C}}(n)$ satisfies a modular linear recurrence relation for each $m \in \mathbb{N}$.*

The proof will be given in section 16.

13.3. Classes of finite Specker or *DU*-index. Using Proposition 13.3 we have seen that Compton-Gessel classes have finite DU-index, and that all classes of connected graphs have finite DU-index. We shall now exhibit a class $\mathcal{C}$ that has DU-index at most 2, but has an infinite Specker index. As stated in Section 7.4, $\mathrm{EQ_2CLIQUE}$ denotes the class of graphs which consist of two disjoint cliques of equal size.

Proposition 13.6. *The class* $\mathrm{EQ_2CLIQUE}$ *has infinite Specker index.*

Proof. We show that for all $i, j \in \mathbb{N}, 1 \leq i < j$, the pairs of cliques K_i and K_j are inequivalent with respect to $\sim_{Su(\mathrm{EQ_2CLIQUE})}$. To see this we look at the pointed structure $(K_j \sqcup K_1, a)$ where the vertex of K_1 is the distinguished point a. Substituting K_j for a gives a disjoint union of two K_j's, whereas substituting K_i, $i \neq j$ for a gives a disjoint union of $K_j \sqcup K_i$. The former is in $\mathrm{EQ_2CLIQUE}$ whereas the latter is not, which proves our claim. □

Let the class $\overline{\mathrm{EQ_2CLIQUE}}$ be the class of all the complement graphs of members of $\mathrm{EQ_2CLIQUE}$. We note that $\overline{\mathrm{EQ_2CLIQUE}}$ contains graphs of arbitrary large degree which are connected.

Corollary 13.7. *The class $\overline{\mathrm{EQ_2CLIQUE}}$ has finite DU-index, but infinite Specker index.*

Proof. As all graphs in $\overline{\mathrm{EQ_2CLIQUE}}$ are connected, $\overline{\mathrm{EQ_2CLIQUE}}$ has DU-index at most 2, using Proposition 13.1. On the other hand, it is not hard to see that similarly to $\mathrm{EQ_2CLIQUE}$ this class has an infinite Specker index. □

It is an easy exercise to show that the class HAM of graphs which contain a Hamiltonian cycle also has infinite Specker index. We have seen in Subsection 3.2

that the classes HAM, $\mathrm{EQ}_2\mathrm{CLIQUE}$ and $\overline{\mathrm{EQ}_2\mathrm{CLIQUE}}$ are not **CMSOL**-definable. So far, our examples of the classes of infinite Specker index were not definable in **CMSOL**. This is no accident. Specker noted that all **MSOL**-definable classes of $\overline{R}$-structures have a finite Specker index. We shall see that this can be improved.

THEOREM 13.8. *If $\mathcal{C}$ is a class of $\overline{R}$-structures (with no restrictions on the arity) which is* **CMSOL**-*definable, then $\mathcal{C}$ has a finite Specker index.*

The proof is given in Section 14. It uses a form of the Feferman-Vaught Theorem for **CMSOL**.

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

DEFINITION 13.5. *Let C_n denote the cycle of size n, i.e. a regular connected graph of degree 2 with n-vertices. Let $A \subseteq \mathbb{N}$ be any set of natural numbers and $Cycle(A) = \{C_n : n \in A\}$.*

PROPOSITION 13.9 (Specker). *$Cycle(A)$ has Specker index at most 5.*

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

COROLLARY 13.10 (Specker). *There is a continuum of classes (of graphs, of $\overline{R}$-structures) of finite Specker index which are not* **CMSOL**-*definable.*

PROOF. Clearly there is continuum of classes of the type $Cycle(A)$, and hence a continuum of classes that are not definable in **CMSOL** (or even in second order logic, **SOL**). $\square$

It is easy to compute $d_{Cycle(A)}$:

$$d_{Cycle(A)}(n) = \begin{cases} 0 & \text{if } n \notin A \\ (n-1)! & \text{otherwise .} \end{cases}$$

Hence it is trivially MC-finite. This does not have to be necessarily the case. Here is a way to modify the above example.

THEOREM 13.11. *Let $\mathcal{C}$ be a class of $\overline{R}$ structures with counting function $d_{\mathcal{C}}(n)$. For every $A \subseteq \mathbb{N}$ there is a class of structures $\mathcal{C}(A)$ (with all classes being different for different choices of A) such that for all $n \in \mathbb{N}$*

$$d_{\mathcal{C}(A)}(n) = d_{\mathcal{C}}(n) + d_{Cycle(A)}(n)$$

Furthermore, if $\mathcal{C}$ is of finite Specker index, then so is $\mathcal{C}(A)$ for every A (but its index is possibly bigger).

Proof. The structures in $\mathcal{C}(A)$ have besides the relation symbols $\overline{R}$ a new unary relation symbol U and a new binary relation symbol E. The interpretations of U^B and E^B in a structure $\mathfrak{B} \in \mathcal{C}(A)$ on the universe $[n]$ are given as follows:

(i) Either $U^B = \emptyset$ and $E^B = \emptyset$ and the underlying $\overline{R}$-structure is in $\mathcal{C}$, or
(ii) $U^B = [n]$ and $\langle [n], E^B \rangle \in Cycle(A)$ and all $S^B = \emptyset$ for all $S \in \overline{R}$.

Clearly, we now have $d_{\mathcal{C}(A)}(n) = d_{\mathcal{C}}(n) + d_{Cycle(A)}(n)$.

To see that the Specker index of $\mathcal{C}(A)$ is finite, we look at two $(\overline{R} \cup \{U, E\})$-structures $\mathfrak{B}_1$ and $\mathfrak{B}_2$ and a pointed $(\overline{R} \cup \{U, E\})$-structures $(\mathfrak{A}, a)$, and put $\mathfrak{C}_i = Subst((\mathfrak{A}, a), \mathfrak{B}_i)$ for $i = 1, 2$. Whether $\mathfrak{C}_1$ and $\mathfrak{C}_2$ are $\mathcal{C}(A)$-equivalent can now be decided by checking whether the interpretations of U, E or $S \in \overline{R}$ are empty or full, and whether the corresponding reducts are $\mathcal{C}$-equivalent or $Cycle(A)$-equivalent. □

Remark 13.2. *This shows that, in contrast to the Myhill-Nerode Theorem, no characterization of the classes of finite Specker index in terms of their definability in* **CMSOL** *is possible.*

13.5. An SOL but not CMSOL-definable class of finite Specker index. Although a classification of all classes of finite Specker index seems unachievable on account of there being a continuum of these, one could still hope to characterize all **SOL**-definable classes of finite index. We show here that definability in **CMSOL** is not such a characterization.

Definition 13.6. *We look at the infinite graph whose vertex set is $\mathbb{Z} \times \mathbb{Z}$ and for which every (i, j) and (i', j') are adjacent if and only if $|i - i'| + |j - j'| = 1$. We say that a graph G is a* grid graph *if it is a (finite) subgraph of the above infinite graph.*

In [**69**] the following was proven:

Theorem 13.12 ([**69**]). *The class of all grid graphs, which is definable in* **SOL**, *is not definable in* **CMSOL**.

On the other hand, this is also a class with a finite Specker index.

Proposition 13.13. *The class of all grid graphs has a finite Specker index; therefore, there exist classes of finite Specker index which are definable in* **SOL** *but not in* **CMSOL**.

Proof. We observe that all graphs with five or more vertices fall into the following two Specker equivalence classes:

(i) Graphs G for which $Subst((S, s), G)$ is a grid graph if and only if S is a grid graph and s is an isolated vertex of S.
(ii) Graphs G for which $Subst((S, s), G)$ is never a grid graph.

All binary structures which are not graphs clearly fall into the second equivalence class above. Thus the index is finite. □

14. The rôle of logic

Although Theorem 5.8 is stated for classes of structures definable in some logic, logic is only used to verify the hypothesis of Specker's Lemma, 13.5. In this section we develop the machinery which serves this purpose. The crucial property needed to prove Theorem 13.8 is a reduction property which says that both for the disjoint union $\mathfrak{A} \sqcup \mathfrak{B}$ and for the substitution $Subst((\mathfrak{A}, a), \mathfrak{B})$ the truth value of a sentence

$\phi \in \mathbf{CMSOL}(\overline{R})$ depends only on the truth values of the sentences of the same quantifier rank in the structures $\mathfrak{A}$ and $\mathfrak{B}$, respectively $\langle \mathfrak{A}, a \rangle$ and $\mathfrak{B}$. For the case of **MSOL** this follows either from the Feferman-Vaught Theorem for disjoint unions together with some reduction techniques, or using Ehrenfeucht-Fraïssé games. The latter is used in [**74**]. We shall use the former, as it is easier to adapt for **CMSOL**. For the Feferman-Vaught Theorem the reader is referred to [**34**], or the monographs [**62, 50**], or the survey [**58**].

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

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

The following is folklore, cf. [**30**].

PROPOSITION 14.1. *There are, up to logical equivalence, only finitely many formulas in* $\mathbf{CMSOL}^q(\overline{R}, \overline{x}, \overline{U})$. *In particular, the equivalence relation* $\sim^q_{\mathbf{CMSOL}}$ *is of finite index.*

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

The following was first proven by E. Beth in 1952 and then generalized by Feferman and Vaught in 1959 for **FOL**. For **MSOL** it is due to Läuchli and Leonhard, 1966 and for **CMSOL** it is due to B. Courcelle, 1990. The respective references are [**11, 33, 34, 55, 24**].

THEOREM 14.2 (Feferman-Vaught-Courcelle).

(i) For every formula $\phi \in \mathbf{CMSOL}^q(\tau)$ *one can compute in* polynomial time *a sequence of formulas*

$$\langle \psi^A_1, \ldots, \psi^A_m, \psi^B_1, \ldots, \psi^B_m \rangle \in \mathbf{CMSOL}^q(\tau)^{2m}$$

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

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

if and only if

$$B_\phi(b^A_1, \ldots b^A_m, b^B_1, \ldots b^B_m) = 1$$

where $b^A_j = 1$ *iff* $\mathfrak{A} \models \psi^A_j$ *and* $b^B_j = 1$ *iff* $\mathfrak{B} \models \psi^B_j$.

A detailed proof is found in [**24**, Lemma 4.5, page 46ff].

14.3. Quantifier-free transductions and CMSOL. Let $\overline{R} = R_1, \ldots, R_s$ where R_i is of arity $\rho(i)$. An $\overline{R}$-translation scheme Φ is a sequence of quantifier-free formulas $\Phi = \langle \theta_0(x), \theta_i(x_1, \ldots, x_{\rho(i)}) : i \leq s \rangle$ with free variables as indicated. With Φ we associate a map Φ^* which maps an $\overline{R}$-structure $\mathfrak{A}$ to an $\overline{R}$-structure where the universe is the subset of the universe of $\mathfrak{A}$ defined by θ_0, and where the interpretations of R_i are replaced by the relations defined by θ_i. Φ^* is called a *quantifier free* $\overline{R}$*-transduction.*

For the general framework of translation schemes and transductions, cf. [**25, 56, 30**]. Note that θ_0 has only one free variable. In the literature this corresponds to *scalar transductions.*

LEMMA 14.3. *Let* Φ^* *be a quantifier free (scalar)* $\overline{R}$*-transduction. Assume* $\mathfrak{A}_1, \mathfrak{A}_2$ *are* $\overline{R}$*-structures and* $\mathfrak{A}_1 \sim^q_{\mathbf{CMSOL}} \mathfrak{A}_2$*. Then* $\Phi^*(\mathfrak{A}_1) \sim^q_{\mathbf{CMSOL}} \Phi^*(\mathfrak{A}_2)$*.*

All we need here is that substitution of pointed structures in pointed structures can be obtained from the disjoint union of the two pointed structures by a scalar transduction. Note that the disjoint union of two pointed structures is, strictly speaking, "doubly pointed", and the two distinguished points play different rôles.

LEMMA 14.4. $Subst((\mathfrak{A}, a), (\mathfrak{B}, b))$ *can be obtained from the doubly pointed disjoint union of* $(\mathfrak{A}, a)$ *and* $(\mathfrak{B}, b)$ *by a quantifier free transduction.*

SKETCH OF PROOF: The universe of the structure is $C = (A \sqcup B) - \{a\}$. For each relation symbol $R \in \overline{R}$ we put

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

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

PROPOSITION 14.5. *Assume* $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{B}_1, \mathfrak{B}_2$ *are* $\overline{R}$*-structures and with contexts* a_1*,* a_2*,* b_1*,* b_2*, respectively, and*

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

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

PROOF. Use Theorem 14.2, Lemma 14.3 and Lemma 14.4. □

14.4. Finite index theorem for CMSOL. Now we can state and prove the Finite Index Theorem:

THEOREM 14.6. *Let* $\mathcal{C}$ *be defined by a* $\mathbf{CMSOL}(\overline{R})$*-sentence* ϕ *of quantifier rank* q*. Then* $\mathcal{C}$ *has finite Specker index, and also finite DU-index, which are both bounded by the number of inequivalent* $\mathbf{CMSOL}^q(\overline{R})$*-sentences. This number is finite by Proposition 14.1.*

PROOF. We have to show that the equivalence relation $\mathfrak{A} \sim^q_{\mathbf{CMSOL}} \mathfrak{B}$ is a refinement of $\mathfrak{A} \sim_{Su(\mathcal{C})} \mathfrak{B}$. But this follows from Proposition 14.5. For the *DU*-index this follows from Proposition 13.3. □

PROBLEM 2. *Are there any logics* $\mathcal{L}$ *on finite structures extending* **CMSOL** *such that the analog of Theorem 14.6 remains true?*

15. Structures of bounded degree

DEFINITION 15.1. *For an* **MSOL** *class* $\mathcal{C}$*, denote by*[6] $f^{(d)}_{\mathcal{C}}(n)$ *the number of structures over* $[n]$ *that are in* $\mathcal{C}$ *and whose degree is at most* d*.*

In this section we prove Theorem 5.10 in the following form:

[6] We change here the notation and use $f^{(d)}$ rather than $d^{(d)}$ to avoid confusion of the counting function with the degree.

THEOREM 15.1. *If $\mathcal{C}$ is a class of $\overline{R}$-structures which has a finite DU-index, then the function $f_{\mathcal{C}}^{(d)}(n)$ is ultimately periodic modulo m, for every $m \in \mathbb{N}$, and therefore is MC-finite.*

Furthermore, if all structures of $\mathcal{C}$ are connected, then $f_{\mathcal{C}}^{(d)}(n)$ is ultimately zero modulo m, and therefore is trivially MC-finite.

LEMMA 15.2. *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 15.1 we define orbits for permutation groups rather than for single permutations.

DEFINITION 15.2. *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 15.3. *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 structures in $\mathcal{C}$ have a degree bounded by d.

OBSERVATION 15.4. *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 $\mathcal{C}_d$.* □

Instead of $\mathcal{C}$ we look at $\mathcal{C}_d$, which by Observation 15.4 also has a finite DU-index. We now note that there is only one equivalence class containing structures whose maximum degree is larger than d, namely 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 exhibit a linear recurrence relation modulo m on the vector function $(d_{\mathcal{E}}(n))_{\mathcal{E}}$ 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 15.3 allows us to prove the following.

LEMMA 15.5. *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*

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

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 15.3 the size of $\mathrm{Orb}_{S_{[tC]}}(\mathfrak{A})$ is divisible by $\binom{tC}{d'}$, and therefore it is divisible by m. Therefore, $d_{\mathcal{D}}(n)$ 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 15.2 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 $\mathfrak{A} \in \mathcal{N}_{\mathcal{C}}^{(d)} \neq \mathcal{D}$). $\square$

Proof of Theorem 15.1: We use Lemma 15.5: Since there is only a finite number of possible values modulo m for the finite dimensional vector $(d_{\mathcal{E}}(n))_{\mathcal{E}}$, the linear recurrence relation in Lemma 15.5 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 $(d_{\mathcal{E}}(n))_{\mathcal{E}}$ 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 15.3. 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. $\square$

16. Structures of unbounded degree

In this section we prove Specker's Lemma 13.5 for structures of unbounded degree. In fact this is a somewhat modified version of Specker's simplified proof for the case where $m = p$ is a prime, as in [**75**].

In order to prove that counting functions of classes of finite Specker index (over unary and binary relation symbols) are ultimately periodic modulo any integer m, it is enough to prove this for any $m = p^k$ where p is a prime number; any other m will then follow by using the Chinese Reminder Theorem. In the following subsections we prove ultimate periodicity for $m = p^k$. First we define a permutation group $G_{p,k}$ which ensures that all structures have large orbits under it, apart from those structures which are "invariant enough" to be represented in terms of a sequence of substitutions in a smaller structure. Then, using this group we show a linear recurrence relation in a vector function that is related to our class $\mathcal{C}$, from which Specker's Lemma follows.

16.1. A permutation group ensuring large orbits.

To deal with the exceptional case $p = 2$ we let $\widetilde{p} = 4$ if $p = 2$, and $\widetilde{p} = p$ otherwise. As our structures have only binary and unary relations, we use the language of graphs, and speak of vertices and edges.

In the following we construct a permutation group $G_{p,k}$.
It acts on the set $\{1,\dots,\widetilde{p}^k\}$ and satisfies the following properties:

(i) The size of $G_{p,k}$ is a power of p and hence the size of an orbit of any structure over $\{1,\dots,n\}$, where $n \geq \widetilde{p}^k$, is also a power of p.
(ii) If the orbit of a binary structure $\mathfrak{A}$ with universe $\{1,\dots,n\}$ has size less than p^k, then $\mathfrak{A}$ is the result of substituting the $\widetilde{p}^{k-1}$ many substructures induced on the sets $\{1+\widetilde{p}i,\dots,\widetilde{p}(i+1)\}$ (a substructure for every $0 \leq i < \widetilde{p}^{k-1}$) into a smaller structure. Equivalently, for any set $v \in \{1+\widetilde{p}i,\dots,\widetilde{p}(i+1)\}$, the relations between v and the vertices outside $\{1+\widetilde{p}i,\dots,\widetilde{p}(i+1)\}$ are invariant with regards to permuting this subset.

To achieve this we define $G_{p,k}$ is as follows:

(i) We relabel the vertices $\{1,\dots,\widetilde{p}^k\}$ of $\mathfrak{A}$ with vectors of $(\mathbb{Z}_{\widetilde{p}})^k$, by relabeling i with $(x_1,\dots,x_k)$, where $x_j \equiv \lfloor (i-1)/\widetilde{p}^{j-1} \rfloor \pmod{\widetilde{p}}$ for $1 \leq j \leq k$.
(ii) We use $\overline{x} = (x_1,\dots,x_k)$ and define σ_i by

$$\sigma(\overline{x}) = \begin{cases} (x_1,\dots,x_{i-1},x_i+1,x_{i+1},\dots,x_k) & \text{if } x_{i+1} = \cdots = x_k = 0 \\ \overline{x} & \text{otherwise} \end{cases}$$

with the addition being modulo $\widetilde{p}$.
(iii) $G_{p,k}$ is the group generated by $\sigma_1,\dots,\sigma_k$.

Observation 16.1. *The following are easy to verify:*

(i) $G_{p,k-1}$ *is a subgroup of* $G_{p,k}$ *in the appropriate sense.*
(ii) The order of $G_{p,k}$ *(and hence of any orbit it induces on structures) is a power of* p *(remember that* $\widetilde{p}$ *is a power of* p*).*

Some additional terminology is needed:

(i) The set of all vertices labeled by $(x_1,\dots,x_k)$ where $x_{i+1} = \cdots = x_k = 0$ is called the *i-origin*.
(ii) The set of all vertices labeled by $(x_1,\dots,x_k)$ where $x_k = r$ for some fixed r is called the *r-shifted $(k-1)$-origin*;
(iii) Similarly, shifts of other origins: the $(r_1,\dots,r_l)$-shifted $(k-l)$-origin is just the set of vertices labeled by $(x_1,\dots,x_k)$ where $x_i + l = r_i$ for $1 \leq i \leq l$.
(iv) The (0-shifted) 1-origin is simply called the *origin*; this is the set which corresponds to $\{1,\dots,\widetilde{p}\}$ before the relabeling above.

Note that the k-origin (for which there exist no shifts) is the whole set which $G_{p,k}$ permutes.

Lemma 16.2. *For a structure* $\mathfrak{A}$ *with the universe* $\{1,\dots,n\}$*, if any of its unary relations is not constant over the origin, then the size of its orbit under* $G_{p,k}$ *is divisible by* p^k.

Proof. The proof is by induction on k, with the case $k=1$ being trivial (note that $\widetilde{p}$ is either p or p^2). Assume now that the lemma is known for $k-1$. We note that either $\mathfrak{A}$ or $\sigma_1(\mathfrak{A})$ has a $(k-1)$-origin that is different from the 1-shifted $(k-1)$-origin of $\mathfrak{A}$ (because of the assumption that there exists a unary relation that is non-constant over the origin). This means that either $\sigma_k(\mathfrak{A})$ or $\sigma_k(\sigma_1(\mathfrak{A}))$ is not a member of $\mathrm{Orb}_{G_{p,k-1}}(\mathfrak{A})$, and so the size of $\mathrm{Orb}_{G_{p,k}}(\mathfrak{A})$ is larger than that of $\mathrm{Orb}_{G_{p,k-1}}(\mathfrak{A})$, while both sizes are powers of p, so the lemma follows. □

LEMMA 16.3. *For a structure $\mathfrak{A}$ with the universe $\{1, \ldots, n\}$, if for any of its binary relations there exists a vertex v not in the origin whose relations with the vertices in the origin are not constant, then the size of its orbit under $G_{p,k}$ is divisible by p^k.*

PROOF. Again by induction, where in the case of $G_{p,1}$ we can just treat the (appropriate direction) of the relation between v and the origin as a unary relation over the origin for the purpose of bounding the orbit size. The induction step splits into three cases.

The first case is where v is outside the k-origin. In this case we just treat the relation from v to the vertices on which $G_{p,k}$ acts as a unary relation for the purpose of bounding the orbit size (remember that since the order of $G_{p,k}$ is a power of p, it is sufficient to show that the orbit size is at least p^k), and use Lemma 16.2.

The second case is where v is inside the $(k-1)$-origin. In this case, similarly to the proof of Lemma 16.2, we note that either $\mathfrak{A}$ or $\sigma_1(\mathfrak{A})$ has a $(k-1)$-origin that is different from the 1-shifted $(k-1)$-origin of $\mathfrak{A}$, and so either $\sigma_k(\mathfrak{A})$ or $\sigma_k(\sigma_1(\mathfrak{A}))$ is different from all members of $\mathrm{Orb}_{G_{p,k-1}}(\mathfrak{A})$.

The third case is where v is in the k-origin but not in the $(k-1)$-origin. Then it is in the j-shifted $(k-1)$-origin for some $0 < j < \widetilde{p}$. In this case either $\mathfrak{A}$ or $\sigma_1(\mathfrak{A})$ is such that the relations between the $(k-1)$-origin and the j-shifted $(k-1)$-origin are different from the relations in $\mathfrak{A}$ between the 1-shifted $(k-1)$-origin and the $j+1$-shifted (modulo $\widetilde{p}$) $(k-1)$-origin. Thus either $\sigma_k(\mathfrak{A})$ or $\sigma_k(\sigma_1(\mathfrak{A}))$ is not a member of $\mathrm{Orb}_{G_{p,k-1}}(\mathfrak{A})$, showing that $\mathrm{Orb}_{G_{p,k}}(\mathfrak{A})$ is larger than $\mathrm{Orb}_{G_{p,k-1}}(\mathfrak{A})$. $\square$

The above lemma is the essence of what we need from this section, but the condition above for not being in a large orbit is problematic in that it is not itself closed under the action of $G_{p,k}$. However, the following corollary gives a condition that is closed under $G_{p,k}$.

COROLLARY 16.4. *If a binary structure over $\{1, \ldots, n\}$ has an orbit under $G_{p,k}$ whose size is not a multiple of p^k, then it is the result of a substitution of the substructures induced by its shifted (and unshifted) 1-origins in an appropriate (smaller) structure.*

PROOF. For every shifted 1-origin apply Lemma 16.3 to $\sigma(\mathfrak{A})$ for an appropriate $\sigma \in G_{p,k}$; since the result of all these applications is an invariance of the relations (apart from those internal to a shifted 1-origin) under any permutation inside the shifted 1-origins, this means that the structure is the result of the appropriate substitutions (the order of substitutions is not important because they are all substitutions of different vertices of the original smaller structure). $\square$

The above is the corollary that we will use to show a modular linear recurrence concerning structures with a finite Specker index.

16.2. Bounded Specker index implies periodicity. We now follow the method of [**75**], only instead of the permutations used there, we use the group $G_{p,k}$ to prove periodicity modulo p^k.

Let $\mathcal{C}_1, \ldots, \mathcal{C}_s$ be the enumeration of all classes residing from $\mathcal{C}$. Given a sequence of integers $\mathbf{a} = (a_1, \ldots, a_l)$, we define $\mathcal{C}_{\mathbf{a}}$ as the class of all structures $\mathfrak{A}$ (over our fixed language with unary and binary relations) with the universe $\{1, \ldots, n\}$, such that $n \geq l$, and if one substitutes the vertex i with $\mathcal{C}_{a_i}$ for every $1 \leq i \leq l$, then one gets a structure in $\mathcal{C}$ (it is not hard to see that the order in which these substitutions are performed is not important). Note in particular that for the sequence ε of size 0 we get $\mathcal{C}_\varepsilon = \mathcal{C}$.

Claim 16.5. *If* $\mathbf{a}$ *is a permutation of* $\mathbf{a}'$*, then* $d_{\mathcal{C}_\mathbf{a}}(n) = d_{\mathcal{C}_{\mathbf{a}'}}(n)$.

Proof. Simple; a one to one correspondence between members of $\mathcal{C}_\mathbf{a}$ and the members of $\mathcal{C}_{\mathbf{a}'}$ is induced by an appropriate permutation of the vertices. □

By virtue of this claim, from now on we focus our attention on sequences $\mathbf{a}$ that are monotone nondecreasing. The ultimate periodicity results from the following two lemmas. Note that $\widetilde{p}$ and $G_{p,k}$ are defined as per the preceding section.

Lemma 16.6. *If* l *is the length of* $\mathbf{a}$ *and* $n \geq l + \widetilde{p}^k$*, then* $d_{\mathcal{C}_\mathbf{a}}(n)$ *is congruent modulo* p^k *to a linear sum (whose coefficients depend only on* $\mathbf{a}$*,* p*,* k *and* $\mathcal{C}$*) of functions of the type* $d_{\mathcal{C}_{\mathbf{a}'}}(n - \widetilde{p}^k + \widetilde{p}^{k-1})$ *where* $\mathbf{a}'$ *ranges over the sequences that are composed from* $\mathbf{a}$ *by inserting* $\widetilde{p}^{k-1}$ *additional values (and in particular are of length* $l + \widetilde{p}^{k-1}$*).*

Proof. We look at the orbits of structures in $\mathcal{C}_\mathbf{a}$ over $\{1, \ldots, n\}$, where the permutation group G is equal to $G_{p,k}$, except that it acts on the vertices $l+1, \ldots, l+\widetilde{p}^k$ (instead of $1, \ldots, \widetilde{p}^k$). Because of Corollary 16.4, to get the number $d_{\mathcal{C}_\mathbf{a}}(n)$ modulo p^k we now need only concern ourselves with structures that result from substituting the substructures, induced by $l + 1 + \widetilde{p}i, \ldots, l + \widetilde{p}(i + 1)$ for every $0 \leq i < \widetilde{p}^{k-1}$, into an appropriate smaller structure.

The number of possibilities for these substitution schemes is finite (depending only on p, k and $\mathcal{C}$), and the count of the smaller structures in which the substitutions take place corresponds to the required linear combination of functions of the type $d_{\mathcal{C}_{\mathbf{a}'}}(n - \widetilde{p}^k + \widetilde{p}^{k-1})$. □

Lemma 16.7. *If* $\mathbf{a}$ *contains at least* $\widetilde{p}^k$ *copies of the same value, then* $d_{\mathcal{C}_\mathbf{a}}(n)$ *is congruent modulo* p^k *to a linear sum (whose coefficients depend only on* $\mathbf{a}$*,* p*,* k *and* $\mathcal{C}$*) of functions of the type* $d_{\mathcal{C}_{\mathbf{a}'}}(n - \widetilde{p}^k + \widetilde{p}^{k-1})$ *where* $\mathbf{a}'$ *ranges over sequences that result from* $\mathbf{a}$ *by removing* $\widetilde{p}^k$ *copies of the recurring value and inserting* $\widetilde{p}^{k-1}$ *new values (some of which may be identical to the removed ones); in particular the length of any possible* $\mathbf{a}'$ *in this sum is* $l - \widetilde{p}^k + \widetilde{p}^{k-1}$*.*

Proof. Somewhat similar to the proof of Lemma 16.6. Let i be such that $a_i, \ldots, a_{i+\widetilde{p}^k-1}$ are identical. We look at orbits of structures in $\mathcal{C}_\mathbf{a}$ over $\{1, \ldots, n\}$, where the permutation group G is equal to $G_{p,k}$, only that it acts on $\{i, \ldots, i+\widetilde{p}^k - 1\}$. Up to congruences modulo p^k we only need to look at structures that result from substituting the $\widetilde{p}^{k-1}$ substructures corresponding to the shifted 1-origins of G into the appropriate structures of order $n - \widetilde{p}^k + \widetilde{p}^{k-1}$. □

With the above two lemmas one can complete the proof of periodicity: The set $\mathbf{A}$ of sequences $\mathbf{a}$ which do not satisfy the requirements for Lemma 16.7 is clearly finite, and it includes ε. We now define a vector function $\mathbf{f}$ from $\mathbb{N}$ to $(\mathbb{Z}_{p^k})^{|\mathbf{A}|}$ by $\mathbf{f}(n) \equiv \langle d_{\mathcal{C}_\mathbf{a}}(n) | \mathbf{a} \in \mathbf{A}\rangle \pmod{p^k}$. Using now Lemma 16.6 and Lemma 16.7 we can obtain a linear recurrence relation between $\mathbf{f}(n)$ and $\mathbf{f}(n-1), \ldots, \mathbf{f}(n-C)$, where C is bounded by $2\widetilde{p}^k$ plus the size of the longest member of $\mathbf{A}$, both of which depend only on p, k, and the Specker index of $\mathcal{C}$. Since $\mathbf{f}(n)$ has a finite number of possible values for any fixed n, ultimate periodicity follows.

Acknowledgements. We are grateful to R. Bar-Yehuda, B. Courcelle, R. Roth, A. Sidi, H. Wilf and I. Yavneh for various suggestions and comments relating to an early draft of this article, and to I. Gessel and M. Grohe for their comments at a later stage. We are particularly grateful to S. Burris and K. Yeats for their careful reading and their suggestions in the last stages of the preparation of this article.

We would also like to thank E. Specker, for his encouragement and interest in our work, and for his various suggestions and clarifications.

References

[1] *The on-line encyclopedia of integer sequences (oeis)*, http://oeis.org/.
[2] M. Ajtai and R. Fagin, *Reachability is harder for directed than for undirected finite graphs*, Journal of Symbolic Logic **55.1** (1990), 113–150.
[3] R. Apéry, *Irrationalité de $\zeta(2)$ et $\zeta(3)$*, Astérisque **61** (1979), 11–13.
[4] József Balogh, Béla Bollobás, and Miklós Simonovits, *The number of graphs without forbidden subgraphs*, J. Comb. Theory, Ser. B **91** (2004), no. 1, 1–24.
[5] József Balogh, Béla Bollobás, and David Weinreich, *The speed of hereditary properties of graphs*, J. Comb. Theory, Ser. B **79** (2000), no. 2, 131–156.
[6] ———, *The penultimate rate of growth for graph properties*, Eur. J. Comb. **22** (2001), no. 3, 277–289.
[7] 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** (2001), 129–153.
[8] J. P. Bell, S.N. Burris, and K. Yeats, *Spectra and systems of equations*, this volume, 2011.
[9] F. Bergeron, G. Labelle, and P. Leroux, *Combinatorial species and tree-like structures*, Cambridge University Press, 1998, Preprint, UQAM Université du Québec à Montréal.
[10] ———, *Introduction to the theory of species of structures*, Preprint, UQAM Université du Québec à Montréal, 2008.
[11] E. W. Beth, *Observations métamathématiques sur les structures simplement ordonneés*, Applications scientifiques de la logique mathématique, Collection de Logique Mathématique, Serie A, vol. 5, Paris and Louvain, 1954, pp. 29–35.
[12] G.D. Birkhoff, *General theory of irregular difference equations*, Acta Mathematica **54** (1930), 205–246.
[13] G.D. Birkhoff and W. J. Trjitzinsky, *Analytic theory of singular difference equations*, Acta Mathematica **60** (1933), 1–89.
[14] C. Blatter and E. Specker, *Le nombre de structures finies d'une th'eorie à charactère fin*, Sciences Mathématiques, Fonds Nationale de la recherche Scientifique, Bruxelles (1981), 41–44.
[15] ———, *Modular periodicity of combinatorial sequences*, Abstracts of the AMS **4** (1983), 313.
[16] B. Bollobás, *Modern graph theory*, Springer, 1999.
[17] M. Bousquet-Mélou, *Counting walks in the quarter plane*, Mathematics and Computer Science: Algorithms, trees, combinatorics and probabilities, Trends in Mathematics, Birkhauser, 2002, pp. 49–67.
[18] S.N. Burris, *Number theoretic density and logical limit laws*, Mathematical Surveys and Monographs, vol. 86, American mathematical Society, 2001.
[19] D. Callan, *A combinatorial derivation of the number of labeled forests*, Journal of Integer Sequences **6** (2003), Article 03.4.7.
[20] N. Chomsky and M.P. Schützenberger, *The algebraic theory of context free languages*, Computer Programming and Formal Systems (P. Brafford and D. Hirschberg, eds.), North Holland, 1963, pp. 118–161.
[21] M. Chudnovsky, N. Robertson, P. Seymour, and R. Thomas, *The strong perfect graph theorem*, Annals of Mathematics **164** (2006), no. 1, 51–229.
[22] Kenneth J. Compton, *Applications of logic to finite combinatorics*, Ph.D. thesis, University of Wisconsin, 1980.
[23] Kevin J. Compton, *Some useful preservation theorems*, J. Symb. Log. **48** (1983), no. 2, 427–440.

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

[25] ———, *Monadic second order graph transductions: A survey*, Theoretical Computer Science **126** (1994), 53–75.

[26] ———, *The expression of graph properties and graph transformations in monadic second-order logic*, Handbook of graph grammars and computing by graph transformations, Vol. 1 : Foundations (G. Rozenberg, ed.), World Scientific, 1997, pp. 313–400.

[27] B. Courcelle and J. Engelfriet, *Graph structure and monadic second-order logic, a language theoretic approach*, Cambridge University Press, 2012, in press.

[28] R. Diestel, *Graph theory*, 3 ed., Graduate Texts in Mathematics, Springer, 2005.

[29] A. Durand, N. Jones, J. Makowsky, and M. More, *Fifty Years of the Spectrum Problem: Survey and New Results*, ArXiv e-prints (2009).

[30] H.-D. Ebbinghaus and J. Flum, *Finite model theory*, Perspectives in Mathematical Logic, Springer, 1995.

[31] P. Erdös and I. Kaplansky, *The asymptotic number of latin rectangles*, Amer. J. Math. **68** (1946), 230–236.

[32] G. Everest, A. van Porten, I. Shparlinski, and T. Ward, *Recurrence sequences*, Mathematical Surveys and Monographs, vol. 104, American Mathematical Society, 2003.

[33] S. Feferman, *Some recent work of Ehrenfeucht and Fraïssé*, Proceedings of the Summer Institute of Symbolic Logic, Ithaca 1957, pp. 201-209, 1957.

[34] S. Feferman and R. Vaught, *The first order properties of algebraic systems*, Fundamenta Mathematicae **47** (1959), 57–103.

[35] E. Fischer, *The Specker-Blatter theorem does not hold for quaternary relations*, Journal of Combinatorial Theory, Series A **103** (2003), 121–136.

[36] E. Fischer and J. A. Makowsky, *The Specker-Blatter theorem revisited*, COCOON, Lecture Notes in Computer Science, vol. 2697, Springer, 2003, pp. 90–101.

[37] P. Flajolet, *On congruences and continued fractions for some classical combinatorial quantities*, Discrete Mathematics **41** (1982), 145–153.

[38] P. Flajolet and R. Sedgewick, *Analytic combinatorics*, Cambridge University Press, 2009.

[39] Haim Gaifman, *On local and non-local properties*, Logic Colloquium '81 (J. Stern, ed.), North-Holland publishing Company, 1982.

[40] F. Gécseg and M. Steinby, *Tree languages*, Handbook of formal languages, Vol. 3 : Beyond words (G. Rozenberg and A. Salomaa, eds.), Springer Verlag, Berlin, 1997, pp. 1–68.

[41] S. Gerhold, *On some non-holonomic sequences*, Electronic Journal of Combinatorics **11** (2004), 1–7.

[42] A. Gertsch and A.M. Robert, *Some congruences concerning the Bell numbers*, Bull. Belg. Math. Soc. **3** (1996), 467–475.

[43] I. Gessel, *Combinatorial proofs of congruences*, Enumeration and design (D.M. Jackson and S.A. Vanstone, eds.), Academic Press, 1984, pp. 157–197.

[44] ———, *Counting latin rectangles*, Bull. Amer. Math. Soc. **16** (1987), 79–83.

[45] ———, *Symmetric functions and p-recursiveness*, J. Comb. Theory, Ser. A **53** (1990), no. 2, 257–285.

[46] M. Grohe and S. Kreutzer, *Methods for algorithmic meta theorems*, this volume, 2011.

[47] G Hansel, *A simple proof of the Skolem-Mahler-Lech theorem*, Theor. Comput. Sci. **43** (1986), 91–98.

[48] F. Harary and E. Palmer, *Graphical enumeration*, Academic Press, 1973.

[49] L. A. Hemaspaandra and H. Vollmer, *The satanic notations: Counting classes beyond $\sharp\mathbf{P}$ and other definitional adventures*, SIGACTN: SIGACT News (ACM Special Interest Group on Automata and Computability Theory) **26** (1995), 3–13.

[50] W. Hodges, *Model theory*, Encyclopedia of Mathematics and its Applications, vol. 42, Cambridge University Press, 1993.

[51] J. E. Hopcroft and J. D. Ullman, *Introduction to automata theory, languages and computation*, Addison-Wesley Series in Computer Science, Addison-Wesley, 1980.

[52] T. Kotek and J. A. Makowsky, *Definability of combinatorial functions and their linear recurrence relations*, Fields of Logic and Computation, 2010, pp. 444–462.

[53] T. Kotek and J.A. Makowsky, *A representation theorem for holonomic sequences based on counting lattice paths*, submitted, 2010.

[54] ———, *A representation theorem for (q-)holonomic sequences*, submitted, 2010.

[55] H. Läuchli and J. Leonhard, *On the elementary theory of order*, Fundamenta Mathematicae **59** (1966), 109–116.
[56] L. Libkin, *Elements of finite model theory*, Springer, 2004.
[57] E. Lucas, *Théorie des fonctions numériques simplement périodiques*, Amer. J. Math. **1** (1878), 184–240, 289–321.
[58] J.A. Makowsky, *Algorithmic uses of the Feferman-Vaught theorem*, Annals of Pure and Applied Logic **126.1-3** (2004), 159–213.
[59] B.D. McKay and E. Rogoyski, *Latin squares of order 10*, Electronic J. Combinatorics **2.1** (1995), N3, 1–4.
[60] B.D. McKay and I.M. Wanless, *On the number of latin squares*, Ann.Comb. **9** (2005), 335–344.
[61] M. Mishna, *Classifying lattice walks restricted to the quarter plane*, Journal of Combinatorial Theory, Series A **116** (2009), 460–477.
[62] J.D. Monk, *Mathematical logic*, Graduate Texts in Mathematics, Springer Verlag, 1976.
[63] E.M. Palmer, R.C. Read, and R.W. Robinson, *Counting claw-free cubic graphs*, Siam J. Discrete Math. **16.1** (2002), 65–73.
[64] M. Petkovsek, H. Wilf, and D. Zeilberger, *A=B*, AK Peters, 1996.
[65] R.C. Read, *The enumeration of locally restricted graphs, i.*, J. London Math. Soc. **34** (1959), 417–436.
[66] ———, *The enumeration of locally restricted graphs, ii.*, J. London Math. Soc. **35** (1960), 344–351.
[67] J.H. Redfield, *The theory of group-reduced distributions*, Amer. J. Math. **49** (1927), 433–455.
[68] G.C. Rota, *The number of partitions of a set*, American Mathematical Monthly **71 (5)** (1964), 498–504.
[69] U. Rotics, *Efficient algorithms for generally intractable graph problems restricted to specific classes of graphs*, Ph.D. thesis, Technion- Israel Institute of Technology, 1998.
[70] G. Rozenberg and A. Salomaa (eds.), *Handbook of formal languages, vol. 3 : Beyond words*, Springer Verlag, Berlin, 1997.
[71] L.A. Rubel, *Some research problems about algebraic differential equations*, Trans. Amer. Math. Soc. **280** (1983), no. 1, 43–52, Problem 16.
[72] E. R. Scheinerman and J. Zito, *On the size of hereditary classes of graphs*, J. Combin. Theory Ser. B **61** (1994), 16–39.
[73] D. Seese, *The structure of the models of decidable monadic theories of graphs*, Annals of Pure Applied Logic **53** (1991), 169–195.
[74] E. Specker, *Application of logic and combinatorics to enumeration problems*, Trends in Theoretical Computer Science (E. Börger, ed.), Computer Science Press, 1988, Reprinted in: Ernst Specker, Selecta, Birkhäuser 1990, pp. 324-350, pp. 141–169.
[75] ———, *Modular counting and substitution of structures*, Combinatorics, Probability and Computing **14** (2005), 203–210.
[76] R. P. Stanley, *Differentiably finite power series*, European Journal of Combinatorics **1** (1980), 175–188.
[77] R.P. Stanley, *Enumerative combinatorics*, vol. II, Cambridge University Press, 1999.
[78] D. S. Stones, *The many formulae for the number of latin rectangles*, The Electronic Journal of Combinatorics **17** (2010), A1.
[79] D. S. Stones and I.M. Wanless, *Divisors of the number of latin rectangles*, J. Combin. Theory Ser. A **117** (2010), 204215.
[80] L. Takács, *On the number of distinct forests*, SIAM J. Disc. Math. **3.4** (1990), 574–581.
[81] H.S. Wilf, *generatingfunctionology*, Academic Press, 1990.
[82] N.C. Wormald, *The number of labelled cubic graphs with no triangles*, Proceedings of the 66 Twelfth Southeastern Conference on Combinatorics, Graph Theory, and Computing, vol. 2, 1981, pp. 359–378.
[83] D. Zeilberger, *A holonomic systems approach to special functions identities*, J. of Computational and Applied Mathematics **32** (1990), 321–368.

Faculty of Computer Science, Technion–Israel Institute of Technology, 32000 Haifa, Israel

E-mail address: {eldar,tkotek,janos}@cs.technion.ac.il

Contemporary Mathematics
Volume **558**, 2011

Spectra and Systems of Equations

Jason P. Bell, Stanley N. Burris, and Karen Yeats

Abstract. Periodicity properties of sets of nonnegative integers, defined by systems $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ of equations, are analyzed. Such systems of set equations arise naturally from equational specifications of combinatorial classes—Compton's equational specification of monadic second order classes of trees is an important example.

In addition to the general theory of set equations and periodicity, with several small illustrative examples, two applications are given:

(1) There is a new proof of the fundamental result of Gurevich and Shelah on the periodicity of monadic second order classes of finite monounary algebras. Also there is a new proof that the monadic second order theory of finite monounary algebras is decidable.
(2) A formula derived for the periodicity parameter $\mathfrak{q}$ is used in the determination of the asymptotics for the coefficients of generating functions defined by well conditioned systems of equations.

1. Introduction

Logicians have developed the subject of finite model theory to study classes of finite *structures* defined by sentences in a formal logic. (In logic, a structure is a set equipped with a selection of functions and/or relations and/or constants.) Combinatorialists have been interested in classes of *objects* defined by equational specifications (the objects are not restricted to the structures studied by logicians). In recent years, there has been an increasing interest in the study of specifications involving several equations, and this article continues these investigations. Background, definitions, and an introduction to the topics are given in §1.1–§1.4. §1.5 is an outline, giving the order of presentation of the topics.

1.1. Combinatorial classes, generating functions and spectra.

A *combinatorial class* $\mathcal{A}$ is a class of objects with a function $\| \;\|$ that assigns a positive integer $\|\mathfrak{a}\|$, the *size* of $\mathfrak{a}$, to each object $\mathfrak{a}$ in the class, such that there are only finitely many objects of each size in the class.[1] (When counting the objects of a

1991 *Mathematics Subject Classification.* Primary 03C13; Secondary 05A15 .
The authors thank NSERC for supporting this project.

[1]This definition is essentially that used in the recently published book *Analytic Combinatorics* by Flajolet and Sedgewick [**19**]—it differs in that we do not allow objects of size 0. These authors have, for several years, kindly made pre-publication drafts of this book available; these drafts have played a singularly important role in our own education and investigations.

given size, it is often the case that one is counting modulo some natural equivalence relation, such as isomorphism.) The objects of $\mathcal{A}$ are said to be *unlabelled*.

Given a combinatorial class $\mathcal{A}$, let the *count function* $a(n)$ for $\mathcal{A}$ be the number of (unlabelled) objects of size n in $\mathcal{A}$. Two combinatorial classes $\mathcal{A}$, $\mathcal{B}$ are *isomorphic* if they have the same count function. Combinatorics is essentially the study of count functions of combinatorial classes, and generating functions provide the main tool for this study. The *(ordinary) generating function* $A(x)$ of $\mathcal{A}$ is the formal power series

$$A(x) \ := \ \sum_{n=1}^{\infty} a(n)x^n.$$

In the following, only unlabelled objects are considered, and thus the only generating functions discussed are ordinary generating functions.

Throughout this article, a power series and its coefficients will use the same letter—for power series the name is upper case, for coefficients the name is lower case, as with $A(x)$ and $a(n)$ above. When a combinatorial class is named by a single letter, say $\mathcal{A}$, the same letter will be used, in the appropriate size and font, to name its count function and its generating function, just as above with $\mathcal{A}$, $a(n)$, $A(x)$. For a combinatorial class with a composite name, say $\mathsf{MSet}(\mathcal{A})$, we can also use $\mathsf{MSet}(\mathcal{A})(x)$ to name its generating function.

The *spectrum* $\mathsf{Spec}(\mathcal{A})$ of a combinatorial class $\mathcal{A}$ is the set of sizes of the objects in $\mathcal{A}$, that is, $\mathsf{Spec}(\mathcal{A}) := \{\|\mathfrak{a}\| : \mathfrak{a} \in \mathcal{A}\}$. The *spectrum* of a power series $A(x)$ is $\mathsf{Spec}(A(x)) := \{n : a(n) \neq 0\}$, the support of the coefficient function $a(n)$. Thus $\mathsf{Spec}(\mathcal{A}) = \mathsf{Spec}(A(x))$.

1.2. The spectrum of a sentence.

In 1952 the *Journal of Symbolic Logic* initiated a section devoted to unsolved problems in the field of symbolic logic. The first problem, posed by Scholz [**28**], concerned the spectra of first order sentences. Given a sentence φ from first order logic, he defined the *spectrum* of φ to be the set of sizes of the finite models of φ. For example, binary trees can be defined by such a φ, and its spectrum is the arithmetical progression $\{1, 3, 5, \ldots\}$. Fields can also be defined by such a φ, with the spectrum being the set $\{2, 4, \ldots, 3, 9, \ldots\}$ of powers of prime numbers. The possibilities for the spectrum $\mathsf{Spec}(\varphi)$ of a first order sentence φ are amazingly complex.[2] The definition of $\mathsf{Spec}(\varphi)$ has been extended to sentences φ in any logic; we will be particularly interested in monadic second order (MSO) logic. If $\mathcal{A}$ is the class of finite models of a sentence φ, then $\mathsf{Spec}(\varphi) = \mathsf{Spec}(\mathcal{A}) = \mathsf{Spec}(A(x))$.

Scholz's problem was to find a necessary and sufficient condition for a set S of natural numbers to be the spectrum of some first order sentence φ. This led to considerable research by logicians—see, for example, the recent survey paper [**17**] of Durand, Jones, Makowsky, and More.

[2] Asser's 1955 conjecture, that *the complement of a first order spectrum is a first order spectrum*, is still open. It is known, through the work of Jones and Selman and Fagin in the 1970s, that this conjecture is equivalent to the question of whether the complexity class NE of problems decidable by a nondeterministic machine in exponential time is closed under complement. Thus the conjecture is, in fact, one of the notoriously difficult questions of computational complexity theory. Stockmeyer [**31**], p. 33, states that if Asser's conjecture is false then NP $\neq$ co-NP, and hence P $\neq$ NP.

1.3. Equational systems.

The study of equational specifications of combinatorial systems and equational systems defining generating functions is well established (see, for example, *Analytic Combinatorics* [**19**]; or [**2**]), but the corresponding study of spectra is new. Equational specifications $\boldsymbol{\mathcal{Y}} = \boldsymbol{\Gamma}(\boldsymbol{\mathcal{Y}})$ of combinatorial classes usually lead to systems of equations $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ defining the generating functions, and as will be seen, either of these usually lead to equational systems $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ defining the spectra. (See §5 for details.) A calculus of sets, to determine the spectrum of a combinatorial class defined by a single equation, was first introduced in 2006 (in [**2**]). That calculus is developed further here, in order to analyze spectra defined by systems of equations.

As an example of how this calculus of sets fits in, consider the class $\mathcal{P}$ of planar binary trees.[3] It is specified by the equation $\mathcal{P} = \{\bullet\} \cup \bullet/\mathsf{Seq}_2(\mathcal{P})$, which one can read as: *the class of planar binary trees is the smallest class $\mathcal{P}$ which has the 1-element tree '$\bullet$', and is closed under taking any sequence of two trees and adjoining a new root '$\bullet$'* . From the specification equation one finds that the generating function $P(x)$ of $\mathcal{P}$ satisfies $P(x) = x + x \cdot P(x)^2$, a simple quadratic equation that can be solved for $P(x)$. One also says that $P(x)$ is a solution to the polynomial equation $y = x + x \cdot y^2$. For the spectrum $\mathsf{Spec}(\mathcal{P})$, one has the equation $\mathsf{Spec}(\mathcal{P}) = \{1\} \cup (1 + 2 * \mathsf{Spec}(\mathcal{P}))$; thus $\mathsf{Spec}(\mathcal{P})$ satisfies the set equation $Y = \{1\} \cup (1 + 2 * Y)$ (see §2 for the notation used here). Solving this set equation gives the periodic spectrum $\mathsf{Spec}(\mathcal{P}) = 1 + 2 \cdot \mathbb{N}$.

If one drops the 'planar' condition, the situation becomes more complicated. The class $\mathcal{T}$ of binary trees has the specification $\mathcal{T} = \{\bullet\} \cup \bullet/\mathsf{MSet}_2(\mathcal{T})$, which one can read as: *the class of binary trees is the smallest class $\mathcal{T}$ which has the 1-element tree '$\bullet$', and is closed under taking any multiset of two trees and adjoining a new root '$\bullet$'.* From this specification equation one finds that the generating function $T(x)$ of $\mathcal{T}$ satisfies $T(x) = x + x \cdot \big(T(x)^2 + T(x^2)\big)/2$. This is not so simple to solve for $T(x)$; however, it gives a recursive procedure to find the coefficients $t(n)$ of $T(x)$, and one can compute the radius of convergence ρ of $T(x)$, and the value of $T(\rho)$, to any desired degree of accuracy (see *Analytic Combinatorics* [**19**], VII.22, p. 477). For the spectrum $\mathsf{Spec}(\mathcal{T})$, one has the equation $\mathsf{Spec}(\mathcal{T}) = \{1\} \cup (1 + 2 * \mathsf{Spec}(\mathcal{T}))$; thus $\mathsf{Spec}(\mathcal{T})$ also satisfies the set equation $Y = \{1\} \cup (1 + 2 * Y)$, and the solution is again $\mathsf{Spec}(\mathcal{T}) = 1 + 2 \cdot \mathbb{N}$.

There is a long history of generating functions defined by a *single* recursion equation $y = G(x, y)$, starting with Cayley's 1857 paper [**8**] on trees, Pólya's 1937 paper (see [**23**]) that gave the form of the asymptotics for several classes of trees associated with classes of hydrocarbons, etc, right up to the present with the thorough treatment in *Analytic Combinatorics*.

Treelike structures have provided an abundance of recursively defined generating functions. Letting $t(n)$ be the number of trees (up to isomorphism) of size n, one has the equation

$$\sum_{n \geq 1} t(n) x^n \;=\; x \cdot \prod_{n \geq 1} (1 - x^n)^{-t(n)},$$

[3] We regard a tree as a certain kind of poset, with a largest element called the root. See §6.2 for a precise definition.

which yields a recursive procedure to calculate the values of $t(n)$. By 1875 Cayley had used this to calculate the first 13 coefficients $t(n)$, that is, the number of trees of size n for $n = 1, \ldots, 13$.

In 1937 Pólya (see [**23**]) would rewrite this equation as

$$T(x) \; = \; x \cdot \exp\left(\sum_{m=1}^{\infty} T(x^m)/m\right).$$

This allowed him to show that the coefficients $t(n)$ have the asymptotic form

$$t(n) \;\sim\; C\rho^{-n}n^{-3/2}, \tag{1.1}$$

for a suitable constant C, where ρ is the radius of convergence of $T(x)$.

Pólya's result has been generalized to show that any *well conditioned* equation $y = G(x, y)$ has a power series solution $y = T(x)$ whose coefficients satisfy the asymptotics in (1.1).[4] The determination of the constant C depends, in part, on knowing the periodicity parameter $\mathfrak{q}$ (defined in §2.4) for the spectrum of $T(x)$—for a generating function defined by a single equation, a formula for $\mathfrak{q}$ was given in [**2**].

The theory of generating functions defined by a system $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ of several equations has built on the successes in the single equation case. If the system is *well conditioned* then the solution $\mathbf{y} = \mathbf{T}(x)$ is such that the $T_i(x)$ have the same radius of convergence $\rho \in (0, \infty)$, and they have the same periodicity parameter $\mathfrak{q}$. Drmota [**14**], [**15**] (see also [**3**]) showed that the coefficients $t_i(n)$ of the $T_i(x)$ satisfy the same asymptotics as in the 1-equation case, namely there are constants C_i such that

$$t_i(n) \;\sim\; C_i\rho^{-n}n^{-3/2}.$$

Such asymptotic expressions are understood to include the restriction *for* $n \in \mathsf{Spec}(T_i(x))$. As in the 1-equation case, the value of C_i depends partly on knowing the parameter $\mathfrak{q}$. The formula (4.2) shows how to find $\mathfrak{q}$ from the $G_i(x, \mathbf{y})$, and §7 shows how this is used to find the constants C_i in many well conditioned systems.

1.4. Monadic second order classes of trees and unary functions.

Let q be a positive integer, and let $\mathcal{T}_1, \ldots, \mathcal{T}_k$ be the minimal (nonempty) classes among the classes of trees defined by MSO sentences of quantifier depth q. The $\mathcal{T}_i$ are pairwise disjoint, and every class of trees defined by a MSO sentence of quantifier depth q is a union of some of the $\mathcal{T}_i$. In the 1980s, Compton [**11**], building on Büchi's 1960 paper [**6**] (on regular languages and MSO classes of m-colored chains), showed that the $\mathcal{T}_i$ have an equational specification $\boldsymbol{\mathcal{Y}} = \boldsymbol{\Gamma}(\boldsymbol{\mathcal{Y}})$. Following standard translation procedures, this gives an equational system $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ that defines the generating functions $T_i(x)$ of the classes $\mathcal{T}_i$. In 1997, Woods [**33**] used the system $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ to prove that the class of trees has a MSO limit law.[5] One can easily convert either of these systems—$\boldsymbol{\mathcal{Y}} = \boldsymbol{\Gamma}(\boldsymbol{\mathcal{Y}})$ and $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$—into an equational system $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ defining the spectra of the $\mathcal{T}_i$. From this one readily proves that every MSO class of trees has an eventually periodic spectrum. An easy

[4] See §7 for a discussion of well conditioned systems.

[5] Recently we have applied Compton's Specification Theorem to prove MSO 0–1 laws for many classes of forests. (See [**4**]).

argument then shows that the same holds for MSO classes of monounary algebras,[6] proving the Gurevich and Shelah result in [**20**]. Additionally, we use Compton's equational specification to give new proofs of the decidability results in [**20**].

1.5. Outline of the presentation.

§2 *Set Operations and Periodicity.* The basic operations $(\cup, +, \cdot, *)$ and laws, for a calculus of subsets of the nonnegative integers $\mathbb{N}$, are introduced; periodicity and the periodicity parameters $\mathfrak{c}$, $\mathfrak{m}$, $\mathfrak{p}$, $\mathfrak{q}$ are defined, and the fundamental results on periodicity are established.

§3 *Systems of Set Equations.* The set calculus of §2 is applied to the study of subsets of $\mathbb{N}$ defined by *elementary* systems $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ of set equations, with the main result being Theorem 3.11. Such systems have unique solutions among subsets of the positive integers $\mathbb{P}$. Conditions for periodicity of the solutions, and formulas for some of the periodicity parameters, are given.

§4 *Elementary Power Series Systems.* The development of §3 is paralleled for *elementary* systems $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ of power series equations (which are used to define generating functions). Elementary systems satisfy a special nonlinear requirement (see Definition 4.5) that guarantees a unique solution. The main result of §4 is Theorem 4.6, which gives criteria for generating functions defined by elementary systems to have periodic spectra. Furthermore, it gives formulas for some of the periodicity parameters.

§5 *Constructions, Operators, and Equational Specifications.* The equational specifications considered in this article use constructions built from compositions of a few basic constructions; it is routine to translate such specifications into systems of equations for the spectra. Also, under suitable conditions, one can translate equational specifications into equational systems defining generating functions.

§6 *Monadic Second Order Classes.* This section gives the aforementioned applications of Compton's Specification Theorem (Theorem 6.7) for monadic second order classes of trees.

§7 *Well Conditioned Systems.* The formulas, for the periodicity parameters in Theorem 4.6, are used to determine the asymptotics for the coefficients of generating functions defined by well conditioned systems.

Appendix A. *Proofs of preliminary material.*

Appendix B. *Büchi's Theorem.* This appendix shows how Büchi used minimal MSO_q classes to prove that MSO classes of m-colored chains can be identified with regular languages over a m-letter alphabet. Implicit in this proof is the fact that each MSO class of m-colored chains has a specification.

Acknowledgment. The authors are indebted to the two referees for detailed suggestions on how to improve the presentation.

2. Set Operations and Periodicity

2.1. Periodic sets.

Definition 2.1. $\mathbb{N}$ is the set of nonnegative integers, $\mathbb{P}$ is the set of positive integers. For $A \subseteq \mathbb{N}$:

[6] A *monounary algebra* $\mathfrak{a} = (A, f)$ is a set A, called the universe of the algebra, with a function $f : A \to A$. If one thinks of f as a binary relation instead of a unary function then one has the equivalent notion of a *functional digraph*.

(a) A is *periodic* if there is a positive integer p such that $p + A \subseteq A$, that is, $a \in A$ implies $p + a \in A$. Such an integer p is a *period* of A.
(b) A is *eventually periodic* if there is a positive integer p such that $p + A$ is eventually in A, that is, there is an m such that for $a \in A$, if $a \geq m$ then $p + a \in A$. Such a p is an *eventual period* of A.

Clearly every arithmetical progression and every cofinite subset of $\mathbb{N}$ is periodic; and every periodic set is eventually periodic. Finite subsets of $\mathbb{N}$ are eventually periodic; the only finite periodic set is $\emptyset$. As will be seen, periodicity seems to be a natural property for the spectra of combinatorial classes specified by a system of equations. The famous Skolem-Mahler-Lech Theorem (see, for example, *Analytic Combinatorics* [**19**], p. 266) says that the spectrum of every rational function $P(x)/Q(x)$ in $\mathbb{Q}(x)$ is eventually periodic, where the spectrum of $P(x)/Q(x)$ is the spectrum of its power series expansion. Consequently, polynomial systems $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$, with rational coefficients, that are linear in the variables y_i, and with a nonsingular Jacobian matrix $\partial\big(\mathbf{y} - \mathbf{G}(x,\mathbf{y})\big)/\partial\mathbf{y}$, have power series solutions $y_i = T_i(x)$ with eventually periodic spectra. However, much simpler methods give this periodicity result for the nonnegative $\mathbf{y}$-linear systems considered here.

If the spectrum of a combinatorial class $\mathcal{A}$ is eventually periodic then one has the hope, as in the case of regular languages and well behaved irreducible systems, that the class $\mathcal{A}$ decomposes into a finite subclass $\mathcal{A}_0$, along with finitely many subclasses $\mathcal{A}_i$, such that the $\mathsf{Spec}(\mathcal{A}_i)$ are arithmetical progressions $a_i + b_i \cdot \mathbb{N}$, and the generating functions $A_i(x)$ have well behaved coefficients (for example, monotone increasing, exponential growth, etc.) on $\mathsf{Spec}(\mathcal{A}_i)$.[7]

2.2. Set operations.

The calculus of set equations (for sets of nonnegative integers) developed in this section was originally extracted from work on the spectra of power series (see §4.2), to analyze the spectra of combinatorial classes. It uses the operations of union ($\cup$, $\bigcup$), addition ($+$), multiplication ($\cdot$), and star ($*$).

DEFINITION 2.2. For $A, B \subseteq \mathbb{N}$ and $n \in \mathbb{N}$ let

$$\begin{array}{rcl|rcl} n + B & := & \{n + b : b \in B\} & A + B & := & \{a + b : a \in A, b \in B\} \\ n \cdot B & := & \{nb : b \in B\} & & & \\ n * B & := & \begin{cases} \{0\} & \text{for } n = 0 \\ \underbrace{B + \cdots + B}_{n \text{ copies of } B} & \text{for } n > 0 \end{cases} & A * B & := & \bigcup_{a \in A} a * B \end{array}$$

The values of these operations, when an argument is the empty set, are: $\emptyset + A = A + \emptyset = \emptyset$, $n \cdot \emptyset = \emptyset$, $\emptyset * B = \emptyset$, and $A * \emptyset = \{0\}$ if $0 \in A$, otherwise $A * \emptyset = \emptyset$.

The obvious definition of $A \cdot B$ is not needed in this study of spectra; only the special case $n \cdot B$ plays a role. The next lemma gives several basic identities needed for the analysis of spectra (all are easily proved).

LEMMA 2.3. *For $A, B, C \subseteq \mathbb{N}$ and $m, n \in \mathbb{N}$*

$$\begin{aligned} A + (B \cup C) &= (A + B) \cup (A + C) \\ (A \cup B) * C &= A * C \cup B * C \end{aligned}$$

[7]The comments in this paragraph are related to Question 7.4 in Compton's 1989 paper [**10**] on MSO logical limit laws. (See [**1**] in this volume).

$$\begin{aligned}
(A+B)*C &= A*C+B*C \\
m*(n*B) &= (m\cdot n)*B \\
n*(B+C) &= n*B+n*C \\
A*\big(B\cup C\big) &= \bigcup_{\substack{j_1,j_2\in\mathbb{N}\\ j_1+j_2\in A}} \big(j_1*B+j_2*C\big).
\end{aligned}$$

It is quite useful that $*$ right distributes over both $\cup$ and $+$. However, neither left distributive law is generally valid; one only has a weak form of the left distributive law of $*$ over $+$, namely $n*(B+C)=n*B+n*C$.

2.3. Periodic and eventually periodic sets.

The following characterizations of periodic and eventually periodic sets are easily proved, if not well known.

LEMMA 2.4. *Let $A\subseteq\mathbb{N}$.*

(a) *A is periodic iff there is a finite set $A_1\subseteq\mathbb{N}$ and a positive integer p (a period for A) such that*

$$A \;=\; A_1+p\cdot\mathbb{N}$$

iff A is the union of finitely many arithmetical progressions.

(b) (Durand, Fagin, Loescher [**16**]; Gurevich and Shelah [**20**]) *A is eventually periodic iff there are finite sets $A_0, A_1\subseteq\mathbb{N}$ and a positive integer p (an eventual period of A) such that*

$$A \;=\; A_0\cup(A_1+p\cdot\mathbb{N})$$

iff A is the union of a finite set and finitely many arithmetical progressions.

REMARK 2.5. An infinite union of arithmetical progressions need not be eventually periodic. Let U be the union of the arithmetical progressions $a\cdot\mathbb{P}$, where a is a composite number. Then U consists of all composite numbers. Given any positive integer p, choose a prime number q that does not divide p. Then, by Dirichlet's theorem, the arithmetical progression $q^2+p\cdot\mathbb{N}$ has an infinite number of primes, thus $q^2+p\cdot\mathbb{N}$ is not a subset of U. Since $q^2\in U$, it follows that p is not an eventual period for U (one can choose q arbitrarily large). Thus U is not eventually periodic.

LEMMA 2.6. *Let $A,B\subseteq\mathbb{N}$.*

(a) *If A,B are eventually periodic, then so are $A\cup B$, $A+B$ and $A*B$.*
(b) *If A,B are periodic, then so are $A\cup B$ and $A+B$.*
(c) *Suppose A is periodic. Then $A*B$ is periodic iff $A*B\neq\{0\}$, which is iff neither $A\neq\emptyset$ and $B=\{0\}$ nor $0\in A$ and $B=\emptyset$ hold.*

PROOF. The results for $A\cup B$ and $A+B$ follow easily from Lemma 2.3 and Lemma 2.4. (The eventually periodic case is discussed in [**20**].)

To show $A*B$ is eventually periodic in (a), there are finite sets A_0 and A_1, and a positive integer p, such that

$$A \;=\; A_0\cup(A_1+p\cdot\mathbb{N}).$$

Using the right distributive laws for $*$ over $+$ and $\cup$, we have

$$A*B \;=\; (A_0*B)\cup\big((A_1*B)+\mathbb{N}*(p\cdot B)\big). \tag{2.1}$$

For M a finite subset of $\mathbb{N}$, note that $M*B$ is eventually periodic because it is either $\emptyset$, or $\{0\}$; or it is a finite union of finite sums of B, and these operations preserve being eventually periodic. Also note that for M any subset of $\mathbb{N}$, $\mathbb{N} * M$ is eventually periodic because $\mathbb{N} * M \supseteq M + (\mathbb{N} * M)$, so either $\mathbb{N} * M$ is $\emptyset$, or $\{0\}$; or there is a positive integer p in M, in which case $p + \mathbb{N} * M \subseteq \mathbb{N} * M$. In the first and third cases, $\mathbb{N} * M$ is actually periodic. The right side of (2.1) results from applying operations, that preserve being eventually periodic, to eventually periodic sets, so $A * B$ is eventually periodic.

For item (c), choose a positive integer p such that $A \supseteq p + A$. Then

$$A * B \ \supseteq \ (p * B) + (A * B).$$

Consequently, one has $A * B$ being either $\emptyset$, or $\{0\}$; or there is a positive integer in B, and thus a positive integer q in $p * B$. $\emptyset$ is periodic, but $\{0\}$ is not. The third case, that there is a positive integer q in $p * B$, implies $A * B \supseteq q + (A * B)$, so $A * B$ is periodic.

□

2.4. Periodicity parameters.

For $A \subseteq \mathbb{N}$, for $n \in \mathbb{N}$, let

$$A - n \quad := \quad \{a - n : a \in A\}.$$

The next definition gives some important parameters for the study of periodicity, with the convention $\gcd(\{0\}) := 0$.

Definition 2.7 (Periodicity parameters). For $A \subseteq \mathbb{N}$, $A \neq \emptyset$, let

(a) $\mathfrak{m}(A) := \min(A)$
(b) $\mathfrak{q}(A) := \gcd\big(A - \mathfrak{m}(A)\big)$.[8]

If A is infinite and eventually periodic:

(c) $\mathfrak{p}(A)$ is the minimum of the eventual periods p of A.
(d) $\mathfrak{c}(A) := \min\big\{a \in A : \mathfrak{p}(A) \text{ is a period for } A \cap [a, \infty)\big\}$.

Remark 2.8. It is useful to note that $\mathfrak{q}(A) = \gcd\{a - b : a, b \in A\}$.

Proposition 2.9. *Let $A_1, A_2 \subseteq \mathbb{N}$ be nonempty, with $\mathfrak{m}_i := \mathfrak{m}(A_i)$, $\mathfrak{q}_i := \mathfrak{q}(A_i)$, for $i = 1, 2$. Then*

Set	$\mathfrak{m}$	$\mathfrak{q}$
$A_1 \cup A_2$	$\min(\mathfrak{m}_1, \mathfrak{m}_2)$	$\gcd(\mathfrak{q}_1, \mathfrak{q}_2, \mathfrak{m}_2 - \mathfrak{m}_1)$
$A_1 + A_2$	$\mathfrak{m}_1 + \mathfrak{m}_2$	$\gcd(\mathfrak{q}_1, \mathfrak{q}_2)$
$A_1 * A_2$	$\mathfrak{m}_1 \mathfrak{m}_2$	$\begin{cases} \{0\} & \text{if } A_1 = \{0\} \\ \gcd(\mathfrak{q}_2,\ \mathfrak{q}_1 \mathfrak{m}_2) & \text{if } A_1 \neq \{0\}, \end{cases}$

where in the last item we assume $\mathfrak{m}_1 \le \mathfrak{m}_2$.

Proof. The calculations for $\mathfrak{m}$ are clear in each case. The calculations for $\mathfrak{q}$ are also elementary, but slightly more delicate—see Appendix A. □

Definition 2.10. For $A \subseteq \mathbb{N}$ and $n \in \mathbb{N}$ let $A|_{\ge n} := A \cap [n, \infty)$. Likewise define $A|_{>n}$, $A|_{\le n}$, and $A|_{<n}$.

---

[8]Although adjoining elements to the beginning of A can decrease $\mathfrak{q}$, they cannot increase it; $\mathfrak{q}(A)$ is always a divisor of $\mathfrak{q}(A|_{\ge \mathfrak{c}})$, the $\mathfrak{q}$ of the periodic part of A. Thus, smallness of $\mathfrak{q}$ gives little information about A (too sensitive to a finite initial segment), but largeness of $\mathfrak{q}$ does give information. Furthermore, $\mathfrak{q}$ is useful as we can obtain a formula for it under quite general conditions, but we can only obtain a formula for $\mathfrak{p}$ by adding constraints so that $\mathfrak{p} = \mathfrak{q}$ (and then using the formula for $\mathfrak{q}$).

LEMMA 2.11. *Suppose $A \subseteq \mathbb{N}$ is infinite and eventually periodic. Letting $\mathfrak{c} := \mathfrak{c}(A)$, $\mathfrak{q} := \mathfrak{q}(A)$, $\mathfrak{p} := \mathfrak{p}(A)$, one has the following:*

(a) (Gurevich and Shelah [**20**], Cor. 3.3) *The set of eventual periods of A is $\mathfrak{p} \cdot \mathbb{P}$.*

(b) *One has $\mathfrak{q} \mid \mathfrak{p}$, and $\mathfrak{p} = \mathfrak{q}$ iff $\mathfrak{p} \mid (A - \mathfrak{m})$.*

(c) *The following are equivalent:*
  - (i) *$A = A_0 \cup (a + b \cdot \mathbb{N})$ for some finite A_0 and some $a, b \in \mathbb{N}$, that is, A is the union of a finite set and a single arithmetical progression.*
  - (ii) *$A = A\big|_{<\mathfrak{c}} \cup (\mathfrak{c} + \mathfrak{p} \cdot \mathbb{N})$.*
  - (iii) *One has $\mathfrak{p} = \mathfrak{q}\big(A|_{\geq \mathfrak{c}}\big)$.*

(d) *The following are equivalent:*
  - (i) *One has $\mathfrak{c} + \mathfrak{p} \cdot \mathbb{N} \subseteq A \subseteq \mathfrak{m} + \mathfrak{p} \cdot \mathbb{N}$.*
  - (ii) *$A = A\big|_{<\mathfrak{c}} \cup (\mathfrak{c} + \mathfrak{p} \cdot \mathbb{N}) \subseteq \mathfrak{m} + \mathfrak{p} \cdot \mathbb{N}$.*
  - (iii) *One has $\mathfrak{p} = \mathfrak{q}$.*

PROOF. The proof of (a) in [**20**] is elementary, as are the proofs for (b)–(d). For completeness, the later can be found in Appendix A. □

REMARK 2.12. The spectra of combinatorial classes, whose generating functions are defined by well conditioned systems of equations (see §7), are quite well behaved; they are periodic sets, and $\mathfrak{p} = \mathfrak{q}$. Thus, by Lemma 2.11 (d), the spectra are cofinal subsets of arithmetical progressions.

The next lemma augments Lemma 2.11 (d), giving a simple condition that is sufficient to guarantee that A is a periodic set with $\mathfrak{p} = \mathfrak{q}$.

LEMMA 2.13. *Suppose $A \subseteq \mathbb{N}$ with $A|_{>0} \neq \emptyset$, and suppose there are integers $r \geq 0$ and $s \geq 2$ such that*

$$A \;\supseteq\; r + s * A.$$

Let $\mathfrak{c} := \mathfrak{c}(A)$, $\mathfrak{m} := \mathfrak{m}(A)$, $\mathfrak{p} := \mathfrak{p}(A)$ and $\mathfrak{q} := \mathfrak{q}(A)$.

Then A is a periodic set with $\mathfrak{p} = \mathfrak{q}$. If $(r, s) = (0, 2)$, that is, if $A \supseteq A + A$, then one has the additional conclusion that $\mathfrak{q} = \gcd(A)$.

PROOF. Choose $a \in A|_{>0}$, and let $b := r + (s-2)a$. Then

$$A \;\supseteq\; b + A + A.$$

Choose $p \in b + A|_{>0}$. Then $A \supseteq p + A$ shows that A is periodic. By Lemma 2.11 (a), $\mathfrak{p} \mid (b + A)$, since all nonzero members of $b + A$ are periods of A. But then $\mathfrak{p} \mid (A - \mathfrak{m})$, so $\mathfrak{p} = \mathfrak{q}$ by Lemma 2.11 (b).

Now suppose $(r, s) = (0, 2)$. Then, for $a \in A$, one has $a + a \in A$. By Remark 2.8, $\mathfrak{q}$ divides their difference, that is, $\mathfrak{q}|a$. Thus $\mathfrak{q}|d := \gcd(A)$. Clearly $d \mid A - \mathfrak{m}$, so $d|\mathfrak{q}$. Thus $\mathfrak{q} = \gcd(A)$. □

3. Systems of Set Equations

For X a set, $\mathrm{Su}(X)$ is *the set of subsets of X*.

We will consider systems of set equations of the form

$$\begin{aligned} Y_1 &= G_1(Y_1, \ldots, Y_k) \\ &\vdots \\ Y_k &= G_k(Y_1, \ldots, Y_k), \end{aligned}$$

written compactly as $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$, with the $G_i(\mathbf{Y})$ having a particular form, namely

$$G_i(\mathbf{Y}) \;=\; \bigcup_{\mathbf{u}\in\mathbb{N}^k} \big(G_{i,\mathbf{u}} + u_1 * Y_1 + \cdots + u_k * Y_k\big), \tag{3.1}$$

where the $G_{i,\mathbf{u}}$ are subsets of $\mathbb{N}$. The system of equations (3.1) is compactly expressed by

$$\mathbf{G}(\mathbf{Y}) \;=\; \bigvee_{\mathbf{u}\in\mathbb{N}^k} \big(\mathbf{G}_{\mathbf{u}} + \mathbf{u} \circledast \mathbf{Y}\big), \tag{3.2}$$

where

$$\mathbf{u} \circledast \mathbf{Y} := u_1 * Y_1 + \cdots + u_k * Y_k.$$

Remark 3.1. $Y_1, \ldots, Y_k$ are variables that range over subsets of $\mathbb{N}$. Let the $G_{i,\mathbf{u}}$ be subsets of $\mathbb{N}$. The formal expression $\bigcup_{\mathbf{u}\in\mathbb{N}^k} \big(G_{i,\mathbf{u}} + u_1 * Y_1 + \cdots + u_k * Y_k\big)$ takes on a value (which is a subset of $\mathbb{N}$) by assigning set values $A_j \subseteq \mathbb{N}$ to the set variables Y_j.

The collection of sets $\mathrm{Su}(\mathbb{N})$ is closed under the familiar operations of union ($\cup$, $\bigcup$) and intersection ($\cap$, $\bigcap$). $\mathrm{Su}(\mathbb{N})^k$ is naturally viewed as a Boolean algebra, namely as the product of k copies of $\mathrm{Su}(\mathbb{N})$, and the inherited operations, corresponding to those just mentioned, are designated by the symbols $\vee$, $\bigvee$ and $\wedge$, $\bigwedge$. Thus when applied to k-tuples of subsets of $\mathbb{N}$, they act coordinatewise as $\cup$, $\bigcup$ and $\cap$, $\bigcap$, for example, $\mathbf{A} \vee \mathbf{B} := (A_1 \cup B_1, \ldots, A_k \cup B_k)$.[9] The notation $\mathbf{A} \le \mathbf{B}$ means $A_i \subseteq B_i$ for $1 \le i \le k$.

The expression $\bigvee_{\mathbf{u}\in\mathbb{N}^k} \big(\mathbf{G}_{\mathbf{u}} + \mathbf{u} \circledast \mathbf{Y}\big)$ is a formal expression such that when $\mathbf{Y}$ is assigned a k-tuple $\mathbf{A}$ from $\mathrm{Su}(\mathbb{N})^k$, the ith coordinate of $\bigvee_{\mathbf{u}\in\mathbb{N}^k} \big(\mathbf{G}_{\mathbf{u}} + \mathbf{u} \circledast \mathbf{A}\big)$ is $\bigcup_{\mathbf{u}\in\mathbb{N}^k} \big(G_{i,\mathbf{u}} + \mathbf{u} \circledast \mathbf{A}\big)$.

The notation $\mathbf{u} \circledast \mathbf{Y}$ is adopted for $u_1 * Y_1 + \cdots + u_k * Y_k$ since a natural definition of $\mathbf{u} * \mathbf{Y}$ would be the k-tuple $(u_1 * Y_1, \cdots, u_k * Y_k)$.

3.1. $\mathsf{Dom}(\mathbf{Y})$ and $\mathsf{Dom}_0(\mathbf{Y})$. This subsection defines the set operators we are interested in, and gives some useful lemmas about their behavior.

Definition 3.2. Let $\mathsf{Dom}(\mathbf{Y})$ be the set of $G(\mathbf{Y})$ of the form

$$G(\mathbf{Y}) \;=\; \bigcup_{\mathbf{u}\in\mathbb{N}^k} \big(G_{\mathbf{u}} + \mathbf{u} \circledast \mathbf{Y}\big),$$

where $G_{\mathbf{u}} \subseteq \mathbb{N}$, and let $\mathsf{Dom}_0(\mathbf{Y})$ be the set of $G(\mathbf{Y}) \in \mathsf{Dom}(\mathbf{Y})$ which map $\mathrm{Su}(\mathbb{P})^k$ into $\mathrm{Su}(\mathbb{P})$.

Lemma 3.3. *Suppose* $\mathbf{G}(\mathbf{Y}) \in \mathsf{Dom}(\mathbf{Y})^k$.

(a) *For* $\mathbf{A} \in \mathrm{Su}(\mathbb{N})^k$ *and* $1 \le i \le k$, *one has*

$$G_i(\mathbf{A}) \;=\; \bigcup_{\mathbf{u}\in\mathbb{N}^k} \Big(G_{i,\mathbf{u}} + \sum_{j\,:\,u_j>0} \big(u_j * A_j\big)\Big),$$

where the summation term is omitted in the case that all $u_j = 0$.

[9]When working with a product of structures, one usually uses the same symbols for the fundamental operations of the product as those used by the factors. For example, when working with rings, say $\mathfrak{r} = \mathfrak{r}_1 \times \mathfrak{r}_2$, one uses the symbols $+, \cdot$ as the fundamental operations for all three rings. However, the situation with products of Boolean algebras of sets is different—the inherited operations are no longer designated by $\cup$, $\bigcup$ and $\cap$, $\bigcap$, because these symbols have been given a fixed meaning in set theory. These fixed meanings lead, for example, to the fact that $\mathbf{A} \cup \mathbf{B}$ and $\mathbf{A} \vee \mathbf{B} := (A_1 \cup B_1, \ldots, A_k \cup B_k)$ are usually different.

(b) *If* $\mathbf{A} \in \mathrm{Su}(\mathbb{P})^k$ *then* $0 \in \mathbf{u} \circledast \mathbf{A} \ \Leftrightarrow\ \mathbf{u} = \mathbf{0}$.
(c) $\mathbf{G}(\mathbf{Y}) \in \mathsf{Dom}_0(\mathbf{Y})^k$ *iff* $\mathbf{G_0} := \big(G_{1,\mathbf{0}}, \ldots, G_{k,\mathbf{0}}\big) \in \mathrm{Su}(\mathbb{P})^k$.

PROOF. (a) follows from the fact that $0 * A_j = \{0\}$, by Definition 2.2.
Given $\mathbf{A} \in \mathrm{Su}(\mathbb{P})^k$, (b) follows from

$$\begin{aligned} 0 \in \mathbf{u} \circledast \mathbf{A} \quad &\Leftrightarrow \quad 0 \in u_i * A_i, \quad \text{for } 1 \le i \le k,\\ &\Leftrightarrow \quad u_i = 0, \quad \text{for } 1 \le i \le k, \end{aligned}$$

the last assertion holding because $0 \notin A_i$, for any i, and Definition 2.2.
For (c), let $\mathbf{A} \in \mathrm{Su}(\mathbb{P})^k$. Then

$$\begin{aligned} \mathbf{G}(\mathbf{A}) \in \mathrm{Su}(\mathbb{P})^k \quad &\Leftrightarrow \quad 0 \notin G_i(\mathbf{A}), \quad \text{for } 1 \le i \le k,\\ &\Leftrightarrow \quad 0 \notin G_{i,\mathbf{u}} + \mathbf{u} \circledast \mathbf{A}, \quad \text{for } 1 \le i \le k,\ \mathbf{u} \in \mathbb{N}^k,\\ &\Leftrightarrow \quad 0 \notin G_{i,\mathbf{u}} \cap \mathbf{u} \circledast \mathbf{A}, \quad \text{for } 1 \le i \le k,\ \mathbf{u} \in \mathbb{N}^k,\\ &\Leftrightarrow \quad 0 \in \mathbf{u} \circledast \mathbf{A} \Rightarrow 0 \notin G_{i,\mathbf{u}}, \quad \text{for } 1 \le i \le k,\ \mathbf{u} \in \mathbb{N}^k,\\ &\Leftrightarrow \quad 0 \notin G_{i,\mathbf{0}}, \quad \text{for } 1 \le i \le k, \end{aligned}$$

the last line by item (b). □

$\mathbf{G}^{(n)}(\mathbf{Y})$ denotes the n-fold composition of $\mathbf{G}(\mathbf{Y})$ with itself, and $G_i^{(n)}(\mathbf{Y})$ is the ith component of this composition. Let

$$\mathbf{G}^{(\infty)}(\mathbf{Y}) \ :=\ \bigvee_{n \ge 0} \mathbf{G}^{(n)}(\mathbf{Y}),$$

that is, the ith component of $\mathbf{G}^{(\infty)}(\mathbf{Y})$ is $\bigcup_n G_i^{(n)}(\mathbf{Y})$. For $\mathbf{A}, \mathbf{B} \in \mathrm{Su}(\mathbb{N})^k$ let,

- $\min \mathbf{A} := (\min A_1, \ldots, \min A_k)$
- $\mathcal{N}(\mathbf{A}) := \{i : A_i = \emptyset\}$,

where $\min(\emptyset) := +\infty$. Recall that $\mathbf{A} \le \mathbf{B}$ expresses $A_i \subseteq B_i$, for $1 \le i \le k$.

LEMMA 3.4. *Given* $\mathbf{G}(\mathbf{Y}) \in \mathsf{Dom}(\mathbf{Y})^k$, *and* $\mathbf{A}, \mathbf{B} \in \mathrm{Su}(\mathbb{N})^k$, *the following hold:*
(a) $\mathbf{A} \le \mathbf{B} \ \Rightarrow\ \mathbf{G}\big(\mathbf{A}\big) \le \mathbf{G}\big(\mathbf{B}\big)$.
(b) $\mathcal{N}(\mathbf{A}) = \mathcal{N}(\mathbf{B}) \ \Rightarrow\ \mathcal{N}\big(\mathbf{G}(\mathbf{A})\big) = \mathcal{N}\big(\mathbf{G}(\mathbf{B})\big)$.
(c) $\mathbf{A} \le \mathbf{B} \ \Rightarrow\ \mathcal{N}\big(\mathbf{G}(\mathbf{A})\big) \supseteq \mathcal{N}\big(\mathbf{G}(\mathbf{B})\big)$.
(d) $\mathcal{N}\big(\mathbf{G}^{(k)}(\boldsymbol{\emptyset})\big) = \mathcal{N}\big(\mathbf{G}^{(k+n)}(\boldsymbol{\emptyset})\big)$, *for* $n \ge 0$.

PROOF. Item (a) follows from the monotonicity of the set operations $\bigcup, +, *$ used in the definition of the $\mathbf{G}(\mathbf{Y})$ in $\mathsf{Dom}(\mathbf{Y})^k$.
Next observe that

$$\mathcal{N}\big(\mathbf{G}(\mathbf{A})\big) \ =\ \Big\{i : \big(\forall \mathbf{u} \in \mathbb{N}^k\big)\Big(G_{i,\mathbf{u}} = \emptyset \text{ or } \big(\exists j\big)\big(u_j > 0 \text{ and } A_j = \emptyset\big)\Big)\Big\}, \tag{3.3}$$

since from (3.2) one has $i \in \mathcal{N}\big(\mathbf{G}(\mathbf{A})\big)$ iff, for every $\mathbf{u} \in \mathbb{N}^k$, one has $G_{i,\mathbf{u}} + \mathbf{u} \circledast \mathbf{A} = \emptyset$, and this holds iff, for every $\mathbf{u} \in \mathbb{N}^k$, one has either $G_{i,\mathbf{u}} = \emptyset$, or for some j, $u_j * A_j = \emptyset$. Note that $u_j * A_j = \emptyset$ holds iff $u_j > 0$ and $A_j = \emptyset$.
Item (b) is immediate from (3.3).
Next note that $\mathbf{A} \le \mathbf{B}$ implies $\mathcal{N}(\mathbf{B}) \subseteq \mathcal{N}(\mathbf{A})$; this and (a) give (c).
To prove (d), note that from $\boldsymbol{\emptyset} \le \mathbf{G}(\boldsymbol{\emptyset})$ and (a) one has an increasing sequence

$$\boldsymbol{\emptyset} \le \mathbf{G}(\boldsymbol{\emptyset}) \le \mathbf{G}^{(2)}(\boldsymbol{\emptyset}) \le \cdots .$$

Then (c) gives the decreasing sequence

$$\{1,\dots,k\} \;=\; \mathcal{N}(\mathbf{Ø}) \;\supseteq\; \mathcal{N}\big(\mathbf{G}(\mathbf{Ø})\big) \;\supseteq\; \mathcal{N}\big(\mathbf{G}^{(2)}(\mathbf{Ø})\big) \;\supseteq\; \cdots.$$

From (b) one sees that once two consecutive members of this sequence are equal, then all members further along in the sequence are equal to them. This shows the sequence must stabilize by the term $\mathcal{N}(\mathbf{G}^{(k)}(\mathbf{Ø}))$. □

Lemma 3.5. *Suppose* $\mathbf{G}(\mathbf{Y}) \in \mathsf{Dom}(\mathbf{Y})^k$ *and* $\mathbf{A} \in \mathrm{Su}(\mathbb{N})^k$, *with* $\mathbf{A} \le \mathbf{G}(\mathbf{A})$. *Then*

$$\min \mathbf{G}^{(\infty)}(\mathbf{A}) \;=\; \min \mathbf{G}^{(k)}(\mathbf{A}).$$

In particular, $\min \mathbf{G}^{(\infty)}(\mathbf{Ø}) \;=\; \min \mathbf{G}^{(k)}(\mathbf{Ø})$.

Proof. Use Lemma 3.4 (a) to show that the minimum stabilizes after at most k steps. The details are in Appendix A. □

3.2. The minimum solution of $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$.

Proposition 3.6. *For* $\mathbf{G}(\mathbf{Y}) \in \mathsf{Dom}(\mathbf{Y})^k$, *the system of set equations* $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ *has a minimum solution* $\mathbf{S}$ *in* $\mathrm{Su}(\mathbb{N})^k$, *namely*

$$\mathbf{S} \;=\; \mathbf{G}^{(\infty)}(\mathbf{Ø}) \;:=\; \bigvee_{n\ge 0} \mathbf{G}^{(n)}(\mathbf{Ø}).$$

Furthermore, for $1 \le i \le k$, *one has* $S_i = \emptyset$ *iff* $G_i^{(k)}(\mathbf{Ø}) = \emptyset$.

Proof. The sequence $\mathbf{G}^{(n)}(\mathbf{Ø})$ is monotone nondecreasing by Lemma 3.4 (a) since $\mathbf{Ø} \le \mathbf{G}(\mathbf{Ø})$. Suppose $a \in G_i^{(\infty)}(\mathbf{Ø})$. Then, for some $n \ge 1$,

$$a \in G_i^{(n)}(\mathbf{Ø}) \;=\; G_i\big(\mathbf{G}^{(n-1)}(\mathbf{Ø})\big) \;\subseteq\; G_i\big(\mathbf{G}^{(\infty)}(\mathbf{Ø})\big).$$

This implies $\mathbf{G}^{(\infty)}(\mathbf{Ø}) \le \mathbf{G}\big(\mathbf{G}^{(\infty)}(\mathbf{Ø})\big)$.

Conversely, suppose $a \in G_i\big(\mathbf{G}^{(\infty)}(\mathbf{Ø})\big)$. Then, for some $\mathbf{u} \in \mathbb{N}^k$,

$$a \in G_{i,\mathbf{u}} + \mathbf{u} \circledast \mathbf{G}^{(\infty)}(\mathbf{Ø}),$$

which in turn implies, for some $\mathbf{u} \in \mathbb{N}^k$ and $n \ge 1$,

$$a \in G_{i,\mathbf{u}} + \mathbf{u} \circledast \mathbf{G}^{(n)}(\mathbf{Ø}) \;\subseteq\; G_i^{(n+1)}(\mathbf{Ø}) \;\subseteq\; G_i^{(\infty)}(\mathbf{Ø}).$$

Thus $\mathbf{G}^{(\infty)}(\mathbf{Ø}) = \mathbf{G}\big(\mathbf{G}^{(\infty)}(\mathbf{Ø})\big)$, so $\mathbf{G}^{(\infty)}(\mathbf{Ø})$ is indeed a solution to $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$.

Now, given any solution $\mathbf{T}$, from $\mathbf{Ø} \le \mathbf{T}$ and Lemma 3.4 (a), it follows that, for $n \ge 0$, one has $\mathbf{G}^{(n)}(\mathbf{Ø}) \le \mathbf{G}^{(n)}(\mathbf{T}) = \mathbf{T}$, and thus $\mathbf{G}^{(\infty)}(\mathbf{Ø}) \le \mathbf{T}$, showing that $\mathbf{G}^{(\infty)}(\mathbf{Ø})$ is the smallest solution to $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$.

The test for $S_i = \emptyset$ is immediate from Lemma 3.5.

□

3.3. The dependency digraph for $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$.

In the study of systems $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ with $\mathbf{G}(\mathbf{Y}) \in \mathsf{Dom}(\mathbf{Y})^k$, it is important to know when Y_i depends on Y_j. This information is succinctly collected in the dependency digraph of the system.

DEFINITION 3.7. The *dependency digraph* D of a system $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ (with k equations) has vertices $1, \ldots, k$, and directed edges given by $i \to j$ iff there is a $\mathbf{u} \in \mathbb{N}^k$ such that $G_{i,\mathbf{u}} \neq \emptyset$ and $u_j > 0$.

The *dependency matrix* M of the system is the matrix of the digraph D.

If $i \to j \in D$ then we say "i depends on j", as well as "Y_i depends on Y_j". The transitive closure of $\to$ is $\to^+$; the notation $i \to^+ j$ is read: "i eventually depends on j". It asserts that there is a directed path in D from i to j. In this case one also says "Y_i eventually depends on Y_j". The reflexive and transitive closure of $\to$ is $\to^*$.

For each vertex i let $[i]$ denote the (possibly empty) *strong component* of i in the dependency digraph, that is,

$$[i] \;:=\; \big\{j \;:\; i \to^+ j \to^+ i\big\}.$$

The system $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ is *irreducible* if the dependency digraph consists of a single strong component, that is, $i \to^+ j \to^+ i$, for all vertices i, j.

For a given system $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$, the following are easily seen to be equivalent:

(a) $i \to^+ j$.
(b) There is an $n \in \{1, \ldots, k\}$ such that $(M^n)_{i,j} > 0$.
(c) The (i, j) entry of $M + \cdots + M^k$ is > 0.

3.4. The main theorem on set equations.

Recall that $\mathbf{u} \circledast \mathbf{Y}$ is $u_1 * Y_1 + \cdots + u_k * Y_k$; and $\mathbf{G}_{\mathbf{u}} + \mathbf{u} \circledast \mathbf{Y}$ is the k-tuple obtained by adding $\mathbf{u} \circledast \mathbf{Y}$ to each component of $\mathbf{G}_{\mathbf{u}}$.

When a solution $\mathbf{S}$ of a system $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ gives the spectra S_i of generating functions, then 0 is excluded from the S_i, so one has the condition $0 \notin G_{i,\mathbf{0}}$, for $1 \leq i \leq k$, that is, $\mathbf{G}(\mathbf{Y}) \in \mathsf{Dom}_0(\mathbf{Y})^k$. One would like to assume that trivial equations $Y_i = Y_j$ have, after suitable substitutions into the other equations, been set aside. A much stronger condition that will appear in the study of spectra of generating functions is: $\mathbf{u} \circledast \mathbf{Y} = Y_j$ implies $0 \notin G_{i,\mathbf{u}}$, for any i. All of these restrictions on $\mathbf{G}(\mathbf{Y})$ are captured in the definition of *elementary* systems of set equations.

DEFINITION 3.8. $\mathbf{G}(\mathbf{Y})$ is *elementary* if $\mathbf{G}(\mathbf{Y}) \in \mathsf{Dom}(\mathbf{Y})^k$ and

$$0 \in G_{i,\mathbf{u}} \;\Rightarrow\; \sum_{j=1}^{k} u_j \geq 2, \qquad \text{for } 1 \leq i \leq k,\ \mathbf{u} \in \mathbb{N}^k.$$

A system $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ of set equations, where $\mathbf{G}(\mathbf{Y}) \in \mathsf{Dom}(\mathbf{Y})^k$, is *elementary* if $\mathbf{G}(\mathbf{Y})$ is elementary. If the system also satisfies $\mathcal{N}\big(\mathbf{G}^{(k)}(\boldsymbol{\emptyset})\big) = \emptyset$, that is, no coordinate of $\mathbf{G}^{(k)}(\boldsymbol{\emptyset})$ is the empty set, then one has a *reduced* elementary system.

One more fact is needed for the main theorem.

PROPOSITION 3.9. *Define* $d_k : \mathrm{Su}(\mathbb{N})^k \times \mathrm{Su}(\mathbb{N})^k \to \mathbb{R}$ *by*

$$d_k(\mathbf{A}, \mathbf{B}) \;:=\; \begin{cases} 2^{-\min \bigcup_{i=1}^{k} \left(A_i \triangle B_i\right)} & \textit{if } \mathbf{A} \neq \mathbf{B} \\ 0 & \textit{if } \mathbf{A} = \mathbf{B}. \end{cases}$$

Then $\big(\mathrm{Su}(\mathbb{N})^k, d_k\big)$ *is a complete metric space, and, for* $\mathbf{A}_1, \mathbf{A}_2, \ldots$ *a Cauchy sequence in this space, one has*

$$\lim_{j\to\infty} \mathbf{A}_j \ = \ \bigvee_{n\ge 1} \bigwedge_{m\ge n} \mathbf{A}_m,$$

that is, the ith coordinate of $\lim_j \mathbf{A}_j$ *is*

$$\bigcup_{n\ge 1} \bigcap_{m\ge n} A_{i,m}.$$

$\big(\mathrm{Su}(\mathbb{P})^k, d_k\big)$ *is a complete subspace.*

PROOF. It is straightforward to verify that $\big(\mathrm{Su}(\mathbb{N}), d_1\big)$ is a metric space. A sequence $A_1, A_2, \ldots$ of subsets of $\mathbb{N}$ is a Cauchy sequence in this space iff, for any $a \in \mathbb{P}$, there is an $b \in \mathbb{P}$ such that for $m, n \ge b$, one has $A_m\big|_{\le a} = A_n\big|_{\le a}$. Then, for $A_1, A_2, \ldots$ a Cauchy sequence in this space, one has

$$\lim_{j\to\infty} A_j \ = \ \bigcup_{n\ge 1} \bigcap_{m\ge n} A_m,$$

so $\big(\mathrm{Su}(\mathbb{N}), d_1\big)$ is a complete metric space.[10] For $\mathbf{A}, \mathbf{B} \in \mathrm{Su}(\mathbb{N})^k$ one has

$$d_k(\mathbf{A}, \mathbf{B}) \ = \ \max\big(d_1(A_1, B_1), \ldots, d_1(A_k, B_k)\big).$$

Thus $\big(\mathrm{Su}(\mathbb{N})^k, d_k\big)$ is also a complete metric space, and for $\mathbf{A}_1, \mathbf{A}_2, \ldots$ a Cauchy sequence in this space,

$$\lim_{j\to\infty} \mathbf{A}_j \ = \ \bigvee_{n\ge 1} \bigwedge_{m\ge n} \mathbf{A}_m \ = \ \Big(\lim_{j\to\infty} A_{1,j}, \ldots, \lim_{j\to\infty} A_{k,j}\Big),$$

where the $\lim_j A_{i,j}$, for $1 \le i \le k$, are calculated in $\big(\mathrm{Su}(\mathbb{N}), d_1\big)$.

Given a Cauchy sequence $\mathbf{A}_n$ from $\mathrm{Su}(\mathbb{P})^k$, it is routine to check that $\lim_n \mathbf{A}_n \in \mathrm{Su}(\mathbb{P})^k$. $\square$

An important collection of Cauchy sequences is given in the following corollary.

COROLLARY 3.10. *Let* $\mathbf{A}_1, \mathbf{A}_2, \ldots$ *be a nondecreasing sequence in* $\mathrm{Su}(\mathbb{N})^k$*, that is,* $\mathbf{A}_1 \le \mathbf{A}_2 \le \cdots$*. Then* $\mathbf{A}_n$ *is a Cauchy sequence in* $\big(\mathrm{Su}(\mathbb{N})^k, d_k\big)$*, and*

$$\lim_{n\to\infty} \mathbf{A}_n \ = \ \bigvee\nolimits_{n\ge 1} \mathbf{A}_n,$$

that is, $\big(\lim_n \mathbf{A}_n\big)_j = \bigcup_n A_{j,n}$*, for* $1 \le j \le k$.

THEOREM 3.11. *Let* $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ *be an elementary system of* k *set equations. Then the following hold:*

---

[10]In descriptive set theory it is well known that the Cantor ternary set, as a topological subspace of the real line, is homeomorphic to the infinite product $2^{\mathbb{N}}$, where 2 has the discrete topology. Consequently the set $2^{\mathbb{N}}$ is called *the Cantor set*, and the topological space $2^{\mathbb{N}}$ is called *the Cantor space*. It is metrizable space using the metric (see Exercise 105 in [**24**])

$$\delta(s,t) \ := \ \begin{cases} 2^{-min\{n\,:\,s(n)\ne t(n)\}} & \text{if } s \ne t \\ 0 & \text{if } s = t. \end{cases}$$

The natural map from the set $2^{\mathbb{N}}$ to $\mathrm{Su}(\mathbb{N})$ converts $\big(2^{\mathbb{N}}, \delta\big)$ into $(\mathrm{Su}(\mathbb{N}), d_1)$.

(a) *There is a unique solution* $\mathbf{T} \in \mathrm{Su}(\mathbb{P})^k$, *and it is given by*

$$\mathbf{T} \;=\; \lim_{n\to\infty} \mathbf{G}^{(n)}(\mathbf{A}),\ \textit{for any } \mathbf{A} \in \mathrm{Su}(\mathbb{P})^k.$$

Setting $\mathbf{A} = \boldsymbol{\emptyset}$ *one has*

$$\mathbf{T} \;=\; \lim_{n\to\infty} \mathbf{G}^{(n)}(\boldsymbol{\emptyset}) \;=\; \mathbf{G}^{(\infty)}(\boldsymbol{\emptyset}).$$

Let $\mathfrak{m} := \mathfrak{m}(\mathbf{T})$ *and* $\mathfrak{q} := \mathfrak{q}(\mathbf{T})$.

(b) $T_i = \emptyset$ *iff* $G_i^{(k)}(\boldsymbol{\emptyset}) = \emptyset$, *that is,* $i \in \mathcal{N}\big(\mathbf{G}^{(k)}(\boldsymbol{\emptyset})\big)$.

For the remaining items, we assume the system is reduced elementary; in particular this means all T_i *are nonempty subsets of* $\mathbb{P}$.

(c) *If* $[i] \neq \emptyset$ *then* T_i *is periodic. If also there is a* $j \in [i]$ *such that, for some* $\mathbf{u} \in \mathbb{N}^k$, *one has* $G_{j,\mathbf{u}} \neq \emptyset$ *and* $\sum_{\ell\in[i]} u_\ell \geq 2$, *then* $\mathfrak{p}_i = \mathfrak{q}_i$, *so* T_i *is the union of a finite set with a single arithmetical progression, in particular*

$$T_i \;=\; T_i|_{<\mathfrak{c}_i} \cup \big(\mathfrak{c}_i + \mathfrak{q}_i\cdot\mathbb{N}\big) \;\subseteq\; \mathfrak{m}_i + \mathfrak{q}_i\cdot\mathbb{N}.$$

(d) *Suppose* $[i] = \emptyset$ *and the ith equation can be written in the form*

$$Y_i \;=\; P_i \;+\; \bigcup_{\mathbf{Q}\in\mathcal{Q}_i} \sum_{j=1}^{k} Q_j * Y_j,$$

with P_i *[eventually] periodic, and* $\mathcal{Q}_i$ *a finite set of* k*-tuples* $\mathbf{Q} = (Q_1,\ldots,Q_k)$ *of [eventually] periodic subsets* Q_j *of* $\mathbb{N}$, *and, for* $i \to j$, *one has* T_j *being [eventually] periodic. Then* T_i *is [eventually] periodic.*

(e) *The periodicity parameters* $\mathfrak{m},\mathfrak{q}$ *of the solution* $\mathbf{T}$ *can be found from* $\mathbf{G}^{(k)}(\boldsymbol{\emptyset})$ *and the* $\mathbf{G}_{\mathbf{u}}$ *via the formulas:*

$$\mathfrak{m}_i := \mathfrak{m}(T_i) \;=\; \min\Big(G_i^{(k)}(\boldsymbol{\emptyset})\Big) \tag{3.4}$$

$$\mathfrak{q}_i := \mathfrak{q}(T_i) \;=\; \gcd\Bigg(\bigcup_{j\,:\,i\to^* j}\ \bigcup_{\mathbf{u}\in\mathbb{N}^k}\Big(G_{j,\mathbf{u}} + \mathbf{u}\circledast\mathfrak{m} - \mathfrak{m}_j\Big)\Bigg). \tag{3.5}$$

Thus, if $[i]\neq\emptyset$, *all* $\mathfrak{q}_j$ *with* $j\in[i]$ *are equal.*

(f) *One has* $\mathfrak{q}_i \mid \mathfrak{q}_j$ *whenever* $i \to j$.

Proof. First observe that since $\mathbf{G}(\mathbf{Y})$ is elementary, the mapping $\mathbf{G} : \mathrm{Su}(\mathbb{P})^k \to \mathrm{Su}(\mathbb{P})^k$ is a contraction map on the metric space $\big(\mathrm{Su}(\mathbb{P})^k, d\big)$. To see this, let $\mathbf{A},\mathbf{B} \in \mathrm{Su}(\mathbb{P})^k$, and let $n \in \mathbb{N}$ be such that $\mathbf{A}|_{\leq n} = \mathbf{B}|_{\leq n}$, that is, $A_j|_{\leq n} = B_j|_{\leq n}$ for $1 \leq j \leq k$. Then, for $\mathbf{u} \in \mathbb{N}^k$ and $1 \leq j \leq k$,

$$\begin{aligned}
\Big(G_{j,\mathbf{u}} + \mathbf{u}\circledast\mathbf{A}\Big)\Big|_{\leq n+1} \;&=\; \Big(G_{j,\mathbf{u}} + \mathbf{u}\circledast(\mathbf{A}|_{\leq n})\Big)\Big|_{\leq n+1}\\
&=\; \Big(G_{j,\mathbf{u}} + \mathbf{u}\circledast(\mathbf{B}|_{\leq n})\Big)\Big|_{\leq n+1}\\
&=\; \Big(G_{j,\mathbf{u}} + \mathbf{u}\circledast\mathbf{B}\Big)\Big|_{\leq n+1},
\end{aligned}$$

since $0 \in G_{j,\mathbf{u}}$ implies $\sum_{j=1}^k u_j \geq 2$, by the elementary property of $\mathbf{G}(\mathbf{Y})$. Thus

$$\mathbf{A}|_{\leq n} = \mathbf{B}|_{\leq n} \;\Rightarrow\; \mathbf{G}(\mathbf{A})|_{\leq n+1} \;=\; \mathbf{G}(\mathbf{B})|_{\leq n+1}.$$

This implies

$$d_k\big(\mathbf{G}(\mathbf{A}),\mathbf{G}(\mathbf{B})\big) \;\leq\; \frac{1}{2} d_k(\mathbf{A},\mathbf{B}),$$

so $\mathbf{G}$ defines a contraction mapping on $\big(\mathrm{Su}(\mathbb{P})^k, d_k\big)$. Since this is a complete metric space, $\mathbf{G}$ has a unique fixpoint $\mathbf{T}$, and $\mathbf{T} = \lim_{n\to\infty} \mathbf{G}^{(n)}(\mathbf{A})$, for any choice of $\mathbf{A}$ in $\mathrm{Su}(\mathbb{P})^k$. $\mathbf{T}$ is the unique solution to $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ in $\mathrm{Su}(\mathbb{P})^k$. With $\mathbf{A} = \boldsymbol{\emptyset}$ one also has $\mathbf{T} = \mathbf{G}^{(\infty)}(\boldsymbol{\emptyset})$, by Lemma 3.10, since the sequence $\mathbf{G}^{(n)}(\boldsymbol{\emptyset})$ is nondecreasing. This proves (a).

Item (b) follows from Proposition 3.6.

Now assume that the system is reduced elementary. Then each T_i is a nonempty set of positive integers.

For (c), first note that given i and $\mathbf{u}$ such that $G_{i,\mathbf{u}} \neq \emptyset$, there is a $q \geq 0$ (any $q \in G_{i,\mathbf{u}}$) such that

$$T_i \ \supseteq \ q + \sum_{\substack{1\leq j\leq k\\ u_j\neq 0}} u_j * T_j.$$

We can choose q to be positive if $G_{i,\mathbf{u}} \neq \{0\}$; if $G_{i,\mathbf{u}} = \{0\}$ then $\sum_j u_j \geq 2$. From this, $i \to j$ implies $T_i \supseteq p + T_j$, for some positive p; hence

$$i \to^+ j \quad \text{implies} \quad T_i \supseteq p + T_j, \quad \text{for some positive } p. \tag{3.6}$$

Now suppose $[i] \neq \emptyset$. Then $i \to^+ i$, so $T_i \supseteq p + T_i$, for some positive p, that is, T_i is periodic.

For the second part of (c), from $i \to^+ j$ follows $T_i \supseteq p_1 + T_j$, for some positive p_1, by (3.6). The hypothesis of (c) gives $T_j \supseteq p_2 + T_a + T_b$, for some a, b (possibly equal) in $[i]$, and some $p_2 \geq 0$. Finally $a \to^+ i$ and $b \to^+ i$ show that $T_a \supseteq p_3 + T_i$ and $T_b \supseteq p_4 + T_i$, for positive p_3, p_4. With $p = p_1+p_2+p_3+p_4$ one has $T_i \supseteq p + 2 * T_i$. Then Lemmas 2.13 and 2.11 (d) give the desired conclusion.

For (d), just apply Lemma 2.6.

Now to prove (e) and (f). The expression (3.4) for $\mathfrak{m}_i$ is given in Lemma 3.5. To derive the formula (3.5) for $\mathfrak{q}_i$, as well as to prove (f), from item (a) one has

$$\mathbf{T} \ = \ \mathbf{G}(\mathbf{T}) \ = \ \mathbf{G}^{(\infty)}(\boldsymbol{\emptyset}) \ := \ \bigvee_{n\geq 0} \mathbf{G}^n(\boldsymbol{\emptyset}). \tag{3.7}$$

Let

$$\mathbf{H}(\mathbf{Y}) \ := \ \bigvee_{\mathbf{u}\in\mathbb{N}^k} \big(\mathbf{H}_{\mathbf{u}} + \mathbf{u} \circledast \mathbf{Y}\big), \quad \text{where} \quad \mathbf{H}_{\mathbf{u}} \ := \ \mathbf{G}_{\mathbf{u}} + \mathbf{u} \circledast \boldsymbol{\mathfrak{m}} - \boldsymbol{\mathfrak{m}}.$$

Then $\mathbf{H}(\mathbf{Y}) \in \mathsf{Dom}(\mathbf{Y})^k$. By Proposition 3.6, the equation $\mathbf{Y} = \mathbf{H}(\mathbf{Y})$ has a minimum solution $\mathbf{S}$; it satisfies

$$\mathbf{S} \ = \ \mathbf{H}^{(\infty)}(\boldsymbol{\emptyset}) \ = \ \mathbf{H}(\mathbf{S}) \ = \ \bigvee_{\mathbf{u}\in\mathbb{N}^k} \big(\mathbf{H}_{\mathbf{u}} + \mathbf{u} \circledast \mathbf{S}\big). \tag{3.8}$$

Since $\mathbf{H}(\mathbf{Y}) \ = \ \mathbf{G}(\mathbf{Y} + \boldsymbol{\mathfrak{m}}) - \boldsymbol{\mathfrak{m}}$,

$$\mathbf{H}^{(n)}(\mathbf{Y}) \ = \ \mathbf{G}^{(n)}(\mathbf{Y} + \boldsymbol{\mathfrak{m}}) - \boldsymbol{\mathfrak{m}}, \quad \text{for } n \geq 0. \tag{3.9}$$

From (3.7), (3.8) and (3.9),

$$\mathbf{S} \ = \ \bigvee_{n\geq 0} \mathbf{H}^{(n)}(\boldsymbol{\emptyset}) \ = \ \bigvee_{n\geq 0} \mathbf{G}^{(n)}(\boldsymbol{\emptyset}) - \boldsymbol{\mathfrak{m}} \ = \ \mathbf{G}^{(\infty)}(\boldsymbol{\emptyset}) - \boldsymbol{\mathfrak{m}} \ = \ \mathbf{T} - \boldsymbol{\mathfrak{m}}.$$

Thus

$$\mathbf{0} \in \mathbf{S} \ = \ \mathbf{T} - \boldsymbol{\mathfrak{m}} \ = \ \mathbf{G}(\mathbf{T}) - \boldsymbol{\mathfrak{m}} \ = \ \mathbf{H}(\mathbf{S}),$$

so, for $1 \le j \le k$,

$$S_j \;=\; \bigcup_{\mathbf{u}\in\mathbb{N}^k} \Big(H_{j,\mathbf{u}} + \mathbf{u} \circledast \mathbf{S}\Big) \;\supseteq\; \bigcup_{\mathbf{u}\in\mathbb{N}^k} H_{j,\mathbf{u}}. \tag{3.10}$$

By definition, $q_j = \gcd(S_j)$, so (3.10) implies

$$\mathfrak{q}_j \,\Big|\, \bigcup_{\mathbf{u}\in\mathbb{N}^k} H_{j,\mathbf{u}} \quad \text{for } 1 \le j \le k. \tag{3.11}$$

For $i \to j$ there is a $\mathbf{u} \in \mathbb{N}^k$ such that $G_{i,\mathbf{u}} \neq \emptyset$ and $u_j > 0$, thus

$$S_i \;\supseteq\; H_{i,\mathbf{u}} + S_j \quad \text{and } H_{i,\mathbf{u}} \neq \emptyset.$$

Then (3.11) and the fact that $q_i \,|\, S_i$ imply $\mathfrak{q}_i \,\big|\, S_j$ whenever $i \to j$. Thus

$$\mathfrak{q}_i \,\big|\, \mathfrak{q}_j \text{ whenever } i \to j, \tag{3.12}$$

which is item (f) of the theorem.

From (3.11) and (3.12),

$$i \to^* j \;\Rightarrow\; \mathfrak{q}_i \,\Big|\, \bigcup_{\mathbf{u}\in\mathbb{N}^k} H_{j,\mathbf{u}},$$

which implies

$$\mathfrak{q}_i \,\Big|\, \mathfrak{q}'_i \;:=\; \gcd\bigg(\bigcup_{j\,:\,i\to^* j} \;\bigcup_{\mathbf{u}\in\mathbb{N}^k} H_{j,\mathbf{u}}\bigg). \tag{3.13}$$

To finish the proof of (e) one needs $\mathfrak{q}'_i \,\big|\, \mathfrak{q}_i$. The key step is to show, by induction on n, that

$$i \to^* j \;\Rightarrow\; \mathfrak{q}'_i \,\Big|\, H_j^{(n)}(\boldsymbol{\emptyset}), \quad \text{for } n \ge 0. \tag{3.14}$$

Ground Case: (n=0)
Clearly $\mathfrak{q}'_i \,\big|\, \emptyset$.
Induction Step:
Assume that $\mathfrak{q}'_i \,\Big|\, H_j^{(n)}(\boldsymbol{\emptyset})$ whenever $i \to^* j$. One has

$$H_j^{(n+1)}(\boldsymbol{\emptyset}) \;=\; \bigcup_{\mathbf{u}\in\mathbb{N}^k} \Big(H_{j,\mathbf{u}} + \mathbf{u} \circledast \mathbf{H}^{(n)}(\boldsymbol{\emptyset})\Big). \tag{3.15}$$

Suppose that $i \to^* j$. Let $\mathbf{u} \in \mathbb{N}^k$. Then $\mathfrak{q}'_i \,\big|H_{j,\mathbf{u}}$ (by the definition of $\mathfrak{q}'_i$ in (3.13)). Clearly

$$H_{j,\mathbf{u}} \;=\; \emptyset \;\Rightarrow\; \mathfrak{q}'_i \,\Big|\, \Big(H_{j,\mathbf{u}} + \mathbf{u} \circledast \mathbf{H}^{(n)}(\boldsymbol{\emptyset})\Big) \;=\; \emptyset. \tag{3.16}$$

Suppose $H_{j,\mathbf{u}} \neq \emptyset$. If $\mathbf{u} = \mathbf{0}$, we have already noted that $\mathfrak{q}'_i \,\big|\, H_{j,\mathbf{0}}$. So suppose $u_m > 0$. Then $j \to m$; and, since $i \to^* j$, one has $i \to^* m$. By the induction hypothesis, $\mathfrak{q}'_i \,\Big|\, H_m^{(n)}(\boldsymbol{\emptyset})$. Consequently $\mathfrak{q}'_i \,\big|\, \big(\mathbf{u} \circledast \mathbf{H}^{(n)}(\boldsymbol{\emptyset})\big)$. Since, as noted above, $\mathfrak{q}'_i \,\big|\, H_{j,\mathbf{u}}$, it follows that

$$H_{j,\mathbf{u}} \;\neq\; \emptyset \;\Rightarrow\; \mathfrak{q}'_i \,\Big|\, \Big(H_{j,\mathbf{u}} + \mathbf{u} \circledast \mathbf{H}^{(n)}(\boldsymbol{\emptyset})\Big). \tag{3.17}$$

Items (3.16) and (3.17) show that for $\mathbf{u} \in \mathbb{N}^k$,

$$i \to^* j \;\Rightarrow\; \mathfrak{q}'_i \,\Big|\, \Big(H_{j,\mathbf{u}} + \mathbf{u} \circledast \mathbf{H}^{(n)}(\boldsymbol{\emptyset})\Big).$$

From this and (3.15) one then has $i \to^* j \;\Rightarrow\; \mathfrak{q}_i' \,\Big|\, H_j^{(n+1)}(\boldsymbol{\emptyset})$, finishing the proof of (3.14). Thus

$$i \to^* j \;\Rightarrow\; \mathfrak{q}_i' \,\Big|\, H_j^{(\infty)}(\boldsymbol{\emptyset}) \;=\; S_j\,.$$

Setting j equal to i gives $\mathfrak{q}_i' \,|\, S_i$, proving that $\mathfrak{q}_i' \,|\, \mathfrak{q}_i = \gcd(S_i)$. □

The following corollary will be used to provide information (see Corollary 4.7) on the spectra of irreducible systems studied in combinatorics.

COROLLARY 3.12. *Suppose* $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ *is an irreducible reduced elementary system of set equations with solution* $\mathbf{T}$*. Let* $\boldsymbol{\mathfrak{m}} := \boldsymbol{\mathfrak{m}}(\mathbf{T})$ *and* $\boldsymbol{\mathfrak{q}} := \boldsymbol{\mathfrak{q}}(\mathbf{T})$*. Then the following hold:*

(a) *All* T_i *are infinite and periodic, with the same parameter* $\mathfrak{q}_i$*, namely* $\mathfrak{q}_i = \mathfrak{q}$*, where*

$$\mathfrak{q} \;:=\; \gcd\left(\bigcup_{\substack{1\le j\le k\\ \mathbf{u}\in\mathbb{N}^k}} \Big(G_{j,\mathbf{u}} + \mathbf{u}\circledast\boldsymbol{\mathfrak{m}} - \mathfrak{m}_j\Big)\right).$$

(b) *If there is a* j *such that* $G_j(\mathbf{Y})$ *has a term* $G_{j,\mathbf{u}} + \mathbf{u}\circledast\mathbf{Y}$*, with* $G_{j,\mathbf{u}} \neq \emptyset$ *and* $\sum_\ell u_\ell \ge 2$*, then* $\mathfrak{p}_i = \mathfrak{q}$*, for* $1 \le i \le k$*, and for all* i*,* T_i *is the union of a finite set and a single arithmetical progression:*

$$T_i \;=\; T_i|_{<\mathfrak{c}_i} \cup (\mathfrak{c}_i + \mathfrak{q}\cdot\mathbb{N}) \;\subseteq\; \mathfrak{m}_i + \mathfrak{q}\cdot\mathbb{N}.$$

PROOF. By Theorem 3.11(c). □

REMARK 3.13. Elementary systems $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ have unique solutions in $\mathrm{Su}(\mathbb{P})^k$, but this need not be the case in $\mathrm{Su}(\mathbb{N})^k$. Consider the single equation elementary system $Y = 2 * Y$. It has infinitely many solutions in $\mathrm{Su}(\mathbb{N})$, namely any $A \subseteq \mathbb{N}$ which satisfies the two conditions: (i) $0 \in A$, and $A + A \subseteq A$. Thus, for example, $p\cdot\mathbb{N}$ is a solution, for $p \in \mathbb{N}$. However, there is a unique solution of $Y = 2 * Y$ in $\mathrm{Su}(\mathbb{P})$, namely the empty set.

4. Elementary Power Series Systems

4.1. General background for power series systems.

Recall that $\mathbb{R}$ is the set of reals, $\mathbb{N}$ the set of nonnegative integers, and $\mathbb{P}$ the set of positive integers. We continue the upper case/lower case convention connecting a power series to its coefficients, for example, $A(x) = \sum_n a(n)x^n$.

The following table gives the notations needed for this section:

$\mathbf{z}$	$=$	$z_1, \ldots, z_k$
$\mathbf{z}^{\mathbf{u}}$	$=$	$z_1^{u_1} \cdots z_k^{u_k}$
$\mathbb{F}$	$=$	a field
$\mathbb{F}[[\mathbf{z}]]$	$=$	set of power series $A(\mathbf{z}) = \sum_{\mathbf{u}} a_{\mathbf{u}} \mathbf{z}^{\mathbf{u}}$ over $\mathbb{F}$
$\mathbb{F}[[\mathbf{z}]]^k$	$=$	$\big\{(A_1(\mathbf{z}), \ldots, A_k(\mathbf{z})) : A_i(\mathbf{z}) \in \mathbb{F}[[\mathbf{z}]]\big\}$
$\mathbb{F}[[\mathbf{z}]]_0$	$=$	$\{A(\mathbf{z}) \in \mathbb{F}[[\mathbf{z}]] : A(\mathbf{0}) = 0\}$
$\big[x^{\le m}\big]\, A(x)$	$=$	$a(0) + a(1)x + \cdots + a(m)x^m$
$J_{\mathbf{G}}(x, \mathbf{y})$	$=$	the Jacobian matrix of $\mathbf{G}(x, \mathbf{y})$ with respect to $\mathbf{y}$
$\mathsf{Spec}(T(x))$	$=$	$\{n \ge 0 : t(n) \ne 0\}$, for $T(x) \in \mathbb{F}[[x]]$
$\mathsf{Spec}(\mathbf{T}(x))$	$=$	$\big(\mathsf{Spec}(T_1(x)), \ldots, \mathsf{Spec}(T_k(x))\big)$, for $\mathbf{T}(x) \in \mathbb{F}[[x]]^k$
The following items assume $\mathbb{F} = \mathbb{R}$, the field of real numbers		
$A(\mathbf{z}) \trianglerighteq B(\mathbf{z})$	says	$a_{\mathbf{u}} \ge b_{\mathbf{u}}$ for all $\mathbf{u}$
$\mathbf{A}(\mathbf{z}) \trianglerighteq \mathbf{B}(\mathbf{z})$	says	$A_i(\mathbf{z}) \trianglerighteq B_i(\mathbf{z})$ for all i
$\mathsf{Dom}[\mathbf{z}]$	$=$	$\{A(\mathbf{z}) \in \mathbb{R}[[\mathbf{z}]] : A(\mathbf{z}) \trianglerighteq 0\}$
$\mathsf{Dom}_0[\mathbf{z}]$	$=$	$\{A(\mathbf{z}) \in \mathsf{Dom}[\mathbf{z}] : A(\mathbf{0}) = 0\}$
$\mathbf{G}(x, \mathbf{y})$	$=$	a member of $\mathsf{Dom}_0[x, \mathbf{y}]^k$
$\mathsf{Dom}_{J0}[x, \mathbf{y}]$	$=$	$\big\{\mathbf{G}(x, \mathbf{y}) \in \mathsf{Dom}_0[x, \mathbf{y}]^k : J_{\mathbf{G}}(0, \mathbf{0}) = \mathbf{0}\big\}$, where $\mathbf{y} = y_1, \ldots, y_k$

For $k \ge 1$, the set $\mathbb{F}[[x]]^k$ becomes a complete metric space (see *Analytic Combinatorics* [**19**], p. 731) when equipped with the metric

$$d_k\big(\mathbf{A}(x), \mathbf{B}(x)\big) \;:=\; \begin{cases} 2^{-\min \operatorname{ldegree}\left(A_i(x) - B_i(x)\,:\,1 \le i \le k\right)} & \text{if } \mathbf{A}(x) \ne \mathbf{B}(x) \\ 0 & \text{if } \mathbf{A}(x) = \mathbf{B}(x), \end{cases}$$

where the *ldegree* of a power series $A(x)$ is its lowest degree, that is, $\operatorname{ldegree}(A(x)) := \min\{n \in \mathbb{N} : a(n) \ne 0\}$. One has $d_k\big(\mathbf{A}_n(x), \mathbf{B}_n(x)\big) \to 0$ as $n \to \infty$ iff for $m \ge 0$ there is an $N \ge 0$ such that $[x^{\le m}]\mathbf{A}_n(x) = [x^{\le m}]\mathbf{B}_n(x)$ for $n \ge N$; that is, for n sufficiently large, the corresponding coordinates of $\mathbf{A}_n(x)$ and $\mathbf{B}_n(x)$ agree on their first $m+1$ coefficients. The subset $\mathbb{F}[[x]]_0^k$ of $\mathbb{F}[[x]]^k$ is, with the same metric, also a complete metric space.

Let $k \ge 1$ be given, and let $\mathbf{y} := y_1, \ldots, y_k$. Given a k-tuple of formal power series $\mathbf{G}(x, \mathbf{y}) \in \mathbb{F}[[x, \mathbf{y}]]_0^k$, and $\mathbf{A}(x) \in \mathbb{F}[[x]]_0^k$, the composition $\mathbf{G}(x, \mathbf{A}(x))$ is a member of $\mathbb{F}[[x]]_0^k$. Such a $\mathbf{G}(x, \mathbf{y})$ can be viewed as a mapping from $\mathbb{F}[[x]]_0^k$ to itself, a mapping whose n-fold composition with itself will be expressed by $\mathbf{G}^{(n)}(x, \mathbf{y})$, also a member of $\mathbb{F}[[x, \mathbf{y}]]_0^k$. More precisely,

$$\begin{aligned} \mathbf{G}^{(0)}(x, \mathbf{y}) &= \mathbf{y}, \\ \mathbf{G}^{(n+1)}(x, \mathbf{y}) &= \mathbf{G}(x, \mathbf{G}^{(n)}(x, \mathbf{y})). \end{aligned}$$

The power series in the ith coordinate of $\mathbf{G}^{(n)}(x,\mathbf{y})$ will be denoted by $G_i^{(n)}(x,\mathbf{y})$, that is,

$$\mathbf{G}^{(n)}(x,\mathbf{y}) \quad = \quad \big(G_1^{(n)}(x,\mathbf{y}),\ldots,G_k^{(n)}(x,\mathbf{y})\big).$$

Basic results, on the existence and uniqueness of power series solutions to systems of equations, hold in the general setting of power series over a field; however, the natural setting for analyzing the generating functions of combinatorial classes is to work with power series over the real field $\mathbb{R}$.

Proposition 4.1. *Let* $\mathbf{G}(x,\mathbf{y}) \in \mathbb{F}[[x,\mathbf{y}]]^k$. *If*

(a) $\mathbf{G}(0,\mathbf{0}) = \mathbf{0}$ *and*
(b) $J_{\mathbf{G}}(0,\mathbf{0}) = \mathbf{0}$

then the equational system $\mathbf{y} = \mathbf{G}(x,\mathbf{y})$

(i) *has a unique solution* $\mathbf{T}(x)$ *in* $\mathbb{F}[[x]]_0^k$, *and,*
(ii) *for any* $\mathbf{A}(x) \in \mathbb{F}[[x]]_0^k$, *one has, in the complete metric space* $\big(\mathbb{F}[[x]]_0^k, d_k\big)$,

$$\mathbf{T}(x) = \lim_{n\to\infty} \mathbf{G}^{(n)}(x,\mathbf{A}(x)).$$

If, furthermore,

(c) $\mathbb{F} = \mathbb{R}$,

then

(iii) $\mathbf{G}(x,\mathbf{y}) \trianglerighteq \mathbf{0} \;\Rightarrow\; \mathbf{T}(x) \trianglerighteq \mathbf{0}$.

Proof. For $\mathbf{A}(x), \mathbf{B}(x) \in \mathbb{F}[[x]]_0^k$, the hypotheses guarantee that

$$[x^{\le n}]\mathbf{A}(x) = [x^{\le n}]\mathbf{B}(x) \quad \Rightarrow \quad [x^{\le n+1}]\mathbf{G}(x,\mathbf{A}(x)) = [x^{\le n+1}]\mathbf{G}(x,\mathbf{A}(x)).$$

This implies that $\mathbf{G}(x,\mathbf{y})$ is a contraction mapping on the complete metric space $\big(\mathbb{F}[[x]]_0^k, d_k\big)$; consequently (i)–(ii) follow. Item (iii) follows from (ii). $\square$

For $T(x) \in \mathbb{F}[[x]]$, let $\mathfrak{m} := \mathfrak{m}\big(\mathsf{Spec}(T(x))\big)$ and $\mathfrak{q} := \mathfrak{q}\big(\mathsf{Spec}(T(x))\big)$, as in Definition 2.7.[11] Assuming $T(x) \neq 0$, one has the following:

(a) The largest power of x dividing $T(x)$ is $x^{\mathfrak{m}}$, since $\mathfrak{m}$ is the smallest index n such that $t(n) \neq 0$.
(b) The monomial $x^{\mathfrak{q}}$ is the largest power of x such that, for $n \ge 0$, $t(n) \neq 0$ implies $\mathfrak{q} \mid (n - \mathfrak{m})$.
(c) There is a (unique) power series $V(x) \in \mathbb{F}[[x]]$ such that $T(x) = x^{\mathfrak{m}} V(x^{\mathfrak{q}})$, and one has $\gcd(\mathsf{Spec}(V(x))) = 1$.
(d) Suppose $T(x) \in \mathsf{Dom}_0[x]$, and it has radius of convergence $\rho \in (0,\infty)$. By Pringsheim's Theorem, ρ is a singularity of $T(x)$, and thus it easily follows that $\rho \cdot \omega^j$, $j = 0,\ldots,\mathfrak{q}-1$, where ω is a primitive $\mathfrak{q}$th root of unity, are among the dominant singularities[12] of $T(x)$. If $T(x)$ is the solution to a well conditioned equation $y = G(x,y)$ then these are the only dominant singularities—this leads to $\mathfrak{q}$ having a prominent role in the expression for the asymptotics of the coefficients $t(n)$ of $T(x)$. (See §7).

[11] In *Analytic Combinatorics* [**19**], $\mathfrak{q}$ is called the *period* of $T(x)$, as well as of the sequence $t(n)$. It is only used in situations where $\mathfrak{p} = \mathfrak{q}$.

[12] Given a power series $A(x)$, the singularities on the circle of convergence are called the *dominant* singularities. A study of $A(x)$ near its dominant singularities is usually the first step, and often the main step, towards determining an asymptotic formula for the coefficients $a(n)$ of $A(x)$.

Under favorable conditions—such as those encountered in [**2**], a study of the solution $T(x)$ to $y = G(x, y)$, a non y-linear single equation system—$\mathsf{Spec}(T(x))$ is the union of a finite set and an arithmetical progression, and the coefficients $t(n)$ of $T(x)$ have 'nice' asymptotics for n on this spectrum. It would be an important achievement to show that any system $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ built from standard components has a solution $\mathbf{T}(x)$ with the $T_i(x)$ exhibiting the positive features just described. A first, and very modest step in this direction, is to show that such systems have spectra of the appropriate kind, namely eventually periodic spectra.

4.2. Nonnegative power series and elementary systems.

There were two stages in our investigation [**2**] of generating functions defined by a single equation. The first considered $y = G(x, y)$, where $G(x, y)$ was a power series with nonnegative coefficients. The second looked at more complex equations $y = \Theta(y)$ involving operators like Multiset, Sequence, and Cycle. The same two stages will be followed in this study of generating functions defined by systems of equations.

Definition 4.2. A power series $A(\mathbf{z}) \in \mathbb{R}[[\mathbf{z}]]$ is *nonnegative* if $A(\mathbf{z}) \trianglerighteq 0$, that is, each coefficient $a_{\mathbf{u}}$ is nonnegative. $\mathbf{A}(\mathbf{z}) \in \mathbb{R}[[\mathbf{z}]]^k$ is nonnegative if each $A_i(\mathbf{z})$ is nonnegative. A system $\mathbf{y} = \mathbf{G}(x, \mathbf{y}) \in \mathbb{R}[[x, \mathbf{y}]]^k$ is nonnegative if $\mathbf{G}(x, \mathbf{y})$ is nonnegative.

When applied to nonnegative power series, the Spec operator acts like a homomorphism, as the next lemma shows. This is used to convert equational systems satisfied by generating functions into equational systems satisfied by spectra.

Lemma 4.3. *Let $c > 0$ and let $A(x), A_i(x), B(x) \in \mathbb{R}[[x]]$ be nonnegative power series. Then*

(a) $\mathsf{Spec}\big(c \cdot A(x)\big) = \mathsf{Spec}(A(x))$
(b) $\mathsf{Spec}\big(A(x) + B(x)\big) = \mathsf{Spec}(A(x)) \cup \mathsf{Spec}(B(x))$
(c) $\mathsf{Spec}\Big(\sum_i A_i(x)\Big) \;=\; \bigcup_i \mathsf{Spec}(A_i(x)),$ *provided* $\sum_i A_i(x) \in \mathbb{R}[[x]]$
(d) $\mathsf{Spec}\big(A(x) \cdot B(x)\big) = \mathsf{Spec}(A(x)) + \mathsf{Spec}(B(x))$
(e) $\mathsf{Spec}\big(A(x) \circ B(x)\big) = \mathsf{Spec}(A(x)) * \mathsf{Spec}(B(x))$, *provided* $B(x) \in \mathbb{R}[[x]]_0$.

Proof. The first four cases (scalar multiplication, addition, summation and Cauchy product) are straightforward, as is composition:

$$\begin{aligned}\mathsf{Spec}\big(A(x) \circ B(x)\big) &= \mathsf{Spec}\Big(\sum_{i \geq 1} a(i) B(x)^i\Big) = \bigcup_{i \in \mathsf{Spec}(A(x))} \mathsf{Spec}\left(B(x)^i\right) \\ &= \bigcup_{i \in \mathsf{Spec}(A(x))} i * \mathsf{Spec}(B(x)) \;=\; \mathsf{Spec}(A(x)) * \mathsf{Spec}(B(x)).\end{aligned}$$

□

One defines the *dependency digraph* for an equational system $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ parallel to the way one defines it for a system of set equations $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$, namely $i \to j$ iff $G_i(x, \mathbf{y})$ depends on y_j, for $1 \leq i, j \leq k$.

Lemma 4.4 (Tests for eventual dependence). *Given a nonnegative system $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$, the following are equivalent:*

(a) *One has $i \to^+ j$.*

(b) *There is an* $m \in \{1, \ldots, k\}$ *such that the* (i, j) *entry of* $J_{\mathbf{G}}(x, \mathbf{y})^m$ *is not* 0.
(c) *The* (i, j) *entry of* $\sum_{m=1}^{k} J_{\mathbf{G}}(x, \mathbf{y})^m$ *is not* 0.

In practice one only works with equational systems that have a connected dependency digraph. Otherwise the system trivially breaks up into several independent subsystems. There has been considerable interest in *irreducible* nonnegative equational systems $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$, where every y_i eventually depends on every y_j, that is, $i \rightarrow^+ j$ holds for all i, j. One can also express this by the condition:

$$\textit{the matrix} \quad \sum_{n=1}^{k} J_{\mathbf{G}}(x, \mathbf{y})^n \quad \textit{has all entries nonzero.}$$

Such systems behave, in many ways, like irreducible 1-equation systems. Clearly an equational system $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ is irreducible if, for some n, no entry of $J_{\mathbf{G}}(x, \mathbf{y})^n$ is zero. In this case the matrix $J_{\mathbf{G}}(x, \mathbf{y})$ (and the system) is said to be *primitive* (or *aperiodic* in *Analytic Combinatorics* [**19**]).

However, even some nonnegative irreducible systems $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ can be easily decomposed into several independent subsystems—this will happen precisely when $J_{\mathbf{G}}(x, \mathbf{y})$ is *imprimitive*, that is, irreducible but not primitive. This case happens precisely when there is a permutation of the indices $1, \ldots, k$ transforming $J_{\mathbf{G}}(x, \mathbf{y})$ into $\widehat{J}_{\mathbf{G}}(x, \mathbf{y})$ such that, for some $n \geq 1$, $\widehat{J}_{\mathbf{G}}(x, \mathbf{y})^n$ has a block diagonal form with at least two blocks. By choosing the permutation to maximize the number of blocks obtainable, one finds that each block gives rise to an irreducible system that is primitive. Awareness of this possibility of decomposing irreducible systems is important for practical computational work.

A power series $G(x, \mathbf{y})$ can be expressed in the form

$$\sum_{\mathbf{u} \in \mathbb{N}^k} G_{\mathbf{u}}(x) \cdot \mathbf{y}^{\mathbf{u}},$$

where $\mathbf{y}^{\mathbf{u}}$ is the monomial $y_1^{u_1} \cdots y_k^{u_k}$. The associated set expression $G(\mathbf{Y})$, where $G_{\mathbf{u}} := \mathsf{Spec}(G_{\mathbf{u}}(x))$, is given by

$$G(\mathbf{Y}) \ := \ \bigcup_{\mathbf{u} \in \mathbb{N}^k} \big(G_{\mathbf{u}} + \mathbf{u} \circledast \mathbf{y}\big).$$

Definition 4.5. A nonnegative power series $\mathbf{G}(x, \mathbf{y})$ is *elementary* if it satisfies the conditions of Proposition 4.1, namely:
(a) $\mathbf{G}(0, \mathbf{0}) = \mathbf{0}$ and
(b) $J_{\mathbf{G}}(0, \mathbf{0}) = \mathbf{0}$.
An equational system $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ is *elementary* iff $\mathbf{G}(x, \mathbf{y})$ is elementary.

This definition is easily seen to be equivalent to requiring: for $\mathbf{u} \in \mathbb{N}^k$ and $1 \leq i \leq k$,

$$G_{i,\mathbf{u}}(0) \neq 0 \ \Rightarrow \ \sum_{j=1}^{k} u_j \geq 2.$$

Consequently a nonnegative power series system $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ is elementary iff the associated system of set equations $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ is elementary.

The next result is the main theorem on power series systems.

Theorem 4.6. *For an elementary system* $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ *the following hold:*

(a) *The system has a unique solution* $\mathbf{T}(x)$ *in* $\mathbb{R}[[x]]_0^k$.
Let $\mathfrak{m} := \mathfrak{m}(\mathsf{Spec}(\mathbf{T}(x)))$ *and* $\mathfrak{q} := \mathfrak{q}(\mathsf{Spec}(\mathbf{T}(x)))$.

(b) $\mathbf{T}(x) \trianglerighteq \mathbf{0}$, *that is, the coefficients of each* $T_i(x)$ *are nonnegative.*

(c) $\mathbf{T}(x) = \lim_{n\to\infty} \mathbf{G}^{(n)}(x, \mathbf{A}(x))$, *for any* $\mathbf{A}(x) \in \mathbb{R}[[x]]_0^k$.

(d) $\mathbf{Y} = \mathsf{Spec}(\mathbf{T}(x))$ *is the unique solution to the elementary system of set equations* $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$, *where*

$$\mathbf{G}(\mathbf{Y}) \ := \ \bigvee_{\mathbf{u}\in\mathbb{N}^k} \big(\mathbf{G}_{\mathbf{u}} + \mathbf{u} \circledast \mathbf{Y}\big).$$

(e) $\mathsf{Spec}(\mathbf{T}(x)) \ = \ \lim_{n\to\infty} \mathbf{G}^{(n)}(\mathbf{A})$, *for any* $\mathbf{A} \in \mathrm{Su}(\mathbb{P})^k$.
Thus $\mathsf{Spec}(\mathbf{T}(x)) \ = \ \mathbf{G}^{(\infty)}(\boldsymbol{\emptyset}) \ := \ \bigvee_{n\ge1} \mathbf{G}^{(n)}(\boldsymbol{\emptyset})$.

(f) $T_i(x) = 0$ *iff* $G_i^{(k)}(x, \mathbf{0}) = 0$ *iff* $\mathsf{Spec}(T_i(x)) = \emptyset$ *iff* $G_i^{(k)}(\boldsymbol{\emptyset}) = \emptyset$ *iff* $\mathfrak{m}_i = \infty$.

Now we assume that the system has been reduced by eliminating all y_i *for which* $T_i(x) = 0$.

(g) $[i] \ne \emptyset$ *implies* $\mathsf{Spec}(T_i(x))$ *is periodic. If also there is a* $j \in [i]$ *such that, for some* $\mathbf{u} \in \mathbb{N}^k$, *one has* $G_{j,\mathbf{u}}(x) \ne 0$ *and* $\sum\{u_\ell : \ell \in [i]\} \ge 2$, *then* $\mathfrak{p}_i = \mathfrak{q}_i$ *and* $\mathsf{Spec}(T_i(x))$ *is the union of a finite set with a single arithmetical progression, namely* $\mathsf{Spec}(T_i(x)) = \mathsf{Spec}(T_i(x))|_{<\mathfrak{c}_i} \cup \big(\mathfrak{c}_i + \mathfrak{q}_i \cdot \mathbb{N}\big) \subseteq \mathfrak{m}_i + \mathfrak{q}_i \cdot \mathbb{N}$.

(h) *If* $[i] = \emptyset$ *and the* ith *set equation can be written in the form*

$$Y_i \ := \ P_i \ + \ \bigcup_{\mathbf{Q}\in\mathcal{Q}_i} \sum_{j=1}^{k} \big(Q_j * Y_j\big),$$

with P_i *[eventually] periodic, and* $\mathcal{Q}_i$ *a finite set of* k*-tuples* $\mathbf{Q} = (Q_1, \ldots, Q_k)$ *of [eventually] periodic subsets* Q_j *of* $\mathbb{N}$, *and if, for* $i \to j$, *one has* $\mathsf{Spec}(T_j(x))$ *being [eventually] periodic, then* $\mathsf{Spec}(T_i(x))$ *is [eventually] periodic.*

(i) *The periodicity parameters* $\mathfrak{m}, \mathfrak{q}$ *of* $\mathsf{Spec}(\mathbf{T}(x))$ *can be found from* $\mathbf{G}^{(k)}(\boldsymbol{\emptyset})$ *and the* $\mathbf{G}_{\mathbf{u}}$ *via the formulas*

$$\mathfrak{m}_i \ := \ \mathfrak{m}(\mathsf{Spec}(T_i(x))) \ = \ \min\Big(G_i^{(k)}(\boldsymbol{\emptyset})\Big) \tag{4.1}$$

$$\mathfrak{q}_i \ := \ \mathfrak{q}(\mathsf{Spec}(T_i(x))) \ = \ \gcd\Bigg(\bigcup_{j\,:\,i\to^* j} \ \bigcup_{\mathbf{u}\in\mathbb{N}^k} \Big(G_{j,\mathbf{u}} + \mathbf{u} \circledast \mathfrak{m} - \mathfrak{m}_j\Big)\Bigg). \tag{4.2}$$

(j) $\mathfrak{q}_i \mid \mathfrak{q}_j$ *whenever* $i \to j$.

Proof. Items (a)–(c) are immediate from Proposition 4.1. For (d) simply apply Spec to both sides of $\mathbf{T}(x) = \mathbf{G}(x, \mathbf{T}(x))$. For (e)–(j) note that the hypotheses of the theorem imply that $\mathbf{G}(\mathbf{Y})$ satisfies the hypotheses of Theorem 3.11. □

The study of irreducible systems $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ has been of interest because of parallels with work on single equation systems, with a particular interest in *non* $\mathbf{y}$*-linear systems*, that is, systems where some $G_i(x, \mathbf{y})$ has a nonvanishing Hessian matrix $\big[\partial^2 G_i(x, \mathbf{y})/\partial y_r \partial y_s\big]$. (See, for example, *Analytic Combinatorics* [**19**], VII.6.)

Corollary 4.7. *Suppose* $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ *is an irreducible reduced elementary system with solution* $\mathbf{T}(x)$. *Let* $\mathfrak{m} := \mathfrak{m}(\mathsf{Spec}(\mathbf{T}(x)))$ *and* $\mathfrak{q} := \mathfrak{q}(\mathsf{Spec}(\mathbf{T}(x)))$.

(a) *All* $\mathsf{Spec}(T_i(x))$ *are infinite and periodic, with the same parameter* $\mathfrak{q}_i$*, namely* $\mathfrak{q}_i = \mathfrak{q}$*, where*

$$\mathfrak{q} \ := \ \gcd\Bigg(\bigcup_{\substack{1\le j\le k\\ \mathbf{u}\in\mathbb{N}^k}} \Big(G_{j,\mathbf{u}} + \mathbf{u}\circledast\boldsymbol{\mathfrak{m}} - \mathfrak{m}_j\Big)\Bigg).$$

(b) *If also there is a* j *such that* $G_j(x,\mathbf{y})$ *is not* $\mathbf{y}$*-linear, then, for* $1 \le i \le k$*,* $\mathfrak{p}_i = \mathfrak{q}$ *and*

$$\mathsf{Spec}(T_i(x)) \ = \ \mathsf{Spec}(T_i(x))\big|_{<\mathfrak{c}_i} \cup (\mathfrak{c}_i + \mathfrak{q}\cdot\mathbb{N}) \ \subseteq \ \mathfrak{m}_i + \mathfrak{q}\cdot\mathbb{N}.$$

Thus each $\mathsf{Spec}(T_i(x))$ *is eventually an arithmetical progression.*

Proof. By Corollary 3.12. □

Systems that arise in combinatorial problems are invariably reduced since the solution gives generating functions for nonempty classes of objects. However, if one should encounter a nonreduced elementary *polynomial* system $\mathbf{y} = \mathbf{G}(x,\mathbf{y})$, Theorem 4.6 (f) provides an efficient way to determine which of the solution components $T_i(x)$ will be 0, namely let $\lambda(0) = 0$, and let λ map any nonzero $A(x) \in \mathsf{Dom}_0[x]$ to its lowest degree term, setting the coefficient to 1; extend this to $\mathsf{Dom}_0[x]^k$ coordinatewise. Then

$$T_i(x) = 0 \quad \text{iff} \quad \Big((\lambda\circ\mathbf{G})^{(k)}(x,\mathbf{0})\Big)_i = 0.$$

4.3. Irreducible elementary y-linear systems.

Irreducible elementary y-linear equations $y = G_0(x) + G_1(x)y$, with solution $y = T(x)$, do not, in general, have the property that $\mathsf{Spec}(T(x))$ is eventually a single arithmetical progression. For example, let $y = T(x)$ be the power series solution to

$$y \ = \ x + x^2 + x^3 y.$$

The periodicity parameters of $\mathsf{Spec}(T(x))$ are $\mathfrak{m} = 1$, $\mathfrak{q} = 1$, $\mathfrak{p} = 3$, and $\mathfrak{c} = 1$. $\mathsf{Spec}(T(x))$ is readily seen to be

$$\{3n+1 : n \ge 0\} \cup \{3n+2 : n \ge 0\},$$

and the set of periods of $\mathsf{Spec}(T(x))$ is the same as the set of eventual periods of $\mathsf{Spec}(T(x))$, namely $3\cdot\mathbb{N}$.

Nonetheless, the periodic spectrum of the solution of an irreducible elementary y-linear equation does have a particularly simple expression.

Proposition 4.8. *Given a 1-equation irreducible elementary* y*-linear system*

$$y \ = \ G(x,\mathbf{y}) \ := \ G_0(x) + G_1(x)\cdot y,$$

the solution is

$$T(x) \ = \ \Big(\sum_{n\ge 0} G_1(x)^n\Big)\cdot G_0(x),$$

the spectral equation is

$$Y \ = \ G(Y) \ := \ G_0 \cup (G_1 + Y),$$

and the spectrum is

$$\mathsf{Spec}(T(x)) \ = \ \Big(\bigcup_{n\ge 0} (n * G_1)\Big) + G_0 \ = \ G_0 + \mathbb{N} * G_1.$$

Let $\mathfrak{m} := \mathfrak{m}(\mathsf{Spec}(T(x)))$ *and* $\mathfrak{q} := \mathfrak{q}(\mathsf{Spec}(T(x)))$. *Then*

$$\mathfrak{m} = \min(G_0) \text{ and } \mathfrak{q} = \gcd\big((G_0 - \mathfrak{m}) \cup G_1\big).$$

The proof of the proposition is straightforward. From the form of the solution for $\mathsf{Spec}(T(x))$, one sees that every periodic subset of $\mathbb{P}$ is the spectrum of the solution to some 1-equation irreducible elementary y-linear system.

The next three examples, of $\mathbf{y}$-linear systems, are staples in the study of systems.

Example 4.9 (Postage Spectra). Given that one has stamps in denominations $d_1, \ldots, d_r$, the associated *postage spectrum* is the set S of amounts of postage one can put on a package using only these sizes of stamps. With $D = \{d_1, \ldots, d_r\}$ being the set of denominations of the stamps, let $D(x) = \sum_{i=1}^r n_i x^{d_i}$, where n_i is the number of distinct stamps of denomination d_i. Let $S(x)$ be the generating function with $s(n)$ giving the number of *ordered* ways to realize the postage amount n, that is, using a *sequence* of such stamps. Then $S(x)$ is the solution to the irreducible elementary y-linear equation

$$y = D(x) + D(x) \cdot y.$$

The spectrum $\mathsf{Spec}(S(x))$ is the solution to the elementary set equation

$$Y = D \cup (D + Y),$$

which, by Proposition 4.8, is $\mathsf{Spec}(S(x)) = \mathbb{P} * D$. (Of course, one easily sees that $\mathbb{P} * D$ must be the spectrum, without the help of Proposition 4.8.)

We need a quick lemma, using results from Section 2.

Lemma 4.10. *Suppose* $B \subseteq \mathbb{N}$ *and* $B \cap \mathbb{P} \neq \emptyset$. *Let* $A = \mathbb{P} * B$. *Then* A *is a periodic set. Let* $\mathfrak{c} := \mathfrak{c}(A)$, $\mathfrak{p} := \mathfrak{p}(A)$, $\mathfrak{q} := \mathfrak{q}(A)$. *Then*

(a) $\mathfrak{p} = \mathfrak{q} = \gcd(A) = \gcd(B)$, *and*
(b) $A = A\big|_{<\mathfrak{c}} \cup \big(\mathfrak{c} + \mathfrak{p} \cdot \mathbb{N}\big) \subseteq \mathfrak{p} \cdot \mathbb{N}$.

Proof. Note that $A \cap \mathbb{P} \supseteq B \cap \mathbb{P} \neq \emptyset$, and

$$A + A = (\mathbb{P} + \mathbb{P}) * B \subseteq \mathbb{P} * B = A,$$

thus Lemma 2.13 applies with $(r, s) = (0, 2)$. Since $\gcd(A) = \gcd(B)$, one has the desired conclusions. $\square$

Returning to the postage spectra, by Lemma 4.10, $\mathsf{Spec}(S(x))$ is periodic, $\mathfrak{q} = \mathfrak{p} = \gcd(S) = \gcd(D)$, and $\mathsf{Spec}(S(x)) = \mathsf{Spec}(S(x))\big|_{<\mathfrak{c}} \cup (\mathfrak{c} + \mathfrak{q} \cdot \mathbb{N})$, where $\mathfrak{c} := \mathfrak{c}(\mathsf{Spec}(S(x)))$, $\mathfrak{q} := \mathfrak{q}(\mathsf{Spec}(S(x)))$, $\mathfrak{p} := \mathfrak{p}(\mathsf{Spec}(S(x)))$.[13]

Example 4.11 (Paths in Labelled Digraphs). The objective in this example is to find a set equation system that determines the set of lengths of the paths going from vertex 1 to vertex 4 in the labelled digraph in Fig. 1.

[13] The number $\gamma(D) := \mathfrak{c}(\mathsf{Spec}(S(x)))$ is called the *conductor* of D by Wilf (see [**32**], §3.15.). $\gamma(D) - 1$ is called the *Frobenius number*, and the problem of finding it is called the Frobenius Problem (or Coin Problem). The problem can easily be reduced to the case that $\gcd(D) = 1$, in which case every number $\geq \gamma(D)$ is in $\mathbb{N} * D$, but $\gamma(D) - 1 \notin \mathbb{N} * D$. For D a finite set of positive integers, considerable effort has been devoted to finding a formula for $\gamma(D)$, for D with few elements. The only known closed forms are for D with 1 or 2 elements. For $D = \{b_1, b_2\}$, with 2 coprime elements, both > 1, the solution is $\gamma(D) = (b_1 - 1)(b_2 - 1)$, found by Sylvester in 1884. Finding $\gamma(D)$ is known to be NP-hard. (See Ramírez-Alfonsín [**26**]; Shallit [**27**].)

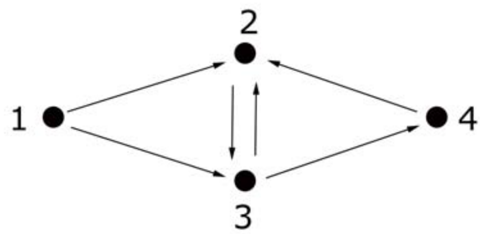


FIGURE 1. A Labelled Digraph

For $1 \le i \le 4$, let $L_i(x) = \sum_{n\ge 1} \ell_i(n)x^n$ be the generating function for the lengths of paths going from vertex i to vertex 4, that is, $\ell_i(n)$ counts the number of paths of length n from vertex i to vertex 4. Then $\mathbf{y} = \mathbf{L}(x)$ satisfies the following system $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$:

$$\begin{array}{rclcrcl} y_1 &=& x\cdot\big(y_2+y_3\big) & \qquad & y_2 &=& x\cdot y_3 \\ y_3 &=& x\cdot\big(y_2+1+y_4\big) & & y_4 &=& x\cdot y_2. \end{array}$$

One has $\mathbf{G}(x, \mathbf{y}) \trianglerighteq 0$ and $\mathbf{G}(0, \mathbf{0}) = J_{\mathbf{G}}(0, \mathbf{0}) = \mathbf{0}$, so the system is elementary. The associated elementary spectral system $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ is:

$$\begin{array}{rclcrcl} Y_1 &=& 1+(Y_2\cup Y_3) & \qquad & Y_2 &=& 1+Y_3 \\ Y_3 &=& \{1\}\cup(1+(Y_2\cup Y_4)) & & Y_4 &=& 1+Y_2. \end{array}$$

To calculate the $\mathfrak{m}_i$ and $\mathfrak{q}_i$ for this system, first

$$\mathbf{G}(\boldsymbol{\varnothing}) = \begin{bmatrix} \varnothing \\ \varnothing \\ \{1\} \\ \varnothing \end{bmatrix} \qquad \mathbf{G}^{(2)}(\boldsymbol{\varnothing}) = \begin{bmatrix} \{2\} \\ \{2\} \\ \{1\} \\ \varnothing \end{bmatrix}$$

$$\mathbf{G}^{(3)}(\boldsymbol{\varnothing}) = \begin{bmatrix} \{2,3\} \\ \{2\} \\ \{1,3\} \\ \{3\} \end{bmatrix} \qquad \mathbf{G}^{(4)}(\boldsymbol{\varnothing}) = \begin{bmatrix} \{2,3,4\} \\ \{2,4\} \\ \{1,3,4\} \\ \{3\} \end{bmatrix},$$

and thus, by (4.1), $\mathfrak{m} = (2, 2, 1, 3)$. For such a simple example, one easily finds the $\mathfrak{m}_i$ by inspection—$\mathfrak{m}_i$ is the length of the shortest path in Fig. 1 from vertex i to vertex 4.

To calculate the $\mathfrak{q}_i$ let

$$S_j := \bigcup_{\mathbf{u}} \big(G_{j,\mathbf{u}} + \mathfrak{m} * \mathbf{u} - \mathfrak{m}_j\big), \quad \text{for } 1 \le j \le 4.$$

Then $S_1 = \{0, 1\}$, $S_2 = \{0\}$, $S_3 = \{0, 2, 3\}$, and $S_4 = \{0\}$. The digraph in Fig. 1 is, conveniently, also the dependency digraph of the system, and $\{2, 3, 4\}$ is a strong component. From (4.2), $\mathfrak{q}_i = \gcd\Big(\bigcup_{i\to^* j} S_j\Big)$, so $\mathfrak{q}_1 = \gcd\big(S_1 \cup S_2 \cup S_3 \cup S_4\big) = \gcd\{0, 1, 2, 3\} = 1$, and $\mathfrak{q}_2 = \mathfrak{q}_3 = \mathfrak{q}_4 = \gcd\big(S_2 \cup S_3 \cup S_4\big) = \gcd\{0, 2, 3\} = 1$.

EXAMPLE 4.12 (Regular Languages). A set $\mathcal{R}$ of words over an m-letter alphabet is a *regular language* if it is precisely the set of words accepted by some finite-state deterministic automaton. A word is accepted by such an automaton if, starting at state 0, one can follow a path to a final state with the labels on the successive edges of the path spelling out the word. Let the states of the automaton be $0, \ldots, k$. We write $i \to j$ if there is an edge in the automaton going from i to j. If $i \to j$, let a_{ij} denote the letter from the alphabet that labels the edge $i \to j$. For each $i \in \{0, \ldots, k\}$ let $\mathcal{U}_i$ be the set of letters a_{ij} for which j is a final state, and let $\mathcal{R}_i$ be the set of words traversed when going from a state i to a final state.

$\mathcal{R}_i$ is the union of $\mathcal{U}_i$ with the classes $a_{ij}\mathcal{R}_j$ for which $i \to j$ is an edge in the automaton. Thus the specification of the $\mathcal{R}_i$ is given by the system of equations

$$\mathcal{R}_i \;=\; \mathcal{U}_i \cup \bigcup_{j\,:\,i\to j} a_{ij}\mathcal{R}_j, \quad 0 \le i \le k.$$

This leads to a system of $\mathbf{y}$-linear equations of a particularly simple form for the generating functions, and for the spectra, namely for $0 \le i \le k$, with $c_i = |\mathcal{U}_i|$,

$$\begin{aligned} R_i(x) &= x \cdot \big(c_i + \sum_{j\,:\,i\to j} R_j(x)\big) \\ \mathsf{Spec}(\mathcal{R}_i) &= A_i \cup \big(1 + \bigcup_{j\,:\,i\to j} \mathsf{Spec}(\mathcal{R}_j)\big), \end{aligned}$$

where $A_i = \emptyset$ if $c_i = 0$, and $A_i = \{1\}$ otherwise.

The theory of the generating functions for regular languages was worked out by Berstel [**5**], 1971 (his results were augmented by Soittola [**30**], 1976). Given a regular language $\mathcal{R}$, one can, as noted above, partition it into classes $\mathcal{R}_i$ such that the generating functions $R_i(x)$ satisfy a system of linear equations $\mathbf{y} = x(C + M \cdot \mathbf{y})$, where C is a column matrix with entries from $\mathbb{N}$, and M is a 0,1-square matrix. The $\mathsf{Spec}(\mathcal{R}_i)$ are eventually periodic, and by Cramer's rule, the generating functions $R_i(x)$ are rational functions; also they are given by $\mathbf{R}(x) = x\big(I - xM\big)^{-1}C$. Berstel showed that each $R_i(x)$ is the sum of a finite number of $R_{ij}(x)$, with each $\mathsf{Spec}(R_{ij}(x))$ being either finite or eventually an arithmetical progression. For those $\mathsf{Spec}(R_{ij}(x))$ which are not finite, there is a polynomial $P_{ij}(x) \ne 0$ and finitely many polynomials $P_{ij\ell}(x)$, a real β_{ij}, and numbers $\beta_{ij\ell}$, with $\beta_{ij} > \max\{|\beta_{ij\ell}|\}$, such that, on the set $\mathsf{Spec}(R_{ij}(x))$, one has the coefficients $r_{ij}(n)$ having an exact polynomial exponential form, and polynomial exponential asymptotics, given by (see *Analytic Combinatorics* [**19**], p. 302):

$$\begin{aligned} r_{ij}(n) &= P_{ij}(n)\beta_{ij}^n + \sum_{\ell} P_{ij\ell}(n)\beta_{ij\ell}^n \quad \text{for } n \in \mathsf{Spec}(R_{ij}(x)) \\ &\sim P_{ij}(n)\beta_{ij}^n \quad \text{on the set } \mathsf{Spec}(R_{ij}(x)). \end{aligned}$$

4.4. Relaxing the conditions on $\mathbf{G}(x, \mathbf{y})$. Recall that a power series system $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ is elementary if (i) $\mathbf{G}(x, \mathbf{y}) \trianglerighteq \mathbf{0}$, (ii) $\mathbf{G}(0, \mathbf{0}) = \mathbf{0}$, and (iii) $J_{\mathbf{G}}(0, \mathbf{0}) = \mathbf{0}$. The 'elementary system' requirement of Theorem 4.6 is usually true for irreducible power series systems $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ arising in combinatorics—see, for example, *Analytic Combinatorics* [**19**], where most of the irreducible examples are such that x is a factor of $\mathbf{G}(x, \mathbf{y})$, a property of $\mathbf{G}(x, \mathbf{y})$ which immediately guarantees that the second and third of the three conditions hold.

Dropping the first requirement, that $\mathbf{G}(x, \mathbf{y}) \trianglerighteq \mathbf{0}$, leads to a difficult area of research, where little is known, even with a single equation $y = G(x, y)$—see the final sections of [**2**] for several remarks on the difficulties mixed signs in $G(x, y)$ pose when trying to determine the asymptotics of the coefficients $t(n)$ of a solution

$y = T(x)$. Such mixed sign situations can arise naturally, for example when dealing with the construction Set, which forms subsets of a given set of objects. The method developed here, for studying the spectra of the solutions $T_i(x)$ of a system $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$, very much depends on $\mathbf{G}(x, \mathbf{y}) \trianglerighteq \mathbf{0}$, in particular, being able to claim that $\mathsf{Spec}\big(\mathbf{G_u}(x) \cdot \mathbf{T}(x)^{\mathbf{u}}\big)$ is equal to $\mathbf{G_u} + \mathbf{u} \circledast \mathbf{T}$. This equality can fail with mixed signs, for example, the spectrum of $G(x) \cdot T(x) = (1 - x) \cdot (x + x^2)$ is not the same as $\mathsf{Spec}(1 - x) + \mathsf{Spec}(x + x^2)$.

The second condition, $\mathbf{G}(0, \mathbf{0}) = \mathbf{0}$, is essential if the solution $\mathbf{T}(x)$ provides generating functions $T_i(x)$ for combinatorial classes $\mathcal{T}_i$ since, in these cases, $\mathsf{Spec}(T_i) \subseteq \mathbb{P}$, so $0 \notin \mathsf{Spec}(T_i)$, for any i. Thus the discussion will be limited to dropping the third requirement, that $J_{\mathbf{G}}(0, \mathbf{0}) = \mathbf{0}$. This simply means that $\mathbf{y}$-linear terms with constant coefficients are permitted to appear in the $G_i(x, \mathbf{y})$, in which case a number of new possibilities can arise when classifying the solutions of such systems:

(a) There may be no (formal power series) solution, for example, $y = x + y$.
(b) There may be a solution, but not $\trianglerighteq \mathbf{0}$, for example, $y = x + 2y$.
(c) There may be infinitely many solutions, for example, $y_1 = y_2$, $y_2 = y_1$.

One can express the system $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ as

$$\mathbf{y} \;=\; \mathbf{G}(x, \mathbf{0}) + J_{\mathbf{G}}(0, \mathbf{0}) \cdot \mathbf{y} + \mathbf{H}(x, \mathbf{y}),$$

where

$$\mathbf{H}(x, \mathbf{y}) \;=\; \sum_{i=1}^{k} y_i \cdot \mathbf{H}_i(x, \mathbf{y}),$$

with each $\mathbf{H}_i(x, \mathbf{y}) \in \mathsf{Dom}_0[x, \mathbf{y}]^k$. The obvious approach to such a system, with $J_{\mathbf{G}}(0, \mathbf{0}) \neq \mathbf{0}$, is to write it in the form

$$\big(I - J_{\mathbf{G}}(0, \mathbf{0})\big) \cdot \mathbf{y} \;=\; \mathbf{G}(x, 0) + \mathbf{H}(x, \mathbf{y})$$

and solve for $\mathbf{y}$.

Definition 4.13 (of $\widehat{\mathbf{G}}(x, \mathbf{y})$). Given $\mathbf{G}(x, \mathbf{y}) \trianglerighteq \mathbf{0}$, with $\mathbf{G}(0, \mathbf{0}) = \mathbf{0}$, if the matrix $I - J_{\mathbf{G}}(0, \mathbf{0})$ has an inverse that is nonnegative then let

$$\widehat{\mathbf{G}}(x, \mathbf{y}) \;:=\; \Big(I - J_{\mathbf{G}}(0, \mathbf{0})\Big)^{-1} \cdot \Big(\mathbf{G}(x, \mathbf{0}) + \mathbf{H}(x, \mathbf{y})\Big).$$

Given a nonnegative square matrix M, let $\Lambda(M)$ denote the largest real eigenvalue of M. (Note: From the Perron-Frobenius theory we know that a nonnegative square matrix M has a nonnegative real eigenvalue; hence there is indeed a largest real eigenvalue $\Lambda(M)$, and it is ≥ 0. The Perron-Frobenius theory also tells us that $\Lambda(M)$ has a nonnegative eigenvector.)

Theorem 4.14. *Let $\mathbf{G}(x, \mathbf{y}) \in \mathbb{R}[[x, \mathbf{y}]]^k$ satisfy the two conditions*

$$\mathbf{G}(x, \mathbf{y}) \trianglerighteq \mathbf{0}, \quad \textit{and } \mathbf{G}(0, \mathbf{0}) = \mathbf{0},$$

that is, $\mathbf{G}(x, \mathbf{y}) \in \mathsf{Dom}_0(x, \mathbf{y})^k$.

(a) *Suppose $I - J_{\mathbf{G}}(0, \mathbf{0})$ has a nonnegative inverse. Then the following hold:*

(i) *The system $\mathbf{y} = \widehat{\mathbf{G}}(x, \mathbf{y})$ is equivalent to the system $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$, that is, they have the same solutions (but not necessarily the same dependency digraph).*

(ii) *$\widehat{\mathbf{G}}(x, \mathbf{y})$ is elementary.*

(iii) *Consequently* $\mathbf{T}(x) := \widehat{\mathbf{G}}^{(\infty)}(x, \mathbf{0})$ *is the unique solution in* $\mathbb{R}[[x]]_0$ *of* $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ *as well as of* $\mathbf{y} = \widehat{\mathbf{G}}(x, \mathbf{y})$. *The periodicity properties of* $\mathbf{T}(x)$ *are as stated in Theorem 4.6, with* $\mathbf{G}$ *replaced by* $\widehat{\mathbf{G}}$.

(b) *Suppose that* $\mathbf{G}^{(k)}(x, \mathbf{0})$ *has all entries nonzero, that is, the associated system* $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ *of set equations is reduced. Then the following are equivalent:*

(i) $I - J_{\mathbf{G}}(0, \mathbf{0})$ *has a nonnegative inverse.*

(ii) *The equation* $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ *has a solution* $\mathbf{T}(x) \in \mathsf{Dom}_0[x]^k$.

(iii) $\Lambda\big(J_{\mathbf{G}}(0, \mathbf{0})\big) < 1$.

Proof. (a): Given that $I - J_{\mathbf{G}}(0, \mathbf{0})$ has a nonnegative inverse, one can transform either of $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ and $\mathbf{y} = \widehat{\mathbf{G}}(x, \mathbf{y})$ into the other by simple operations that preserve solutions. It is routine to check that $\widehat{\mathbf{G}}(x, \mathbf{y})$ is elementary; then use Theorem 4.6.

(b): Assume $\mathbf{G}^{(k)}(x, \mathbf{0})$ has all entries nonzero. (i) $\Rightarrow$ (ii) follows from (a). If (ii) holds then

$$\mathbf{T}(x) \;=\; \mathbf{G}^{(k)}\big(x, \mathbf{T}(x)\big) \;\trianglerighteq\; \mathbf{G}^{(k)}(x, \mathbf{0}).$$

Let $\mathbf{v} \geq 0$ be a left eigenvector of $\Lambda\big(J_{\mathbf{G}^{(k)}}(0, \mathbf{0})\big)$. From

$$\mathbf{T}(x) \;=\; \mathbf{G}^{(k)}(x, \mathbf{0}) \,+\, J_{\mathbf{G}^{(k)}}(0, \mathbf{0}) \cdot \mathbf{T}(x) \,+\, \widetilde{\mathbf{H}}\big(x, \mathbf{T}(x)\big),$$

one has

$$\mathbf{v} \cdot \mathbf{T}(x) \;=\; \mathbf{v} \cdot \mathbf{G}^{(k)}(x, \mathbf{0}) \,+\, \Lambda\big(J_{\mathbf{G}^{(k)}}(0, \mathbf{0})\big) \cdot \mathbf{v} \cdot \mathbf{T}(x) \,+\, \mathbf{v} \cdot \widetilde{\mathbf{H}}\big(x, \mathbf{T}(x)\big). \tag{4.3}$$

Since all $T_i(x)$ are nonnegative and nonzero, one has $\mathbf{v} \cdot \mathbf{T}(x)$ and $\mathbf{v} \cdot \mathbf{G}^{(k)}(x, \mathbf{0}) + \mathbf{v} \cdot \widetilde{\mathbf{H}}\big(x, \mathbf{T}(x)\big)$ are nonzero power series with nonnegative coefficients; consequently (4.3) implies $\Lambda\big(J_{\mathbf{G}^{(k)}}(0, \mathbf{0})\big) < 1$. From $J_{\mathbf{G}^{(k)}}(0, \mathbf{0}) = J_{\mathbf{G}}(0, \mathbf{0})^k$ it follows that $\Big(\Lambda\big(J_{\mathbf{G}}(0, \mathbf{0})\big)\Big)^k$ is an eigenvalue of $J_{\mathbf{G}^{(k)}}(0, \mathbf{0})$, and thus also < 1. This clearly implies $\Lambda\big(J_{\mathbf{G}}(0, \mathbf{0})\big) < 1$, so (ii) $\Rightarrow$ (iii).

If (iii) holds, then by Neumann's expansion theorem (see [**22**], p. 201), one knows that $I - J_{\mathbf{G}}(0, \mathbf{0})$ has an inverse, and $\big(I - J_{\mathbf{G}}(0, \mathbf{0})\big)^{-1} = \sum_{n \geq 0} J_{\mathbf{G}}(0, \mathbf{0})^n$, a nonnegative matrix. Thus (iii) $\Rightarrow$ (i). $\square$

The condition that $\mathbf{G}^{(k)}(x, \mathbf{0})$ has all entries nonzero is usual for power series systems in combinatorics since the $T_i(x)$ in the solution of $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ are generating functions for nonempty classes $\mathcal{T}_i$.

A somewhat tedious calculation shows that one can use the formulas (4.1) and (4.2) of Theorem 4.6 to calculate the $\mathfrak{m}_i$ and $\mathfrak{q}_i$ with the original system $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ as well as with the derived system $\mathbf{y} = \widehat{\mathbf{G}}(x, \mathbf{y})$. It can be useful to note that if the two hypotheses of Theorem 4.14 hold, then the condition that $\mathbf{G}^{(k)}(x, \mathbf{0})$ has all entries nonzero is equivalent to requiring that $\mathbf{G}^{(j)}(x, \mathbf{0})$ has all entries nonzero, for some j, $1 \leq j \leq k$.

Remark 4.15. The uniqueness of the solution $\mathbf{T}(x)$ in $\mathbb{R}[[x]]_0^k$, for a power series system $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ satisfying the two hypotheses of Theorem 4.14 and the hypothesis of part (a), does not in general extend to the associated spectral system $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ when $J_{\mathbf{G}}(0, \mathbf{0}) \neq \mathbf{0}$. For example, consider the consistent single equation system $y = G(x, y)$, where $G(x, y) = x^2 + (1/2)y + xy$. The spectral system $Y = G(Y)$ is $Y = \{2\} \cup Y \cup (1 + Y)$, which has three solutions: $\mathbb{N}$, $1 + \mathbb{N}$, and $2+\mathbb{N}$. On the other hand, the elementary system $y = \widehat{G}(x, y)$ is $y = 2x^2 + 2xy$; its spectral system is $Y = \{2\} \cup (1 + Y)$, which has the unique solution $2 + \mathbb{N}$.

4.5. An example of the tools.

The following simple example uses all the tools developed so far.

EXAMPLE 4.16. Consider the class $\mathcal{T}$ of planar trees with blue and red nodes, defined by the conditions:

(i) Every blue node that is not a leaf has exactly three nodes immediately below it, but not all of the same color.
(ii) Every red node that is not a leaf has exactly two nodes immediately below it.

Let $\mathcal{B}$ be the collection of trees in $\mathcal{T}$ with blue roots, and $\mathcal{R}$ the collection of those with red roots. Then, letting $\bullet_B$ be a blue node and $\bullet_R$ a red node, one has the equational specification (see §5)

$$\begin{aligned}
\mathcal{B} &= \{\bullet_B\} \cup \frac{\bullet_B}{\mathcal{B}+\mathcal{R}+\mathcal{R}} \cup \frac{\bullet_B}{\mathcal{R}+\mathcal{B}+\mathcal{R}} \cup \frac{\bullet_B}{\mathcal{R}+\mathcal{R}+\mathcal{B}} \cup \\
&\qquad \frac{\bullet_B}{\mathcal{B}+\mathcal{B}+\mathcal{R}} \cup \frac{\bullet_B}{\mathcal{B}+\mathcal{R}+\mathcal{B}} \cup \frac{\bullet_B}{\mathcal{R}+\mathcal{B}+\mathcal{B}} \\
\mathcal{R} &= \{\bullet_R\} \cup \frac{\bullet_R}{\mathcal{T}+\mathcal{T}} \\
\mathcal{T} &= \mathcal{B} \cup \mathcal{R}.
\end{aligned}$$

The three generating functions, $B(x)$ for $\mathcal{B}$, $R(x)$ for $\mathcal{R}$, and $T(x)$ for $\mathcal{T}$, are related by the system of equations:

$$\begin{aligned}
B(x) &= x + 3x \cdot B(x) \cdot R(x)^2 + 3x \cdot B(x)^2 \cdot R(x) \\
R(x) &= x + x \cdot T(x)^2 \\
T(x) &= B(x) + R(x).
\end{aligned}$$

Thus $(B(x), R(x), T(x))$ gives a solution (y_1, y_2, y_3) for the system of polynomial equations:

$$\begin{aligned}
y_1 &= x + 3x \cdot y_1 \cdot y_2^2 + 3x \cdot y_1^2 \cdot y_2 \\
y_2 &= x + x \cdot y_3^2 \\
y_3 &= y_1 + y_2.
\end{aligned}$$

The spectra $\mathsf{Spec}(\mathcal{B}), \mathsf{Spec}(\mathcal{R}), \mathsf{Spec}(\mathcal{T})$ are related by the set equations

$$\begin{aligned}
\mathsf{Spec}(\mathcal{B}) &= \{1\} \cup (1 + \mathsf{Spec}(\mathcal{B}) + 2 * \mathsf{Spec}(\mathcal{R})) \cup (1 + 2 * \mathsf{Spec}(\mathcal{B}) + \mathsf{Spec}(\mathcal{R})) \\
\mathsf{Spec}(\mathcal{R}) &= \{1\} \cup (1 + 2 * \mathsf{Spec}(\mathcal{T})) \\
\mathsf{Spec}(\mathcal{T}) &= \mathsf{Spec}(\mathcal{B}) \cup \mathsf{Spec}(\mathcal{R}),
\end{aligned}$$

so $(\mathsf{Spec}(\mathcal{B}), \mathsf{Spec}(\mathcal{R}), \mathsf{Spec}(\mathcal{T}))$ is a solution to the system of set equations

$$\begin{aligned}
Y_1 &= \{1\} \cup (1 + Y_1 + 2 * Y_2) \cup (1 + 2 * Y_1 + Y_2) \\
Y_2 &= \{1\} \cup (1 + 2 * Y_3) \\
Y_3 &= Y_1 \cup Y_2.
\end{aligned}$$

Next,

$$\mathbf{G}(x, y_1, y_2, y_3) = \begin{bmatrix} x + 3x \cdot y_1 \cdot y_2^2 + 3x \cdot y_1^2 \cdot y_2 \\ x + x \cdot y_3^2 \\ y_1 + y_2 \end{bmatrix},$$

so

$$\mathbf{G}^{(2)}(x,0,0,0) \;=\; \begin{bmatrix} 6x^4 + x \\ x \\ 2x \end{bmatrix}$$

has all entries nonzero—this implies $\mathbf{G}^{(3)}(x,\mathbf{0})$ has all entries nonzero.

The Jacobian matrix $J_{\mathbf{G}}(x,\mathbf{y})$ is

$$J_{\mathbf{G}}(x,y_1,y_2,y_3) \;=\; \begin{bmatrix} 3xy_2^2 + 6xy_1y_2 & 6xy_1y_2 + 3xy_1^2 & 0 \\ 0 & 0 & 2xy_3 \\ 1 & 1 & 0 \end{bmatrix}$$

so

$$J_{\mathbf{G}}(x,0,0,0) \;=\; \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}.$$

The eigenvalues of $J_{\mathbf{G}}(0,0,0,0)$ are the roots of $\det\big(\lambda I - J_{\mathbf{G}}(0,0,0,0)\big) = 0$, that is, $\lambda^3 = 0$. Thus $\Lambda\big(J_{\mathbf{G}}(0,0,0,0)\big) = 0 < 1$, so the system $\mathbf{y} = \mathbf{G}(x,\mathbf{y})$ has a solution $\mathbf{T}(x) \in \mathbb{R}[[x]]_0^3$, and the solution has all entries nonzero. The inverse of $I - J_{\mathbf{G}}(0,0,0,0)$ is a nonnegative matrix:

$$\big(I - J_{\mathbf{G}}(0,0,0,0)\big)^{-1} \;=\; \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

Thus

$$\widehat{\mathbf{G}}(x,\mathbf{y}) \;=\; \begin{bmatrix} x + 3xy_1y_2^2 + 3xy_1^2y_2 \\ x + xy_3^2 \\ 2x + 3xy_1y_2^2 + 3xy_1^2y_2 + xy_3^2 \end{bmatrix},$$

so $\mathbf{y} = \widehat{\mathbf{G}}(x,\mathbf{y})$ is an irreducible reduced non $\mathbf{y}$-linear elementary system. Thus Corollary 4.7 applies.

The spectral system $\mathbf{Y} = \widehat{\mathbf{G}}(\mathbf{Y})$ is

$$\begin{aligned} Y_1 &= \{1\} \cup \big(1 + Y_1 + 2 * Y_2\big) \cup \big(1 + 2 * Y_1 + Y_2\big) \\ Y_2 &= \{1\} \cup \big(1 + 2 * Y_3\big) \\ Y_3 &= \{1\} \cup \big(1 + Y_1 + 2 * Y_2\big) \cup \big(1 + 2 * Y_1 + Y_2\big) \cup \big(1 + 2 * Y_3\big). \end{aligned}$$

One readily calculates that $\mathfrak{m}_i = \mathfrak{p}_i = \mathfrak{q}_i = 1$, for $1 \le i \le 3$.

5. Constructions, Operators and Equational Specifications

The general theory of setting up equational specifications $\boldsymbol{\mathcal{Y}} = \boldsymbol{\Gamma}(\boldsymbol{\mathcal{Y}})$ for combinatorial classes $\mathcal{A}_1,\ldots,\mathcal{A}_k$ is developed by forming the $\Gamma_i(\boldsymbol{\mathcal{Y}})$ from compositions of basic constructions, and then translating the specifications into systems of equations

(a) $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$, for the spectra $\mathsf{Spec}(\mathcal{A}_1),\ldots,\mathsf{Spec}(\mathcal{A}_k)$, and
(b) $\mathbf{y} = \mathbf{G}(x,\mathbf{y})$, for the generating functions $A_1(x),\ldots,A_k(x)$,

There are two important approaches to constructions: one is to have constructions that may be influenced by the nature of the objects in combinatorial classes, and may have a limited range of application; the other is to only have constructions that apply to all combinatorial classes. §5.1–§5.2 look at the first approach, based on Compton's papers [**9**], [**10**], from the late 1980s. The second approach, following *Analytic Combinatorics* [**19**], is sketched in §5.3.

5.1. Constructions for relational classes: $\cup, +, \mathsf{MSet}$.

Many combinatorial classes of interest are classes of relational structures such as digraphs, functional digraphs, graphs, posets, forests or planar forests.[14] The count function for a class of relational structures counts up to isomorphism. A class $\mathcal{A}$ of relational structures will always be assumed to belong to a fixed finite relational language, for example, $\mathcal{A}$ could be a class of binary relational structures (also known as a class of digraphs).

For $\mathcal{A}$ a class of relational structures, a *k-ary construction Θ for $\mathcal{A}$* is a mapping $\Theta : \mathrm{Su}(\mathcal{A})^k \to \mathrm{Su}(\mathcal{A})$ from k-tuples of subclasses of $\mathcal{A}$ to subclasses of $\mathcal{A}$. The following gives three constructions that apply to any class of relational structures that belong to a given relational language. Recall that $A(x) \trianglelefteq B(x)$ means $a(n) \le b(n)$ for all n.

(a) *Union* $(\cup)$

Given two classes $\mathcal{A}, \mathcal{B}$ of relational structures, the *union* construction is exactly as in set theory. One has:

$$\begin{array}{rcl} (\mathcal{A} \cup \mathcal{B})(x) & \trianglelefteq & A(x) + B(x) \\ \mathsf{Spec}(\mathcal{A} \cup \mathcal{B}) & = & \mathsf{Spec}(\mathcal{A}) \cup \mathsf{Spec}(\mathcal{B}). \end{array} \tag{5.1}$$

One has equality in (5.1) iff $\mathcal{A}$ and $\mathcal{B}$ are (up to isomorphism) disjoint.

(b) *Relational Sum* $(+)$

$\mathfrak{a} + \mathfrak{b}$ is defined for a pair of relational structures $\mathfrak{a}, \mathfrak{b}$ by first replacing $\mathfrak{a}$ and $\mathfrak{b}$ by isomorphic structures $\mathfrak{a}'$ and $\mathfrak{b}'$ whose universes are disjoint; then define $\mathfrak{a} + \mathfrak{b}$ to be $\mathfrak{a}' + \mathfrak{b}'$, the relational structure whose universe is the union of the universes of $\mathfrak{a}'$ and $\mathfrak{b}'$, and for each relational symbol R, one has $R_{\mathfrak{a}'+\mathfrak{b}'} := R_{\mathfrak{a}'} \cup R_{\mathfrak{b}'}$. Clearly $\|\mathfrak{a} + \mathfrak{b}\| = \|\mathfrak{a}\| + \|\mathfrak{b}\|$.

A useful defining property of $\mathfrak{a} + \mathfrak{b}$ is as follows. For any connected structure $\mathfrak{c}$, let $\nu_{\mathfrak{c}}(\mathfrak{a})$ be the number of components of $\mathfrak{a}$ that are isomorphic to $\mathfrak{c}$. Then $\nu_{\mathfrak{c}}(\mathfrak{a} + \mathfrak{b}) = \nu_{\mathfrak{c}}(\mathfrak{a}) + \nu_{\mathfrak{c}}(\mathfrak{b})$, for all $\mathfrak{c}$.

One has the following for the generating function and spectrum:

$$\begin{array}{rcl} (\mathcal{A} + \mathcal{B})(x) & \trianglelefteq & A(x) \cdot B(x) \\ \mathsf{Spec}(\mathcal{A} + \mathcal{B}) & = & \mathsf{Spec}(\mathcal{A}) + \mathsf{Spec}(\mathcal{B}). \end{array} \tag{5.2}$$

One has equality in (5.2) iff

(*) given any $\mathfrak{c} \in \mathcal{A}+\mathcal{B}$, there are, up to isomorphism, unique $\mathfrak{a} \in \mathcal{A}$, $\mathfrak{b} \in \mathcal{B}$ such that $\mathfrak{c} = \mathfrak{a} + \mathfrak{b}$.

(c) *Relational Multiset* (MSet)

For $\mathcal{A}$ a class of relational structures let

$$\mathsf{MSet}_M(\mathcal{A}) \quad := \quad \bigcup_{m \in M} \underbrace{\mathcal{A} + \cdots + \mathcal{A}}_{m}.$$

Then

$$\begin{array}{rcl} \mathsf{MSet}_M(\mathcal{A})(x) & \trianglelefteq & \mathsf{MSet}_M(A(x)) \\ & := & \displaystyle\sum_{m \in M} \sigma_m\big(\mathfrak{s}_m, A(x), \ldots, A(x^m)\big) \\ \mathsf{Spec}\big(\mathsf{MSet}_M(\mathcal{A})\big) & = & M * \mathsf{Spec}(\mathcal{A}), \end{array} \tag{5.3}$$

[14]Each of these classes, except planar forests, is closed under sum and the extraction of components, and thus can be viewed as an *additive number system* (see [**7**]).

where $\sigma_m(\mathfrak{s}_m, y_1, \ldots, y_m)$ is the Pólya cycle index polynomial on m variables, for the symmetric group $\mathfrak{s}_m$.

If $\mathcal{A}$ is a class of connected structures then one has equality in (5.3).

REMARK 5.1. MSet_M is written simply as MSet when $M = \mathbb{P}$. Let Θ be a unary construction that acts on classes $\mathcal{A}$ by performing constructions on finitely many objects from $\mathcal{A}$. Given a nonempty set M of positive integers, let $\Theta_M(\mathcal{A})$ be the class of all objects that one can construct by applying Θ to m-many objects from $\mathcal{A}$, for $m \in M$. (Repeats are allowed, and they are counted with multiplicity when determining if the count is in M). Θ_M is called a *restriction* of Θ.

5.2. Specifications for m-colored forests: $\bullet_i/, \cup, +, \mathsf{MSet}$.

For an illustration of how one derives systems of equations for spectra from equational specifications of classes of relational structures, we choose the setting of m-colored forests.

Let $\bullet_i$ be the 1-element forest with color i, for $1 \le i \le m$. Then

$$\begin{aligned} \{\bullet_i\}(x) &= x \\ \mathsf{Spec}(\{\bullet_i\}) &= \{1\}. \end{aligned}$$

In addition to the three constructions in § 5.1, there is the construction $\bullet_i/$, which adds a root with color i to a forest. Clearly $\| \bullet_i /F\| = 1 + \|F\|$. Extend this construction to classes $\mathcal{F}$ of m-colored forests by $\bullet_i/\mathcal{F} = \{\bullet/F : F \in \mathcal{F}\}$. Then

$$\begin{aligned} \big(\bullet_i /\mathcal{F}\big)(x) &= x \cdot F(x) \\ \mathsf{Spec}\big(\bullet_i /\mathcal{F}\big) &= 1 + \mathsf{Spec}(\mathcal{F}). \end{aligned}$$

Given a k-tuple $\boldsymbol{\mathcal{F}}$ of classes of m-colored forests, a *specification* for $\boldsymbol{\mathcal{F}}$ is a system of equations $\boldsymbol{\mathcal{Y}} = \boldsymbol{\Gamma}(\boldsymbol{\mathcal{Y}})$, where each $\Gamma_i(\boldsymbol{\mathcal{Y}})$ is a composition of the basic constructions for m-colored forests, and $\boldsymbol{\mathcal{F}}$ is the minimum solution $\boldsymbol{\Gamma}^{(\infty)}(\boldsymbol{\emptyset})$ of this system. There is a routine method to translate the specification into a system $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ of equations for the spectra.

5.3. The standard constructions of *Analytic Combinatorics*. The set calculus that has been developed in this article can also be applied to specifications $\boldsymbol{\mathcal{Y}} = \boldsymbol{\Gamma}(\boldsymbol{\mathcal{Y}})$, where the Γ_i are composed from the class $\mathcal{Z}$, consisting of a single object of size 1, and the following *standard constructions* found in *Analytic Combinatorics* [**19**]:[15]

(a) *Disjoint Union* $(\sqcup)$[16] Given any two combinatorial classes $\mathcal{A}$ and $\mathcal{B}$, let $\mathcal{A}', \mathcal{B}'$ be isomorphic copies that are disjoint. Let

$$\mathcal{A} \sqcup \mathcal{B} \;:=\; \mathcal{A}' \cup \mathcal{B}'.$$

(b) *Combinatorial Product* $(\times)$ Given any two combinatorial classes $\mathcal{A}$ and $\mathcal{B}$, let $\mathcal{A} \times \mathcal{B}$ be the Cartesian product $\mathcal{A} \times \mathcal{B}$, with $\|(\mathfrak{a}, \mathfrak{b})\| = \|\mathfrak{a}\| + \|\mathfrak{b}\|$.

(c) *Combinatorial Multiset* (MSET) Given any combinatorial class $\mathcal{A}$ and nonempty $M \subseteq \mathbb{P}$, let $\mathrm{MSET}_M(\mathcal{A})$ be the class of formal unordered sums $\mathfrak{a}_1 + \cdots + \mathfrak{a}_m$ of m objects from $\mathcal{A}$, for $m \in M$, with size equal to $\|\mathfrak{a}_1\| + \cdots + \|\mathfrak{a}_m\|$.

[15] The constructions of the previous sections have standard analogs: the analog of $\{\bullet\}$ is $\mathcal{Z}$, of $\cup$ is $\sqcup$, of $+$ is $\times$, and of MSet is MSET.

[16] The *disjoint union* construction in [**19**] uses *sum* for its primary name, and the symbol used is $+$ instead of $\sqcup$.

(d) *Combinatorial Sequence* (Seq) Given any combinatorial class $\mathcal{A}$ and nonempty $M \subseteq \mathbb{P}$, let $\mathrm{Seq}_M(\mathcal{A})$ be the class of formal ordered sums $\mathfrak{a}_1 + \cdots + \mathfrak{a}_m$ of m objects from $\mathcal{A}$, for $m \in M$, with size equal to $\|\mathfrak{a}_1\| + \cdots + \|\mathfrak{a}_m\|$.
(e) *Combinatorial Cycle* (Cyc) Given a combinatorial class $\mathcal{A}$ and a nonempty set $M \subseteq \mathbb{P}$, $\mathrm{Cyc}_M(\mathcal{A})$ is the class of formal cyclic arrangements of m members of $\mathcal{A}$, where $m \in M$, with the size of the cycle being the sum of the sizes of the members.
(f) *Combinatorial Directed Cycle* (DCyc) Given a combinatorial class $\mathcal{A}$ and a nonempty set $M \subseteq \mathbb{P}$, $\mathrm{DCyc}_M(\mathcal{A})$ is the class of formal directed cyclic arrangements of m members of $\mathcal{A}$, where $m \in M$, with the size of the directed cycle being the sum of the sizes of the members.

These constructions are not relational constructions (as defined in §5.1). For example, letting $\mathcal{D}$ be the class of digraphs one has $(\mathcal{D} \sqcup \mathcal{D})(x) = 2D(x)$, so $\mathcal{D} \sqcup \mathcal{D}$ has twice as many objects of each size as $\mathcal{D}$. Thus $\mathcal{D} \sqcup \mathcal{D}$ cannot be viewed as a class of digraphs.

Nonetheless, given a specification $\boldsymbol{\mathcal{Y}} = \mathbf{\Gamma}(\boldsymbol{\mathcal{Y}})$, where $\mathbf{\Gamma}(\boldsymbol{\mathcal{Y}})$ is composed of relational constructions as in the previous sections, if one lets $\widehat{\mathbf{\Gamma}}(\boldsymbol{\mathcal{Y}})$ be the expression obtained by replacing the relational constructions in $\mathbf{\Gamma}(\boldsymbol{\mathcal{Y}})$ by their standard analogs (see Footnote 15), then one usually finds that *the combinatorial classes specified by* $\boldsymbol{\mathcal{Y}} = \mathbf{\Gamma}(\boldsymbol{\mathcal{Y}})$ *have the same generating functions and spectra as that of the classes specified by* $\boldsymbol{\mathcal{Y}} = \widehat{\mathbf{\Gamma}}(\boldsymbol{\mathcal{Y}})$.[17]

The standard constructions have two important properties which make the analysis of classes defined by specifications proceed smoothly in *Analytic Combinatorics*:

(1) Standard constructions $\Theta(\mathbf{Y})$ are *total*, that is, they are defined for all arguments $\boldsymbol{\mathcal{A}}$ of combinatorial classes.
(2) Standard constructions $\Theta(\mathbf{Y})$ are *admissible*, that is, they are total, and furthermore, given combinatorial classes $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ with $\boldsymbol{\mathcal{A}}(x) = \boldsymbol{\mathcal{B}}(x)$, then
$$\Theta(\boldsymbol{\mathcal{A}})(x) \;=\; \Theta(\boldsymbol{\mathcal{B}})(x).$$

To this list we can add the following:

(3) Standard constructions $\Theta(\mathbf{Y})$ are *spectrally admissible*, that is, given combinatorial classes $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ with $\mathsf{Spec}(\boldsymbol{\mathcal{A}}) = \mathsf{Spec}(\boldsymbol{\mathcal{B}})$, then
$$\mathsf{Spec}(\Theta(\boldsymbol{\mathcal{A}})) \;=\; \mathsf{Spec}(\Theta(\boldsymbol{\mathcal{B}})).$$

Each admissible k-ary construction Θ induces a k-ary operator on $\mathbb{N}[[x]]_0$, which we also call Θ, defined as follows: For $\mathbf{A}(x) \in \mathbb{N}[[x]]_0{}^k$, let $\Theta(\mathbf{A}(x))$ be $\Theta(\boldsymbol{\mathcal{A}})(x)$, where $\boldsymbol{\mathcal{A}}$ is any combinatorial class with $\boldsymbol{\mathcal{A}}(x) = \mathbf{A}(x)$. If Θ is a standard construction then the induced operator is called a *standard* operator. For the standard operators discussed in this section, one finds the following formulas in [**19**]:

Lemma 5.2. *For combinatorial classes $\mathcal{A}$, $\mathcal{B}$ one has:*

(a) Disjoint Union ($\sqcup$)
$$A(x) \sqcup B(x) \;=\; A(x) + B(x),$$
thus $\mathsf{Spec}(\mathcal{A} \sqcup \mathcal{B}) = \mathsf{Spec}(\mathcal{A}) \cup \mathsf{Spec}(\mathcal{B})$.

[17]The conditions, under which these two specifications lead to the same generating functions and spectra, do not seem to be detailed in the literature. For simple examples, like the specification for trees, or planar trees, the reader will readily see that the italicized claim holds.

(b) Product ($\times$)

$$A(x) \times B(x) \ = \ A(x) \cdot B(x),$$

thus $\mathsf{Spec}(\mathcal{A} \times \mathcal{B}) = \mathsf{Spec}(\mathcal{A}) + \mathsf{Spec}(\mathcal{B})$.

(c) Combinatorial Multiset (MSet)

$$\mathrm{MSET}_M(A(x)) \ = \ \sum_{m \in M} \sigma_m(\mathfrak{s}_m, A(x), \ldots, A(x^m)),$$

where $\sigma_m(\mathfrak{s}_m, y_1, \ldots, y_m)$ *is the Pólya cycle index polynomial on* m *variables for the symmetric group* $\mathfrak{s}_m$. *Thus* $\mathsf{Spec}\big(\mathrm{MSET}_M(\mathcal{A})\big) = M * \mathsf{Spec}(\mathcal{A})$.

(d) Combinatorial Sequence (Seq)

$$\mathrm{SEQ}_M(A(x)) \ = \ \sum_{m \in M} A(x)^m,$$

thus $\mathsf{Spec}\big(\mathrm{SEQ}_M(\mathcal{A})\big) = M * \mathsf{Spec}(\mathcal{A})$.

(e) Combinatorial Cycle (Cyc)

$$\mathrm{CYC}_M(A(x)) \ = \ \sum_{m \in M} \sigma_m(\mathfrak{d}_m, A(x), \ldots, A(x^m)),$$

where $\sigma_m(\mathfrak{d}_m, y_1, \ldots, y_m)$ *is the Pólya cycle index polynomial on* m *variables for the dihedral group* $\mathfrak{d}_m$ *of order* $2m$. *Thus* $\mathsf{Spec}\big(\mathrm{CYC}_M(\mathcal{A})\big) = M * \mathsf{Spec}(\mathcal{A})$.

(f) Combinatorial Directed Cycle (DCyc)

$$\mathrm{DCYC}_M(\mathcal{A}(x)) \ = \ \sum_{m \in M} \sigma_m(\mathfrak{c}_m, A(x), \ldots, A(x^m)),$$

where $\sigma_m(\mathfrak{c}_m, y_1, \ldots, y_m)$ *is the Pólya cycle index polynomial on* m *variables for the cyclic group* $\mathfrak{c}_m$ *of order* m. *Thus* $\mathsf{Spec}\big(\mathrm{DCYC}_M(\mathcal{A})\big) = M * \mathsf{Spec}(\mathcal{A})$.

Recall that

$$\begin{aligned} \mathsf{Dom}_0[x] &= \big\{A(x) \in \mathbb{R}[[x]] : A(0) = 0, A(x) \trianglerighteq 0\big\} \\ \mathsf{Dom}_0[x, \mathbf{y}] &= \big\{G(x, \mathbf{y}) \in \mathbb{R}[[x, \mathbf{y}]] : G(0, \mathbf{0}) = 0, G(x, \mathbf{y}) \trianglerighteq 0\big\} \\ \mathsf{Dom}_{J0}[x, \mathbf{y}] &= \{\mathbf{G}(x, \mathbf{y}) \in \mathsf{Dom}_0[x, \mathbf{y}]^k : J_{\mathbf{G}}(0, \mathbf{0}) = \mathbf{0}\}. \end{aligned}$$

The $A(x), B(x)$ in Lemma 5.2 range over the power series in $\mathbb{N}[[x]]_0$. After noting that the right sides of the equations in (a)–(f) of Lemma 5.2 are defined for all $A(x), B(x) \in \mathsf{Dom}_0[x]$, we will assume that these formulas define the standard operators on $\mathsf{Dom}_0[x]$.

Definition 5.3. A k-ary operator Θ on $\mathsf{Dom}_0[x]$ is *spectrally admissible* if for $\mathbf{A}(x), \mathbf{B}(x) \in \mathsf{Dom}_0[x]^k$ one has

$$\mathsf{Spec}(\mathbf{A}(x)) = \mathsf{Spec}(\mathbf{B}(x)) \ \Rightarrow \ \mathsf{Spec}(\Theta(\mathbf{A}(x))) = \mathsf{Spec}(\Theta(\mathbf{B}(x))).$$

An operator $\mathbf{\Theta}$ on $\mathsf{Dom}_0[x]^k$ is *spectrally admissible* if each Θ_i is spectrally admissible.

Lemma 5.4. *Each* $\mathbf{G}(x, \mathbf{y}) \in \mathsf{Dom}_{J0}[x, \mathbf{y}]^k$ *defines an operator on* $\mathsf{Dom}_0[x]^k$, *namely* $\mathbf{A}(x) \mapsto \mathbf{G}(x, \mathbf{A}(x))$, *that is spectrally admissible. Such operators are called* elementary operators.[18] *As a spectrally admissible operator,* $\mathbf{G}(x, \mathbf{y})$ *induces a set*

[18] Many natural operators are not elementary operators. For example, the operator $\Theta : A(x) \mapsto A(x^2)$ is not elementary. There is no power series $G(x, y) \in \mathsf{Dom}[x, y]$ such that $A(x^2) = \sum_n G_n(x) A(x)^n$ for $A(x) \in \mathsf{Dom}_0[x]$. Likewise, the operator $\Theta : A(x) \mapsto A^{(2)}(x)$ is not elementary.

operator on $\mathrm{Su}(\mathbb{P})^k$*, namely*

$$\mathbf{G} : \mathbf{A} \mapsto \bigvee_{\mathbf{u}\in\mathbb{N}^k} \big(\mathbf{G}_{\mathbf{u}} + \mathbf{u} \circledast \mathbf{A}\big).$$

THEOREM 5.5 (Systems based on Spectrally Admissible Operators).

(a) *Elementary operators* $\mathbf{G}(x,\mathbf{y})$ *and standard operators* Θ *are spectrally admissible.*

(b) *Each standard unary operator* Θ_M *is spectrally equivalent to the elementary operator* $\sum_{j\in M} y^j$*, and* $\mathsf{Spec}\big(\Theta_M\big(A(x)\big)\big) = M * \mathsf{Spec}(A(x))$*.*

(c) *The sum* $\Theta_1 + \Theta_2$*, product* $\Theta_1 \cdot \Theta_2$ *and composition* $\Theta_1 \circ \Theta_2$ *of spectrally admissible operators are spectrally admissible.*

(d) *Any combination of elementary operators and standard operators—using the operations of sum, product and composition—yields an operator that is spectrally admissible and spectrally equivalent to an elementary operator.*

(e) *Suppose* $\mathbf{\Theta}$ *and* $\mathbf{\Theta}'$ *have components* Θ_i, Θ_i' *that are combinations as described in* (d)*. If* $\mathbf{\Theta}$ *and* $\mathbf{\Theta}'$ *are spectrally equivalent then*

$$\mathsf{Spec}\big(\mathbf{\Theta}^{(\infty)}(\mathbf{0})\big) \;=\; \mathsf{Spec}\big(\mathbf{\Theta}'^{(\infty)}(\mathbf{0})\big).$$

(f) *Let* $\mathbf{y} = \mathbf{\Theta}(\mathbf{y})$ *be a system with solution* $\mathbf{T}(x) \in \mathsf{Dom}_0[x]^k$*, where the operators* Θ_i *are combinations as described in item* (d)*. By* (d)*,* $\mathbf{\Theta}$ *is spectrally equivalent to an elementary operator* $\mathbf{G}(x,\mathbf{y})$*. Let* $\mathbf{U}(x)$ *be the unique solution to* $\mathbf{y} = \mathbf{G}(x,\mathbf{y})$ *guaranteed by Theorem 4.6. Then* $\mathsf{Spec}(\mathbf{T}(x)) = \mathsf{Spec}(\mathbf{U}(x))$*.*

Thus periodicity properties for the $T_i(x)$ *can be deduced by applying Theorem 4.6 to* $\mathbf{y} = \mathbf{G}(x,\mathbf{y})$*.*

PROOF. Items (a) through (e) are straightforward. For item (f), one has

$$\mathsf{Spec}(\mathbf{T}(x)) \;=\; \mathsf{Spec}\big(\mathbf{\Theta}\big(\mathbf{T}(x)\big)\big) \;=\; \mathsf{Spec}\big(\mathbf{G}\big(x,\mathbf{T}(x)\big)\big) \;=\; \mathbf{G}(\mathsf{Spec}(\mathbf{T}(x))),$$

where $\mathbf{G}$ is the set operator corresponding to $\mathbf{G}(x,\mathbf{y})$. So $\mathsf{Spec}(\mathbf{T}(x))$ is a solution of $\mathbf{Y} = \mathbf{G}\big(\mathbf{Y}\big)$. Now $\mathbf{U}(x) = \mathbf{G}(\mathbf{U}(x))$ implies that $\mathsf{Spec}(\mathbf{U}(x))$ is also a solution of $\mathbf{Y} = \mathbf{G}\big(\mathbf{Y}\big)$. Theorem 3.11 says that the elementary system $\mathbf{Y} = \mathbf{G}(\mathbf{Y})$ has a unique solution, so $\mathsf{Spec}(\mathbf{U}(x)) = \mathsf{Spec}(\mathbf{T}(x))$. Consequently the periodicity properties of $\mathsf{Spec}(\mathbf{T}(x))$ are those of $\mathsf{Spec}(\mathbf{U}(x))$, and thus Theorem 4.6 can be used to analyze $\mathsf{Spec}(\mathbf{T}(x))$. □

6. Monadic Second Order Classes

When working with relational structures like graphs and trees, logicians have been able to strengthen some first order logic results to *monadic second order logic* (MSO logic) results.[19] A primary reason for the success with MSO logic is the powerful connection between Ehrenfeucht-Fraïssé games and sentences of quantifier depth at most q.[20] These games, although very combinatorial in nature, have not

[19]This is first order logic augmented with unary predicates U as variables— this means that one can quantify over subsets as well as individual elements, and say that an element belongs to a subset. The fact that the U are predicates and not domain elements make the logic *second order*, and the fact that these predicates have only one argument (for example, $U(x)$) makes the logic *monadic*.

[20]The connection with Ehrenfeucht-Fraïssé games fails if one has quantification over more general relations, like binary relations.

been widely used in the combinatorics community. (For the following discussion, one can find the background material needed on MSO logic and Ehrenfeucht-Fraïssé games in Chapter VI of [**7**].)

6.1. MSO classes of m-colored chains.

One of the early success stories in the study of MSO classes of structures was Büchi's analysis of MSO classes of m-colored chains, connecting them to the regular languages discussed in Example 4.12.

THEOREM 6.1 (Büchi [**6**], 1960). *MSO classes of colored chains are precisely the regular languages.*

For more detail on this result, the precursor to Compton's Specification Theorem in §6.3, see Appendix B. Combining this with the results stated in Example 4.12 for regular languages, one immediately sees that MSO classes of m-colored chains are eventually periodic, their generating functions are determined by systems of particularly simple $\mathbf{y}$-linear equations, and there is Berstel's penetrating analysis of the nature of the coefficients of the generating functions.

6.2. Trees and forests.

Compton extended parts of Büchi's analysis of m-colored chains to the setting of m-colored trees, to show that MSO classes of m-colored trees have an equational specification.

When speaking of structures, in particular the models of a sentence φ, it will be understood that *only finite structures are being considered.* All results stated for trees and forests can easily be extended to trees and forests with finitely many unary predicates, or, if one prefers, to finitely colored trees and forests.

A *tree* $\mathfrak{t} = (T, <)$ is a poset such that: (i) there is a unique maximal element r called the *root* of the tree, and (ii) every interval $[t, r]$ is linear. A *forest* $\mathfrak{f} = (F, <)$ is a poset whose components are trees. Let TREES be the class of (finite) trees, and let $\mathsf{FORESTS}$ be the class of (finite) forests.

The class $\mathsf{FORESTS}$ is defined by the following first order sentence φ_F of quantifier depth 3:

$$\begin{aligned}
&(\forall x)[\neg(x < x)] \\
&\wedge (\forall x)(\forall y)[(x < y) \rightarrow \neg(y < x)] \\
&\wedge (\forall x)(\forall y)(\forall z)[(x < y) \ \wedge \ (y < z) \ \rightarrow \ (x < z)] \\
&\wedge (\forall x)(\forall y)(\forall z)[(x < y) \ \wedge \ (x < z) \ \rightarrow \ (y < z) \vee (y = z) \vee (z < y)]
\end{aligned}$$

The class TREES is defined by the following first order sentence φ_T, also of quantifier depth 3:

$$\varphi_F \wedge (\exists x)(\forall y)[(y = x) \vee (y < x)].$$

A forest $\mathfrak{f}$ is determined (up to isomorphism) by the number of each (isomorphism type of) tree appearing in it, thus by its component counting function $\nu_{\mathfrak{f}} : \mathsf{TREES} \rightarrow \mathbb{N}$. One can combine two forests $\mathfrak{f}_1$ and $\mathfrak{f}_2$ into a single forest $\mathfrak{f}_1 + \mathfrak{f}_2$ which is determined up to isomorphism by $\nu_{\mathfrak{f}} = \nu_{\mathfrak{f}_1} + \nu_{\mathfrak{f}_2}$. Extend this operation to classes $\mathcal{F}$ of forests by $\mathcal{F}_1 + \mathcal{F}_2 = \{\mathfrak{f}_1 + \mathfrak{f}_2 : \mathfrak{f}_i \in \mathcal{F}_i\}$. The *ideal class* $\mathcal{O}$ of forests is introduced with the properties $\mathcal{O} \cup \mathcal{F} = \mathcal{O} + \mathcal{F} = \mathcal{F}$ (it is introduced solely as a notational device to smooth out the presentation).

Define the operation $*$ between nonempty subsets A of $\mathbb{N}$ and nonempty classes $\mathcal{F}$ of forests by

$$\begin{aligned} n * \mathcal{F} &= \begin{cases} \mathcal{O} & \text{if } n = 0 \\ \underbrace{\mathcal{F} + \cdots + \mathcal{F}}_{n-\text{fold}} & \text{if } n \geq 1 \end{cases} \\ A * \mathcal{F} &= \bigcup_{a \in A} (a * \mathcal{F}). \end{aligned}$$

$A * \mathcal{F}$ is just a minor extension of the construction $\mathsf{MSet}_A(\mathcal{F})$.

6.3. Compton's specification of MSO classes of trees.

Definition 6.2. Given a positive integer q, a MSO_q sentence φ is a MSO sentence of quantifier depth q, and a MSO_q class is a class defined by a MSO_q sentence.[21] $\mathfrak{f} \equiv_q^{\mathrm{MSO}} \mathfrak{f}'$ means that $\mathfrak{f}$ and $\mathfrak{f}'$ satisfy the same MSO_q sentences.

The binary relation $\equiv_q^{\mathrm{MSO}}$ is an equivalence relation of finite index on $\mathsf{FORESTS}$. Let $\mathcal{F}_1, \ldots, \mathcal{F}_\ell$ be the $\equiv_q^{\mathrm{MSO}}$-classes of $\mathsf{FORESTS}$, the equivalence classes of forests under the relation $\equiv_q^{\mathrm{MSO}}$. These classes are MSO_q classes—the $\mathcal{F}_i$ are the minimal nonempty MSO_q classes of forests—and every MSO_q class $\mathcal{F}$ of forests is a union of some of the $\mathcal{F}_i$, namely the $\mathcal{F}_i \subseteq \mathcal{F}$. Two forests satisfy the same MSO_q sentences iff they are in the same $\mathcal{F}_i$.

In the following, when given a MSO_q class $\mathcal{F}$ of forests, it will be assumed that $q \geq 3$, so that one can use the MSO_q sentence φ_F, resp. φ_T, to express "is a forest", resp. "is a tree". Consequently, if one member of an $\mathcal{F}_i$ is a tree, then so is every member of $\mathcal{F}_i$. Let $\mathcal{T}_1, \ldots, \mathcal{T}_k$ be the $\mathcal{F}_i$ whose members are trees. Each $\mathcal{T}_i$ is a MSO_q class of trees—the $\mathcal{T}_i$ are the minimal nonempty MSO_q classes of trees—and every MSO_q class $\mathcal{T}$ of trees is a union of some of the $\mathcal{T}_i$, namely the $\mathcal{T}_i \subseteq \mathcal{T}$.

If $\mathcal{T}_i$ has a 1-element tree in it then, up to isomorphism, no other tree is in $\mathcal{T}_i$ (since every member of $\mathcal{T}_i$ will satisfy $(\forall x)(\forall y)(x = y)$). Let $\bullet$ denote the 1-element tree, and assume $\mathcal{T}_1 = \{\bullet\}$. This is the only $\mathcal{T}_i$ with a 1-element member.

Given a tree $\mathfrak{t}$ with more than one element, let $\partial\mathfrak{t}$ be the forest that results from removing the root from $\mathfrak{t}$; and for $\mathcal{F}$ a class of forests let

$$\partial\mathcal{F} \quad := \quad \big\{\partial\mathfrak{t} : \mathfrak{t} \in \mathcal{F}, \|\mathfrak{t}\| \geq 2\big\}.$$

Recall from §5.2 that, given any forest $\mathfrak{f}$, $\bullet/\mathfrak{f}$ is the tree that results by adding a root to the forest; and for $\mathcal{F}$ a class of forests, $\bullet/\mathcal{F} := \{\bullet/\mathfrak{f} : \mathfrak{f} \in \mathcal{F}\}$.

Lemma 6.3. *Let q be a positive integer ≥ 3.*

(a) *In the class $\mathsf{FORESTS}$, the operations of addition and adding a root preserve $\equiv_q^{\mathsf{MSO}}$, that is,*

$$\mathfrak{f}_i \equiv_q^{\mathsf{MSO}} \mathfrak{f}_i' \ \Rightarrow \ \sum_i \mathfrak{f}_i \equiv_q^{\mathsf{MSO}} \sum_i \mathfrak{f}_i', \text{ and}$$
$$\mathfrak{f} \equiv_q^{\mathsf{MSO}} \mathfrak{f}' \ \Rightarrow \ \bullet/\mathfrak{f} \equiv_q^{\mathsf{MSO}} \bullet/\mathfrak{f}'.$$

(b) *There is a constant C_q such that, for all trees $\mathfrak{t}$ and all $n \geq C_q$, one has $n * \mathfrak{t} \equiv_q^{\mathsf{MSO}} C_q * \mathfrak{t}$.*

(c) *The relation $\equiv_q^{\mathsf{MSO}}$ is decidable.*

[21] Since every MSO sentence of quantifier depth $\leq q$ is logically equivalent to one of quantifier depth $\mathfrak{q}$, any class defined by a MSO sentence of quantifier depth $\leq q$ is a MSO_q class.

Proof. One can find a discussion of the first item of (a) in [**7**], based on E-F games. Use E-F games for (b) and (c) as well. (a)–(c) are basic tools of Gurevich and Shelah ([**20**], 2003). □

Definition 6.4. With C_q as in Lemma 6.3(b), let $\mathcal{S}_q$ be the collection of k-tuples $\mathbf{S} := (S_1, \ldots, S_k)$ such that each S_j is either a singleton $\{n_j\}$, with $0 \le n_j < C_q$, or the cofinite set $\mathbb{N}|_{\ge C_q}$. For $\mathbf{S} \in \mathcal{S}_q$, define the class of forests $\mathcal{F}_{\mathbf{S}}$ by

$$\mathcal{F}_{\mathbf{S}} \;:=\; \sum_{j=1}^{k} \big(S_j * \mathcal{T}_j\big).$$

Clearly there are $(1 + C_q)^k$ distinct choices for $\mathbf{S} \in \mathcal{S}_q$, every forest is in some $\mathcal{F}_{\mathbf{S}}$, and given two distinct members $\mathbf{S}$ and $\mathbf{S}'$ of $\mathcal{S}_q$, the classes $\mathcal{F}_{\mathbf{S}}$ and $\mathcal{F}_{\mathbf{S}'}$ are disjoint.

The next lemma gives the crucial structure result for MSO_q classes of forests.

Lemma 6.5.

(a) *For $\mathbf{S} \in \mathcal{S}_q$ there is an $r \in \{1, \ldots, \ell\}$ such that $\mathcal{F}_{\mathbf{S}} \subseteq \mathcal{F}_r$, that is, all forests in $\mathcal{F}_{\mathbf{S}}$ are equivalent modulo $\equiv_q^{\mathrm{MSO}}$.*

(b) *Let $\mathcal{F}$ be a MSO_q class of forests. Then there is an $\mathbb{S} \subseteq \mathcal{S}_q$ such that*

$$\mathcal{F} \;=\; \bigcup_{\mathbf{S} \in \mathbb{S}} \mathcal{F}_{\mathbf{S}}.$$

Proof. For (a), let $\mathbf{n}$ and $\mathbf{n}'$ be two k-tuples of nonnegative integers satisfying the condition $n_j \ne n_j'$ implies $n_j, n_j' \ge C_q$, for $1 \le j \le k$. Then Lemma 6.3 shows that every forest in $\sum_{j=1}^{k} \big(n_j * \mathcal{T}_j\big)$ is equivalent modulo $\equiv_q^{\mathrm{MSO}}$ to every forest in $\sum_{j=1}^{k} \big(n_j' * \mathcal{T}_j\big)$. Thus, given $\mathbf{S} \in \mathcal{S}_q$, all members of $\mathcal{F}_{\mathbf{S}}$ are equivalent modulo $\equiv_q^{\mathrm{MSO}}$, and thus they are in some $\equiv_q^{\mathrm{MSO}}$ class $\mathcal{F}_r$.

For (b), since $\mathcal{F}$ is an $\equiv_q^{\mathrm{MSO}}$ class of forests, for each $\mathbf{S} \in \mathcal{S}_q$ one has either $\mathcal{F}_{\mathbf{S}} \subseteq \mathcal{F}$ or $\mathcal{F}_{\mathbf{S}} \cap \mathcal{F} = \emptyset$. Consequently $\mathcal{F}$ is a union of some of the $\mathcal{F}_{\mathbf{S}}$. □

Lemma 6.6. *For $2 \le i \le k$, the class of forests $\partial\mathcal{T}_i$ is definable by a MSO_q sentence.*

Proof. Suppose $\mathfrak{f} \in \partial\mathcal{T}_i$ and $\mathfrak{f} \equiv_q^{\mathrm{MSO}} \mathfrak{f}'$. By Lemma 6.3(a), $\bullet/\mathfrak{f} \equiv_q^{\mathrm{MSO}} \bullet/\mathfrak{f}'$. Since $\bullet/\mathfrak{f} \in \mathcal{T}_i$, and $\mathcal{T}_i$ is a $\equiv_q^{\mathrm{MSO}}$ class, it follows that $\bullet/\mathfrak{f}' \in \mathcal{T}_i$, so $\partial\mathcal{T}_i$ is closed under $\equiv_q^{\mathrm{MSO}}$, proving that $\partial\mathcal{T}_i$ is a MSO_q class. □

Theorem 6.7 (Compton, see [**33**]). *Let $\mathcal{T}$ be a class of trees defined by a MSO_q sentence. Then:*

(a) *$\mathcal{T}$ is a union of some of the $\mathcal{T}_i$, and*

(b) *the $\mathcal{T}_i$ satisfy a system of equations*

$$\Sigma_q : \left\{ \begin{array}{rcl} \mathcal{T}_1 & = & \Phi_1\big(\mathcal{T}_1, \ldots, \mathcal{T}_k\big) \\ & \vdots & \\ \mathcal{T}_k & = & \Phi_k\big(\mathcal{T}_1, \ldots, \mathcal{T}_k\big), \end{array} \right.$$

where $\Phi_1\big(\mathcal{T}_1, \ldots, \mathcal{T}_k\big)$ is $\{\bullet\}$, and, for $2 \le i \le k$, there is an $\mathbb{S}_i \subseteq \mathcal{S}_q$ such that

$$\Phi_i\big(\mathcal{T}_1, \ldots, \mathcal{T}_k\big) \;:=\; \bullet\Big/ \bigcup_{\mathbf{S} \in \mathbb{S}_i} \mathcal{F}_{\mathbf{S}} \;=\; \bullet\Big/ \bigcup_{\mathbf{S} \in \mathbb{S}_i} \sum_{j=1}^{k} \big(S_j * \mathcal{T}_j\big). \tag{6.1}$$

PROOF. (a) is obviously true. For (b) note that for $2 \le i \le k$, $\mathcal{T}_i = \bullet/\partial\mathcal{T}_i$. Lemma 6.6 says $\partial\mathcal{T}_i$ is definable by a MSO_q sentence. Then Lemma 6.5(b) shows that $\partial\mathcal{T}_i$ can be expressed as $\bigcup_{\mathbf{S}\in\mathbb{S}_i} \mathcal{F}_{\mathbf{S}}$, for a suitable $\mathbb{S}_i \subseteq \mathcal{S}_q$. One only needs to attach the root $\bullet$ to have (6.1). $\square$

Compton's specification leads to a system of equations that define generating functions.

COROLLARY 6.8. *For $\mathcal{T}$ as in Compton's Theorem, one has:*

$$T(x) = \sum_{\mathcal{T}_i\subseteq\mathcal{T}} T_i(x) \tag{6.2}$$

$$T_i(x) = \begin{cases} x & \text{for } i=1 \\ x\cdot \displaystyle\sum_{\mathbf{S}\in\mathbb{S}_i}\sum_{u_1\in S_1}\cdots\sum_{u_k\in S_k} \mathbf{T}(x)^{\mathbf{u}} & \text{for } 2\le i\le k, \end{cases} \tag{6.3}$$

where $\mathbf{T}(x)^{\mathbf{u}} := T_1(x)^{u_1}\cdots T_k(x)^{u_k}$.

Applying Spec to Σ_q gives a system of set equations for the spectra.

COROLLARY 6.9. *For $\mathcal{T}$ as in Compton's Theorem,*

$$\mathsf{Spec}(\mathcal{T}) = \bigcup_{\mathcal{T}_i\subseteq\mathcal{T}} \mathsf{Spec}(\mathcal{T}_i) \tag{6.4}$$

$$\mathsf{Spec}(\mathcal{T}_i) = \begin{cases} \{1\} & \text{for } i=1 \\ 1+\displaystyle\bigcup_{\mathbf{S}\in\mathbb{S}_i}\sum_{j=1}^{k}\big(S_j * \mathsf{Spec}(\mathcal{T}_j)\big) & \text{for } 2\le i\le k. \end{cases} \tag{6.5}$$

REMARK 6.10. Compton [**11**] described his equational specification for the minimal classes $\mathcal{T}_1,\ldots,\mathcal{T}_k$ of trees defined by MSO_q sentences to Alan Woods, during a visit to Yale in 1986; at the time Woods was a PostDoc at Yale. A careful analysis of the Jacobian of the system (6.3) by Woods [**33**] (1997) led to a proof that the class of (m-colored) trees has a MSO limit law. (It is not a 0–1 law.)

REMARK 6.11. To handle m-colored trees and forests, let $\bullet_i$ be a 1-element tree with the color i, for $1 \le i \le m$. Let $\mathcal{T}_1,\ldots,\mathcal{T}_k$ be the $\equiv_q^{\mathrm{MSO}}$ classes of m-colored trees, where $\mathcal{T}_i = \{\bullet_i\}$, for $1 \le i \le m$. Note that the roots of members of any class $\mathcal{T}_i$, $1 \le i \le k$, all have the same color, say i'. Thus for $m < i \le k$, $\mathcal{T}_i = \bullet_{i'}/\partial\mathcal{T}_i$. With this, the development of Compton's results for m-colored trees and forests proceeds exactly as before.

6.4. The spectrum of a MSO class of trees is eventually periodic.

The dependency digraph D_q for Σ_q is defined parallel to the definition for systems of set equations. D_q has vertices $1,\ldots,k$ and, referring to (6.5), directed edges given by $i \to j$ iff there is a $\mathbf{S} \in \mathbb{S}_i$ such that $S_j \ne \{0\}$. One defines a *height function* on D_q by setting $h(1) = 0$, and then, for $2 \le i \le k$, use the inductive definition $h(i) = 1 + \max\{h(j) : i \to^+ j, \text{ but not } j \to^+ i\}$.

COROLLARY 6.12. *The spectrum of a MSO class $\mathcal{T}$ of trees is eventually periodic.*

PROOF. It suffices to prove this result for the $\equiv_q^{\mathrm{MSO}}$ classes $\mathcal{T}_i$, $1 \le i \le k$, in view of Lemma 2.6 (which guarantees that eventual periodicity is preserved by finite union). $\mathsf{Spec}(\mathcal{T}_1) = \{1\}$ is eventually periodic. So suppose $2 \le i \le k$. Note that $i \to j$ implies $\mathsf{Spec}(\mathcal{T}_i) \supseteq p_{ij} + \mathsf{Spec}(\mathcal{T}_j)$ for some positive integer p_{ij}, by (6.5). Thus $i \to^+ j$ implies the same conclusion.

If $[i] \ne \emptyset$ then $i \to^+ i$, so $\mathsf{Spec}(\mathcal{T}_i) \supseteq p{+}\mathsf{Spec}(\mathcal{T}_i)$, for some $p \in \mathbb{P}$. Consequently $\mathsf{Spec}(\mathcal{T}_i)$ is actually periodic.

If $[i] = \emptyset$ then one argues, by induction on the height $h(i)$, that $\mathsf{Spec}(\mathcal{T}_i)$ is eventually periodic. The ground case, $h(i) = 0$, holds precisely for $i = 1$, a case discussed above. Now suppose the result holds for all $\mathcal{T}_i$ with $h(i) \le n$. If $h(i) = n + 1$ then $2 \le i \le k$, and one has

$$\mathsf{Spec}(\mathcal{T}_i) \quad = \quad 1 + \bigcup_{\mathbf{S} \in \mathbb{S}_i} \sum_{j=1}^{k} \big(S_j * \mathsf{Spec}(\mathcal{T}_j)\big).$$

For the j such that there is an $\mathbf{S}$ with $S_j \ne \{0\}$ (there is at least one such j since $i > 1$) one has $i \to j$, so $h(j) < h(i)$, implying $\mathsf{Spec}(\mathcal{T}_j)$ is eventually periodic (by the induction hypothesis). The S_j are either singleton or cofinite subsets of $\mathbb{N}$, and therefore eventually periodic. Then Lemma 2.6 shows $\mathsf{Spec}(\mathcal{T}_i)$ is eventually periodic, since being eventually periodic is preserved by finite unions, (finite) sums, and $*$, proving the result. We have the additional information that those $\mathsf{Spec}(\mathcal{T}_i)$ with i belonging to a strong component of the dependency digraph are actually periodic. □

Corollary 6.13. *The spectrum of a MSO class $\mathcal{F}$ of forests is eventually periodic.*

Proof. Since $\bullet/\mathcal{F}$ is a MSO class of trees, one has $\{1\} + \mathsf{Spec}(\mathcal{F})$ eventually periodic; hence so is $\mathsf{Spec}(\mathcal{F})$. □

One can view unary functions as multisets of directed cycles of trees. Since forests are posets, they carry the structure of a digraph with up directed edges; hence they are partial unary functions. In order to complete such digraphs to unary functions, one only needs to add to each tree in the forest an edge directed from the root of the tree to a member of the tree. Recall that a monounary algebra is an algebra (A, f), where f is a unary function on A.

Theorem 6.14 (Gurevich and Shelah [**20**], 2003). *Let $\mathcal{M}$ be a MSO class of monounary algebras. Then the spectrum $\mathsf{Spec}(\mathcal{M})$ is eventually periodic.*

Proof. It suffices to show that one can find a MSO class of forests with the same spectrum. Let $\mathcal{F}$ be the class of forests defined as follows:

> for each forest in the class there exists a subset V of the forest, with exactly one element from each tree in the forest, such that if one adds a directed edge from the root of each tree in the forest to the unique node of the tree in V, then one has a functional digraph which satisfies a sentence defining $\mathcal{M}$.

Clearly this condition can be expressed by a MSO sentence, so $\mathsf{Spec}(\mathcal{F})$ is eventually periodic; hence so is $\mathsf{Spec}(\mathcal{M})$. □

This theorem is almost best possible for MSO classes—for example, one cannot replace 'monounary algebra' with 'digraph' or 'graph' as one can easily find classes of such structures where the theorem fails to hold. For example, the class $\mathcal{G}$ of finite graphs that look like rectangular grids is a MSO class, and since $\mathsf{Spec}(\mathcal{G})$ is the set of composite numbers, by Remark 2.5 we see that this is not an eventually periodic set.

The converse, that every eventually periodic set $S \subseteq \mathbb{P}$ can be realized as the spectrum of a MSO sentence for monounary algebras, is easy to prove. Although the proof presented here of the Gurevich and Shelah Theorem comes after considerable development of the theory of spectra defined by equations, actually all that is needed for this proof, beyond Compton's Specification Theorem, is Lemma 2.6.

In a related direction, one has the following:

COROLLARY 6.15. *A MSO class of graphs with bounded defect has an eventually periodic spectrum.*

PROOF. A connected graph has defect d if $d+1$ is the minimum number of edges that need to be removed in order to have an acyclic graph. Thus trees have defect $= -1$. A graph has defect d if the maximum defect of its components is d. For graphs of defect at most d, introduce $d+2$ colors, one to mark a choice of a root in each component, and the others to mark the endpoints of edges which, when removed, convert the graph into a forest. For a MSO class of m-colored graphs of defect at at most d, carrying out this additional coloring in all possible ways gives a MSO class of $m+d+2$ colored graphs. Then removing the marked edges from each graph converts this into a MSO class of $m+d+2$ colored forests with the same spectrum. $\square$

This can be easily generalized further to MSO classes of digraphs with bounded defect, giving a slight generalization of the Gurevich-Shelah result (since trees have defect -1, functional digraphs have defect 0). These examples suffice to indicate the value of knowing that MSO classes of trees have eventually periodic spectra. The method of showing that MSO spectra are eventually periodic, by reducing them to the spectra of MSO classes of trees, has been successfully pursued by Fischer and Makowsky in [**18**] (2004)—they prove that a MSO class that is contained in a class of bounded patch-width has an eventually periodic spectrum. In the same year Shelah [**29**] proved that MSO classes having a certain recursive constructibility property have eventually periodic spectra, and in 2007 Doron and Shelah [**13**] showed that the bounded patch-width result is a consequence of the constructibility property.

6.5. Effective tree procedures.

What follows is a program, given q, to effectively find a value for C_q and representatives of the $\equiv_q^{\mathsf{MSO}}$ classes of TREES and of $\mathsf{FORESTS}$, with applications to the decidability results mentioned in Gurevich and Shelah ([**20**], 2003), and an effective procedure to construct Compton's system of equations for trees. The particular classes of trees constructed in the WHILE loop of this program are similar to the classes $\mathcal{T}_k^m$ used in 1990 by Compton and Henson (see [**12**], p. 38) to prove lower bounds on computational complexity.

Program Steps	Comments
$\mathsf{FindReps}:= \mathrm{PROC}(q)$	q is the quantifier depth
$\widehat{\mathcal{F}}_0 := \emptyset$	Initialize collection of forests
$\widehat{\mathcal{T}}_0 := \{\bullet\}$	Initialize collection of trees
$t(0) := 1$	cardinality of $\widehat{\mathcal{T}}_0/\equiv_q^{\mathsf{MSO}}$
$d(0) := 1$	initialize $d(n)$
$n := 0$	initialize n

WHILE $t(n-1) < t(n)$ OR $d(n) > 0$ DO
$n := n+1$ — augment the value of n
$\widehat{\mathcal{F}}_n := \Big\{ \sum_{\mathfrak{t}\in\widehat{\mathcal{T}}_{n-1}} m_{\mathfrak{t}} * \mathfrak{t} : m_{\mathfrak{t}} \le n \Big\}$ — make forests using at most n copies of each tree in $\widehat{\mathcal{T}}_{n-1}$
$\widehat{\mathcal{T}}_n := \widehat{\mathcal{T}}_{n-1} \cup \bullet/\widehat{\mathcal{F}}_n$
$t(n) := \big|\widehat{\mathcal{T}}_n/\equiv_q^{\mathsf{MSO}}\big|$, — # of $\equiv_q^{\mathsf{MSO}}$ classes represented by $\widehat{\mathcal{T}}_n$
$d(n) := \big|\big\{\mathfrak{t}\in\widehat{\mathcal{T}}_{n-1} : (n+1)*\mathfrak{t} \not\equiv_q n*\mathfrak{t}\big\}\big|$ — # of failures of $(n+1)*\mathfrak{t} \equiv_q^{\mathsf{MSO}} n*\mathfrak{t}$, for $\mathfrak{t}\in\widehat{\mathcal{T}}_{n-1}$
END WHILE
Define $N := n$.
Choose a maximal set $\mathsf{REP}_{\mathsf{TREES}} := \{\mathfrak{t}_1,\ldots,\mathfrak{t}_k\}$ of $\equiv_q^{\mathsf{MSO}}$ distinct trees $\mathfrak{t}_i$ from $\widehat{\mathcal{T}}_{N-1}$.
Choose a maximal set $\mathsf{REP}_{\mathsf{FORESTS}} := \{\mathfrak{f}_1,\ldots,\mathfrak{f}_\ell\}$ of $\equiv_q^{\mathsf{MSO}}$ distinct forests $\mathfrak{f}_j$ from $\widehat{\mathcal{F}}_{N-1}$.
RETURN $(N, \mathsf{REP}_{\mathsf{TREES}}, \mathsf{REP}_{\mathsf{FORESTS}})$
END PROC

THEOREM 6.16. *The procedure* $\mathsf{FindReps}(q)$ *halts for all* $q\in\mathbb{N}$*, giving an effective procedure to find a set* $\mathsf{REP}_{\mathsf{TREES}}$ *of representatives for the* $\equiv_q^{\mathsf{MSO}}$ *equivalence classes of finite trees, a set* $\mathsf{REP}_{\mathsf{FORESTS}}$ *of representatives for the* $\equiv_q^{\mathsf{MSO}}$ *equivalence classes of finite forests, and a number* N *such that, for any finite tree* $\mathfrak{t}$ *and* $n \ge N$*, one has* $n*\mathfrak{t} \equiv_q^{\mathsf{MSO}} (N-1)*\mathfrak{t}$.

PROOF. Define $\widehat{\mathcal{F}}_n$ and $\widehat{\mathcal{T}}_n$ recursively for all $n\in\mathbb{N}$ by changing the WHILE loop in the program into an unconditional DO loop. Then the classes $\widehat{\mathcal{F}}_n$ and $\widehat{\mathcal{T}_n}$ are finite and monotone nondecreasing; furthermore every finite forest is in some $\widehat{\mathcal{F}}_n$, and every finite tree is in some $\widehat{\mathcal{T}}_n$.

From Lemma 6.3(b), one knows that eventually $d(n) = 0$ (indeed, for $n \ge C_q$). Also, since $\equiv_q^{\mathsf{MSO}}$ has finite index, and the $\mathcal{T}_n$ are monotone nondecreasing, for n sufficiently large, one has $t(n-1) = t(n)$. Thus the program will eventually exit the WHILE loop, returning the triple $(N, \mathsf{REP}_{\mathsf{TREES}}, \mathsf{REP}_{\mathsf{FORESTS}})$.

From the definition of N, one has $t(N-1) = t(N)$ and $d(N) = 0$. Thus

$$\Big(\widehat{\mathcal{T}}_{N-1}/\equiv_q^{\mathsf{MSO}}\Big) = \Big(\widehat{\mathcal{T}}_N/\equiv_q^{\mathsf{MSO}}\Big)$$
$$(N+1)*\mathfrak{t} \equiv_q^{\mathsf{MSO}} N*\mathfrak{t} \quad \text{for } \mathfrak{t}\in\widehat{\mathcal{T}}_{N-1}.$$

We will prove, by induction, for $n \ge N$,

$$t(n-1) = t(n),\ d(n) = 0 \text{ and } \Big(\widehat{\mathcal{F}}_{n-1}/\equiv_q^{\mathsf{MSO}}\Big) = \Big(\widehat{\mathcal{F}}_n/\equiv_q^{\mathsf{MSO}}\Big). \tag{6.6}$$

We have already noted that it is true for $n = N$. Suppose this is true for some $n \ge N$. Then

$$\Big(\widehat{\mathcal{T}}_{n-1}/\equiv_q^{\mathsf{MSO}}\Big) = \Big(\widehat{\mathcal{T}}_n/\equiv_q^{\mathsf{MSO}}\Big) \tag{6.7}$$

and

$$(n+1)*\mathfrak{t} \equiv_q^{\mathsf{MSO}} n*\mathfrak{t} \quad \text{for } \mathfrak{t}\in\widehat{\mathcal{T}}_{n-1}. \tag{6.8}$$

Then

$$\begin{aligned}
\widehat{\mathcal{F}}_{n+1}/\equiv_q^{\mathsf{MSO}} &:= \Big\{\sum_{\mathfrak{t}\in\widehat{\mathcal{T}}_n}\big(m_{\mathfrak{t}}*\mathfrak{t}\big) : m_{\mathfrak{t}}\le n+1\Big\}\Big/\equiv_q\\
&\subseteq \Big\{\sum_{\mathfrak{t}\in\widehat{\mathcal{T}}_{n-1}}\big(m_{\mathfrak{t}}*\mathfrak{t}\big) : m_{\mathfrak{t}}\in\mathbb{N}\Big\}\Big/\equiv_q \quad \text{by (6.7)}\\
&= \Big\{\sum_{\mathfrak{t}\in\widehat{\mathcal{T}}_{n-1}}\big(m_{\mathfrak{t}}*\mathfrak{t}\big) : m_{\mathfrak{t}}\le n\Big\}\Big/\equiv_q \quad \text{by (6.8)}\\
&= \widehat{\mathcal{F}}_n/\equiv_q^{\mathsf{MSO}},
\end{aligned}$$

so $\big(\widehat{\mathcal{F}}_{n+1}/\equiv_q^{\mathsf{MSO}}\big) = \big(\widehat{\mathcal{F}}_n/\equiv_q^{\mathsf{MSO}}\big)$.

Next,

$$\begin{aligned}
\widehat{\mathcal{T}}_{n+1}/\equiv_q^{\mathsf{MSO}} &= \big(\bullet/\widehat{\mathcal{F}}_{n+1}\big)\Big/\equiv_q^{\mathsf{MSO}}\\
&= \big(\bullet/\widehat{\mathcal{F}}_{n}\big)\Big/\equiv_q^{\mathsf{MSO}} \quad \text{by Lemma 6.3(a)}\\
&= \widehat{\mathcal{T}}_n/\equiv_q^{\mathsf{MSO}},
\end{aligned}$$

so $t(n) = t(n+1)$.

Given $\mathfrak{t}\in\widehat{\mathcal{T}}_{n+1}$, choose $\mathfrak{t}'\in\widehat{\mathcal{T}}_n$ such that $\mathfrak{t}\equiv_q^{\mathsf{MSO}}\mathfrak{t}'$. Then

$$(n+2)*\mathfrak{t} \;\equiv_q^{\mathsf{MSO}}\; (n+2)*\mathfrak{t}' \;\equiv_q^{\mathsf{MSO}}\; (n+1)*\mathfrak{t}' \;\equiv_q^{\mathsf{MSO}}\; (n+1)*\mathfrak{t},$$

so $d(n+1) = 0$. This finishes the proof of (6.6).

Thus for $n\ge N$,

$$\begin{aligned}
\Big(\widehat{\mathcal{T}}_n/\equiv_q^{\mathsf{MSO}}\Big) &= \Big(\widehat{\mathcal{T}}_{N-1}/\equiv_q^{\mathsf{MSO}}\Big)\\
n*\mathfrak{t} &\equiv_q^{\mathsf{MSO}} (N-1)*\mathfrak{t}, \text{ for } \mathfrak{t}\in\widehat{\mathcal{T}}_n\\
\Big(\widehat{\mathcal{F}}_n/\equiv_q^{\mathsf{MSO}}\Big) &= \Big(\widehat{\mathcal{F}}_{N-1}/\equiv_q^{\mathsf{MSO}}\Big).
\end{aligned}$$

Consequently $\widehat{\mathcal{T}}_{N-1}$ has representatives for all $\equiv_q^{\mathsf{MSO}}$ classes of finite trees, $n*\mathfrak{t}\equiv_q^{\mathsf{MSO}}(N-1)*\mathfrak{t}$, for any finite tree $\mathfrak{t}$, and $\widehat{\mathcal{F}}_{N-1}$ has representatives for all $\equiv_q^{\mathsf{MSO}}$ classes of finite forests. Thus one can choose $C_q = N-1$.

The procedures for constructing the classes $\widehat{\mathcal{F}}_n$ and $\widehat{\mathcal{T}}_n$ are effective, as are the calculations of the functions $t(n)$ and $d(n)$. □

Further Conclusions (all can be extended to m-colorings, or adding m unary predicates to the language):

(a) The trees in $\widehat{\mathcal{T}}_n$ are all of height $\le n$.

(b) One can effectively find MSO_q sentences φ_i, $1\le i\le k$, such that φ_i defines $\mathfrak{t}_i/\equiv_q^{\mathsf{MSO}}$, the $\equiv_q^{\mathsf{MSO}}$ equivalence class of finite trees with the representative $\mathfrak{t}_i$ in it.

[For $1\le i\le k$, start enumerating the MSO_q sentences φ and test each one in turn until one finds one such that $\mathfrak{t}_j\models\varphi$ iff $j=i$. Then let $\varphi_i := \varphi$.]

(c) Likewise for $1\le j\le \ell$ one can effectively find MSO_q sentences ψ_j defining $\mathfrak{f}_j/\equiv_q^{\mathsf{MSO}}$, the $\equiv_q^{\mathsf{MSO}}$ equivalence class of finite forests with $\mathfrak{f}_j$ in it.

(d) The MSO theory of FORESTS is decidable.[22] (Given ψ of quantifier depth q, it will be true of all finite forests iff it is true of each $\mathfrak{f}_j$ in $\mathsf{REP}_{\mathsf{FORESTS}}$.)

(e) Finite satisfiability for the MSO theory of monounary algebras is decidable. (This can be proved directly, by interpretation into FORESTS.)

(f) One can effectively find the Compton Equations Σ_q for the $\equiv_q^{\mathsf{MSO}}$ equivalence classes of finite trees, namely, with $[\mathfrak{t}_i]_q := \mathfrak{t}_i / \equiv_q^{\mathsf{MSO}}$, one has

$$\begin{aligned}
[\mathfrak{t}_i]_q &= \{\bullet\} \quad \text{if } \mathfrak{t}_i = \{\bullet\}; \text{ otherwise} \\
[\mathfrak{t}_i]_q &= \bigcup \Big\{ \bullet \Big/ \sum_{j=1}^{k} \big(G_j * [\mathfrak{t}_j]_q\big) \;:\; G_j \in \big\{\{0\},\{1\},\ldots,\{N-2\},\mathbb{N}|_{\geq N-1}\big\} \text{ and} \\
&\qquad\qquad \mathfrak{t}_i \equiv_q^{\mathsf{MSO}} \bullet \Big/ \sum_{j=1}^{k} \big(G_j * \mathfrak{t}_j\big) \Big\}.
\end{aligned}$$

To test the last condition (concerning $\equiv_q^{\mathsf{MSO}}$) one replaces any $G_j = \mathbb{N}|_{\geq N-1}$ by $N-1$, so one is deciding $\equiv_q^{\mathsf{MSO}}$ between two finite trees.

(g) One can effectively find the dependency digraph of Σ_q (immediate from the previous step).

(h) One can effectively find the periodicity parameters $\mathfrak{m}$, $\mathfrak{q}$ for the spectra of the $[\mathfrak{t}_i]_q$ (using Theorem 3.11(e)).

Question 1. One question stands out concerning §6.5, namely can one find an explicit bound (in terms of known functions, like exponentiation) for the value $n = N$ for which the WHILE loop halts? This would give the maximum height of the trees in the set of representatives $\mathsf{REP}_{\mathsf{TREES}}$ of the $\equiv_q^{\mathsf{MSO}}$ classes of trees.

Using Compton's equations, we have carried out a detailed study [**4**] of MSO classes $\mathcal{T}$ of finite trees whose generating functions $T(x)$ have radius of convergence $\rho = 1$. One conclusion obtained is that if the class of forests $\partial\mathcal{T}$ is closed under addition, and under extraction of trees (thus forming an additive number system as described in [**7**]), then $\mathcal{T}$ has a MSO 0–1 law.

7. Well Conditioned Systems

In the following, a method for determining the asymptotics of generating functions defined by well conditioned systems of equations is described, as an application of the formula (4.2) for $\mathfrak{q}$. Details for how to apply the method to well conditioned 2-equation systems follows the general discussion. First some terminology from [**3**] is needed:

[22] This item, as well as the next one, are easy corollaries of results in Rabin's 1969 paper [**25**] on the decidability of the MSO theory of two successors. Idziak and Idziak [**21**] use Rabin's result to prove that the MSO theory of finite trees is decidable. Essentially the same proof works for m-colored finite forests as well.

Monounary algebras can be interpreted into finite forests, or one can use Rabin's Theorem 2.4 and the subsequent paragraph from [**25**], which show that the MSO theory of countable monounary algebras is decidable, even when the language is augmented by variables for finite subsets. By relativizing quantifiers to a finite subset closed under the unary function, one has the fact that the MSO theory of finite monounary algebras is decidable. A more elementary proof, not depending on the Rabin result, was given by Gurevich and Shelah [**20**]—the proof given here seems to be yet more elementary.

DEFINITION 7.1. A system of equations

$$\begin{array}{rcl} y_1 & = & G_1(x, y_1, \ldots, y_k) \\ & \vdots & \\ y_k & = & G_k(x, y_1, \ldots, y_k), \end{array}$$

abbreviated $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$, is *well conditioned* if the following hold:[23]

(a) The $G_i(x, \mathbf{y})$ are power series in $\mathsf{Dom}_0[x, \mathbf{y}]$.
(b) $\mathbf{G}(x, \mathbf{y})$ is holomorphic in a neighborhood of the origin.
(c) $\mathbf{G}(0, \mathbf{y}) = \mathbf{0}$.
(d) For all i, $G_i(x, \mathbf{0}) \neq 0$.
(e) The system is irreducible (as defined in §4.2).
(f) For some i, j, k, $\dfrac{\partial^2 G_i(x, \mathbf{y})}{\partial y_j \partial y_k} \neq 0$.

The *characteristic system* of $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ is

$$\left\{ \begin{array}{rcl} \mathbf{y} & = & \mathbf{G}(x, \mathbf{y}) \\ 0 & = & \det\big(I - J_{\mathbf{G}}(x, \mathbf{y})\big), \end{array} \right.$$

where $J_{\mathbf{G}}(x, \mathbf{y})$ is the Jacobian matrix $\dfrac{\partial \mathbf{G}}{\partial \mathbf{y}}$. The positive solutions of the characteristic system, that is, the solutions $(a, \mathbf{b})$ with $a, b_1, \ldots, b_k > 0$, are called *characteristic points*. A characteristic point $(a, \mathbf{b})$ is an *eigenpoint* if the largest real eigenvalue of the matrix $J_{\mathbf{G}}(a, \mathbf{b})$ is 1.

PROPOSITION 7.2. *For a well conditioned system $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$, one has the following:*

(a) *The system $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ is a reduced elementary system.*
(b) *There is a unique solution $\mathbf{y} = \mathbf{T}(x)$ in $\mathsf{Dom}_0[x]^k$.*
(c) *The $T_i(x)$ have the same radius of convergence $\rho \in (0, \infty)$, and they converge at ρ. Let $\boldsymbol{\tau} = \mathbf{T}(\rho)$.*
(d) *If $(a, \mathbf{b})$ is an eigenpoint of $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ then it is equal to $(\rho, \boldsymbol{\tau})$.*
(e) *If $(\rho, \boldsymbol{\tau})$ is a characteristic point then, among the characteristic points $(a, \mathbf{b})$, it is the one with the largest value of a.*
(f) *The $\mathsf{Spec}(T_i(x))$ are infinite and periodic, with the same parameter $\mathfrak{q}_i$, namely $\mathfrak{q}_i = \mathfrak{q}$, where*

$$\begin{array}{rcl} \mathfrak{q} & = & \gcd\left(\bigcup\limits_{\substack{1 \leq j \leq k \\ \mathbf{u} \in \mathbb{N}^k}} \Big(G_{j,\mathbf{u}} + \mathbf{u} \circledast \mathfrak{m} - \mathfrak{m}_j \Big) \right) \\ \mathfrak{m}_i & = & \min\Big(G_i^{(k)}(\boldsymbol{\emptyset}) \Big). \end{array}$$

(g) *For all i, the minimum eventual period $\mathfrak{p}_i$ of $\mathsf{Spec}(T_i(x))$ is equal to $\mathfrak{q}$, so $\mathsf{Spec}(T_i(x))$ is eventually an arithmetical progression, in particular*

$$\mathsf{Spec}(T_i(x)) \;=\; \mathsf{Spec}(T_i(x))|_{<\mathfrak{c}_i} \cup (\mathfrak{c}_i + \mathfrak{q} \cdot \mathbb{N}) \;\subseteq\; \mathfrak{m}_i + \mathfrak{q} \cdot \mathbb{N}.$$

(h) *Each $T_i(x)$ can be written in the form $x^{\mathfrak{m}_i} V_i(x^{\mathfrak{q}})$, where $V_i(x) \trianglerighteq 0$, $V_i(0) \neq 0$ and $\mathsf{Spec}(V_i(x))$ is a cofinal subset of $\mathbb{N}$.*

---

[23] The condition (e) from the definition of *well conditioned* in §2 of [**3**] is redundant; hence it is omitted here.

(i) *The dominant singularities of* $T_i(x)$ *are* $\rho, \rho \cdot \omega, \ldots, \rho \cdot \omega^{\mathfrak{q}-1}$*, where* $\omega = \exp(2\pi i/\mathfrak{q})$*, a primitive* $\mathfrak{q}$*th root of unity.*

Proof. (a) follows from Definition 7.1 (b),(c),(d). Use Proposition 4.1 for (b). Item (c) is Proposition 3 (iv) of [**3**]. (d) follows from Theorem 21 (d),(e) of [**3**]. (e) is Theorem 14 (a) of [**3**]. (f) and (g) follow from Theorem 4.6 and Corollary 4.7. (h) follows from (g) and the definition of $\mathfrak{m}_i$. (i) follows from $\mathbf{T}(x) = \mathbf{G}(x, \mathbf{T}(x))$ and (h). $\square$

Theorem 7.3. *Suppose* $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$ *is well conditioned, with an eigenpoint in the interior of the domain of* $\mathbf{G}(x, \mathbf{y})$*. Let* $\mathbf{T}(x)$*,* $(\rho, \boldsymbol{\tau})$ *and* $\mathfrak{q}$ *be as in Proposition 7.2. Then one has the following:*

(a) $\mathfrak{c}_i + \mathfrak{q} \cdot \mathbb{N} \subseteq \mathsf{Spec}(T_i(x)) \subseteq \mathfrak{m}_i + \mathfrak{q} \cdot \mathbb{N}$.

(b) *There are well conditioned single equation systems* $y_i = \widehat{G}_i(x, y_i)$*, with the unique solution in* $\mathsf{Dom}_0[x]$ *being* $y_i = T_i(x)$*, and with* (ρ, τ_i) *in the interior of the domain of* $\widehat{G}_i(x, y_i)$*.*

(c) *The coefficient sequence* $t_i(n)$ *of each* $T_i(x)$ *has the asymptotics*

$$t_i(n) \sim C_i \rho^{-n} n^{-3/2} \quad \text{on } \mathsf{Spec}(T_i(x)),$$

where (letting $\widehat{G}_{ix}(x, y_i)$ *denote* $\partial \widehat{G}_i(x, y_i)/\partial x$*, etc.),*

$$C_i \;=\; \mathfrak{q} \cdot \sqrt{\frac{\rho}{2\pi}} \cdot \sqrt{\frac{\widehat{G}_{ix}(\rho, \tau_i)}{\widehat{G}_{iy_iy_i}(\rho, \tau_i)}}.$$

(d) *The quotients* $\widehat{G}_{ix}(x, y_i)/\widehat{G}_{iy_iy_i}(x, y_i)$ *are rational functions.*

Proof. (a) is from item (g) of Proposition 7.2. (b) is from Drmota's Theorem, as presented in [**3**] (where it is Theorem 22). (c) is from Theorem 28 of [**2**] (the factor of $\mathfrak{q}$ in C_i is a consequence of (h) and (i) in Proposition 7.2). The method to prove item (d) is detailed in the case of two equations in §7.1 below. $\square$

Next, this theorem is applied to find the asymptotics for the $t_i(n)$ in the case $k = 2$.

7.1. Systems of two equations.

Assume the system[24]

$$\begin{aligned} y &= E(x, y, z) \\ z &= F(x, y, z) \end{aligned} \tag{7.1}$$

is well conditioned, and there is an eigenpoint in the interior of the domain of $(E(x, y, z), F(x, y, z))$. In view of Proposition 7.2, let the solution in $\mathsf{Dom}_0[x]^2$ be $(y, z) = (S(x), T(x))$, let $\rho \in (0, \infty)$ be the radius of convergence of both $S(x)$ and $T(x)$, and let $(\tau_1, \tau_2) = (S(\rho), T(\rho))$.

As in the first step of the proof of Drmota's Theorem (see Theorem 22 in [**3**]), solve the first equation for y as a function of x, z, say

$$y \;=\; Y(x, z) \in \mathsf{Dom}_0(x, z).$$

[24]Instead of the usual $\mathbf{y} = \mathbf{G}(x, \mathbf{y})$, with subscripted variables and functions, this discussion of a 2-equation system will use single letters for different variables and functions—it is more reader friendly when using subscripted variables for partial derivatives. We replace y_1, y_2 by y, z, etc.

Then (again from the proof of Drmota's Theorem), (ρ, τ_2) is in the interior of the domain of $Y(x, z)$, and

$$\begin{aligned} Y(x,z) &\trianglerighteq 0 \\ Y(x,z) &= E(x, Y(x,z), z) \\ Y(\rho,\tau_2) &= \tau_1. \end{aligned} \tag{7.2}$$

Differentiating (7.2) gives:

$$\begin{aligned} Y_x(x,z) &= E_x(x,Y(x,z),z) + E_y(x,Y(x,z),z) \cdot Y_x(x,z) \\ Y_z(x,z) &= E_z(x,Y(x,z),z) + E_y(x,Y(x,z),z) \cdot Y_z(x,z) \\ Y_{zz}(x,z) &= E_{zz}(x,Y(x,z),z) + 2E_{yz}(x,Y(x,z),z) \cdot Y_z(x,z) \\ &\quad + E_{yy}(x,Y(x,z),z) \cdot Y_z(x,z)^2 + E_y(x,Y(x,z),z) \cdot Y_{zz}(x,z). \end{aligned}$$

Solving these equations for the partial derivatives of $Y(x, z)$, and and evaluating at $(x, z) = (\rho, \tau_2)$ gives

$$\begin{aligned} Y_x(\rho,\tau_2) &= \frac{E_x(\rho,\boldsymbol{\tau})}{1 - E_y(\rho,\boldsymbol{\tau})} \\ Y_z(\rho,\tau_2) &= \frac{E_z(\rho,\boldsymbol{\tau})}{1 - E_y(\rho,\boldsymbol{\tau})} \\ Y_{zz}(\rho,\tau_2) &= \frac{E_{zz}(\rho,\boldsymbol{\tau}) + 2E_{yz}(\rho,\boldsymbol{\tau}) \cdot Y_z(\rho,\tau_2) + E_{yy}(\rho,\boldsymbol{\tau}) \cdot Y_z(\rho,\tau_2)^2}{1 - E_y(\rho,\boldsymbol{\tau})}. \end{aligned}$$

Substituting $Y(x, z)$ for y in (7.1) gives a well conditioned 1-equation system solved by $z = T(x)$, namely:

$$z = \widehat{F}(x,z) := F(x, Y(x,z), z).$$

From Theorem 7.3 (c),

$$t(n) \sim C_T \rho^{-n} n^{-3/2},$$

where

$$C_T = \mathfrak{q} \cdot \sqrt{\frac{\rho}{2\pi}} \cdot \sqrt{\frac{\widehat{F}_x(\rho,\tau_2)}{\widehat{F}_{zz}(\rho,\tau_2)}}.$$

We have

$$\begin{aligned} \widehat{F}_x(x,z) &= F_x(x,Y(x,z),z) + F_y(x,Y(x,z),z) \cdot Y_x(x,z) \\ \widehat{F}_z(x,z) &= F_z(x,Y(x,z),z) + F_y(x,Y(x,z),z) \cdot Y_z(x,z) \\ \widehat{F}_{zz}(x,z) &= F_{zz}(x,Y(x,z),z) + 2F_{yz}(x,Y(x,z),z) \cdot Y_z(x,z) \\ &\quad + F_{yy}(x,Y(x,z),z) \cdot Y_z(x,z)^2 + F_y(x,Y(x,z),z) \cdot Y_{zz}(x,z). \end{aligned}$$

Evaluating these at (ρ, τ_2) gives

$$\begin{aligned} \widehat{F}_x(\rho,\tau_2) &= F_x(\rho,\boldsymbol{\tau}) + F_y(\rho,\boldsymbol{\tau}) \cdot Y_x(\rho,\tau_2) \\ \widehat{F}_{zz}(\rho,\tau_2) &= F_{zz}(\rho,\boldsymbol{\tau}) + 2F_{yz}(\rho,\boldsymbol{\tau}) \cdot Y_z(\rho,\tau_2) \\ &\quad + F_{yy}(\rho,\boldsymbol{\tau}) \cdot Y_z(\rho,\tau_2)^2 + F_y(\rho,\boldsymbol{\tau}) \cdot Y_{zz}(\rho,\tau_2). \end{aligned}$$

This information suffices to determine C_T. A similar procedure gives C_S.

For well conditioned *polynomial* systems, $(\rho, \boldsymbol{\tau})$ is a characteristic point in the interior of the domain of $\mathbf{G}(x, \mathbf{y})$, thus the method just described determines the constants C_i for such systems.

EXAMPLE 7.4. Consider the polynomial system

$$\begin{array}{rcl} y &=& E(x,y,z) \ := \ x\cdot(x+x^5y^5+x^3z^5) \\ z &=& F(x,y,z) \ := \ x\cdot(1+y^3z^8). \end{array}$$

We solve for C_T, where $(S(x),T(x))$ is the solution.

By Proposition 7.2 (f), $\mathfrak{m}=(2,1)$ and $\mathfrak{q}=(7,7)$. There are two characteristic points:

$$x=.4275279509\ldots, y=3.5297999379\ldots, z=.4886125984\ldots$$
$$x=.7580667215\ldots, y=.7485529361\ldots, z=.8289799201\ldots$$

The second point has the largest first coordinate; by Proposition 7.2 (e) it must be $(\rho,\boldsymbol{\tau})$. From the formulas above, one has

$$\begin{array}{rcl} Y_x(\rho,\tau_2) &=& 3.6339912586\ldots \\ Y_z(\rho,\tau_2) &=& 1.1106860072\ldots \\ Y_{zz}(\rho,\tau_2) &=& 8.1565981501\ldots \\ \widehat{F}_x(\rho,\tau_2) &=& 2.1263292470\ldots \\ \widehat{F}_{zz}(\rho,\tau_2) &=& 15.1259723598\ldots \end{array}$$

Then

$$C_T \ = \ 0.9116233215\ldots,$$

so

$$t(n) \ \sim \ (0.9116233215\ldots)\cdot(0.7580667215\ldots)^{-n}n^{-3/2}.$$

Likewise one finds C_S and the asymptotics for $s(n)$.

Appendix A. Routine proofs of preliminary material

PROOF OF THE $\mathfrak{q}$ FROM PROPOSITION 2.9. Let $a\in\mathbb{N}$, $U,V\subseteq\mathbb{N}$. Then

$$\begin{array}{rcl} 0\in U+V &\Rightarrow& \gcd(U+V)=\gcd(\gcd(U),\gcd(V)) \\ a>0\in V &\Rightarrow& \gcd(a*V)=\gcd(V). \end{array}$$

For $A:=A_1\cup A_2$: Let $\mathfrak{m}:=\mathfrak{m}(A)$, $\mathfrak{q}:=\mathfrak{q}(A)$. Then $\mathfrak{m}=\mathfrak{m}_1$, so

$$\begin{array}{rcl} \mathfrak{q} &=& \gcd\Big((A_1-\mathfrak{m}_1)\cup(A_2-\mathfrak{m}_1)\Big) \\ &=& \gcd\Big((A_1-\mathfrak{m}_1)\cup\big((A_2-\mathfrak{m}_2)+(\mathfrak{m}_2-\mathfrak{m}_1)\big)\Big) \\ &=& \gcd\big(q_1,q_2,\mathfrak{m}_2-\mathfrak{m}_1\big). \end{array}$$

For $A:=A_1+A_2$: Let $\mathfrak{m}:=\mathfrak{m}(A)$, $\mathfrak{q}:=\mathfrak{q}(A)$. Then

$$\begin{array}{rcl} \mathfrak{q} &:=& \gcd\big(A-\mathfrak{m}\big) \ = \ \gcd\Big((A_1-\mathfrak{m}_1)+(A_2-\mathfrak{m}_2)\Big) \\ &=& \gcd\big(q_1,q_2\big). \end{array}$$

For $A:=A_1*A_2$: Let $\mathfrak{m}:=\mathfrak{m}(A)$, $\mathfrak{q}:=\mathfrak{q}(A)$. If $A_1=\{0\}$ then $A=\{0\}$, so $\mathfrak{q}=0$. Now suppose $A_1\neq\{0\}$. Then

$$\begin{array}{rcl} \mathfrak{q} &:=& \gcd\big(A-\mathfrak{m}\big) \ = \ \gcd\big(A_1*A_2-\mathfrak{m}_1\mathfrak{m}_2\big) \ = \ \gcd\Big(\bigcup_{a_1\in A_1}\big(a_1*A_2-\mathfrak{m}_1\mathfrak{m}_2\big)\Big) \\ &=& \gcd\Big(\bigcup_{a_1\in A_1}\big(a_1*(A_2-\mathfrak{m}_2)+(a_1-\mathfrak{m}_1)\mathfrak{m}_2\big)\Big) \end{array}$$

$$\begin{aligned}
&= \operatorname{gcd}\Big\{ \operatorname{gcd}\Big(\operatorname{gcd}\big(a_1 * (A_2 - \mathfrak{m}_2)\big),\ (a_1 - \mathfrak{m}_1)\mathfrak{m}_2 \Big) : a_1 \in A_1 \Big\} \\
&= \operatorname{gcd}\Big\{ \operatorname{gcd}\Big(\operatorname{gcd}\big(a_1 * (A_2 - \mathfrak{m}_2)\big),\ (a_1 - \mathfrak{m}_1)\mathfrak{m}_2 \Big) : a_1 \in A_1, a_1 \neq 0 \Big\} \\
&= \operatorname{gcd}\Big\{ \operatorname{gcd}\big(\mathfrak{q}_2,\ (a_1 - \mathfrak{m}_1)\mathfrak{m}_2\big) : a_1 \in A_1, a_1 \neq 0 \Big\} \\
&= \operatorname{gcd}\Big(\mathfrak{q}_2,\ \operatorname{gcd}\big((A_1 - \mathfrak{m}_1)\mathfrak{m}_2\big)\Big) \;=\; \operatorname{gcd}\big(\mathfrak{q}_2,\ \mathfrak{q}_1\mathfrak{m}_2\big).
\end{aligned}$$

□

Proof of (b)–(d) of Lemma 2.11. (b): Since $\mathfrak{c}, \mathfrak{c} + \mathfrak{p} \in A$, by Remark 2.8 we have $\mathfrak{q} | \mathfrak{p}$. Clearly $\mathfrak{p} = \mathfrak{q}$ implies $\mathfrak{p} \,\big|\, (A - \mathfrak{m})$. Conversely, if $\mathfrak{p} \,\big|\, (A - \mathfrak{m})$ then $\mathfrak{p} \mid \mathfrak{q} = \operatorname{gcd}(A - \mathfrak{m})$; and since $\mathfrak{q} \mid \mathfrak{p}$, one has $\mathfrak{p} = \mathfrak{q}$.

(c): (ii)⇒(i) is obvious. Assume (i) holds. Clearly $p = b$, so $A = A_0 \cup (a + p \cdot \mathbb{N})$. Let $x \in A|_{\geq \mathfrak{c}}$. For n sufficiently large, $x + pn \in A \smallsetminus A_0$, so $x + pn \in a + p \cdot \mathbb{N}$. From this we have $x \in A|_{\geq \mathfrak{c}} \Rightarrow p | (x - a)$, so $x \in A|_{\geq \mathfrak{c}} \Rightarrow p | (x - \mathfrak{c})$. Thus $A|_{\geq \mathfrak{c}} = \mathfrak{c} + \mathfrak{p} \cdot \mathbb{N}$, proving (i)⇒(ii).

(ii)⇒(iii) is obvious. Assume (iii) holds. For $x \in A|_{\geq \mathfrak{c}}$ one has $\mathfrak{q}\big(A|_{\geq \mathfrak{c}}\big) \,\big|\, (x - c)$, thus $\mathfrak{p} \,\big|\, (x - c)$. As before, $A|_{\geq \mathfrak{c}} = \mathfrak{c} + \mathfrak{p} \cdot \mathbb{N}$, proving (i)⇒(ii).

(d): Assume (i) holds. Then $A \subseteq \mathfrak{m} + \mathfrak{p} \cdot \mathbb{N} \Rightarrow \mathfrak{p} \,\big|\, (A - \mathfrak{m}) \Rightarrow \mathfrak{p} = \mathfrak{q}$, proving (iii).

Next assume (iii) holds. Then $\mathfrak{q}\big(A|_{\geq \mathfrak{c}}\big) \,\big|\, \mathfrak{p}\big(A|_{\geq \mathfrak{c}}\big) = \mathfrak{p} = \mathfrak{q} \,\big|\, \mathfrak{q}\big(A|_{\geq \mathfrak{c}}\big)$, so $\mathfrak{p} = \mathfrak{q}\big(A|_{\geq \mathfrak{c}}\big)$. By (b), $A = A\big|_{< \mathfrak{c}} \cup (\mathfrak{c} + \mathfrak{p} \cdot \mathbb{N})$. Also $\mathfrak{p} = \mathfrak{q} \,\big|\, (A - \mathfrak{m}) \Rightarrow A \subseteq \mathfrak{m} + \mathfrak{p} \cdot \mathbb{N}$. Thus (iii)⇒(ii). Finally, (ii)⇒(i) is obvious. □

Proof of Lemma 3.5. From

$$\mathbf{G}(\mathbf{Y}) \;:=\; \bigvee_{\mathbf{u} \in \mathbb{N}^k} \big(\mathbf{G}_{\mathbf{u}} + \mathbf{u} \circledast \mathbf{Y}\big),$$

one has, by Lemma 3.3 (a), for $1 \leq i \leq k$,

$$G_i^{(n+1)}(\mathbf{A}) \;=\; \bigcup_{\mathbf{u} \in \mathbb{N}^k} \Big(G_{i,\mathbf{u}} + \sum_{j\,:\,u_j > 0} u_j * G_j^{(n)}(\mathbf{A})\Big).$$

Let

$$\mathbf{b}_{\mathbf{u}} \;:=\; \min \mathbf{G}_{\mathbf{u}} \quad \text{and} \quad \mathbf{b}^{(n)} \;:=\; \min \mathbf{G}^{(n)}(\mathbf{A}),$$

that is, for $1 \leq i \leq k$,

$$b_{i,\mathbf{u}} \;=\; \min G_{i,\mathbf{u}} \quad \text{and} \quad b_i^{(n)} \;:=\; \min G_i^{(n)}(\mathbf{A}).$$

Then, for $n \geq 0$,

$$\mathbf{b}^{(n+1)} \leq \mathbf{b}^{(n)}, \tag{A.1}$$

since $\mathbf{A} \leq \mathbf{G}(\mathbf{A})$ implies $\mathbf{G}^{(n)}(\mathbf{A}) \leq \mathbf{G}^{(n+1)}(\mathbf{A})$, by repeated application of Lemma 3.4 (a).

From the above,

$$\begin{aligned}
b_i^{(n+1)} &:= \min G_i^{(n+1)}(\mathbf{A}) \\
&= \min \bigcup_{\mathbf{u} \in \mathbb{N}^k} \Big(G_{i,\mathbf{u}} + \sum_{j\,:\,u_j > 0} u_j * G_j^{(n)}(\mathbf{A})\Big) \quad \text{by (3.1)},
\end{aligned}$$

so

$$b_i^{(n+1)} = \min\Big\{b_{i,\mathbf{u}} + \sum_{j\,:\,u_j>0} u_j b_j^{(n)} : \mathbf{u} \in \mathbb{N}^k\Big\}. \tag{A.2}$$

For $n \geq 1$ let

$$I_n := \Big\{j : b_j^{(n)} < b_j^{(n-1)}\Big\}.$$

Then

$$(\forall n \geq 1)(\forall i \in I_{n+1})(\exists r \in I_n)\Big(b_i^{(n+1)} \geq b_r^{(n)}\Big), \tag{A.3}$$

which says that if some b_i decreases in round $n+1$, then it is because it depends on some b_r which decreased in round n; hence $b_i \geq b_r$. In more detail, suppose $n \geq 1$ and $i \in I_{n+1}$, that is,

$$b_i^{(n+1)} < b_i^{(n)}.$$

From (A.2), let $\mathbf{u} \in \mathbb{N}^k$ be such that

$$b_i^{(n+1)} = b_{i,\mathbf{u}} + \sum_{j\,:\,u_j>0} u_j b_j^{(n)}. \tag{A.4}$$

Let $r \in \{1,\ldots,k\}$ be such that $u_r > 0$ and $r \in I_n$. Such an r must exist, for otherwise $u_j > 0$ would imply $j \notin I_n$, that is, $b_j^{(n)} = b_j^{(n-1)}$; then, from (A.4), and from (A.2) with $n-1$ substituted for n,

$$b_i^{(n+1)} = b_{i,\mathbf{u}} + \sum_{j\,:\,u_j>0} u_j b_j^{(n-1)} \geq b_i^{(n)},$$

contradicting the assumption that $i \in I_n$, that is, $b_i^{(n+1)} < b_i^{(n)}$. For this choice of $\mathbf{u}$ and r, (A.4) implies

$$b_i^{(n+1)} \geq b_r^{(n)},$$

establishing (A.3).

Now suppose $I_n \neq \emptyset$ for some $n \geq k+1$. Then (A.3) says one can choose a sequence $i_n, \ldots, i_{n-k}$ of indices from $\{1,\ldots,k\}$ such that

$$b_{i_n}^{(n)} \geq b_{i_{n-1}}^{(n-1)} \geq \cdots \geq b_{i_{n-k}}^{(n-k)}, \tag{A.5}$$

and $i_j \in I_j$ for $n-k \leq j \leq n$. By the pigeonhole principle, there are two j such that the indices i_j are the same, say $\ell = i_p = i_q$, where $n-k \leq p < q \leq n$. Then $b_\ell^{(q)} \geq b_\ell^{(p)}$ by (A.5). But from $\ell \in I_q$ and (A.1) one has $b_\ell^{(q)} < b_\ell^{(q-1)} \leq \cdots \leq b_\ell^{(p)}$, giving a contradiction. Thus $I_n = \emptyset$ for $n > k$, completing the proof of the lemma. □

Appendix B. Büchi's Theorem

Given a finite alphabet $A = \{a_1, \ldots, a_m\}$, a *word* $w = a_{i_1} \cdots a_{i_\ell}$ over the alphabet is a string of letters from the alphabet. One can associate with w a structure $\mathfrak{c}(w) := (\{1,\ldots,\ell\}, <, U_1, \ldots, U_m)$, called an *$m$-colored chain*. The U_n are unary predicates, called the colors, and $<$ is a linear order on the universe of $\mathfrak{c}(w)$, namely one has $1 < \cdots < \ell$. Define $U_n(j)$ to hold in $\mathfrak{c}(w)$ (that is, the element j of the chain $\mathfrak{c}(w)$ has the color U_n) iff $n = i_j$, that is, the jth letter of the word w is a_n. The mapping $\mu : w \mapsto \mathfrak{c}(w)$ is a bijection between words over the alphabet A and m-colored chains, with the property that the length ℓ of the word w is the size of the m-colored chain $\mathfrak{c}(w)$. Thus if $\mathcal{L}$ is a set of words over the alphabet A

then $\mathcal{L}$ has the same generating function as the collection $\mu(\mathcal{L})$ of m-colored chains. Büchi proved that a set $\mathcal{L}$ of words over A is a regular language iff $\mu(\mathcal{L})$ is (up to isomorphism) a MSO class of m-colored chains, the result which we stated briefly in Theorem 6.1: *MSO classes of colored chains are precisely the regular languages.*

An important step in Büchi's proof was, given a MSO class of m-colored chains $\mathcal{C}$, to show how to find a regular language $\mathcal{R}$ such that $\mu(\mathcal{R}) = \mathcal{C}$. Knowledge of this procedure would be the starting point for Compton's discovery that every MSO class of m-colored trees has an equational specification (see §6.2). One finds $\mathcal{R}$ as follows.

Let $\mathcal{C}$ be defined by a MSO sentence φ of quantifier depth $\leq q$, where $q \geq 3$. There are finitely many MSO classes of m-colored chains defined by MSO sentences of quantifier depth at most q, and they are closed under union, intersection and complement; so they form a finite Boolean algebra. Among these, let $\mathcal{C}_1, \ldots, \mathcal{C}_k$ be the minimal ones, the atoms of the Boolean algebra. Then the $\mathcal{C}_i$ are pairwise disjoint, and their union is the class of all m-colored chains. Every class of m-colored chains defined by a MSO sentence of quantifier depth at most q is a union of some of the $\mathcal{C}_i$.

For $1 \leq j \leq m$, let $\bullet_j$ be the 1-element chain of color U_j. Then the class $\{\bullet_j\}$ is (when closed under isomorphism) one of the $\mathcal{C}_n$, since the property that 'the size of a chain is 1 and the color of the single node is U_j' can be expressed by a MSO sentence of quantifier depth 3. We can assume $\mathcal{C}_1 = \{\bullet_1\}$, …, $\mathcal{C}_m = \{\bullet_m\}$. Then all m-colored chains in the minimal classes $\mathcal{C}_{m+1}, \ldots, \mathcal{C}_k$ have size ≥ 2.

In each of the classes $\mathcal{C}_n$, $1 \leq n \leq k$, all chains have first elements of the same color. This is because the property that the first element of a chain has a given color U_i can be expressed by a MSO sentence φ_i of quantifier depth 3. Likewise, all chains in each $\mathcal{C}_n$ have the last element of the same color. Let the first element of members of $\mathcal{C}_n$ have the color $U_{\alpha(n)}$, and the last element of members of $\mathcal{C}_n$ have the color $U_{\omega(n)}$.

For an m-colored chain $\mathfrak{c}$ of size ≥ 2, let $\partial\mathfrak{c}$ be the chain that results from removing the last element from $\mathfrak{c}$; and, for any m-chain $\mathfrak{c}$, let $\mathfrak{c}\bullet_i$ be the result of adding a new last element, of color U_i, to $\mathfrak{c}$. For $j > m$ define $\partial\mathcal{C}_j := \{\partial\mathfrak{c} : \mathfrak{c} \in \mathcal{C}_j\}$; and for $j \geq 1$ define $\mathcal{C}_j\bullet_i := \{\mathfrak{c}\bullet_i : \mathfrak{c} \in \mathcal{C}_j\}$. For $j > m$ one then has $\mathcal{C}_j = \big(\partial\mathcal{C}_j\big)\bullet_{\omega(j)}$, that is, by removing and then adding back the last element, with the correct color, in each member of $\mathcal{C}_j$, one has the original class $\mathcal{C}_j$.

For $j > m$, $\partial\mathcal{C}_j$ is clearly a collection of m-colored chains. Using Ehrenfeucht-Fraïssé games, one can prove that $\partial\mathcal{C}_j$ is actually a union of some of the minimal classes $\mathcal{C}_i$, that is,

$$\partial\mathcal{C}_j \;=\; \bigcup_{\mathcal{C}_i \subseteq \partial\mathcal{C}_j} \mathcal{C}_i,$$

and thus

$$\mathcal{C}_j \;=\; \bigcup_{\mathcal{C}_i \subseteq \partial\mathcal{C}_j} \mathcal{C}_i \bullet_{\omega(j)}\,. \tag{B.1}$$

Now, for $1 \leq n \leq k$, define a finite-state automaton $\mathfrak{A}_n$ that accepts a regular language $\mathcal{R}_n$, with $\mu(\mathcal{R}_n) = \mathcal{C}_n$, as follows. The states of $\mathfrak{A}_n$ are $0, 1, \ldots, k$, the initial state is 0, and the unique final state is n. There is an edge from 0 to $\alpha(n)$, labelled with the letter $a_{\alpha(n)}$. For $1 \leq i, j \leq k$, there is an edge from i to j iff $\mathcal{C}_i\bullet_{\omega(j)} \subseteq \mathcal{C}_j$, in which case the label on the edge is $a_{\omega(j)}$.

Let $\mathcal{R}_n$ be the regular language accepted by $\mathfrak{A}_n$. It is not difficult to see that $\mu(\mathcal{R}_n) = \mathcal{C}_n$. Since a union of regular languages is a regular language, a union of some of the $\mathcal{C}_i$ also corresponds to a regular language. This finishes the sketch of how to prove, for each MSO class $\mathcal{C}$ of m-colored chains, there is a regular language $\mathcal{R}$ with $\mu(\mathcal{R}) = \mathcal{C}$.

Büchi's Theorem shows that Berstel's detailed analysis of the generating functions for regular languages (see Example 4.12) applies to the generating functions of MSO classes of m-colored chains. This is the *Berstel Paradigm* that we would like to see paralleled in the study of all MSO classes of m-colored trees. In particular, can one show that the generating functions $T(x)$ of such classes decompose into a polynomial and finitely many "nice" functions $T_i(x)$, with each spectrum $\mathsf{Spec}(T_i(x))$ being an arithmetical progression?

References

1. Jason P. Bell, Stanley N. Burris, *Compton's method for proving logical limit laws.* This volume.
2. ———, ———, and Karen A. Yeats, *Counting rooted trees: The universal law $t(n) \sim C \cdot \rho^{-n} \cdot n^{-3/2}$.* Electron. J. Combin. **13** (2006), #R63 [64pp.]
3. ———, ———, ———, *Characteristic points of recursive systems.* Electron. J. Combin. **17** (2010), #R121 [34pp.]
4. ———, ———, ———, *Monadic second-order classes of trees of radius 1.* (Submitted.)
5. Jean Berstel, *Sur les pôles et le quotient de Hadamard de séries n-rationnelles.* C.R. Acad. Sci. Paris, **272**, Sér. A-B (1971), 1079–1081.
6. J. Richard Büchi, *Weak second-order arithmetic and finite automata.* Z. Math. Logik Grundlagen Math. **6** 1960, 66–92.
7. Stanley N. Burris, *Logical Limit Laws and Number Theoretic Density.* Mathematical Surveys and Monographs, Vol. **86**, Amer. Math. Soc., 2001.
8. A. Cayley, *On the theory of the analytical forms called trees.* Phil. Magazine **13** (1857), 172–176.
9. Kevin J. Compton, *A logical approach to asymptotic combinatorics I: First-order properties.* Adv. Math. **65** (1987), 65–96.
10. ———, *A logical approach to asymptotic combinatorics. II. Monadic second-order properties.* J. Combin. Theory, Ser. A **50** (1989), 110–131.
11. ———, *Private communication*, July, 2009.
12. ——— and C. Ward Henson, *A uniform method for proving lower bounds on the computational complexity of logical theories.* Ann. Pure Appl. Logic **48** (1990), 1–79.
13. Mor Doron and Saharon Shelah, *Relational structures constructible by quantifier free definable operations.* J. Symbolic Logic, **72** (2007), 1283–1298.
14. Michael Drmota, *Systems of functional equations.* Random Structures and Algorithms **10** (1997), 103–124.
15. Michael Drmota, *Random Trees.* Springer, 2009.
16. Arnaud Durand, Ronald Fagin and Bernd Loescher, *Spectra with only unary function symbols.* Proceedings of the 1997 Annual Conference of the European Association for Computer Science Logic (CSL97). [The manuscript can be found at http://www.almaden.ibm.com/cs/people/fagin/]
17. A. Durand, N.D. Jones, J.A. Makowsky, and M. More, *Fifty years of the spectrum problem.* (Preprint, July, 2009).
18. E. Fischer and J.A. Makowsky, *On spectra of sentences of monadic second order logic with counting.* J. Symbolic Logic **69** (2004), no. 3, 617–640.
19. Philippe Flajolet and Robert Sedgewick, *Analytic Combinatorics.* Cambridge University Press, 2009.
20. Yuri Gurevich and Saharon Shelah, *Spectra of monadic second-order formulas with one unary function.* 18th Annual IEEE Symposium on Logic in Computer Science, June 22–25, 2003, Ottawa, Canada.
21. Katarzyna Idziak and Pawel M. Idziak, *Decidability problem for finite Heyting Algebras.* J. Symbolic Logic **53** (1988), 729–735.

22. James M. Ortega, *Matrix Theory. A Second Course.* Plenum Press, 1987.
23. G. Pólya and R.C. Read, *Combinatorial Enumeration of Groups, Graphs and Chemical Compounds.* Springer Verlag, New York, 1987.
24. Charles Chapman Pugh, *Real Mathematical Analysis.* Undergraduate Texts in Mathematics. Springer-Verlag, 2002.
25. Michael O. Rabin, *Decidability of second-order theories and automata on infinite trees.* Trans. Amer. Math. Soc. **141** (1969), 1–35.
26. J.L. Ramírez-Alfonsín, *The Diophantine Frobenius Problem.* Oxford Lecture Series in Mathematics and its Applications, **30** (2005), Oxford University Press.
27. Jeffrey Shallit, *The Frobenius Problem and its generalizations.* Developments in Language Theory, 72–83, Lecture Notes in Comput. Sci., **5257** (2008), Springer-Verlag.
28. Heinrich Scholz, *Ein ungelöstes Problem in der Symbolischen Logik.* J. Symbolic Logic, **17**, No. 2 (1952), p. 160.
29. Saharon Shelah, *Spectra of monadic second order sentences.* Sci. Math. Jap., **59**, No. 2, (2004), 351–355.
30. Matti Soittola, *Positive rational sequences.* Theoret. Comput. Sci. **2** (1976), 317-322.
31. Larry Stockmeyer, *Classifying the computational complexity of problems.* J. Symbolic Logic, **52**, No. 1 (1987), 1–43.
32. Herbert S. Wilf, *generatingfunctionology.* 2nd ed., Academic Press, 1994.
33. Alan R. Woods, *Coloring rules for finite trees, probabilities of monadic second order sentences.* Random Structures Algorithms **10** (1997), 453–485.

Department of Mathematics, Simon Fraser University, 8888 University Dr., Burnaby, BC,V5A 1S6, Canada

E-mail address: `jpb@math.sfu.ca`

Department of Pure Mathematics, University of Waterloo, Waterloo, ON, N2L 3G1, Canada

E-mail address: `snburris@math.uwaterloo.ca`

Department of Mathematics, Simon Fraser University, 8888 University Dr., Burnaby, BC,V5A 1S6, Canada

E-mail address: `karen.yeats@math.sfu.ca`

Contemporary Mathematics
Volume **558**, 2011

Compton's Method for Proving Logical Limit Laws

Jason P. Bell and Stanley N. Burris

Abstract. Developments in the study of logical limit laws for both labelled and unlabelled structures, based on the methods of Compton (1987/1989), are surveyed, and a sandwich theorem is proved for multiplicative systems.

1. Introduction

All structures in this paper will be finite structures for a finite language, unless stated otherwise. Likewise, by the models of a sentence φ we mean the finite models.

In two papers ([**23**] 1987, [**26**] 1989) Compton introduced a new method to prove logical limit laws for classes $\mathcal{A}$ of relational structures, a method that, in retrospect, depended solely on the function $a(n)$ that counted the number of structures of size n for $n \geq 1$. He treated two different count functions $a(n)$ in parallel, namely:

- the *unlabelled* count function that counts the number of isomorphism types of size n, and
- the *labelled* count function that counts the number of labelled structures of size n, that is, the number of structures in $\mathcal{A}$ on the universe $\{1, \ldots, n\}$.

The study of logical limit laws started with the results of Glebskiĭ, Kogan, Liogon'kiĭ, and Talanov [**33**] 1969, and Liogon'kiĭ [**39**] 1970, where they proved: *the class $\mathcal{A}$ of all relational structures, for a finite language $\mathcal{L}$, has both a labelled and an unlabelled first-order 0–1 law.* Later Fagin [**31**], 1976, independently found a simpler proof in the interesting cases where at least one relation in $\mathcal{L}$ is not unary—here is an outline of the steps of his proof:

(a) Let Φ be a collection of first-order sentences consisting of: (i) axioms for the class $\mathcal{A}$ being considered, and (ii) for any finite structure $\mathfrak{S}_1$ and a one-element extension $\mathfrak{S}_2$, both in $\mathcal{A}$, there is a sentence in Φ that asserts: a substructure isomorphic to $\mathfrak{S}_1$ can be likewise extended to a substructure isomorphic to $\mathfrak{S}_2$. The sentences described in (ii) are called *extension axioms.*

(b) Prove that Φ axiomatizes a complete theory.

(c) For $\varphi \in \Phi$, prove that φ has labelled asymptotic probability $= 1$.

1991 *Mathematics Subject Classification.* Primary 03C13 ; Secondary 05A16, 11N45, 11N80, 11R45, 11U99, 60F20.

The authors thank NSERC for supporting this project.

(d) Conclude that every φ in the theory axiomatized by Φ has labelled asymptotic probability $= 1$.
(e) Conclude that $\mathcal{A}$ has a labelled first-order 0–1 law.
(f) Show that the asymptotic probability of a structure in $\mathcal{A}$ being automorphism-rigid (the only automorphism is the identity map) is $= 1$. [For graphs this was proved by Pólya.]
(g) Conclude that $\mathcal{A}$ has an unlabelled first-order 0–1 law. Furthermore, given a first-order sentence φ, the labelled and unlabelled asymptotic probabilities that φ holds are the same.

This method applied to a few other situations, for example, the class $\mathcal{A}$ of graphs. The first-order sentences about graphs that *almost always hold*, that is, have asymptotic probability $= 1$, are precisely the sentences true of the famous Erdős random graph, where the edges occur with probability 1/2. Aside from these and similar examples, all the ingredients needed for Fagin's method of proving 0–1 laws are seldom available for classes $\mathcal{A}$.

Compton's 1987 paper, on first-order 0–1 laws for admissible classes $\mathcal{A}$, used the above idea of finding a set Φ of axioms for a complete theory that extended the first-order theory of $\mathcal{A}$, and then showing that the sentences in Φ had asymptotic probability $= 1$, in both the labelled and unlabelled settings. The axioms he added to the theory of $\mathcal{A}$ simply say: for any given connected structure $\mathfrak{S}$ in $\mathcal{A}$ and any $n \geq 0$, it is not the case that a structure has exactly n components isomorphic to $\mathfrak{S}$.

With this choice of Φ for an admissible class $\mathcal{A}$, Compton found a simple property that sufficed to obtain an unlabelled first-order 0–1 law, namely $a(n-1)/a(n) \to 1$ as $n \to \infty$. In the labelled setting, a more complicated condition, expressed as $\mathcal{A} \to \infty$, was used. When such a condition held, he could first show that another notion of probability, similar to one Dirichlet used to analyze the density of prime numbers in an arithmetical progression, gave a first-order 0–1 law. Then a Tauberian theorem was applied, to show that the same 0–1 law held for the asymptotic probability. The automorphism-rigid aspect (f) of Fagin's argument, which allowed one to deduce the unlabelled 0–1 law from the labelled 0–1 law, was not available for such admissible classes—indeed, the probability of being automorphism-rigid was 0.

Compton's 1989 paper lifted the first-order 0–1 laws of the 1987 paper to monadic second-order 0–1 laws (see Theorem 1.8 below). However, parts of the method of proof used in the 1987 paper for first-order 0–1 laws did not carry over to the MSO setting—the axiomatization Φ of the sentences that were almost always true was replaced by a structure theorem for the class $\mathcal{A}_\varphi$ of members of $\mathcal{A}$ that satisfied φ, where φ was any given MSO sentence. As will be described below, the structure theorem said that $\mathcal{A}_\varphi$ was a finite union of pairwise disjoint partition classes—this was proved using Ehrenfeucht-Fraïssé games. Then followed a proof of the existence of (an extended version of the previous) Dirichlet density for $\mathcal{A}_\varphi$, and then the application of a Tauberian theorem to obtain the asymptotic probability $\mathsf{Prob}(\varphi)$. Parallel and similar arguments gave a MSO 0–1 law in both the labelled and the unlabelled cases.

Compton also gave sufficient conditions for simply having a MSO limit law (see Theorem 1.9 below), where the asymptotic probability of each MSO sentence φ exists, but is not necessarily 0 or 1. In the labelled case he provided a single application of this theorem, namely to the class of permutations. In the unlabelled case, no examples were given where these conditions applied.

1.1. Counting functions, generating series. Let $\mathcal{A}$ be a class of relational structures. For labelled structures one uses the *exponential generating series*

$$\mathbf{A}(x) := \sum_{n\geq 1} \frac{a(n)}{n!} x^n.$$

If $a(n)$ counts unlabelled structures of size n then one uses the *ordinary generating series* for $\mathcal{A}$ given by

$$\mathbf{A}(x) := \sum_{n\geq 1} a(n) x^n.$$

In either case ρ_A denotes the *radius of convergence* of the series.

1.2. Compton's admissible classes. The class of all structures for a given relational language $\mathcal{L}$ is closed under the operation of disjoint union, as are many others—for the language of one binary relation, the class of graphs and the class of forests give two examples. As is well known, every graph is uniquely a disjoint union of connected graphs (its components), and connected graphs are indecomposable under disjoint union. Thus graphs have a unique decomposition into indecomposables.

Compton gave a natural generalization of the definition of 'connected', as used in graph theory, to relational structures, and noted that every relational structure is uniquely a disjoint union of connected structures, called the *components* of the structure.

The class of all $\mathcal{L}$-structures, the class of graphs and the class of forests are examples of classes closed under both disjoint union and extraction of components. Compton was perhaps the first to realize that classes of relational structures with these two closure properties enjoy a privileged status in combinatorics, and he introduced a name for them.

Definition 1.1. Given a finite relational language $\mathcal{L}$, a class $\mathcal{A}$ of $\mathcal{L}$-structures (that is closed under isomorphism) is *admissible* if

(a) it is closed under disjoint union, and
(b) it is closed under extraction of components.

Note that the classes for which the first important logical limit laws were proved—namely for each $\mathcal{L}$ the class all $\mathcal{L}$-structures, as well as the class of graphs—are admissible classes $\mathcal{A}$ with $\rho_A = 0$.

Given $\mathcal{A}$, let $\mathcal{P}$ denote the subclass of connected members, with counting function $p(n)$ and generating series $\mathbf{P}(x)$. The bonus one has when working with an admissible class is called the *fundamental identity* in [**15**]. In the labelled case it is

$$1 + \mathbf{A}(x) = \exp\big(\mathbf{P}(x)\big), \tag{1.1}$$

and in the unlabelled case

$$1 + \mathbf{A}(x) = \prod_{n\geq 1} (1 - x^n)^{-p(n)} = \exp\Big(\sum_{m\geq 1} \mathbf{P}(x^m)/m\Big). \tag{1.2}$$

Clearly either of $\mathbf{P}(x)$ and $\mathbf{A}(x)$ determines the other. The fundamental identities are key tools when applying analytic methods (for example, the Cauchy Integral Theorem) to study the count functions $a(n)$ and $p(n)$.

Compton developed the theory of logical limit laws for the labelled and unlabelled cases in parallel. For φ an $\mathcal{L}$-sentence, $a_\varphi(n)$ denotes the (labelled or unlabelled) count function for $\mathcal{A}_\varphi$, the class of members of $\mathcal{A}$ which satisfy φ.

Let $\mathbb{L}$ be a logic for a finite relational language $\mathcal{L}$—we only consider first-order (FO) and monadic second-order (MSO) logic. A class $\mathcal{A}$ of $\mathcal{L}$-structures has an $\mathbb{L}$ *limit law* if, for every $\mathbb{L}$ sentence φ, the probability $\mathsf{Prob}(\varphi)$ that 'a randomly selected structure from $\mathcal{A}$ satisfies φ' is defined. Compton used *asymptotic probability* and denoted the probability of φ by $\mu(\varphi)$ in the labelled case, by $\nu(\varphi)$ in the unlabelled case. For the labelled case this means:

$$\mathsf{Prob}(\varphi) \;:=\; \mu(\varphi) \;:=\; \lim_{n\to\infty} \frac{a_\varphi(n)}{a(n)};$$

and similarly one has the unlabelled case by changing μ to ν, and using the unlabelled count functions. This definition assumes that $\mathsf{Spec}(\mathcal{A})$, the spectrum[1] of $\mathcal{A}$, is cofinite, that is, $a(n)$ is eventually positive. For an arbitrary class $\mathcal{A}$ of $\mathcal{L}$-structures, Compton ([**23**] p. 87, 1987) introduced the *generalized* asymptotic probability

$$\mu^*(\varphi) \;:=\; \lim_{\substack{n\to\infty \\ a(n)\neq 0}} \frac{a_\varphi(n)}{a(n)}$$

in the labelled case; change the μ to ν for the unlabelled case.

Admissible classes $\mathcal{A}$ have a periodic spectrum, so the theory developed for the cases when $\mathcal{A}$ has a cofinite spectrum easily lifts to arbitrary admissible classes $\mathcal{A}$. In particular, given an admissible class $\mathcal{A}$, let $d = \gcd\{n : a(n) > 0\}$. Then the spectrum of $\mathcal{A}$ eventually agrees with $d \cdot \mathbb{N}$, the multiples of d. Thus, for a sentence φ, one has the labelled asymptotic probability of φ expressed by

$$\mathsf{Prob}(\varphi) \;:=\; \mu^*(\varphi) \;=\; \lim_{n\to\infty} \frac{a_\varphi(nd)}{a(nd)},$$

provided the limit exists; and similarly for the unlabelled asymptotic probability. *The standing assumption is that* $d = 1$, unless explicitly stated otherwise. (For details on why it suffices to consider only the case $d = 1$, see [**15**], Chapter 3.)

In the 1987 paper [**23**], Compton used another notion of probability, but did not identify it as such until his 1989 paper ([**26**], p. 117) on MSO limit laws. In the labelled case he defined

$$\overline{\mu}(\varphi) \;:=\; \lim_{x\to\rho_A} \frac{\mathbf{A}_\varphi(x)}{\mathbf{A}(x)}, \tag{1.3}$$

provided the limit exists, where $x \to \rho_A$ means that x approaches ρ_A from the left. (In order for the limit to exist one must have $\rho_A > 0$.) Likewise one can define $\overline{\nu}(\varphi)$, using ordinary generating functions—however Compton did not follow this obvious parallel. Instead he generalized $\overline{\mu}$ to the *extended asymptotic density* (defined after the next theorem), and then gave a parallel definition of $\overline{\nu}$. For the next theorem we will use the obvious definition of $\overline{\nu}$ that is parallel to (1.3).

THEOREM 1.2 (Compton [**26**] Proposition 3.1, 1989). *Suppose $\mathcal{A}$ is an admissible class with $\rho_A > 0$ and $\mathbf{A}(\rho_A) = \infty$. Then $\mu \subseteq \overline{\mu}$, that is, if φ is a MSO sentence and $\mu(\varphi)$ is defined, then $\overline{\mu}(\varphi)$ is defined and equals $\mu(\varphi)$. More generally, one has $\mu^* \subseteq \overline{\mu}$. Similar assertions hold for ν and $\overline{\nu}$.*

[1]See [**9**], this volume, for a detailed discussion of basic properties of spectra.

There are admissible classes where $\rho_A > 0$ but $\mathbf{A}(\rho_A) < \infty$, for example the class of forests. To have the possibility of an approach to such cases, Compton immediately modified the definition of $\overline{\mu}$ ([**26**] p. 118, 1989) to the *extended asymptotic probability* in the labelled case:

$$\overline{\mu}(\mathcal{A}_\varphi) \;:=\; \lim_{j\to\infty}\lim_{x\to\rho_A}\frac{\mathbf{A}_\varphi^{(j)}(x)}{\mathbf{A}^{(j)}(x)}, \tag{1.4}$$

where $\mathbf{A}^{(j)}(x)$ denotes the jth derivative of $\mathbf{A}(x)$ with respect to x, etc. Compton points out that this is indeed an extension of the $\overline{\mu}$ defined in (1.3). With this new definition of $\overline{\mu}$ the next theorem shows: (i) $\mu \subseteq \overline{\mu}$ under weaker hypotheses than those in Theorem 1.2; and (ii) with the hypotheses of Theorem 1.2, the extended asymptotic probability $\overline{\mu}$ is always defined and is given by (1.3). Likewise, one has the parallel results for ν and $\overline{\nu}$.

Theorem 1.3 (Compton [**26**] Proposition 3.2, Theorem 4.3, 1989). *Suppose $\mathcal{A}$ is an admissible class with $\rho_A > 0$. Then $\mu \subseteq \overline{\mu}$. If furthermore $\mathbf{A}(\rho_A) = \infty$ then $\overline{\mu}(\varphi)$ exists for any MSO sentence φ and it is given by* (1.3). *Similar assertions hold for ν and $\overline{\nu}$.*

It now seems clear that the more general definition of $\overline{\mu}$ given in (1.4) does not offer any advantages over the original definition given in (1.3) when it comes to proving MSO limit laws—see §2 for details.

A key concept regarding admissible classes $\mathcal{A}$ is $\mathcal{A} \to \lambda$, defined as follows (Compton [**23**] 1987, p. 74):

(a) (Labelled): For $0 \le \lambda \le \infty$, $\mathcal{A} \to \lambda$ holds iff $\displaystyle\lim_{n\to\infty}\frac{a(n-j)/(n-j)!}{a(n)/n!} = \lambda^j$, for any $j \in \mathsf{Spec}(\mathcal{P})$, where $a(n)$ is the labelled count function.

(b) (Unlabelled): For $0 \le \lambda \le 1$, $\mathcal{A} \to \lambda$ holds iff $\displaystyle\lim_{n\to\infty}\frac{a(n-j)}{a(n)} = \lambda^j$, for any $j \in \mathsf{Spec}(\mathcal{P})$, where $a(n)$ is the unlabelled count function.

This concept simplifies considerably for most values of λ.

Lemma 1.4 (Compton [**23**] Proposition 4.1, 1987). *Suppose $\mathcal{A}$ is an admissible class. Then $\mathcal{A} \to \lambda$ implies $\rho_A = \lambda$. Furthermore,*

(a) (Labelled): *For $0 < \lambda < \infty$, $\mathcal{A} \to \lambda$ holds iff* $\displaystyle\lim_{n\to\infty}\frac{na(n-1)}{a(n)} = \lambda.$

(b) (Unlabelled): *For $0 < \lambda \le 1$, $\mathcal{A} \to \lambda$ holds iff* $\displaystyle\lim_{n\to\infty}\frac{a(n-1)}{a(n)} = \lambda.$

The first question posed by Compton at the end of [**23**] 1987, presented in §1.8 below as (Q1), was whether or not these results could be extended—for item (a) to the cases $\lambda \in \{0, \infty\}$, and for item (b) to the case $\lambda = 0$. The direction ($\Leftarrow$) is true in both items (a) and (b) for these extreme values of λ, so the question was whether or not ($\Rightarrow$) holds. The case $\lambda = 0$, that is, $\mathcal{A} \to 0$, occurs in the 'classical' examples of 0–1 laws mentioned in the beginning, namely graphs, etc.

A key observation concerning an admissible class $\mathcal{A}$ with component class $\mathcal{P}$ is that, for any choice of components $\mathfrak{S}_1, \ldots, \mathfrak{S}_r$ and any choice $j_1, \ldots, j_r$ of nonnegative integers, there is a first-order sentence $\theta(\vec{\mathfrak{S}}, \vec{j})$ which says: *there are exactly j_i components isomorphic to $\mathfrak{S}_i$, for $1 \le i \le r$.* For $r = 1$ this becomes simply $\theta(\mathfrak{S}, j)$, expressing: *there are exactly j components isomorphic to $\mathfrak{S}$.* Compton gives explicit formulas for $\mathsf{Prob}(\theta(\vec{\mathfrak{S}}, \vec{j}))$ in both the labelled and unlabelled settings ([**23**] 1987, Theorem 5.6). From this follows:

PROPOSITION 1.5. *Suppose $\mathcal{A}$ is an admissible class with a first-order limit law. Then $\mathcal{A} \to \rho_A$. If $\mathcal{A}$ has a first-order 0–1 law then:*

(a) (Labelled):
- $\rho_A \in \{0, \infty\}$.
- *If $\rho_A = 0$ then for any connected $\mathfrak{S}$ from $\mathcal{A}$,* $\mathsf{Prob}(\theta(\mathfrak{S}, 0)) = 1$, *and* $\mathsf{Prob}(\theta(\mathfrak{S}, j)) = 0$ *for $j \geq 1$. Thus the asymptotic probability that there is a component isomorphic to $\mathfrak{S}$ is* 0.
- *If $\rho_A = \infty$ then for any connected $\mathfrak{S}$ from $\mathcal{A}$,* $\mathsf{Prob}(\theta(\mathfrak{S}, j)) = 0$ *for $j \geq 0$. Thus given any n, the asymptotic probability that there are at most n components isomorphic to $\mathfrak{S}$ is* 0.

(b) (Unlabelled):
- $\rho_A \in \{0, 1\}$.
- *If $\rho_A = 0$ then for any connected $\mathfrak{S}$ from $\mathcal{A}$,* $\mathsf{Prob}(\theta(\mathfrak{S}, 0)) = 1$, *and* $\mathsf{Prob}(\theta(\mathfrak{S}, j)) = 0$ *for $j \geq 1$. Thus the asymptotic probability that there is a component isomorphic to $\mathfrak{S}$ is* 0.
- *If $\rho_A = 1$ then for any connected $\mathfrak{S}$ from $\mathcal{A}$,* $\mathsf{Prob}(\theta(\mathfrak{S}, j)) = 0$ *for $j \geq 0$. Thus given any n, the asymptotic probability that there are at most n components isomorphic to $\mathfrak{S}$ is* 0.

Suppose an admissible class $\mathcal{A}$ has a first-order 0–1 law, and $\rho_A = 0$. In both the labelled and unlabelled cases one has $\mathcal{A} \to 0$, and for any connected $\mathfrak{S}$ in $\mathcal{A}$, the asymptotic probability of $\mathfrak{S}$ appearing as a component is 0. Compton's papers offer no further general facts regarding admissible classes $\mathcal{A}$ with $\rho_A = 0$ and a first-order 0–1 law.

The development of a substantial theory by Compton was based on assuming $\rho_A > 0$. For a first-order 0–1 law that meant $\rho_A = \infty$ in the labelled case, and $\rho_A = 1$ in the unlabelled case. For $\rho_A = 1$ in the unlabelled case the conclusion $\mathcal{A} \to 1$ can be expressed simply as $\lim\limits_{n\to\infty} \dfrac{a(n-1)}{a(n)} = 1$. For $\rho_A = \infty$ in the labelled case, no simplification is currently known for $\mathcal{A} \to \infty$.

The method of Compton's 1989 paper [**26**], as stated at the beginning of §6 of [**26**], was for the case $\rho_A > 0$, and was: (i) to find conditions to show that the extended asymptotic probability $\overline{\mu}(\varphi)$, respectively $\overline{\nu}(\varphi)$, exists, and then (ii) to find conditions that showed the asymptotic probability $\mu(\varphi)$, respectively $\nu(\varphi)$, was defined. This method was first used for individual MSO sentences φ.

THEOREM 1.6 (Compton [**26**] Theorem 6.1, 1989). *Suppose $\mathcal{A}$ is an admissible class with $\rho_A > 0$, $\mathbf{A}(\rho_A) = \infty$ and $\mathcal{A} \to \rho_A$. Let φ be a MSO sentence. Then:*

(a) (Labelled): *If $\overline{\mu}(\varphi) \in \{0, 1\}$ then $\mu(\varphi)$ exists and $= \overline{\mu}(\varphi)$.*

(b) (Unlabelled): *If $\overline{\nu}(\varphi) \in \{0, 1\}$ then $\nu(\varphi)$ exists and $= \overline{\nu}(\varphi)$.*

It is easy to see that one can express 'is connected' by a sentence in MSO logic.

COROLLARY 1.7 (Compton [**26**] Corollary 6.2, 1989). *Given the hypotheses of Theorem 1.6, let φ be a MSO sentence that expresses 'is connected'. Then in the labelled case $\mu(\varphi) = 0$, and in the unlabelled case $\nu(\varphi) = 0$. In either case this says that the probability of being connected is* 0.

Next Compton proves the fundamental results on MSO 0–1 laws for admissible classes $\mathcal{A}$ with $\rho_A > 0$.

Theorem 1.8 (Compton [**26**] Theorems 6.3 and 6.4, 1989). *Suppose $\mathcal{A}$ is an admissible class with $\rho_A > 0$. Then:*

(a) (Labelled): *$\mathcal{A}$ has a MSO 0–1 law iff $\mathcal{A}$ has a FO 0–1 law iff $\mathcal{A} \to \infty$. If (any of) these conditions hold then $\rho_A = \infty$ and $\mathsf{Prob}(\varphi) = \overline{\mu}(\mathcal{A}_\varphi)$, for φ a MSO sentence.*

(b) (Unlabelled): *$\mathcal{A}$ has a MSO 0–1 law iff $\mathcal{A}$ has a FO 0–1 law iff $\mathcal{A} \to 1$. If (any of) these conditions hold then $\rho_A = 1$ and $\mathsf{Prob}(\varphi) = \overline{\nu}(\mathcal{A}_\varphi)$ for φ a MSO sentence.*

This is not the end of the story about 0–1 laws for admissible classes with $\rho_A > 0$, for the simple reason that the last of the equivalent conditions in each of (a) and (b), concerning $\mathcal{A} \to$, is not so easy to verify in practice. With admissible classes it is usual that one knows a great deal more about the count function $p(n)$ for the components of the class than for the entire class; consequently much effort has been expended to find conditions on $p(n)$ that imply the desired condition on $a(n)$.

The more difficult, and even harder to apply, result was for MSO limit laws that were not 0–1 laws.

Theorem 1.9 (Compton [**26**] Theorem 6.6, 1989). *Suppose $\mathcal{A}$ is an admissible class with $\rho_A > 0$, $\mathbf{A}(\rho_A) = \infty$ and $\mathcal{A} \to \rho_A$.*

(a) (Labelled): *If $\rho_A < \infty$ and for some $C, N > 0$ one has, for $n \geq N$, $\dfrac{a(n-i)/(n-i)!}{a(n)/n!} \leq C{\rho_A}^i$, for $0 \leq i \leq n$, then $\mathcal{A}$ has a labelled MSO limit law, that is, $\mu(\varphi)$ exists for all MSO sentences φ.*

(b) (Unlabelled): *If $\rho_A < 1$ and for some $C, N > 0$ one has for $n \geq N$, $\dfrac{a(n-i)}{a(n)} \leq C{\rho_A}^i$, for $0 \leq i \leq n$, then $\mathcal{A}$ has an unlabelled MSO limit law, that is, $\nu(\varphi)$ exists for all MSO sentences φ.*

Except for a result of Woods [**56**] (discussed in §3.2 below) that covers some interesting cases concerning a single unary function, after 20 years one finds that, for admissible classes $\mathcal{A}$, Theorem 1.9 is still the foundation result for proving MSO limit laws that are not 0–1 laws. However it is not an easy theorem to apply—Compton gave one labelled example, namely permutations, and no unlabelled examples based on it.

In the following examples and results, *trees* are rooted trees and *forests* are forests of rooted trees.

1.3. Compton's examples. Compton's examples were stated in his 1987 paper [**23**] for first-order logic, and then upgraded in the 1989 paper [**26**] to MSO logic.

An important tool for the labelled case is *Hayman-admissible functions*—this use of the word 'admissible' is not directly related to Compton's admissible classes. These functions have a complicated definition, and it can be difficult to establish that a function is Hayman admissible; however a few nice examples are known—see, for example, items (c) and (d) in §1.4.

THEOREM 1.10 (Compton [**23**] Theorem 4.2, 1987). *If* $\mathbf{B}(x)$ *is Hayman admissible then* $\lim_{n\to\infty} \dfrac{b(n-1)}{b(n)} = \rho_B$.

Compton's applications of Theorem 1.8 above relied primarily on the following two propositions. First there is the case where $\mathcal{A}$ is a finitely generated admissible class, that is, $\mathcal{A}$ has only finitely many indecomposable members. In the unlabelled case this is connected to the study of the famous Coin Problem.[2]

PROPOSITION 1.11 (Compton [**23**] Example 7.15, 1987). *Suppose* $\mathcal{P}$ *is finite (counting up to isomorphism), say* $\mathcal{P}$ *is represented by structures* $\mathfrak{S}_1, \ldots, \mathfrak{S}_r$*, with sizes* $s_1, \ldots, s_r$*. Then:*

(a) *In the labelled case one has* $\mathcal{A} \to \infty$.

(b) *In the unlabelled case the count function is asymptotic to a polynomial, namely*

$$a(n) \ \sim \ \frac{1}{(r-1)!}\frac{n^{r-1}}{s_1\cdots s_r}\ ,$$

thus $\dfrac{a(n-1)}{a(n)} \to 1$ *as* $n \to \infty$.

Consequently $\mathcal{A}$ *has both a labelled and an unlabelled MSO 0–1 law.*

Compton gave no examples of admissible classes with only finitely many components, likely because it was trivial to do so. Simple examples would be the class of graphs with components of bounded size; and forests with trees having bounded height and width.

Next there is a famous result on partitions:[3]

PROPOSITION 1.12 (Bateman and Erdős [**1**], 1956). *Suppose* $p(n)$ *is a nonnegative integer for* $n \geq 1$ *with* $\gcd(\{n : p(n) > 0\}) = 1$*, and*

$$1 + \mathbf{A}(x) \ = \ \prod_{n\geq 1}(1-x^n)^{-p(n)}.$$

If $p(n) \leq 1$*, for* $n \geq 1$*, then* $\lim_{n\to\infty} \dfrac{a(n-1)}{a(n)} = 1$.

1.4. Four examples with an unlabelled 0–1 law. Theorem 1.8 and Proposition 1.12 show that the following four examples of Compton have an unlabelled MSO 0–1 law—the status of a labelled law varies:

(a) *Permutations.*[4]

In the labelled setting one has $a(n) = n!$, so $1+\mathbf{A}(x) = 1/(1-x)$, which gives

---

[2]Suppose one has coins with values $d_1, \ldots, d_k$, where $\gcd(d_1, \ldots, d_k) = 1$. Sylvester knew that the possible values of combinations of such coins formed a cofinite subset of the positive integers; he used partial fractions over the complex numbers to find the asymptotics for the number of ways that one could select coins to create a total value $= n$. The Coin Problem (a.k.a. Frobenius's Problem) was to find a formula for the largest n such that one could not realize the value n with the given coins. Sylvester found the formula for $k = 2$. The problem is still open for $k \geq 3$.

[3]The topic being studied by Bateman and Erdős was: how many ways can one partition numbers into parts when the parts come from a fixed subset of the positive integers? In particular they were looking at the kth difference function of the count function $a(n)$, and succeeded in finding a simple necessary and sufficient condition for the kth difference to be eventually monotonic.

[4]In the 1987 paper Compton states that permutations do not have a labelled first-order 0–1 law. In the 1989 paper one finds that they do have a labelled MSO limit law.

$\mathcal{A} \to 1$. Thus permutations do not have a first-order labelled 0–1 law. But Theorem 1.9 above applies to show that permutations indeed have a labelled MSO limit law.

(b) *Forests of height* 1.
Note that the labelled $p(n) = n$, so $1 + \mathbf{A}(x) = \exp\big(xe^x\big)$, a Hayman-admissible function with $\rho_A = \infty$. Thus $\mathcal{A} \to \infty$, so $\mathcal{A}$ has a labelled MSO 0–1 law.

(c) *Forests whose trees are linear.* For the labelled case note that $p(n) = n!$, so $1 + \mathbf{A}(x) = \exp\big(x/(1-x)\big)$. Compton states that one can prove this is a Hayman-admissible function, and clearly $\rho_A = 1$. Thus $\mathcal{A} \to 1$, so $\mathcal{A}$ does not have a labelled first-order 0–1 law.[5]

(d) *Equivalence relations*, or *partitions*. In the labelled case $p(n) = 1$, so $1 + \mathbf{A}(x) = \exp\big(e^x - 1\big)$, a Hayman-admissible function with $\rho_A = \infty$. Thus $\mathcal{A} \to \infty$, so $\mathcal{A}$ has a labelled MSO 0–1 law.

Remark 1.13. Since the unlabelled count function for $\mathcal{P}$ is $p(n) = 1$ in each of the four cases above, one could also use the famous 1917 result of Hardy and Ramanujan [**34**] on the number of partitions of an integer n, that is, the number of ways n can be expressed as a sum of positive integers. The fundamental identity in this case is

$$1 + \mathbf{A}(x) \;=\; \prod_{n\geq 1} \big(1 - x^n\big)^{-1},$$

from which they prove

$$a(n) \;\sim\; \frac{\exp\big(\pi\sqrt{2n/3}\,\big)}{4\sqrt{3}n}.$$

It easily follows that $a(n-1)/a(n) \to 1$ as $n \to \infty$. However the result of Bateman and Erdős is much easier to prove, and leads to far more unlabelled 0–1 laws.

1.5. Partitions with selected subsets. The final example where Compton uses his methods to obtain a MSO 0–1 law is the class $\mathcal{A}$ of *partitions with selected subsets*.

In the labelled case one has $p(n) = 2^n - 1$, so $\mathbf{A}(x) = \exp\big(e^x(e^x - 1)\big)$, a Hayman admissible function with $\rho_A = \infty$. Thus $\mathcal{A} \to \infty$, giving a labelled MSO 0–1 law.

For the unlabelled case $p(n) = n$. Compton appeals to an application of a difficult theorem of Meinardus (a theorem whose goal is to obtain results like those of Hardy and Ramanujan on partitions) to obtain asymptotics for $a(n)$ which show that $a(n-1)/a(n) \to 1$ as $n \to \infty$. Thus $\mathcal{A} \to 1$, so $\mathcal{A}$ has an unlabelled MSO 0–1 law as well.

1.6. First-order 0–1 laws that cannot be obtained by Compton's method. In his list of examples Compton states that the following admissible classes $\mathcal{A}$ have both labelled and unlabelled first-order 0–1 laws: *all $\mathcal{L}$-structures, directed graphs, oriented graphs* and *posets*. But these cannot be obtained by his method since in each case $\mathcal{A} \to 0$.

[5]The claim in [**23**] that the class of linear forests has a labelled first-order 0–1 law is incorrect.

1.7. Admissible classes without a first-order limit law. Compton also gives examples of admissible classes $\mathcal{A}$ that fail to have a first-order limit law, namely: the class of *binary forests* has neither a labelled nor an unlabelled limit law since $\mathcal{A}\to\rho_A$ fails in both cases; and likewise for the class of *binary forests with various orders* (pre-, post-, inorder).

1.8. Compton's status report and questions. In the 1987 paper Compton noted that the status of limit laws for several well-known classes was open. The following lists those examples, and updates what is known about them in the bulleted items.

Admissible classes for which the status of either the labelled or unlabelled first-order limit laws was not known at the time of publication of [**23**]:

(a) *Unlabelled* structures consisting of *a single unary function*, a.k.a. *functional digraphs*. (For the labelled case, Lynch [**40**], 1985, had proved a first-order limit law.)
  - [Woods [**56**] Corollaries 1.1 and 1.2, 1997, proved that a single m-colored unary function has a MSO limit law in both the labelled and unlabelled cases.]

(b) *Labelled or unlabelled Forests.*
  - [Woods [**55**] Theorems 1.1 and 6.5, 1997, proved that the class of m-colored trees (which is not an admissible class) has both labelled and unlabelled MSO limit laws (but not 0–1 laws). A MSO class of forests naturally corresponds to a MSO class of trees by simply adding a root to each forest. Using this, one sees that the class of m-colored forests also has both labelled and unlabelled MSO limit laws.]

(c) *Labelled or unlabelled Unary-Binary Forests.* This means every non-leaf has one or two immediate descendants.

(d) *Labelled or unlabelled Forests with pre- or post-order.*

(e) *Labelled or unlabelled Unary-Binary Forests with pre-, post-, or in-order.*

(f) *Labelled or unlabelled Acyclic Graphs.*
  - [McColm [**41**] 2002, proved that the class of *connected* acyclic graphs (also known as free trees) has a labelled MSO 0–1 law; and McColm [**42**] Corollary 2.1, 2004, proved that the same class has an unlabelled MSO 0–1 law. The situation for the class of all acyclic graphs is open.]

Compton also posed the following five questions about admissible classes $\mathcal{A}$ and first-order laws in [**23**]:

(Q1) For the labelled case there were two sub-questions, namely does:
$\lim_{n\to\infty}\dfrac{a(n-j)/(n-j)!}{a(n)/n!}=0$, for $j\in\mathsf{Spec}(\mathcal{P})$, imply $\lim_{n\to\infty}\dfrac{na(n-1)}{a(n)}=0$?
$\lim_{n\to\infty}\dfrac{a(n-j)/(n-j)!}{a(n)/n!}=\infty$, for $j\in\mathsf{Spec}(\mathcal{P})$, imply $\lim_{n\to\infty}\dfrac{na(n-1)}{a(n)}=\infty$?
In the unlabelled case suppose $\lim_{n\to\infty}\dfrac{a(n-j)}{a(n)}=0$, for $j\in\mathsf{Spec}(\mathcal{P})$. Does this imply $\lim_{n\to\infty}\dfrac{a(n-1)}{a(n)}=0$?
  - [There is no progress to report on (Q1).]

(Q2) In the labelled case, does $\rho_A=\infty$ imply $\mathcal{A}\to\infty$? (Compton conjectured the answer was "No".)

In the unlabelled case, does $\rho_A = 1$ imply $\mathcal{A} \to 1$? (Compton conjectured the answer was "No".)

- [A counterexample for the unlabelled case appeared in §4.4 of [**15**].]

(Q3) Find easily verifiable conditions for $\lim_{n\to\infty} \dfrac{a(n-1)}{a(n)} = 1$. Does $p(n) = \mathrm{O}(n^k)$, for some k, suffice?

- [Answered in the affirmative by Bell [**4**] Theorem 1.8, 2001.]

Is there a similar result one could use in the labelled case?

- [Burris and Yeats [**19**] Theorem 7, 2008, show that $p(n) = \mathrm{O}\Big(n^{\theta\cdot n}\Big)$, $0 < \theta < 1$, guarantees a labelled MSO 0–1 law.]

(Q4) Are there natural examples of admissible classes with a labelled 0–1 law but not an unlabelled 0–1 law? (Permutations give an example with an unlabelled 0–1 law, but not a labelled 0–1 law.)

- [Forests of brooms with 2-colored handles provide an example, in Burris and Yeats [**19**], 2008.]

(Q5) If the radius of convergence of the exponential generating series (for labelled structures) is 0, does it follow that there is a labelled 0–1 law iff there is an unlabelled 0–1 law?

- [There is no progress to report on (Q5).]

More questions were added in his 1989 paper:

(Q6) Is it true that every MSO sentence φ has a labelled probability iff it has an unlabelled probability?

- [There is no progress to report on (Q6).]

(Q7) If $\rho_A > 0$ and $\mathbf{A}(\rho_A) < \infty$, does it follow that the extended asymptotic probability $\overline{\mu}(\varphi)$ $[\overline{\nu}(\varphi)]$ of every MSO sentence φ exists?

- [There is no progress to report on (Q7).]

(Q8) If $\mathcal{A} \to \rho_A > 0$ and $\mathbf{A}(\rho_A) = \infty$, does it follow that the asymptotic probability $\mu(\varphi)$ $[\nu(\varphi)]$ of every MSO sentence φ exists?

- [There is no progress to report on (Q8).]

(Q9) Suppose $\mathcal{A} \to \rho_A > 0$, $\mathbf{A}(\rho_A) < \infty$ and φ is a MSO sentence. In the labelled, or unlabelled, case: Is there a modulus $m > 0$ such that

$$\lim_{q\to\infty} \frac{a_\varphi(mq+r)}{a(mq+r)} \quad \text{exists, for any } r \geq 0?$$

- [There is no progress to report on (Q9).]

In a second 1989 paper [**27**], Compton surveyed various approaches to proving logical limit laws and posed several questions, the following being relevant to the topic discussed here:

(Q10) Develop techniques for proving FO and MSO limit laws for classes $\mathcal{A}$ whose generating series $\mathbf{A}(x)$ converges at its radius of convergence. Two specific examples mentioned, where the existence of a logical limit laws was not known, were forests and unit interval graphs.

- [As noted above, Woods [**55**] 1997 proved that m-colored trees (hence forests) have both labelled and unlabelled MSO limit laws.]

(Q11) Prove a 0–1 law for graphs in a logic strong enough to express Hamiltonicity.

- [There is no progress to report on (Q11).]

(Q12) Develop a theory of asymptotic probability for classes where direct product is an appropriate operation (such as in the class of groups).
• [In the unlabelled setting, the theory of *first-order* logical limit laws for multiplicatively admissible classes has been developed to the point where it completely parallels the work on additively admissible classes—these results are discussed in this article.
In the labelled setting it seems there is simply no literature on multiplicatively admissible classes, not even a statement of a fundamental identity.]

(Q13) Investigate asymptotic probability and 0–1 laws for classes of regular graphs.
• [There is no progress to report on (Q13).]

(Q14) Show that all first-order sentences have labelled asymptotic probability in a class of directed graphs with the amalgamation property and closed under substructures.
• [There is no progress to report on (Q14).]

2. The presentation in "Number Theoretic Density and Logical Limit Laws", 2001

The interest in Compton's suggestion in (Q12) above, to develop a theory of asymptotic probability for classes where direct product is a natural operation for combining and decomposing structures, developed rapidly in the mid 1990s, following a lecture at the University of Waterloo by Compton on a first-order limit law for finite Abelian groups. In 2001 Burris published a book [**15**] where the first half was essentially an exposition of Compton's 1980s work on *additive* classes, that is, where the operation of combination is disjoint union, and the second half gave a parallel development for *multiplicative* classes, where the operation of combination is direct product. The book *only treated the unlabelled case*, mainly because a basic theory of labelled structures in the context of direct products did not (and still does not) exist.

Although Compton never quite said that his method was

> *to find criteria for logical limit laws for admissible classes $\mathcal{A}$ that depend only on the behavior of the generating series $\mathbf{A}(x)$,*

nonetheless this conclusion is strongly suggested by the fact that his results are of this form.[6] This view of Compton's method was adopted in [**15**], and it had strong consequences, which will be discussed in more detail below in §3.6. First it reduced Compton's method to showing that all partition classes $\vec{\gamma}\star\vec{\mathcal{P}}$, defined in §2.3, have asymptotic density in $\mathcal{A}$. This condition implies:

(a) $\displaystyle\lim_{n\to\infty}\frac{a(n-1)}{a(n)} = \rho_A$,

(b) the class of components $\mathcal{P}$ has density 0,

(c) $\rho_A > 0$, and

[6]This is not to say that Compton restricted his search for classes $\mathcal{A}$ with logical limit laws to those for which the laws were determined by the generating series $\mathbf{A}(x)$—in [**25**] he proved that posets have both a labelled and an unlabelled first-order 0–1 law by arguments that required more information than just knowing the generating series. Nonetheless, the remarkable feature of his logical limit laws, proved in the two papers ([**23**] 1987, [**26**] 1989) featured above, is that indeed they depend solely on the behavior of the generating series $\mathbf{A}(x)$, that is, solely on the number $a(n)$ of structures in $\mathcal{A}$ of each size n.

(d) $\mathbf{A}(\rho_A) = \infty$.

The last item was not known at the time of publication of [**15**]. The notation used in the following is mainly based on that of [**15**]. For $\mathcal{B} \subseteq \mathcal{A}$ let $B(x) := \sum_{n \le x} b(n)$. Then:

notation	meaning	name
$\delta(\mathcal{B})$	$\displaystyle\lim_{n\to\infty} \frac{b(n)}{a(n)}$	*local asymptotic density*
$\Delta(\mathcal{B})$	$\displaystyle\lim_{x\to\infty} \frac{B(x)}{A(x)}$	*global asymptotic density*
$\partial(\mathcal{B})$	$\displaystyle\lim_{x\to\rho_A} \frac{\mathbf{B}(x)}{\mathbf{A}(x)}$	*Dirichlet density*
$a(n)$ [or $\mathbf{A}(x)$] $\in \mathsf{RT}_\rho$	$\displaystyle\lim_{n\to\infty} \frac{a(n-1)}{a(n)} = \rho$	

Compton's method can only be used on admissible classes $\mathcal{A}$ that exhibit a medium to slow growth rate, since the radius of convergence of the generating series $\mathbf{A}(x)$ for the class $\mathcal{A}$ needs to be positive. For such classes the generating series (a formal power series) defines an analytic function in a neighborhood of 0. Fast growing classes, like the class of graphs (which has a first-order 0–1 law), require other means to establish their logical limit laws.

The 'Dirichlet density' ∂ is based on the original form of $\overline{\nu}$ used by Compton in [**26**] (see Theorem 1.2 above), not the extended asymptotic probability that requires differentiation. That is because item (d) above shows that there is no need to differentiate—admissible classes such as forests, which have $\mathbf{A}(\rho_A) < \infty$, simply cannot be handled by the method of Compton (as described above).

2.1. Classes closed under direct product and directly indecomposable factors. Multiplicative classes occur most naturally when one studies algebraic systems; for example, every finite group factors uniquely into a direct product of indecomposable groups. Likewise every finite lattice factors uniquely into a direct product of indecomposable lattices. When dealing with multiplicative classes, there were some key differences from the previous work on additive classes to take into account. First, finite structures may not have a unique factorization into indecomposables.[7] The second item was that the quotient $b(n)/a(n)$ was difficult to analyze—to obtain interesting results a certain smoothing out was needed, and this was done by replacing $b(n)$ by $B(x) := \sum_{n \le x} b(n)$, etc., when defining asymptotic density. The third item was that one needs Dirichlet series, not power series, for generating functions. The fourth item was that some multiplicative classes behaved like additive classes, which tended to be rather different from other multiplicative classes.

One has the definition of an admissible class in this context by replacing 'disjoint union' by 'direct product', and requiring unique factorization.

Definition 2.1. Given a finite language $\mathcal{L}$, a class $\mathcal{A}$ of $\mathcal{L}$-structures is *multiplicatively admissible* (or *admissible with respect to direct product*) if

(a) it is closed under direct product, and

[7]The question of unique factorization under direct product was studied intensely by Tarski, Jónsson and McKenzie (see [**44**]).

(b) every member of $\mathcal{A}$ of size at least 2 can be uniquely expressed as a direct product of members of $\mathcal{P}$, where $\mathcal{P}$ is the collection of directly indecomposable members of $\mathcal{A}$.

Multiplicatively admissible classes $\mathcal{A}$ split into two distinct types: (i) $\mathcal{A}$ is *discrete* if there is a positive integer λ such that the sizes of members of $\mathcal{A}$ are powers of λ, and (ii) $\mathcal{A}$ is *strictly multiplicative* if it is not discrete. Multiplicatively admissible classes that are discrete can be viewed as additively admissible classes by simply changing the notion of size—for such classes the previous results apply.

The generating series used for $\mathcal{B} \subseteq \mathcal{A}$ in the multiplicative case is the Dirichlet series $\mathbf{B}(x) := \sum_{n\geq 1} b(n)n^{-x}$. Thus for $\mathcal{A}$ a multiplicatively admissible class we have

$$\begin{aligned}\mathbf{A}(x) &:= \sum_{n\geq 1} a(n)n^{-x}, \quad a(1)=1,\\ \mathbf{P}(x) &:= \sum_{n\geq 2} p(n)n^{-x},\end{aligned}$$

and the fundamental identity for $\mathcal{A}$ is

$$\sum_{n\geq 1} a(n)n^{-x} \;=\; \prod_{n\geq 2}\left(1-n^{-x}\right)^{-p(n)} \;=\; \exp\Big(\sum_{m\geq 1}\mathbf{P}(mx)/m\Big). \tag{2.1}$$

$\mathbf{A}(x)$ is often called the *zeta function* for $\mathcal{A}$, and the product in the middle of (2.1) is called the *Euler product* for $\mathcal{A}$.

For $\mathcal{B} \subseteq \mathcal{A}$, α_B is the abscissa of convergence of $\mathbf{B}(x)$. The appropriate choice of asymptotic density for the multiplicative setting is

$$\Delta(\mathcal{B}) \;:=\; \lim_{x\to\alpha_A}\frac{B(x)}{A(x)},$$

provided the limit exists, where $x \to \alpha_A$ means that x approaches α_A from the right. This of course requires that $\alpha_A < \infty$. The notion of Dirichlet density $\partial(\mathcal{B})$ in the multiplicative setting is the obvious analog of the same in the additive setting, namely:

$$\partial(\mathcal{B}) \;:=\; \lim_{x\to\alpha_A}\frac{\mathbf{B}(x)}{\mathbf{A}(x)}.$$

The multiplicative analog of RT is RV, the well-known concept of *regular variation.* The notations $B(x) \in \mathsf{RV}_\alpha$ and $\mathbf{B}(x) \in \mathsf{RV}_\alpha$ both mean that $B(x)$ has *regular variation of index α at infinity*, that is,

$$\lim_{t\to\infty}\frac{B(tx)}{B(t)} \;=\; x^\alpha.$$

$B(x)$ is *slowly varying at infinity* if $B(x) \in \mathsf{RV}_0$.

A guiding principle in the post-1997 work with multiplicatively admissible classes was the belief that every local result for additive classes would have a global analog in the strictly multiplicative setting. Based on this principle, there were three conjectures for multiplicative classes posed in [**15**]:[8]

(C1) Does $\Delta(\mathcal{P}) = 0$ imply $\mathbf{A}(\alpha_A) = \infty$?
(C2) Does $P(x) \in \mathsf{RV}_0$ imply $A(x) \in \mathsf{RV}_0$?

[8]In [**15**] the conjectures were labelled Conjectures 9.70, 10.7 and 11.26.

(C3) Do $\alpha_A > 0$, $P_0(x) := (\log x)x^{-\alpha_A}P(x) \in \mathsf{RV}_0$ and $\liminf_{x\to\infty} P_0(x) > 1$ imply $(\star)$ holds?

All three were subsequently verified (conjecture C3 needed to be revised)—see §3 below.

2.2. Possible generating series for $\mathcal{P}$. In the unlabelled setting there is a tidy description of the possible generating series $\mathbf{P}(x)$ for the class of indecomposables $\mathcal{P}$ of admissible classes $\mathcal{A}$ in the additive [multiplicative] case, provided $\mathbf{P}(x)$ has a positive radius of convergence [a finite abscissa of convergence].

PROPOSITION 2.2. *Let $q(n)$ be a sequence of non-negative integers.*

(a) (Additive Case) *Suppose $\mathbf{Q}(x) := \sum_{n\geq 1} q(n)x^n$ has a positive radius of convergence ρ_Q. Then there is an admissible class $\mathcal{A}$ with $\mathbf{Q}(x)$ being the generating series for the subclass $\mathcal{P}$ of indecomposables.*

(b) (Multiplicative Case) *Suppose $\mathbf{Q}(x) := \sum_{n\geq 2} q(n)n^{-x}$ has a finite abscissa of convergence α_Q. Then there is an admissible class $\mathcal{A}$ with $\mathbf{Q}(x)$ being the generating series for the subclass $\mathcal{P}$ of indecomposables.*

PROOF. For (a) one has $\rho_Q > 0$ iff there is a positive integer M such that $q(n) \leq M^n$ for all $n \geq 1$. There are M^n M-colored chains (as posets) of size n. Let $\mathcal{P}$ be a subclass of M-colored chains with, up to isomorphism, exactly $q(n)$ members of size n, and let $\mathcal{A}$ be the closure of $\mathcal{P}$ under disjoint union.

For (b) one has $\alpha_Q < \infty$ iff there is a positive integer M such that $q(2) + \cdots + q(n) \leq n^M$ for all $n \geq 2$. There are n^M chains (as lattices) of size n with M constants. Let $\mathcal{P}$ be a subclass of such augmented chains with, up to isomorphism, exactly $q(n)$ members of size n, and let $\mathcal{A}$ be the closure of $\mathcal{P}$ under direct product. □

2.3. Partition Classes. Compton's method revolves around one key concept, that of a partition class. First, define $n \star \mathcal{B}$ in the additive [multiplicative] case to mean the collection of disjoint unions [direct products] of n structures from $\mathcal{B}$, repeats allowed among the n structures. Then for $\gamma \subseteq \mathbb{N}$ define $\gamma \star \mathcal{B}$ to be the union of the $n \star \mathcal{B}$ for $n \in \gamma$.

DEFINITION 2.3. A *partition class* of an admissible class $\mathcal{A}$ is a subclass of $\mathcal{A}$ of the form

$$\vec{\gamma} \star \vec{\mathcal{P}} \;:=\; \begin{cases} \gamma_1 \star \mathcal{P}_1 + \cdots + \gamma_k \star \mathcal{P}_k & \text{additive case} \\ \gamma_1 \star \mathcal{P}_1 \times \cdots \times \gamma_k \star \mathcal{P}_k & \text{multiplicative case,} \end{cases}$$

where $\mathcal{P}_1, \ldots, \mathcal{P}_k$ is a partition of $\mathcal{P}$, the class of indecomposable members of $\mathcal{A}$, and the γ_i are subsets of $\mathbb{N}$ which are either finite or co-finite.

The importance of partition classes is made clear in the next two propositions—this is the part of Compton's method that belongs to model theory.

PROPOSITION 2.4. *Let $\mathcal{A}$ be an admissible class.*

($\mathfrak{a}$) *In the additive case, $\mathcal{A}_\varphi$ is always a union of finitely many pairwise disjoint partition classes $\vec{\gamma} \star \vec{\mathcal{P}}$, for φ a MSO sentence.*

[Compton **[26]** (in proof of) Lemma 4.1, 1989; for a presentation in the notation used here, see **[15]** Proposition 6.28, 2001]

($\mathfrak{m}$) *In the multiplicative case, $\mathcal{A}_\varphi$ is always a union of finitely many pairwise disjoint partition classes $\vec{\gamma} \star \vec{\mathcal{P}}$, for φ a FO sentence.*

[Burris and Sárközy [**17**] Theorem 3.4(a), 1997, following results of Burris and Idziak [**16**] Lemmas 10 and 11, 1996; for a presentation in the notation used here, see [**15**] Proposition 12.17, 2001]

Once φ is given, one can add a further restriction concerning which γ_i are actually needed for this proposition, namely there is a positive integer c_φ such that one can assume each γ_i is either one of $0, 1, \ldots, c_\varphi - 1$, or it is the set $\{n \geq c_\varphi\}$.

PROPOSITION 2.5. *Suppose $\mathcal{A}$ is an admissible class of structures.*

($\mathfrak{a}$) *In the additive case, if some partition class of $\mathcal{A}$ does not have asymptotic density* [*in* $\{0, 1\}$] *then there is an admissible class $\widehat{\mathcal{A}}$ such that (i) $\widehat{\mathbf{A}}(x) = \mathbf{A}(x)$ and (ii) $\widehat{\mathcal{A}}$ does not have a MSO limit* [0–1] *law.*

[Burris [**15**] Proposition 6.33, 2001].

($\mathfrak{m}$) *In the multiplicative case, if there is a partition class of $\mathcal{A}$ that does not have asymptotic density* [*in* $\{0, 1\}$] *then there is an admissible class $\widehat{\mathcal{A}}$ such that (i) $\widehat{\mathbf{A}}(x) = \mathbf{A}(x)$ and (ii) $\widehat{\mathcal{A}}$ does not have a FO limit* [0–1] *law.*

[Burris and Sárközy [**17**] Theorem 3.4(b), 1997; Burris, Compton, Odlyzko and Richmond [**18**] Theorem 2.1, 1997].

Compton's method depends on:

(1) showing that $\mathcal{A}_\varphi$ is a finite union of partition classes of $\mathcal{A}$, and then
(2) finding conditions on $\mathbf{A}(x)$ and/or $\mathbf{P}(x)$ that guarantee
  ($\star$) every partition class of $\mathcal{A}$ has asymptotic density.[9]

The model-theoretic part of establishing Compton's method is item (1), which has been successfully completed (Propositions 2.4, 2.5 above)—in the additive case item (1) holds for MSO sentences; in the multiplicative case (1) holds for FO sentences. The proof of item (1) is based on the next two lemmas. For q a positive integer the expression $\mathfrak{S}_1 \equiv_q^{\mathsf{MSO}} \mathfrak{S}_2$ means that $\mathfrak{S}_1$ and $\mathfrak{S}_2$ satisfy the same MSO sentences of quantifier depth q.

LEMMA 2.6 (Compton [**26**], 1989). *Let $\mathcal{A}$ be an admissible class of relational structures. Given a non-negative integer q, let $\equiv_q^{\mathsf{MSO}}$ partition $\mathcal{P}$ into the classes $\mathcal{P}_1, \ldots, \mathcal{P}_k$. There is a positive integer c_q such that:*

(a) *If $m_1, \ldots, m_k$ and $m'_1, \ldots, m'_k$ are such that for each i one has either $m_i = m'_i$ or $m_i, m'_i \geq c_q$ then*

$$\sum_{i=1}^{k} m_i \star \mathcal{P}_i \ \equiv_q^{\mathsf{MSO}} \ \sum_{i=1}^{k} m'_i \star \mathcal{P}_i,$$

meaning that one has the relation $\equiv_q^{\mathsf{MSO}}$ holding between any member $\mathfrak{S}$ of the left side and any member $\mathfrak{S}'$ of the right side.

[9] The property ($\star$) for the multiplicative case was expressed in Burris and Sárközy [**17**], 1997, by saying that $\mathcal{A}$ was 'loaded'. To say that every partition set had asymptotic density either 0 or 1 was expressed in Burris, Compton, Odlyzko and Richmond [**18**], 1997, by the phrase 'front loaded'. This terminology is no longer used.

(b) *Let C_q be the set of k-tuples $\vec{\gamma} := (\gamma_1, \ldots, \gamma_k)$ where each γ_i is one of the coefficients* $0, 1, \ldots, c_q - 1, (\geq c_q)$. *For $\vec{\gamma} \in C_q$ let*

$$\vec{\gamma} \star \vec{\mathcal{P}} \ := \ \sum_{i=1}^{k} \gamma_i \star \mathcal{P}_i,$$

and for φ a MSO sentence of quantifier depth q let

$$S_\varphi \ : = \{\vec{\gamma} \in C_q : \vec{\gamma} \star \vec{\mathcal{P}} \subseteq \mathcal{A}_\varphi\}.$$

Then

$$\mathcal{A}_\varphi \ = \ \bigcup_{\vec{\gamma} \in S_\varphi} \vec{\gamma} \star \vec{\mathcal{P}},$$

a union of finitely many disjoint partition classes.

Proof. A routine application of Ehrenfeucht-Fraïssé games. □

The basic lemma for multiplicative classes, the analog of the above Lemma 2.6 (of Compton), was stated and proved in 1996 by Burris and Idziak ([**16**], Lemmas 10, 11) in the special context of directly representable equational classes. In 1997 it was used in the general multiplicative setting by Burris and Sárközy ([**17**], Lemma 3.1 and Theorem 3.4). For a clear statement and proof in the general setting see Burris [**15**], Proposition 12.17.

Lemma 2.7. (Burris and Idziak [**16**], 1996) *Let $\mathcal{A}$ be a multiplicatively admissible class of structures. Given a first-order sentence φ, there is a positive integer c_φ and a partition of $\mathcal{P}$ into classes $\mathcal{P}_1, \ldots, \mathcal{P}_k$ such that $\mathcal{A}_\varphi$ is a finite union of disjoint partition classes $\vec{\gamma} \star \vec{\mathcal{P}}$, and each $\vec{\gamma}_i$ is one of* $0, 1, 2, \ldots, c_\varphi - 1, (\geq c_\varphi)$.

Key Idea of Proof. Choose a Feferman-Vaught sequence $\langle \Phi; \varphi_1, \ldots, \varphi_r \rangle$ for φ, and then choose Feferman-Vaught sequences $\langle \Phi_i; \varphi_{i,1}, \ldots, \varphi_{i,r_i} \rangle$ for the φ_i. Define the partition of $\mathcal{P}$ by letting two structures $\mathfrak{S}_1$ and $\mathfrak{S}_2$ be equivalent if they satisfy the same $\varphi_{i,j}$. □

Thus when searching for conditions on admissible classes $\mathcal{A}$ which allow one to conclude, just by examining $\mathbf{A}(x)$ and/or $\mathbf{P}(x)$, that $\mathcal{A}$ has a logical limit law, the focus shifts entirely to item (2) above, namely showing that all partition classes have asymptotic density. This requires tools from analysis and combinatorics. When using Compton's method, it is Proposition 2.5 that says one needs to know *every* partition class has asymptotic density, even though it is clear that only countably many such classes can be defined by sentences φ.

The formulation of Compton's approach to asymptotic density for unlabelled structures, as simply requiring that $(\star)$ holds, was used for both the additive and multiplicative cases in Burris's book [**15**], after being introduced for the multiplicative case by Burris and Sárközy [**17**]. In the additive case the main results are for local asymptotic density $\delta(\mathcal{A}_\varphi)$; in the multiplicative case they are for global asymptotic density $\Delta(\mathcal{A}_\varphi)$.

2.4. Number systems. Since Compton's method depends solely on the partition classes of $\mathcal{A}$ having asymptotic density in $\mathcal{A}$, it is convenient in the unlabelled setting to switch the focus from admissible classes to number systems. This is essentially what Compton did in the additive case by selecting representatives for the isomorphism classes. One can view these representatives as the 'numbers' of an additive number system, a system closed under addition with each member having a size and a unique decomposition into indecomposables. The size of the sum of two members is the sum of their sizes, justifying the name 'additive number system'. Likewise one has multiplicative number systems.

Number systems have fundamental identities of the forms we have seen, and, in a natural sense, they are determined by their fundamental identities. Generalized (or abstract) number systems have been studied intensely with regard to the prime number theorem. The goal was to analyze the classical prime number theorem for the integers—and more generally, Landau's prime ideal theorem, for the ideals of the integers of an arithmetic number field—to determine just how little was needed from the properties of the classical number systems in order to prove a prime number theorem. The pioneer in this work was Beurling[10] [**13**] 1937. In 1975 Knopfmacher [**37**] published his first book on the subject, *Abstract Analytic Number Theory*, with a strong emphasis on multiplicative number systems derived from well-known classes of structures such as groups and rings. Burris's book [**15**] adopts the Beurling-Knopfmacher framework of abstract number systems, but replaces the goal of proving a prime number theorem with the goal of proving that all partition sets have asymptotic density in the number system. It turned out that the conditions that had been found for proving a prime number theorem also sufficed to prove all partition sets have asymptotic density. (See Corollary 3.14.)

3. Further results regarding Compton's method

This section presents the main results on limit laws (up to June, 2011) for admissible classes, following the publication of Compton's papers in 1987/1989.

3.1. Directly representable equational classes. Burris and Idziak [**16**], 1996, published the first paper on logical limit laws for multiplicative systems. They showed that a finitely generated directly representable equational class[11] $\mathcal{A}$ has a *discrete* logical limit law, that is, there is a positive integer m such that the probability of a FO sentence φ holding in $\mathcal{A}$ is one of $\{0, 1/m, 2/m, \ldots, 1\}$. Given such an $\mathcal{A}$, let m_0 be the smallest choice of m. Then one always has a sentence φ such that $\mathsf{Prob}_{\mathcal{A}}(\varphi) = 1/m_0$. Furthermore, $\mathcal{A}$ has a 0–1 law iff $m_0 = 1$ iff $\mathcal{A}$ has unique factorization.

Thus, for example, the class of Boolean algebras has a FO 0–1 law. On the other hand, consider the equational class of Abelian groups of exponent 2 with two arbitrary constants; the smallest choice of m is $m_0 = 5$, so one could say that this class has a FO $0\text{–}\frac{1}{5}\text{–}\frac{2}{5}\text{–}\frac{3}{5}\text{–}\frac{4}{5}\text{–}1$ law.

This paper was novel in that some of the classes it dealt with do not have unique factorization, and hence are not admissible. However, thanks to the detailed study of such equational classes by McKenzie [**43**], it is known that the possible factorizations of a finite algebra in such a class are 'well-behaved'.

[10]During WWII the Swedish mathematician Beurling did brilliant work deciphering German codes—later he became a member of the Institute for Advanced Study in Princeton.

[11]These are finitely generated equational classes with only finitely many finite directly indecomposable members. Unfortunately they often lack the unique factorization property, and hence are not admissible classes—but some of these classes, such as Boolean algebras, are indeed admissible.

3.2. Unary functions. In response to Compton's work and questions on unary functions, Woods proved the following.

THEOREM 3.1 (Woods [**56**] Theorems 1.1 and 1.2, 1997). *Let $\mathcal{A}$ be an additively admissible class for which there are constants $C > 0$, $\beta > 0$ and $\alpha < 1$ such that:*

(a) (Labelled):

$$\frac{a(n)}{n!} \sim C\beta^n n^{-\alpha}, \quad \textit{and} \quad \frac{p(n)}{n!} = \mathrm{O}\big(\beta^n/n\big)$$

(b) (Unlabelled):

$$a(n) \sim C(1+\beta)^n n^{-\alpha}, \quad \textit{and} \quad p(n) = \mathrm{O}\big((1+\beta)^n/n\big).$$

Then $\mathcal{A}$ has a MSO limit law, and indeed the asymptotic probability of a MSO sentence φ is given by $\mathsf{Prob}(\varphi) = \partial(\mathcal{A}_\varphi)$.

Woods noted that one could replace the constant C by $L(n)$ where $L(x)$ is slowly varying at infinity. In addition to finding suitable hypotheses, the key step of Woods was to prove appropriate Tauberian theorems, to convert the existence of $\overline{\mu}(\varphi)$ to the existence of $\mu(\varphi)$; and likewise for ν. This theorem was applied to the class $\mathcal{A}$ of a single [partial] unary function with m colors.

3.3. Trees and forests. In a second paper in 1997, Woods [**55**] answered another question posed by Compton.

THEOREM 3.2. *Let $\mathcal{A}$ be the class of m-colored trees. Then $\mathcal{A}$ has both a labelled and an unlabelled MSO limit law.*

The proof started from Compton's observation that, given a positive integer q, the equivalence classes $\mathcal{A}_i$ of m-colored trees under the equivalence relation $\equiv_q^{\mathsf{MSO}}$ satisfy a fairly simple system of equations $\mathcal{A}_i = \Phi_i(\mathcal{A}_1, \ldots, \mathcal{A}_k)$. This gives a system of equations $y_i = \mathbf{G}_i(x, y_1, \ldots, y_k)$ that is solved by $y_i = \mathbf{A}_i(x)$, where the $\mathbf{A}_i(x)$ are generating series for the classes $\mathcal{A}_i$. A combination of extraordinary factors make it possible to show that the asymptotic density of the $\mathcal{A}_i$ exist—in particular the fact that the dependency digraph of the system has a single strong component that immediately dominates all other nodes in the digraph, and the fact that the Jacobian of the $\mathbf{G}_i$ with respect to the y_j is a stochastic matrix. This combination of properties is very rare, and allows for a precision attack using the Perron-Frobenius results on the dominant eigenvalue of a non-negative matrix. Finally, given a MSO sentence φ of quantifier depth q, one has $\mathcal{A}_\varphi$ equal to a union of some of the $\mathcal{A}_i$, so it also has asymptotic density in $\mathcal{A}$.

COROLLARY 3.3. *Let $\mathcal{A}$ be the class of m-colored forests. Then $\mathcal{A}$ has both a labelled and an unlabelled MSO limit law.*

This follows easily, as noted by Compton. Extend the mapping $F \mapsto \bullet/F$, which attaches a root to a forest, to a mapping from a class $\mathcal{F}$ of forests to a class $\mathcal{T} = \bullet/\mathcal{F}$ of trees. If $\mathcal{F}$ is defined by a MSO sentence, then so is $\bullet/\mathcal{F}$. With this simple device one sees that the MSO limit law for trees gives a MSO limit law for forests. This argument lifts to m-colored trees and forests.

Woods' analysis shows that $\mathsf{Prob}(\varphi)$ is positive iff the radius of convergence of $\mathbf{A}_\varphi(x)$ is equal to that of $\mathbf{A}(x)$. At the end of the paper Woods noted that this leads to a labelled MSO limit law for connected acyclic graphs, and that McColm had informed him that the results on trees actually gave a labelled MSO 0–1 law for these graphs—McColm published his proof in [**41**] 2002. Woods ended his paper by asking if there is also an unlabelled MSO limit law for connected acyclic graphs, and if so, is it a 0–1 law. McColm published a positive answer, that indeed there is a 0–1 law, in [**42**] Corollary 2.1, 2004.

3.4. A presentation convention. Much of the research on finding ways to prove $(\star)$ was motivated by the expectation that for every local result in the additive case there would be a corresponding global result in the strictly multiplicative case. So far this expectation has been well-founded—consequently for the rest of this section (on main results) each theorem has two parts, the $(\mathfrak{a})$ part and the $(\mathfrak{m})$ part:

- Item $(\mathfrak{a})$ is for the (unlabelled) additive case (with disjoint union);
- Item $(\mathfrak{m})$ is for the (unlabelled) strictly multiplicative case (with direct product);

and rarely there is a third part, the $L\mathfrak{a}$ part:

- Item $(L\mathfrak{a})$ is for the labelled additive case (with disjoint union).

To avoid being overly repetitious in stating conclusions, let it be noted that every result stated below, which concludes with the existence of a logical limit law for an admissible class $\mathcal{A}$ via Compton's method, can be strengthened to say:

furthermore the probability of each φ is equal to the Dirichlet density $\partial(\mathcal{A}_\varphi)$.

This can be quite useful when explicitly calculating $\mathsf{Prob}(\varphi)$.

3.5. Two consequences of $\mathcal{P}$ having density 0. While Burris was writing the book [**15**], he was fortunate to have Bell, a graduate student at UC San Diego, and Warlimont, a retired German professor living in South Africa, proving key results on the fundamental consequences of the condition $(\star)$. It was easy to prove Theorem 3.7 below, that $(\star)$ implies $\mathcal{P}$ has density 0; but to show that $\mathcal{P}$ having density 0 leads to important restrictions on the generating series was quite challenging.

THEOREM 3.4. *Let $\mathcal{A}$ be an admissible class. Then:*

$(\mathfrak{a})$ $\delta(\mathcal{P}) = 0$ *implies* $\rho_A > 0$.

[Bell [**2**] 2000, Theorem 1(a)]

$(\mathfrak{m})$ $\Delta(\mathcal{P}) = 0$ *implies* $\alpha_A < \infty$.

[Warlimont [**52**], 2001]

Bell actually proved a stronger result than that stated in $(\mathfrak{a})$, namely $\delta(\mathcal{P}) = 0$ implies $\limsup_{n\to\infty} \dfrac{p(n)}{a(n)} = 1$. Consequently, if an admissible class $\mathcal{A}$ with $\rho_A = 0$ has a MSO limit law then, since $\mathcal{P}$ is definable by a MSO sentence, it must be the case that $\delta(\mathcal{P}) = 1$, that almost all members of $\mathcal{A}$ are connected.

THEOREM 3.5. *Let $\mathcal{A}$ be an admissible class. Then:*

$(\mathfrak{a})$ $\delta(\mathcal{P}) = 0$ *implies* $\mathbf{A}(\rho_A) = \infty$.

[Bell, Bender, Cameron and Richmond [**12**] Theorem 1, 2000]

$(\mathfrak{m})$ $\Delta(\mathcal{P}) = 0$ *implies* $\mathbf{A}(\alpha_A) = \infty$.

[Warlimont [**53**], 2003]

REMARK 3.6. In an unpublished note ca. 2002, I. Ruzsa showed that one can easily derive the multiplicative results of the previous two theorems from the additive results.

3.6. Two consequences of the condition $(\star)$**.** The paper [**17**] of Burris and Sárközy is the foundation paper on adapting Compton's method to multiplicative systems, followed by the paper [**18**] of Burris, Compton, Odlyzko and Richmond that treats the case of 0–1 laws in the multiplicative setting.

THEOREM 3.7. *Let $\mathcal{A}$ be an admissible class for which $(\star)$ holds. Then:*

($\mathfrak{a}$) $\delta(\mathcal{P}) = 0$.

[Burris [**15**] Proposition 3.28, 2001]

($\mathfrak{m}$) $\Delta(\mathcal{P}) = 0$.

[Burris and Sárközy [**17**] 1997, Proposition 5.7(c)]

THEOREM 3.8. *Let $\mathcal{A}$ be an admissible class for which $(\star)$ holds. Then:*

($\mathfrak{a}$) $\mathbf{A}(x) \in \mathsf{RT}_{\rho_A}$.

[Burris [**15**] Corollary 3.30, 2001]

($\mathfrak{m}$) $\mathbf{A}(x) \in \mathsf{RV}_{\alpha_A}$.

[Burris and Sárközy [**17**] 1997, Corollary 5.10]

Summarizing the last four theorems gives the following corollary.

COROLLARY 3.9. *Let $\mathcal{A}$ be an admissible class for which $(\star)$ holds. Then:*

($\mathfrak{a}$) $\delta(\mathcal{P}) = 0$, $\rho_A > 0$, $\mathbf{A}(\rho_A) = \infty$, *and* $\mathbf{A}(x) \in \mathsf{RT}_{\rho_A}$.

($\mathfrak{m}$) $\Delta(\mathcal{P}) = 0$, $\alpha_A < \infty$, $\mathbf{A}(\alpha_A) = \infty$, *and* $\mathbf{A}(x) \in \mathsf{RV}_{\alpha_A}$.

3.7. Conditions for 0–1 laws.

LEMMA 3.10.

($\mathfrak{a}$) $\mathbf{A}(x) \in \mathsf{RT}_\rho$ *implies* $\rho_A = \rho$.

($\mathfrak{m}$) $\mathbf{A}(x) \in \mathsf{RV}_\alpha$ *implies* $\alpha_A = \alpha$.

For comparison and symmetry of presentation, the ($\mathfrak{a}$) items in the next two theorems repeat previously stated results of Compton.

THEOREM 3.11. *Let $\mathcal{A}$ be an admissible class.*

($\mathfrak{a}$) *If $\rho_A > 0$ then $\mathcal{A}$ has a MSO 0–1 law iff it has a FO 0–1 law iff* $\mathbf{A}(x) \in \mathsf{RT}_1$.

[Compton [**23**] Theorem 5.9, 1987, and Compton [**26**] Theorem 6.4, 1989]

($\mathfrak{m}$) *If $\alpha_A < \infty$ then $\mathcal{A}$ has a FO 0–1 law iff* $\mathbf{A}(x) \in \mathsf{RV}_0$.

[Burris [**15**] Theorem 10.2, 2001, following on Burris, Compton, Odlyzko and Richmond [**18**] Theorem 2.1(a), 1997].

3.8. Conditions for other laws.

THEOREM 3.12. *Let $\mathcal{A}$ be an admissible class.*

($\mathfrak{a}$) *Suppose $\mathbf{A}(x) \in \mathsf{RT}_\rho$ and there are $C, K > 0$ such that*

$$\frac{a(n-k)}{a(n)} \ \leq \ C\rho^k \text{ for } K \leq k \leq n.$$

Then $(\star)$ holds, so $\mathcal{A}$ has a MSO limit law.

[Compton [**26**] Theorem 6.6 (ii), 1989; Burris [**15**] Theorem 6.31(c), 2001]

($\mathfrak{m}$) *Suppose $A(x) \in \mathsf{RV}_\alpha$, and there are $C, K > 0$ such that*

$$\frac{A(x/k)}{A(x)} \leq Ck^{-\alpha} \text{ for } K \leq k \leq n.$$

Then ($\star$) holds, so $\mathcal{A}$ has a FO limit law.

[Burris and Sárközy [**17**] Theorem 6.3, 1997; Burris [**15**] Proposition 12.19(b), 2001]

After finishing the paper [**17**] on multiplicative classes with Sárközy, Burris initiated a thorough study of what had been achieved in Compton's work on additive classes. It came as a bit of a shock to discover that Compton's result, Theorem 3.12 ($\mathfrak{a}$) for additive classes, looked remarkably similar to the Burris and Sárközy result ($\mathfrak{m}$) for multiplicative classes. Several useful corollaries to Theorem 3.12 ($\mathfrak{m}$) had obvious analogs in additive systems—these analogs had not been noted by Compton. It was at this point that a parallel development of logical limit laws for additive and multiplicative classes seemed possible, and work was started on the book [**15**]—Chapters 1–6 are on additive classes, Chapters 7–12 on multiplicative classes, with Chapter 1 corresponding to Chapter 7, etc.

Corollary 3.13. *Let $\mathcal{A}$ be an admissible class.*

($\mathfrak{a}$) *Suppose there is a $\beta \geq 1$ such that $a(n) \sim \beta^n \cdot \widehat{a}(n)$, $\widehat{a}(n) \in \mathsf{RT}_1$, and $\widehat{a}(n)$ is eventually non-decreasing. Then ($\star$) holds, so $\mathcal{A}$ has a MSO limit law.*

[Burris [**15**] Corollary 5.15, 2001]

($\mathfrak{m}$) *Suppose there is an $\alpha \geq 0$ and $A(x) \sim x^\alpha \cdot \widehat{A}(x)$, where $\widehat{A}(x) \in \mathsf{RV}_0$ and $\widehat{A}(x)$ is eventually non-decreasing. Then ($\star$) holds, so $\mathcal{A}$ has a FO limit law.*

[Burris and Sárközy [**17**] Corollary 6.5, 1997; Burris [**15**] Corollary 11.17, 2001]

The first part of this corollary to be proved was item ($\mathfrak{m}$), for multiplicative classes, by Burris and Sárközy. Later, working on the parallel development for the book [**15**], Burris made the routine translation of ($\mathfrak{m}$) into ($\mathfrak{a}$) to have the corresponding additive result, and realized that this gave a tool to find the first examples of MSO laws for unlabelled classes based on Theorem 3.12 ($\mathfrak{a}$) proved by Compton a decade earlier—these examples were additive classes with non 0–1 MSO limit laws. Item ($\mathfrak{a}$) of Corollary 3.13, combined with results of Knopfmacher, Knopfmacher and Warlimont [**38**], showed that if one had positive constants a, b, C with $1, b < a$ and $p(n) = Ca^n + \mathrm{O}(b^n)$, then $a(n)$ satisfies the hypotheses of ($\mathfrak{a}$). (In this case $a = 1/\rho_A$.) The admissible class $\mathcal{A}$ of 2-colored linear forests has $p(n) = 2^n$, so $\mathcal{A}$ satisfies ($\star$), and thus has a MSO limit law. Since $\rho_A = 1/2$, it cannot be a 0–1 law.

As a special case of Corollary 3.13 we have:

Corollary 3.14. *Let $\mathcal{A}$ be an admissible class.*

($\mathfrak{a}$) *Suppose there is a $\beta \geq 1$ and a constant $C > 0$ such that $a(n) \sim C\beta^n$. Then ($\star$) holds, so $\mathcal{A}$ has a MSO limit law.*

($\mathfrak{m}$) *Suppose there is an $\alpha \geq 0$ and a constant $C > 0$ such that $A(x) \sim Cx^\alpha$. Then ($\star$) holds, so $\mathcal{A}$ has a FO limit law.*

[Burris and Sárközy [**17**] Corollary 6.6, 1997]

The simple condition in item ($\mathfrak{m}$) shows that the requirements of Beurling as well as of Knopfmacher, for a multiplicative prime number theorem, suffice to guarantee that all partition sets have asymptotic density. The class $\mathcal{A}$ of Abelian groups, the original multiplicative example studied by Compton, is covered by Corollary 3.14 ($\mathfrak{m}$) since Erdős and Szekeres [**30**] proved that $A(x) \sim Cx$, where $C := \prod_{n=2}^{\infty} \zeta(n)$.

3.9. Combining admissible classes with a 0–1 law.

THEOREM 3.15. *Let $\mathcal{A}_1, \ldots, \mathcal{A}_n$ be admissible classes.*

($\mathfrak{a}$) *If each $\mathcal{A}_i$ has a MSO 0–1 law and a positive radius of convergence, then $\mathcal{A}_1 + \cdots + \mathcal{A}_n$ has a MSO 0–1 law.*

[Stewart, see [**15**] Theorem 4.15, 2001]

($\mathfrak{m}$) *If each $\mathcal{A}_i$ has a FO 0–1 law and a finite abscissa of convergence, then $\mathcal{A}_1 \times \cdots \times \mathcal{A}_n$ has a FO 0–1 law.*

[Odlyzko, see [**15**] Theorem 10.5, 2001]

3.10. More conditions for 0–1 laws. The search continued for user-friendly ways to apply Theorem 3.11, resp. Theorem 3.12, to obtain interesting examples of admissible classes that had logical 0–1 laws, resp. logical limit laws.

THEOREM 3.16. *Let $\mathcal{A}$ be an admissible class. Then:*

($\mathfrak{a}$) $p(n) = \mathrm{O}(n^c)$ *implies* $\mathbf{A}(x) \in \mathsf{RT}_1$*, showing that $\mathcal{A}$ has a MSO 0–1 law.*[12]

[Bell [**4**] Theorem 1.5, 2001]

($L\mathfrak{a}$) $p(n) = \mathrm{O}\big(n^{\theta n}\big)$ *for some* $0 < \theta < 1$ *implies* $\mathcal{A} \to \infty$*, and thus $\mathcal{A}$ has a labelled MSO 0–1 law.*

[Burris and Yeats [**19**] Theorem 7, 2008]

($\mathfrak{m}$) $P(x) = \mathrm{O}\big((\log x)^c\big)$ *implies* $\mathbf{A}(x) \in \mathsf{RV}_0$*, and thus $\mathcal{A}$ has a FO 0–1 law.*

[Bell [**4**] Theorem 1.8, 2001]

Compton's question (Q3) above is completely answered by Theorem 3.16. After being informed of item ($L\mathfrak{a}$), Compton asked in an e-mail if, in view of his Theorem 1.10 above, the condition on $p(n)$ actually implies $\mathbf{A}(x)$ is Hayman admissible. This is still open.

THEOREM 3.17. *Let $\mathcal{A}$ be an admissible class. Then:*

($\mathfrak{a}$) $\mathbf{P}(x) \in \mathsf{RT}_1$ *implies* $\mathbf{A}(x) \in \mathsf{RT}_1$*, and thus $\mathcal{A}$ has a MSO 0–1 law.*[13]

[Bell and Burris [**6**] Theorem 9.1, 2003]

($\mathfrak{m}$) $\mathbf{P}(x) \in \mathsf{RV}_0$ *implies* $\mathbf{A}(x) \in \mathsf{RV}_0$*, and thus $\mathcal{A}$ has a FO 0–1 law.*

[Bell [**5**] Theorem 17, 2004]

[12]Bell's proof followed in part from his previous study of the Bateman and Erdős paper, as an undergraduate at the University of Waterloo. Shortly after proving that polynomially bounded $p(n)$ led to MSO 0–1 laws, he proved the nearly 50-year old conjecture (in the Bateman-Erdős paper) about an error term for $a^{(k+1)}(n)/a^{(k)}(n)$, where $a^{(k)}(n)$ is the kth difference of $a(n)$. This result was in turn considerably generalized in [**8**], to give estimates for the error term when $p(n)$ is polynomially bounded.

[13]One lemma in the proof of this result showed that $\mathbf{P}(x) \in \mathsf{RT}_1 \Rightarrow e^{\mathbf{P}(x)} \in \mathsf{RT}_1$, which turned out to answer a conjecture of Durrett, Granovsky and Gueron [**29**] connected with problems in coagulation and fragmentation.

In particular this result shows that there are admissible classes $\mathcal{A}$ with a 0–1 law that have $a(n)$ growing much faster than in the cases covered by Theorem 3.16. This is most easily seen by viewing the fundamental identity as a mapping Θ taking the generating series $\mathbf{P}(x)$ to the generating series $\mathbf{A}(x)$. Theorem 3.17 says Θ preserves RT_1 [RV_0] in the additive [multiplicative] setting. Consider the sequence of series $\mathbf{P}(x), \Theta(\mathbf{P}(x)), \Theta^2(\mathbf{P}(x)), \ldots$ From Proposition 2.2, one can regard each of these series as the generating series for the indecomposables of an admissible class with a 0–1 law.

In the additive case, for example, if one starts with $p(n) = 1$ for all n, then the $\Theta^n(\mathbf{P}(x))$ give the partition hierarchy, with the case $n = 1$ being the generating series for partitions of integers (see [**7**]). One can realize this sequence of generating series by looking at the admissible classes $\mathcal{F}_n$ of forests of height at most n. Examining the asymptotics for this hierarchy quickly led to the fact that conditions like $p(n) \sim a \exp\left(bn^c\right)$, where $a, b, c > 0$ and $c < 1$, imply a MSO 0–1 law.

After proving Theorem 3.17, there was the question as to just how far one could have $p(n)$ deviate from RT_1 and still have a 0–1 law. In the additive case this was to a certain extent answered by the sandwich theorems in the 2004 paper of Bell and Burris. The first sandwich theorem in the additive case, and its recently proved multiplicative analog, are stated in the next theorem.

Theorem 3.18 (Sandwich Theorem). *Let $\mathcal{A}_0$ and $\mathcal{A}$ be admissible classes.*

($\mathfrak{a}$) *Suppose $a_0(n) \in \mathsf{RT}_1$ and $p_0(n) \le p(n) = \mathrm{O}(a_0(n))$. Then $\mathbf{A}(x) \in \mathsf{RT}_1$, so $\mathcal{A}$ has a MSO 0–1 law.*

[Bell and Burris [**7**] Theorem 4.4, 2004]

($\mathfrak{m}$) *Suppose $A_0(x) \in \mathsf{RV}_0$ and $P_0(x) \le P(x) = \mathrm{O}(A_0(x))$. Then $\mathbf{A}(x) \in \mathsf{RV}_0$, so $\mathcal{A}$ has a FO 0–1 law.*

[This is a new result—the proof is in §4]

This gives a great deal of freedom to $p(n)$ as n increases. For example, if in the additive setting one has $p_0(n) = 1$ for all n then $a_0(n) \in \mathsf{RT}_1$, indeed,

$$a_0(n) \ \sim \ \frac{\exp\left(\pi\sqrt{2n/3}\right)}{4\sqrt{3}n}.$$

Thus for any constant $C > 0$, if $1 \le p(n) \le C \cdot \mathsf{part}(n)$, where $\mathsf{part}(n)$ is the number of partitions of n, then one has $a(n) \in \mathsf{RT}_1$, and any associated admissible class $\mathcal{A}$ will have a MSO 0–1 law. A natural example that has a MSO 0–1 law based on ($\mathfrak{a}$) is the class $\mathcal{G}_n$ of acyclic graphs of diameter at most n.

There is a second sandwich theorem in the additive case that allows the condition $p_0(n) \le p(n)$ to fail for finitely many values of n, that is, $p(n)$ is eventually greater or equal to $p_0(n)$. It requires a considerable strengthening of the other hypotheses.

Theorem 3.19 (Eventual Sandwich Theorem). *Let $\mathcal{A}$ be an admissible class. Then:*

($\mathfrak{a}$) *Suppose $p_0(n) \in \mathsf{RT}_1$, $p_0(n) \le p(n)$ for n sufficiently large, $p(n) = \mathrm{O}(a_0(n))$, and $\sum_n (p(n) - p_0(n)) \ge 0$. Then $\mathbf{A}(x) \in \mathsf{RT}_1$, so $\mathcal{A}$ has a MSO 0–1 law.*

[Bell and Burris [**7**] Theorem 5.3, 2004]

($\mathfrak{m}$) (*So far there is no analog for the multiplicative case.*)

Thus in the example for the previous theorem, with $p_0(n) = 1$, one can allow $p(n) = 0$ for finitely many values of n provided the places where $p(n)$ exceeds 1 compensates for the places where it took on the value 0.

3.11. Combining admissible classes with general limit laws. Compton's method for proving logical limit laws, based solely on properties of the generating functions, succeeds precisely when $(\star)$ holds, that is, all partition sets have asymptotic density. What happens if one combines admissible classes $\mathcal{A}_i$ which satisfy $(\star)$? (Combining admissible classes with a 0–1 law was considered in Theorem 3.15.)

The definitive study on this topic is due to Yeats. Unlike other results in this article, this cannot be reduced to the study of admissible classes $\mathcal{A}_i$ that satisfy our standing assumption that the $d_i = 1$, where $d_i := \gcd \mathsf{Spec}(\mathcal{A}_i)$. For the next theorem (and only for the next theorem), the d_i are unrestricted in the $(\mathfrak{a})$ part; and in the $(\mathfrak{m})$ part one allows discrete as well as the usual strictly multiplicative classes.

THEOREM 3.20. *Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be admissible classes satisfying $(\star)$. For $i = 1, 2$ let $d_i = \gcd \mathsf{Spec}(\mathcal{A}_i)$.*

$(\mathfrak{a})$ *Let ρ_i be the radius of convergence of $\mathcal{A}_i$, with $\rho_1 \le \rho_2$. Then $\mathcal{A}_1 + \mathcal{A}_2$ satisfies $(\star)$ iff $d_1 \,\big|\, d_2$ or $\rho_1 = \rho_2$.*

[Yeats [**57**] Theorem 57, 2002]

$(\mathfrak{m})$ *Let α_i be the abscissa of convergence of $\mathcal{A}_i$, with $\alpha_1 \ge \alpha_2$. Then $\mathcal{A}_1 \times \mathcal{A}_2$ satisfies $(\star)$ iff*

- *$\alpha_1 = \alpha_2$, or*
- *$\alpha_1 > \alpha_2$ and $\mathcal{A}_1$ is strictly multiplicative, or*
- *$\alpha_1 > \alpha_2$ and $\mathcal{A}_1, \mathcal{A}_2$ are both discrete and λ_2 is a power of λ_1.*

[Yeats [**57**] Theorem 68, 2002]

3.12. More conditions for general limit laws.

THEOREM 3.21. *Let $\mathcal{A}$ be an admissible class. Then:*

$(\mathfrak{a})$ *$\rho_A > 0$, $p(n) \in \mathsf{RT}_{\rho_A}$ and $\liminf\limits_{n\to\infty} n{\rho_A}^n p(n) > 1$ imply $(\star)$ holds, so $\mathcal{A}$ has a MSO limit law.*

[Bell and Burris [**6**] Theorem 9.3, 2003]

$(\mathfrak{m})$ *$\alpha_A > 0$, $P(x) \sim x^{\alpha_A} P_0(x) \big/ \log x$ for some non-decreasing $P_0(x) \in \mathsf{RV}_0$, and $\lim\limits_{x\to\infty} P_0(x) \in (1/\alpha_A, \infty)$ imply $(\star)$ holds, so $\mathcal{A}$ has a FO 0–1 law.*

[Bell [**5**] Theorem 18, 2004]

This theorem is a favorite for finding logical limit laws that are not 0–1 laws—in the additive setting it is more general than our original method for finding such laws using the asymptotic results of Knopfmacher, Knopfmacher and Warlimont mentioned in §3.8. An interesting example is the class of forests of planted plane trees of height at most h.

4. The multiplicative sandwich theorem

The multiplicative analogue of the additive sandwich theorem has not appeared in the literature. In this chapter we fill this gap by proving the following result.

THEOREM 4.1. *Suppose that*

$$\mathbf{A}_0(s) = \sum_{n=1}^{\infty} a_0(n) n^{-s} = \prod_{j \geq 2} (1 - 1/j^s)^{-p_0(j)}$$

has the property that $p_0(n)$ is a nonnegative integer for all n and $A_0(x)$ is slowly varying at infinity. If

$$\mathbf{A}(s) = \sum_{n=1}^{\infty} a(n) n^{-s} = \prod_{j \geq 2} (1 - 1/j^s)^{-p(j)}$$

has the property that

$$P_0(x) \leq P(x) = \mathrm{O}(A_0(x))$$

for all $x \geq 1$, then $A(x)$ is slowly varying at infinity. Thus any multiplicatively admissible class $\mathcal{A}$ with generating series $\mathbf{A}(x)$ will have a first-order 0–1 law.

We give a streamlined proof that uses Theorem 3.16 ($\mathfrak{m}$). It is possible to give a slightly longer proof that does not use this result. Let $\mathbf{R}(s) = \sum_{n \geq 1} r(n) n^{-s}$ be a Dirichlet series with non-negative coefficients. Note that

$$-\frac{d}{ds}\mathbf{R}(s) = \sum_{n \geq 1} r(n) \log(n) n^{-s}.$$

Let $\overline{R}(x)$ denote the global count function of $-\dfrac{d}{ds}\mathbf{R}(s)$; that is,

$$\overline{R}(x) = \sum_{n \leq x} r(n) \log n. \tag{4.1}$$

We need a few simple estimates about the global count functions of products of Dirichlet series. These results are not optimal.

LEMMA 4.2. *Let $\mathbf{A}(s)$ be a Dirichlet series with nonnegative coefficients and suppose that $A(x)$ is slowly varying at infinity. If*

$$\mathbf{B}(s) = \mathbf{A}(s) \prod_{j=1}^{r} (1 - 1/m_j^s)^{-1},$$

then we have the following:

(i) *$B(x)$ is slowly varying at infinity and $A(x) = \mathrm{o}(B(x))$,*

(ii) *if $r = 1$, $\overline{B}(x) \geq A(x)(\log x)/2$ for x sufficiently large;*

(iii) *if $r > 1$, $A(x)\log(x) = \mathrm{o}(\overline{B}(x))$.*

PROOF. We first consider (i) in the case that $\mathbf{B}(s) = \mathbf{A}(s)(1 - 1/m^s)^{-1}$. In this case $B(x) = \sum_{j \leq \log x/\log m} A(x/m^j)$, and so $B(x) - B(x/m) = A(x)$. To show that $B(x)$ is slowly varying at infinity, it is therefore sufficient to show that $A(x) = \mathrm{o}(B(x))$. Note, however, that $A(x/m)/A(x) \to 1$ as $x \to \infty$; hence if k is fixed we see that

$$\liminf_{x \to \infty} \frac{B(x)}{A(x)} \;\geq\; \lim_{x \to \infty} \frac{A(x) + A(x/m) + \cdots + A(x/m^{k-1})}{A(x)} \;=\; k.$$

Thus $A(x) = \mathrm{o}(B(x))$. This demonstrates (i) in the case that $r = 1$. The general case follows by induction on r.

To prove (ii), we have $r = 1$. Again assume that

$$\mathbf{B}(s) = \mathbf{A}(s)(1 - 1/m^s)^{-1}.$$

Fix a positive number x. For $n \le x$, let c_n denote the largest nonnegative integer satisfying $m^{c_n} \le x/n$. Then $x/m < nm^{c_n} \le x$. Moreover

$$b(nm^{c_n}) \ \ge \ a(nm^{c_n}) + a(nm^{c_n-1}) + \cdots + a(n).$$

Note that

$$\begin{aligned}
\overline{B}(x) &= \sum_{n \le x} b(n) \log n \\
&\ge \sum_{\substack{n \le x \\ m \nmid n}} b(nm^{c_n}) \log(nm^{c_n}) \\
&\ge \sum_{\substack{n \le x \\ m \nmid n}} b(nm^{c_n}) \log(x/m) \\
&\ge (\log(x/m)) \sum_{\substack{n \le x \\ m \nmid n}} (a(n) + a(nm) + \cdots + a(nm^{c_n})) \\
&\ge \frac{\log(x)}{2} \sum_{k \le x} a(k) \quad \text{(for } x \text{ sufficiently large)} \\
&= \log(x) A(x)/2.
\end{aligned}$$

For (iii), suppose that $r \ge 2$. Let $\mathbf{A}_0(s) = \prod_{j=1}^{r-1}(1 - 1/m_j^s)^{-1}\mathbf{A}(s)$. By (i), $\mathbf{A}_0(s)$ is slowly varying at infinity. By (ii), we have $\overline{B}(x) \ge \log(x)A_0(x)/2$ for x sufficiently large. But by (i), $A(x) = \mathrm{o}\big(A_0(x)\big)$. Thus we obtain (iii). $\square$

For functions $\mathbf{B}(s) = \prod_{j=1}^{r}(1 - 1/m_j^s)^{-1}$, the global count functions are well understood (Bell [**3**] Theorem 3.5, 2001):

$$\mathbf{B}(s) = \prod_{j=1}^{r}(1 - 1/m_j^s)^{-1} \quad \Rightarrow \quad B(x) = C \cdot (\log x)^r + \mathrm{O}\big((\log x)^{r-1}\big) \tag{4.2}$$

where

$$C \ = \ \frac{1}{r!(\log 2)\cdots(\log r)}. \tag{4.3}$$

We introduce the concept of *domination*, which will be used in giving a criterion for the product of two Dirichlet series to have a global count function that is slowly varying at infinity.

Definition 4.3. *Let* $\mathbf{F}(s) = \sum_{n\ge1} f(n)/n^s$ *and* $\mathbf{G}(s) = \sum_{n\ge1} g(n)/n^s$ *be two Dirichlet series with nonnegative coefficients. We say* $\mathbf{F}(s)$ *is* dominated by $\mathbf{G}(s)$ *if* $F(x) = \mathrm{o}(G(x))$.

Lemma 4.4. *Suppose* $\mathbf{R}(s)$ *is a Dirichlet series with nonnegative coefficients. Then either* $R(x)$ *is bounded or* $\mathbf{R}(s)$ *is dominated by* $-\mathbf{R}'(s)$.

Proof. Let $\varepsilon > 0$. Then

$$\overline{R}(x) \ = \ \sum_{n \le x} r(n) \log n$$

$$
\begin{aligned}
&\geq \sum_{e^{1/\varepsilon} \leq n \leq x} r(n) \log n \\
&\geq \frac{1}{\varepsilon} \sum_{e^{1/\varepsilon} \leq n \leq x} r(n) \\
&\geq \frac{1}{\varepsilon} \left(R(x) - R(\exp(1/\varepsilon))\right) \\
&= \frac{1}{\varepsilon} \cdot R(x) + \mathrm{O}(1).
\end{aligned}
$$

It follows that either $R(x)$ is bounded or

$$
\liminf_{x \to \infty} \frac{\overline{R}(x)}{R(x)} \;\geq\; \frac{1}{\varepsilon}.
$$

□

LEMMA 4.5. *Let*

$$
\mathbf{A}(s) \;=\; \sum_{n \geq 1} a(n) n^{-s} \text{ and } \mathbf{B}(s) \;=\; \sum_{n \geq 1} b(n) n^{-s}
$$

be two Dirichlet series with nonnegative coefficients. If $\mathbf{A}(s)$ is slowly varying at infinity and $\mathbf{B}(s)$ is dominated by $\mathbf{C}(s) = \mathbf{A}(s)\mathbf{B}(s)$, then $\mathbf{A}(s)\mathbf{B}(s)$ is also slowly varying at infinity.

PROOF. Fix $\varepsilon > 0$. Then we have $C(x) = \sum_{n \leq x} A(x/n) b(n)$. Since $\mathbf{A}(s)$ is slowly varying at infinity, we can find a positive number $M \geq 1$ such that $A(2x) - A(x) < \varepsilon A(2x)$ for $x > M$. Furthermore, as $\mathbf{B}(s)$ is dominated by $\mathbf{C}(s)$ we have

$$
2A(2M)B(2x) \leq \varepsilon C(2x)
$$

for all x sufficiently large. Consequently,

$$
\begin{aligned}
C(2x) - C(x) &= \sum_{n \leq 2x} A(2x/n) b(n) - \sum_{n \leq x} A(x/n) b(n) \\
&= \sum_{n \leq x} \left(A(2x/n) - A(x/n)\right) b(n) + \sum_{x < n \leq 2x} A(2x/n) b(n) \\
&\leq \sum_{x/M \leq n \leq x} A(2x/n) b(n) + \varepsilon \sum_{n < x/M} A(2x/n) b(n) + A(1)B(2x) \\
&\leq A(2M)B(x) + \varepsilon \sum_{n \leq 2x} A(2x/n) b(n) + A(1)B(2x) \\
&\leq \varepsilon C(2x) + 2A(2M)B(2x) \\
&\leq 2\varepsilon C(2x),
\end{aligned}
$$

for all x sufficiently large. The result follows. □

LEMMA 4.6. *Let $\mathbf{H}(s) = \sum_{n \geq 1} h(n) n^{-s}$ be a Dirichlet series with nonnegative coefficients such that $h(1) = 0$ and let $\mathbf{A}(s) = \exp(\mathbf{H}(s))$. If $-\mathbf{H}'(s)$ is dominated by $\mathbf{B}(s)$, then either $\mathbf{A}(s)$ is dominated by $\mathbf{C}(s) = \mathbf{B}(s)\mathbf{A}(s)$ or $A(x)$ is uniformly bounded.*

PROOF. Let $\varepsilon > 0$. Since $-\mathbf{H}'(s)$ is dominated by $\mathbf{B}(s)$, there is $M > 0$ such that $\overline{H}(x) \le \varepsilon B(x)$ for $x > M$.

We note that $-\mathbf{A}'(s) = -\mathbf{H}'(s)\mathbf{A}(s)$ and hence

$$\begin{aligned}
\overline{A}(x) &= \sum_{n\le x} \overline{H}(x/n)a(n) \\
&= \sum_{n<x/M} \overline{H}(x/n)a(n) + \sum_{x/M\le n\le x} \overline{H}(x/n)a(n) \\
&\le \varepsilon \sum_{n<x/M} B(x/n)a(n) + \sum_{x/M\le n\le x} \overline{H}(x/n)a(n) \\
&= \varepsilon C(x/M) + \sum_{x/M\le n\le x} \overline{H}(x/n)a(n) \\
&\le \varepsilon C(x/M) + \overline{H}(M)A(x) \\
&\le \varepsilon C(x/M) + \varepsilon B(M)A(x) \\
&\le \varepsilon C(x) + \varepsilon B(M)A(x).
\end{aligned}$$

By Lemma 4.4, either $\overline{A}(x)/A(x) \to \infty$ or $A(x)$ is uniformly bounded. If $A(x)$ is uniformly bounded, we are done. Otherwise, we have $C(x)/A(x) \to \infty$, so $\mathbf{A}(s)$ is dominated by $\mathbf{C}(s)$. □

PROOF OF THEOREM 4.1. Let $\mathbf{P}_0(s)$ and $\mathbf{P}(s)$ denote respectively the Dirichlet generating series for $p_0(n)$ and $p(n)$. If $\sum_n p_0(n) < \infty$, then $A_0(x)$ is polylog bounded, by Lemma 4.2, and hence $\overline{P}(x)$ is polylog bounded. This case follows from Theorem 3.16 ($\mathfrak{m}$).

Now consider the case that $\sum_n p_0(n) = \infty$. Then there exist distinct integers n_1 and n_2 with $p_0(n_1) + p_0(n_2) \ge 2$. Let

$$\mathbf{P}_1(s) = \sum_n p_1(n)/n^s := \mathbf{P}_0(s) - 1/n_1^s - 1/n_2^s \tag{4.4}$$

and

$$\mathbf{P}_2(s) = \sum_n p_2(n)/n^s := \mathbf{P}(s) - \mathbf{P}_1(s). \tag{4.5}$$

Similarly, we define

$$\mathbf{A}_1(s) = \sum_n a_1(n)/n^s := \prod_j (1 - 1/j^s)^{-p_1(j)} \tag{4.6}$$

and

$$\mathbf{A}_2(s) = \sum_n a_2(n)/n^s := \prod_j (1 - 1/j^s)^{-p_2(j)}. \tag{4.7}$$

Since $\mathbf{P}(s) = \mathbf{P}_1(s) + \mathbf{P}_2(s)$, we have

$$\mathbf{A}(s) = \mathbf{A}_1(s)\mathbf{A}_2(s).$$

Our goal is to show that $A_1(x)$ is slowly varying at infinity and that $\mathbf{A}_2(s)$ is dominated by $\mathbf{A}(s)$. Once we show this, we can use Lemma 4.5 to infer that $A(x)$ is also slowly varying at infinity.

We note that $\mathbf{A}_1(s) = \mathbf{A}_0(s)(1 - 1/n_1^s)^{-1}(1 - 1/n_2^s)^{-1}$. Since $A_0(x)$ is slowly varying at infinity, $A_1(x)$ is also slowly varying at infinity by part (i) of Lemma

4.2. Furthermore, $A_0(x)\log(x) = \mathrm{o}(\overline{A}_1(x))$ by part (iii) of Lemma 4.2. Note that $\mathbf{A}_2(s) = \exp(\mathbf{H}_2(s))$, where

$$\mathbf{H}_2(s) \;=\; \sum_{n\geq 1} h_2(n)n^{-s}$$

and

$$h_2(n) \;=\; \sum_{j^\ell=n} p_2(j)/\ell.$$

It follows that

$$\begin{aligned}
\overline{H}_2(x) &= \sum_{n\leq x} h_2(n)\log n \\
&= \sum_{n\leq x} \log(n) \sum_{j^\ell=n} p_2(j)/\ell \\
&\leq \sum_{2\leq j\leq x}\ \sum_{\ell\leq \log x/\log j} \log(j^\ell)p_2(j)/\ell \\
&= \sum_{2\leq j\leq x}\ \sum_{\ell\leq \log x/\log j} (\log j)p(j) \\
&\leq \log(x)P(x) \\
&\leq C\log(x)A_0(x) \quad \text{for some } C>0 \\
&= \mathrm{o}(\overline{A}_1(x)).
\end{aligned}$$

Thus we see that $-\mathbf{H}_2'(s)$ is dominated by $\mathbf{A}_1(s)$. It follows that $\mathbf{A}_2(s)$ is dominated by $\mathbf{A}_1(s)\mathbf{A}_2(s) = \mathbf{A}(s)$. The result follows. □

References

1. P.T. Bateman and P. Erdős, *Monotonicity of partition functions*, Mathematika **3** (1956), 1–14.
2. Jason P. Bell, *When structures are almost surely connected,* Electron. J. Combin. **7** (2000), R36 (8 pp).
3. ———, *A proof of a partition conjecture of Bateman and Erdős,* J. Number Theory **87** (2001), 144–153.
4. ———, *Sufficient conditions for zero-one laws,* Trans. Amer. Math. Soc. **354** (2001), no. 2, 613–630.
5. ———, *Dirichlet series whose partial sums of coefficients have regular variation*, Israel J. Math. **144** (2004), 343–365.
6. Jason P. Bell and Stanley N. Burris, *Asymptotics for logical limit laws: when the growth of the components is in an RT class*, Trans. Amer. Math. Soc. **355** (2003), 3777–3794.
7. ———, *Partition identities I. Sandwich theorems and logical 0-1 laws,* Electron. J. Combin. **11**(1) (2004), RS49 (25 pp).
8. ———, *Partition identities II. The results of Bateman and Erdős,* J. Number Theory **117** No. 1 (2006), 160–190.
9. Jason Bell, Stanley N. Burris and Karen Yeats, *Spectra and systems of equations,* this volume.
10. Jason P. Bell, Stanley N. Burris and Karen A. Yeats, *Counting Rooted Trees: The Universal Law* $t(n) \sim C\cdot\rho^{-n}\cdot n^{-3/2}$, Electron. J. Combin. **13** (2006), R63 (64 pp.)
11. ———, *Monadic Second Order Classes of Trees of Radius 1*, (Prepint.)
12. ———, Edward A. Bender, Peter J. Cameron and L. Bruce Richmond, *Asymptotics for the probability of connectedness and the distribution of number of components,* Electron. J. Combin. **7** (2000), R33 (22 pp).
13. Arne Beurling, *Analyse de la loi asymptotique de la distribution des nombres premiers généralisés, I*, Acta Math. **68** (1937), 255–291.

14. Stanley Burris, *Spectrally determined first-order limit laws,* Logic and Random Structures (New Brunswick, NJ, 1995), 33–52, ed. by Ravi Boppana and James Lynch. DIMACS Ser. Discrete Math. Theoret. Comput. Sci. **33**, Amer. Math. Soc., Providence, RI, 1997.
15. ______, *Logical Limit Laws and Number Theoretic Density*, Math. Surveys Monogr., Vol. **86**, Amer. Math. Soc., 2001.
16. ______ and Paweł Idziak, *A directly representable variety has a discrete first-order law,* Internat. J. Algebra Comput. **6** (1996), 269–276.
17. ______ and András Sárközy, *Fine spectra and limit laws I. First-order laws,* Canad. J. Math. **49** (1997), 468–498.
18. ______, K. Compton, A. Odlyzko and B. Richmond, *Fine spectra and limit laws II. First-order 0–1 laws,* Canad. J. Math. **49** (1997), 641–652.
19. ______ and Karen A. Yeats, *Sufficient conditions for a labelled 0–1 law,* Discrete Math. Theor. Comput. Sci. **10**, no. 1 (2008), 147–156.
20. Peter J. Cameron, *On the probability of connectedness,* 15th British Combinatorics Conference (Stirling 1995), Discrete Math. **167/168** (1997), 175–187.
21. A. Cayley, *On the theory of the analytical forms called trees*, Phil. Magazine **13** (1857), 172–176.
22. K. J. Compton, *Application of a Tauberian theorem to finite model theory,* Arch. Math. Logik Grundlag. **25** (1985), 91–98.
23. ______, *A logical approach to asymptotic combinatorics I: first order properties,* Adv. Math. **65** (1987), 65–96.
24. ______, C. W. Henson and S. Shelah, *Nonconvergence, undecidability, and intractability in asymptotic problems*, Ann. Pure Appl. Logic, **36** (1987), 207–224.
25. ______, *The computational complexity of asymptotic problems I: Partial orders,* Inform. and Comput., **78** (1988), 108–123.
26. ______, *A logical approach to asymptotic combinatorics II: monadic second-order properties,* J. Combin. Theor. Ser. A, **50** (1989), 110–131.
27. ______, *Laws in logic and combinatorics,* Algorithms and Order (Ottawa, ON, 1987), 353–383, ed. by I. Rival. Kluwer Acad. Publ., 1989.
28. Arnaud Durand, N.D. Jones, J.A. Makowsky and M. More, *Fifty years of the spectrum problem,* (Preprint, July, 2009).
29. R. Durrett, B. Granovsky and S. Gueron, *The equilibrium behavior of reversible coagulation-fragmentation processes,* J. Theoret. Probab. **12** (1999), 447–474.
30. P. Erdős and G. Szekeres, *Über die Anzahl der Abelschen Gruppen gegebener Ordnung und über ein verwandtes zahlentheoretisches Problem,* Acta Litt. Sci. Reg. Univ. Hungar. Fr.–Jos. Sect. Sci. Math. **7** (1934), 94–103.
31. R. Fagin, *Probabilities on finite models,* J. Symbolic Logic, **41** (1976), 50–58.
32. J.L. Geluk and L. de Haan, *Regular variation, extensions and Tauberian theorems,* CWI Tract **40**, Centre for Mathematics and Computer Science, 1987.
33. Y. V. Glebskiĭ, D. I. Kogan, M. I. Liogon'kiĭ and V. A. Talanov, *Volume and fraction of satisfiability of formulas of the lower predicate calculus,* (Russian) Kibernetika (Kiev), **5** (1969), 17–27. English Translation in Cybernetics **5** (1972), 142–154.
34. G.H. Hardy and S. Ramanujan, *Asymptotic formulae for the distribution of integers of various types,* Proc. Lond. Math. Soc. (2) **16** (1917), 112–132.
35. G.H. Hardy and E.M. Wright, *An Introduction to the Theory of Numbers,* Oxford Press, 5th ed., 1980.
36. Paweł M. Idziak and Jerzy Tyszkiewicz, *Monadic second order probabilities in algebra. Directly representable varieties and groups,* Logic and random structures (New Brunswick, NJ, 1995), 79–107, DIMACS Ser. Discrete Math. Theoret. Comput. Sci., **33**. Amer. Math. Soc., Providence, RI, 1997.
37. J. Knopfmacher, *Abstract Analytic Number Theory,* North-Holland Mathematical Library, Vol. 12, North-Holland Publishing Co., Amsterdam-Oxford; American Elsevier Publishing Co., Inc., New York, (1975).
38. Arnold Knopfmacher, John Knopfmacher and Richard Warlimont, *"Factorisatio numerorum" in arithmetical semigroups,* Acta Arith. **61** (1992), 327–336.
39. M.I. Liogon'kii, M. I. *On the question of quantitative characteristics of logical formulas,* Kibernetika (Kiev) 1970, No. 3, 16–22; English translation, Cybernetics **6**, 205-211.

40. J.F. Lynch, *Probabilities of first-order sentences about unary functions,* Trans. Amer. Math. Soc. **287** (1985), 543–568.
41. Gregory L. McColm, *MSO zero-one laws on random labelled acyclic graphs,* Discrete Math. **254** (2002), 331–347.
42. ______, *On the structure of random unlabelled acyclic graphs,* Discrete Math. **277** (2004), 147–170.
43. R.N. McKenzie, *Narrowness implies uniformity,* Algebra Universalis **15** (1983), 543–568.
44. Ralph N. McKenzie, George F. McNulty and Walter F. Taylor, *Algebras, Lattices, and Varieties. I.* The Wadsworth and Brooks/Cole Mathematics Series, Wadsworth and Brooks/Cole, 1987.
45. Günter Meinardus, *Asymptotische Aussagen über Partitionen,* Math. Z. **59** (1954), 388–398.
46. A. Meir and J.W. Moon, *Some asymptotic results useful in enumeration problems,* Aequationes Math. **33** (1987), 260–268.
47. A.M. Odlyzko, *Asymptotic enumeration methods,* Handbook of Combinatorics Vol. II, R.L. Graham, M. Grötschel and L. Lovász, eds., 1063–1230, Elsevier, 1995.
48. G. Pólya and R.C. Read, *Combinatorial enumeration of groups, graphs and chemical compounds*, Springer Verlag, New York, 1987.
49. G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis. I,* Springer-Verlag, 1970.
50. Irme Ruzsa, *Two results on density*, 2002 preprint.
51. Richard Warlimont, *About the radius of convergence of the zeta function of an additive arithmetical semigroup. Dedicated to the memory of John Knopfmacher,* Quaest. Math. **24** (2001), no. 3, 355–362.
52. ______, *About the abscissa of convergence of the zeta function of a multiplicative arithmetical semigroup. Dedicated to the memory of John Knopfmacher*, Quaest. Math. **24** (2001), no. 3, 363–371.
53. ______, *On the zeta function of an arithmetical semigroup*, Math. Z. **245** (2003), no. 3, 419–434.
54. Herbert S. Wilf, *Generatingfunctionology,* 2nd ed., Academic Press, Inc., 1994.
55. Alan R. Woods, *Coloring rules for finite trees, probabilities of monadic second-order sentences*, Random Structures Algorithms **10** (1997), 453–485.
56. ______, *Counting finite models,* J. Symbolic Logic **62** (1997), 925–949.
57. Karen Yeats, *Asymptotic density in combined number systems,* New York J. Math. **8** (2002), 63–83.
58. Karen Yeats, *A multiplicative analog of Schur's Tauberian theorem,* Can. Math. Bull. **46** (2003), 473–480.

Department of Mathematics, Simon Fraser University, 8888 University Dr., Burnaby, BC,V5A 1S6, Canada
E-mail address: jpb@math.sfu.ca

Department of Pure Mathematics, University of Waterloo, Waterloo, ON, N2L 3G1, Canada
E-mail address: snburris@math.uwaterloo.ca

Contemporary Mathematics
Volume **558**, 2011

Logical complexity of graphs: a survey

Oleg Pikhurko and Oleg Verbitsky

Abstract. We discuss the definability of finite graphs in first-order logic with two relation symbols for adjacency and equality of vertices. The *logical depth* $D(G)$ of a graph G is equal to the minimum quantifier depth of a sentence defining G up to isomorphism. The *logical width* $W(G)$ is the minimum number of variables occurring in such a sentence. The *logical length* $L(G)$ is the length of a shortest defining sentence. We survey known estimates for these graph parameters and discuss their relations to other topics (such as the efficiency of the Weisfeiler-Lehman algorithm in isomorphism testing, the evolution of a random graph, quantitative characteristics of the zero-one law, or the contribution of Frank Ramsey to the research on Hilbert's *Entscheidungsproblem*). Also, we trace the behavior of the descriptive complexity of a graph as the logic becomes more restrictive (for example, only definitions with a bounded number of variables or quantifier alternations are allowed) or more expressible (after powering with counting quantifiers).

Contents

1991 *Mathematics Subject Classification.* Primary 03C13, 68Q19; Secondary 05C60, 05C80.

The first author was partially supported by the National Science Foundation (Grant DMS-0758057) and the Alexander von Humboldt Foundation.

The second author was partially supported by the Alexander von Humboldt Foundation.

1. Introduction

1.1. Basic notions and examples. We consider the first-order language of graph theory whose vocabulary contains two relation symbols $\sim$ and $=$, respectively for adjacency and equality of vertices. The term *first-order* imposes the condition that the variables represent vertices and hence the quantifiers apply to vertices only. Without quantification over sets of vertices, we are unable to express by a single formula some basic properties of graphs, such as being bipartite, being connected, etc. (see, e.g., [**72**, Theorems 2.4.1 and 2.4.2]). However, first-order logic is powerful enough to define any *individual* graph. How succinctly this can be done is the subject of this article.

As a starting example, let us say in the first-order language that vertices x and y are at distance at most n from one another. A possible formula $\Delta_n(x, y)$ can look as follows:

$$(1)\quad \Delta_1(x,y) \stackrel{\text{def}}{=} x \sim y \vee x = y,$$

$$(2)\quad \Delta_n(x,y) \stackrel{\text{def}}{=} \exists z_1 \ldots \exists z_{n-1}\Big(\Delta_1(x,z_1) \wedge \bigwedge_{i=1}^{n-2} \Delta_1(z_i,z_{i+1}) \wedge \Delta_1(z_{n-1},y)\Big).$$

By a *sentence* we mean a first-order formula where every variable is bound by a quantifier. If we specify a graph G, a sentence Φ is either *true* or *false* on it. If H is a graph isomorphic to G, then Φ is either true or false on G and H simultaneously. In other words, first-order logic cannot distinguish between isomorphic graphs. In general, we say that a sentence Φ *distinguishes* a graph G from another graph H if Φ is true on G but false on H.

For example, sentence $\forall x \forall y\, \Delta_1(x, y)$ distinguishes a complete graph K_n from any graph H that is not complete. The sentence $\forall x \forall y\, \Delta_{n-1}(x, y)$ distinguishes P_n, the path with n vertices, from any longer path P_m, $m > n$.

Throughout this survey we consider only graphs whose vertex set is finite and non-empty. We say that a sentence Φ *defines* a graph G (up to isomorphism) if Φ distinguishes G from every non-isomorphic graph H.

For example, the single-vertex graph P_1 is defined by sentence $\forall x \forall y\,(x = y)$. If $n \geq 2$, then the path P_n is defined by

$$
\begin{aligned}
&\forall x \forall y \Delta_{n-1}(x,y) \wedge \neg \forall x \forall y \Delta_{n-2}(x,y) \\
&\qquad\qquad \texttt{to say that the diameter equals } n-1 \\
&\wedge\, \forall x \neg \exists y_1 \exists y_2 \exists y_3 \left(\bigwedge_{i=1,2,3} x \sim y_i \wedge \bigwedge_{i \neq j} \neg(y_i = y_j) \right) \\
&\qquad\qquad \texttt{to say that the maximum degree } \leq 2 \\
&\wedge\, \exists x \neg \exists y_1 \exists y_2 \left(\bigwedge_{i=1,2} x \sim y_i \wedge \neg(y_1 = y_2) \right) \\
&\qquad\qquad \texttt{to say that the minimum degree } \leq 1 \texttt{ (thereby} \\
&\qquad\qquad \texttt{distinguishing from cycles } C_{2n-2} \texttt{ and } C_{2n-1}\texttt{)}
\end{aligned} \tag{3}
$$

We have already mentioned the following basic fact: *Every finite graph G is definable.*[1] Indeed, let $V(G) = \{v_1, \ldots, v_n\}$ be the vertex set of G and $E(G)$ be its edge set. A sentence defining G could read:

$$
\begin{aligned}
&\exists x_1 \ldots \exists x_n\ (\mathrm{Distinct}(x_1, \ldots, x_n) \wedge \mathrm{Adj}(x_1, \ldots, x_n)) \\
&\qquad \wedge\, \forall x_1 \ldots \forall x_{n+1}\ \neg\, \mathrm{Distinct}(x_1, \ldots, x_{n+1}),
\end{aligned} \tag{4}
$$

where, for the notational convenience, we use the following shorthands

$$
\mathrm{Distinct}(x_1, \ldots, x_k) \stackrel{\mathrm{def}}{=} \bigwedge_{1 \leq i < j \leq k} \neg\,(x_i = x_j),
$$

$$
\mathrm{Adj}(x_1, \ldots, x_n) \stackrel{\mathrm{def}}{=} \bigwedge_{\{v_i, v_j\} \in E(G)} x_i \sim x_j \ \wedge \bigwedge_{\{v_i, v_j\} \notin E(G)} \neg\,(x_i \sim x_j).
$$

In other words, we first specify that there are n distinct vertices, list the adjacencies and the non-adjacencies between them, and then state that we cannot find $n+1$ distinct vertices.

The sentence (4) is an exhaustive description of G and seems rather wasteful. We want to know if there is a more succinct way of defining a graph on n vertices. The following natural succinctness measures of a first-order formula Φ are of interest:

- the *length* $L(\Phi)$ which is the total number of symbols in Φ (each variable symbol contributes 1);
- the *quantifier depth* $D(\Phi)$ which is the maximum length of a chain of nested quantifiers in Φ;
- the *width* $W(\Phi)$ which is the number of variables used in Φ (different occurrences of the same variable are not counted).[2]

[1] This fact, though very simple, highlights a fundamental difference between the finite and the infinite: There are non-isomorphic countable graphs satisfying precisely the same first-order sentences (see, e.g., [**72**, Theorem 3.3.2]).

[2] Grädel [**33**] defines the width of a formula Φ as the maximum number of free variables in a subformula of Φ. Denote this version by $W'(\Phi)$. Clearly, $W'(\Phi) \leq W(\Phi)$ and the inequality can be strict. Nevertheless, the two parameters are closely related: Φ can be rewritten by renaming bound variables in an equivalent form Φ' so that $W(\Phi') = W'(\Phi)$; see [**33**, Lemma 3.1.4].

Formula Δ_n in (2) was intentionally written in a non-optimal way. Note that $L(\Delta_n) = \Theta(n)$, $D(\Delta_n) = n-1$, and $W(\Delta_n) = n+1$. The same distance restriction can be expressed more succinctly with respect to the latter two parameters, namely

$$\begin{aligned} \Delta'_1(x,y) &\stackrel{\text{def}}{=} \Delta_1(x,y), \\ \Delta'_n(x,y) &\stackrel{\text{def}}{=} \exists z \left(\Delta'_{\lfloor n/2 \rfloor}(x,z) \wedge \Delta'_{\lceil n/2 \rceil}(z,y)\right), \end{aligned} \tag{5}$$

where $\lceil x \rceil$ (resp. $\lfloor x \rfloor$) stands for the integer nearest to x from above (resp. from below). Now $D(\Delta'_n) = \lceil \log_2 n \rceil$, giving an exponential gain for the quantifier depth! The width can be reduced even more drastically: by recycling variables we can write Δ'_n with only 3 variables in total, achieving $W(\Delta'_n) = 3$.

We now come to the central concepts of our survey. Let us define $L(G)$ (resp. $D(G)$, $W(G)$) to be the minimum of $L(\Phi)$ (resp. $D(\Phi)$, $W(\Phi)$) over all sentences Φ defining a graph G. We will call these graph invariants, respectively, the *logical length, depth*, and *width* of G.

EXAMPLE 1.1.

1. Using Δ'_n in place of Δ_n in (3), we see that $D(P_n) < \log_2 n + 3$ and $W(P_n) \le 4$. The reader is encouraged to improve the latter to $W(P_n) \le 3$.
2. The generic defining sentence (4) shows that $L(G) = O(n^2)$ and $D(G) \le n+1$ for every graph G on n vertices.
3. The *complement* of G, denoted by $\overline{G}$, is the graph on the same vertex set $V(G)$ whose edges are those pairs that are not in $E(G)$. One can easily prove that $D(\overline{G}) = D(G)$ and $W(\overline{G}) = W(G)$.

The logical length, depth, and width of a graph satisfy the following inequalities:

$$W(G) \le D(G) < L(G).$$

The latter relation follows from an obvious fact that $D(\Phi) < L(\Phi)$ for any first-order formula Φ. The former follows from a bit less obvious fact that for any first-order formula Φ there is a logically equivalent formula Ψ with $W(\Psi) \le D(\Phi)$.

1.2. Variations of logic.

1.2.1. *Fragments.* Suppose that we put some restrictions on the structure of a defining sentence. This may cause an increase in the resources (length, depth, width) that we need in order to define a graph in the straitened circumstances. These effects will be one of our main concerns in this survey. We will deal with restrictions of the following two sorts. We may be allowed to make only a small (constant) number of quantifier alternations or to use only a bounded number of variables. The former is commonly used in logic and complexity theory to obtain hierarchical classifications of various problems. The latter is in the focus of *finite-variable logics* (see, e.g, Grohe [**34**]). Moreover, the number of variables has relevance to the computational complexity of the graph isomorphism problem, see Section 4.

Bounded number of quantifier alternations. A first-order formula Φ with connectives $\{\neg, \wedge, \vee\}$ is in a *negation normal form* if all negations apply only to relations (one can think that we now do not have negation at all but introduce instead two new relation symbols, for inequality and non-adjacency). It is well known that this structural restriction actually does not make first-order logic weaker: We can always move negations in front of relation symbols without increasing the formula's length more than twice and without changing the quantifier depth and the width.

Given such a formula Φ and a sequence of nested quantifiers in it, we count the number of *quantifier alternations*, that is, the number of successive pairs $\forall\exists$ and $\exists\forall$ in the sequence. The *alternation number* of Φ is the maximum number of quantifier alternations over all such sequences. The *a-alternation logic* consists of all first-order formulas in the negation normal form whose alternation number does not exceed a. We will adhere to the following notational convention: a subscript a will always indicate that at most a quantifier alternations are allowed. For example, $D_a(G)$ is the minimum quantifier depth of a sentence in the a-alternation logic that defines a graph G.

For any graph G on n vertices we have

$$D(G) \le \ldots \le D_{a+1}(G) \le D_a(G) \le \ldots \le D_1(G) \le D_0(G) \le n+1,$$

where the last bound is due to the defining sentence (4).

Bounded number of variables. The *k-variable logic* is the fragment of first-order logic where only k variable symbols are available, that is, the formula width is bounded by k. The restriction of defining sentences to the k-variable logic will be always indicated by a superscript k. To make this notation always applicable, we set $D^k(G) = \infty$ if the k-variable logic is too weak to define G. If $k \ge W(G)$ for a graph G of order n, then we have

$$D(G) \le D^{k+1}(G) \le D^k(G) < n^{k-1} + k,$$

where the last bound will be established in Theorem 4.7 below. Note that the bounds in Example 1.1.1 can be strengthened to $D^3(P_n) < \log_2 n + 3$.

1.2.2. *An extension with counting quantifiers.* We will also enrich first-order logic by allowing one to use expressions of the type $\exists^m \Psi$ in order to say that there are at least m vertices with property Ψ. Those are called *counting quantifiers* and the extended logic will be referred to as *counting logic*. A counting quantifier $\exists^m$ contributes 1 in the quantifier depth irrespectively of the value of m. For the counting logic we will use the "sharp-notation", thus denoting the logical depth and width of a graph G in this logic, respectively, by $D_\#(G)$ and $W_\#(G)$. Clearly, $D_\#(G) \le D(G)$ and $W_\#(G) \le W(G)$. The counting quantifiers often allow us to define a graph much more succinctly. For example, $D_\#(K_n) = W_\#(K_n) = 2$ as this graph is defined by

$$\forall x \forall y\, (x \sim y \vee x = y) \wedge \exists^n x\, (x = x) \wedge \neg \exists^{n+1} x\, (x = x).$$

This is in sharp contrast with the fact that $D(K_n) = W(K_n) = n+1$, where the lower bound follows from the simple observation that n variables are not enough to distinguish between K_n and K_{n+1}.

1.3. Outline of the survey. Section 2 specifies notation and proves a couple of basic facts about first-order sentences. The latter are applied to establish an upper bound on the logical length $L(G)$ of a graph in terms of its logical depth $D(G)$ and to estimate from above the number of graphs whose logical depth is bounded by a given parameter k. The existence of such bounds is more important than the bounds themselves that are huge, involving the tower function. Furthermore, we define $D(G, H)$ to be the smallest quantifier depth sufficient to distinguish between non-isomorphic graphs G and H. We will observe that the obvious inequality

$D(G,H) \le D(G)$ gives the sharp lower bound on $D(G)$. Thus estimating $D(G)$ reduces to estimating $D(G,H)$ for all $H \not\cong G$

The value of $D(G,H)$ is characterized in Section 3 as the length of the *Ehrenfeucht game* on G and H. Moreover, the logical width admits a characterization in terms of another parameter of the game. Thus, the determination of the logical depth and width of a graph reduces to designing optimal strategies in the Ehrenfeucht game.

In Section 4, the logical width and the logical depth are also characterized, respectively, as the minimum *dimension* and the minimum *number of rounds* such that the so-called *Weisfeiler-Lehman algorithm* returns the correct answer. The algorithm tries to decide whether two input graphs are isomorphic; its one-dimensional version is just the well-known color-refining procedure. Thus, an analysis of the algorithm can give us information on the logical complexity of the input graphs. This relationship is even more advantageous in the other direction: Once we prove that all graphs in some class C have low logical complexity, we immediately obtain an efficient isomorphism test for C.

This paradigm is successful for graphs with bounded treewidth and planar graphs, with good prospects for covering all classes of graphs with an excluded minor. In Section 5.1 we report strong upper bounds for the logical depth/width of graphs in these classes. In Section 5.2 we survey the bounds known in the general case. In particular, if a graph G on n vertices has no *twins*, i.e., no two vertices have the same adjacency to the rest of the graph, then $D(G) < \frac{1}{2}n + 3$. The factor of $\frac{1}{2}$ can be improved for graphs with bounded vertex degrees. Here we have to content ourselves with linear bounds in view of a linear lower bound by Cai, Fürer, and Immerman [**15**]. They constructed examples of graphs with maximum degree 3 such that $W_{\#}(G) > c\,n$ for a positive constant c.

Section 6 discusses the logical complexity of a random graph. We obtain rather close lower and upper bounds for *almost all* graphs. Furthermore, we trace the behavior of the logical depth in the evolutional random graph model $G_{n,p}$ where p is a function of n.

While in Sections 5 and 6 we deal with, respectively, worst case and average case bounds, Section 7 is devoted to the *best case*. More specifically, we define *succinctness function* $q(n)$ to be equal to the minimum of $D(G)$ over all G on n vertices. Since only finitely many graphs are definable with a fixed quantifier depth, $q(n)$ goes to infinity as n increases. It turns out that its growth is inconceivably slow: We show a superrecursive gap between the values of $q(n)$ and n. This phenomenon disappears if we "smoothen" $q(n)$ by considering the least monotonic upper bound for this function: the smoothed succinctness function is very close to the log-star function. Furthermore, the succinctness function can be considered in any logic. Let $q_0(n)$ be its variant for the logic with no quantifier alternation. We can determine $q_0(n)$ with rather high precision: It is also related to the log-star function. The lower bound for $q_0(n)$ implies a superrecursive gap between the graph parameters $D(G)$ and $D_0(G)$, yet another evidence of the weakness of the 0-alternation logic. The tight upper bound for $q_0(n)$ shows that, nevertheless, there are graphs whose definitions, even if quantifiers are not allowed to alternate, can have surprisingly low quantifier depth. We give several methods of explicit constructions of such graphs. These constructions have another interesting aspect. They allow us to show that the previously mentioned tower-function bounds from Section 2 cannot be improved substantially.

Some of the most interesting open questions are collected in Section 8.

1.4. Other structures. Some of the results presented in the survey generalize to relational structures over a fixed vocabulary. Such generalizations are often straightforward. For example, the upper bounds on succinctness functions hold true if the vocabulary contains at least one relation symbol of arity more than 1 (since any graph can be trivially represented as a structure over this vocabulary). Extension of the worst case bounds to general structures is also possible but requires essential additional efforts; see [**65**].

Various definability parameters were investigated also for special structures: colored graphs (Immerman and Lander [**47**], Cai, Fürer, and Immerman [**15**]), digraphs and hypergraphs (Pikhurko, Veith, and Verbitsky [**64**]), bit strings and ordered trees (Spencer and St. John [**73**]), linear orders (Grohe and Schweikardt [**41**]).

2. Preliminaries

2.1. Notation: Arithmetic and graphs. We define the *tower function* by $Tower(0) = 1$ and $Tower(i) = 2^{Tower(i-1)}$ for each subsequent integer i. Given a function f, by $f^{(i)}(x)$ we will denote the i-fold composition of f. In particular, $f^{(0)}(x) = x$. By $\log n$ we always mean the logarithm base 2. The "inverse" of the tower function, the *log-star* function $\log^* n$, is defined by $\log^* n = \min\{\, i : Tower(i) \ge n\}$. We use the standard asymptotic notation. For example, $f(n) = \Omega(g(n))$ means that there is a constant $c > 0$ such that $f(n) \ge c\, g(n)$ for all sufficiently large n.

The number of vertices in a graph G is called the *order* of G and is denoted by $v(G)$. The *neighborhood* $N(v)$ of a vertex v consists of all vertices adjacent to v. The *degree* of v is defined by $\deg v = |N(v)|$. The *maximum degree* of a graph G is defined by $\Delta(G) = \max_{v \in V(G)} \deg v$.

The *distance* between vertices u and v in a graph G is defined to be the minimum length of a path from u to v and denoted by $dist(u, v)$. If u and v are in different connectivity components, then we set $dist(u, v) = \infty$. The *eccentricity* of a vertex v is defined by $e(v) = \max_{u \in V(G)} dist(v, u)$.

Let $X \subset V(G)$. The subgraph induced by G on X is denoted by $G[X]$. We denote $G \setminus X = G[V(G) \setminus X]$, which is the result of the removal of all vertices in X from G. If a single vertex v is removed, we write $G - v = G \setminus \{v\}$. A set of vertices X is called *homogeneous* if $G[X]$ is a complete or an empty graph.

A graph is *k-connected* if it has at least $k + 1$ vertices and remains connected after removal of any $k - 1$ vertices. 2-connected graphs are also called *biconnected.*

A graph is *asymmetric* if it admits no non-trivial automorphism.

2.2. A length-depth relation. We have already mentioned the trivial relation $D(G) < L(G)$. Now we aim at bounding $L(G)$ from above in terms of $D(G)$. We write $G \equiv_k H$ to say that graphs G and H cannot be distinguished by any sentence with quantifier depth k. As it is easy to see, $\equiv_k$ is an equivalence relation. Its equivalence classes will be referred to as *$\equiv_k$-classes.* We say that a sentence Φ *defines* a $\equiv_k$-class α if Φ is true on all graphs in α and false on all other graphs.

LEMMA 2.1.

1. *The number of* $\equiv_k$*-classes is finite and does not exceed* $Tower(k+\log^* k+2)$.
2. *Every* $\equiv_k$*-class is definable by a sentence* Φ *with* $D(\Phi) = k$ *and* $L(\Phi) < Tower(k+\log^* k+2)$.

PROOF. The case of $k = 1$ is easy: There is only one $\equiv_1$-class (consisting of all graphs), which is definable by $\forall x(x = x)$.

Let $k \geq 2$ and $0 \leq s \leq k$. When we write $\bar{z}$, we will mean an s-tuple $(z_1, \ldots, z_s)$ (if $s = 0$, the sequence is empty). If $\bar{u} \in V(G)^s$ and Φ is a formula with s free variables $x_1, \ldots, x_s$, then notation $G, \bar{u} \models \Phi(\bar{x})$ will mean that $\Phi(\bar{x})$ is true on G with each x_i being assigned the respective u_i as its value.

A formula $\Phi(x_1, \ldots, x_s)$ of quantifier depth $k - s$ is *normal* if Φ is built from variables $x_1, \ldots, x_k$ and every maximal sequence of nested quantifiers in Φ has length $k - s$ and quantifies the variables $x_{s+1}, \ldots, x_k$ exactly in this order. A simple inductive syntactic argument shows that any $\Phi(x_1, \ldots, x_s)$ has an equivalent normal formula $\Phi'(x_1, \ldots, x_s)$ of the same quantifier depth as Φ.

We write $G, \bar{u} \equiv_{k,s} H, \bar{v}$ to say that $G, \bar{u} \models \Phi(\bar{x})$ exactly when $H, \bar{v} \models \Phi(\bar{x})$ for every normal formula Φ of quantifier depth $k-s$. A normal formula $\Phi(\bar{x})$ *defines* a $\equiv_{k,s}$-class α if $G, \bar{u} \models \Phi(\bar{x})$ exactly when $G, \bar{u}$ belongs to α. The $\equiv_{k,s}$-equivalence class of $G, \bar{u}$ will be denoted by $[G, \bar{u}]_{k,s}$.

Let $f(k, s)$ denote the number of all $\equiv_{k,s}$-classes and $l(k, s)$ denote the minimum l such that every $\equiv_{k,s}$-class is definable by a normal formula of depth at most $k-s$ and length at most l. Note that relations $\equiv_k$ and $\equiv_{k,0}$ coincide. Thus, our goal is to estimate the numbers $f(k, 0)$ and $l(k, 0)$ from above.

We use the backward induction on s. A $\equiv_{k,k}$-class can be determined by specifying, for each pair of the k elements, whether they are equal and, if not, whether they are adjacent or non-adjacent. There are at most three choices per pair. It easily follows that

$$f(k, k) \leq 3^{\binom{k}{2}} \quad \text{and} \quad l(k, k) < 9k^2. \tag{6}$$

We are now going to estimate $f(k, s)$ and $l(k, s)$ in terms of $f(k, s+1)$ and $l(k, s+1)$. Suppose that each $\equiv_{k,s+1}$-class β is defined by a formula $\Phi_\beta(x_1, \ldots, x_s, x_{s+1})$ whose length is bounded by $l(k, s+1)$.

Define $S(G, \bar{u}) = \{\, [G, \bar{u}, u]_{k,s+1} : u \in V(G)\}$, the set of $\equiv_{k,s+1}$-classes obtainable from $G, \bar{u}$ by specifying one extra vertex. Note that

$$G, \bar{u} \equiv_{k,s} H, \bar{v} \ \text{ if and only if } \ S(G, \bar{u}) = S(H, \bar{v}).$$

Indeed, suppose that $S(G, \bar{u}) \neq S(H, \bar{v})$, say, $\beta = [G, \bar{u}, u]_{k,s+1}$ is not in $S(H, \bar{v})$ for some $u \in V(G)$. Then $G, \bar{u} \not\equiv_{k,s} H, \bar{v}$ because formula $\exists x_{s+1}\Phi_\beta$ is true for $G, \bar{u}$ but false for $H, \bar{v}$. Suppose now that $G, \bar{u}$ and $H, \bar{v}$ are distinguishable by a normal formula of quantifier depth $k - s$. As it is easily seen, they are distinguishable by such a formula of the form $\exists x_{s+1}\Phi$. Without loss of generality, assume that the formula $\exists x_{s+1}\Phi$ is true for $G, \bar{u}$ but false for $H, \bar{v}$. Let $u \in V(G)$ be such that $G, \bar{u}, u \models \Phi$. Since Φ distinguishes $G, \bar{u}, u$ from all $H, \bar{v}, v$ with $v \in V(H)$, the class $[G, \bar{u}, u]_{k,s+1}$ is not in $S(H, \bar{v})$ and, hence, $S(G, \bar{u}) \neq S(H, \bar{v})$.

Thus, for a $\equiv_{k,s}$-class α we can correctly define the set of $\equiv_{k,s+1}$-classes accessible from α by $S(\alpha) = S(G, \bar{u})$ for some (in fact, arbitrary) $G, \bar{u}$ in α. It follows from what we have proved that for arbitrary $\equiv_{k,s}$-classes α and α', we have

$$\alpha = \alpha' \text{ if and only if } S(\alpha) = S(\alpha').$$

As an immediate consequence,

$$f(k, s) \le 2^{f(k,s+1)}. \tag{7}$$

Since $2\cdot\binom{k}{2} \le 2^k$ for every integer $k \ge 1$, we have $f(k,k) \le 2^{2^k} \le \mathit{Tower}(\log^* k + 2)$. By the above recursion, we conclude that $f(k, 0) \le \mathit{Tower}(k+\log^* k+2)$, which proves Part 1 of the lemma.

Another conclusion is that any $\equiv_{k,s}$-class α can be defined by a normal formula[3]

$$\Phi_\alpha(\bar{x}) \stackrel{\text{def}}{=} \bigwedge_{\beta\in S(\alpha)} \exists x_{s+1}\, \Phi_\beta(\bar{x}, x_{s+1}) \;\wedge\; \forall x_{s+1} \bigwedge_{\beta\notin S(\alpha)} \neg\, \Phi_\beta(\bar{x}, x_{s+1}).$$

Looking at the length of $\Phi_\alpha(\bar{x})$, we have the recurrence

$$l(k, s) \le f(k, s + 1)(l(k, s + 1) + 9). \tag{8}$$

Using (8), (7), and (6), after routine calculations (omitted) we obtain

$$l(k, 0) < \mathit{Tower}(k + \log^* k + 2)$$

for every $k \ge 2$, which gives us Part 2. □

Lemma 2.1.2 gives us a bound for the logical length of a graph in terms of its logical depth. It suffices to notice that each single graph G constitutes a $\equiv_k$-class for $k = D(G)$.

THEOREM 2.2 (Pikhurko, Spencer, and Verbitsky [**61**]).

$$L(G) < \mathit{Tower}(D(G) + \log^* D(G) + 2).$$

Lemma 2.1.1 gives the following result.

THEOREM 2.3. *The number of graphs with logical depth at most k does not exceed $\mathit{Tower}(k + \log^* k + 2)$.*

Notice two further consequences of Lemma 2.1.

THEOREM 2.4.

1. *There are at most $\mathit{Tower}(k + \log^* k + 3)$ pairwise inequivalent sentences about graphs of quantifier depth k.*
2. *Every sentence Φ about graphs of quantifier depth k has an equivalent sentence Φ' with the same quantifier depth and length less than $3\,\mathit{Tower}(k + \log^* k + 2)^2$.*

PROOF. Note that, if a sentence Φ has quantifier depth k, then the set of all graphs on which Φ is true is the union of some $\equiv_k$-classes. Therefore, there are $2^{f(k)}$ and no more pairwise inequivalent sentences of quantifier depth k, where $f(k)$ is the number of $\equiv_k$-classes. Part 1 now follows from Lemma 2.1.1. By the same reason every sentence Φ of quantifier depth k is equivalent to the disjunction of sentences defining some $\equiv_k$-classes. By Lemma 2.1.2, such disjunction does not need to be longer than $(f(k) + 3)\,\mathit{Tower}(k + \log^* k + 2)$. This proves Part 2. □

[3] This is a variant of Hintikka's formula, cf. [**24**, Definition 2.2.5].

2.3. Distinguishability vs. definability. Given two non-isomorphic graphs G and H, we define $D(G,H)$ (resp. $W(G,H)$) to be the minimum of $D(\Phi)$ (resp. $W(\Phi)$) over all sentences Φ distinguishing G from H. Thus, $D(G,H) > k$ if and only if $G \equiv_k H$. Obviously, $D(G,H) = D(H,G)$. Also, $D(G,H) \le D(G)$ and $W(G,H) \le W(G)$. It turns out that these inequalities are tight in the following sense.

LEMMA 2.5.

1. $D(G) = \max_{H \not\cong G} D(G,H)$.
2. $W(G) = \max_{H \not\cong G} W(G,H)$.

PROOF. *1.* For each H non-isomorphic to G fix a sentence Φ_H that distinguishes G from H and has the minimum possible quantifier depth, i.e., $D(\Phi_H) = D(G,H)$. Consider the sentence $\Phi \stackrel{\text{def}}{=} \bigwedge_{H \not\cong G} \Phi_H$. It distinguishes G from each non-isomorphic H and has quantifier depth $\max_H D(\Phi_H)$. Therefore, $D(G) \le \max_H D(G,H)$ as wanted. An obvious drawback of this argument is that the above conjunction over H in Φ is actually infinite. However, we have $D(\Phi_H) \le D(G)$ and there are only finitely many pairwise inequivalent first-order sentences about graphs of bounded quantifier depth, see Theorem 2.4 above. Thus we can obtain a legitimate finite sentence defining G by removing from Φ duplicates up to logical equivalence.

2. Running the same argument, we have to "prune" the infinite conjunction $\bigwedge_{H \not\cong G} \Phi_H$, where $W(\Phi_H) = W(G,H)$. Here we encounter a complication because there are infinitely many inequivalent sentences of the same width. (Consider e.g. the sentences from Example 1.1.1.) However, Theorem 4.7.1 in Section 4 implies that for every H we can additionally require that the depth of Φ_H is at most, for example, $n^n + n$, where n is the order of G. Now we can proceed as in Part 1 of the lemma. □

Lemma 2.5 stays true in any finite-variable logic, any logic with bounded number of quantifier alternations, the logic with counting quantifiers, and any hybrid thereof. We set $D^k(G,H) = \infty$ if k variables do not suffice to distinguish G from H.

3. Ehrenfeucht games

Let G and H be graphs with disjoint vertex sets. The *r-round k-pebble Ehrenfeucht game on G and H*, denoted by $\textsc{Ehr}_r^k(G,H)$, is played by two players, Spoiler and Duplicator, to whom we may refer as he and she respectively. The players have at their disposal k pairwise distinct pebbles $p_1, \ldots, p_k$, each given in duplicate. A *round* consists of a move of Spoiler followed by a move of Duplicator. At each move Spoiler takes a pebble, say p_i, selects one of the graphs G or H, and places p_i on a vertex of this graph. In response Duplicator should place the other copy of p_i on a vertex of the other graph. It is allowed to move previously placed pebbles to other vertices and place more than one pebble on the same vertex.

After each round of the game, for $1 \le i \le k$ let x_i (resp. y_i) denote the vertex of G (resp. H) occupied by p_i, irrespectively of who of the players placed the pebble on this vertex. If p_i is off the board at this moment, x_i and y_i are undefined. If after every of r rounds the component-wise correspondence $(x_1, \ldots, x_k)$ to $(y_1, \ldots, y_k)$ is a partial isomorphism from G to H, this is a win for Duplicator. Otherwise the winner is Spoiler. The following example should provide the reader with a hint for the solution of the exercise suggested in Example 1.1.1.

EXAMPLE 3.1. *Spoiler wins* $\text{EHR}_4^3(P_n, H)$ *if* $\Delta(H) \geq 3$. Assume that H contains no triangle because otherwise Spoiler wins by pebbling its vertices. Let v be a vertex in H of degree at least 3. Spoiler pebbles 3 neighbors of v. Duplicator should pebble 3 distinct pairwise non-adjacent vertices in P_n for otherwise she loses the game. The distance between any two vertices pebbled in H is equal to 2. Unlike to this, some two vertices pebbled in P_n (say, by pebbles p_1 and p_2) are at a larger distance. Spoiler moves p_3 to v. Duplicator is forced to violate the adjacency relation.

The particular case of $\text{EHR}_r^k(G, H)$ in which the number of pebbles is the same as the number of rounds, i.e., $k = r$, deserves a special attention. In this case, the outcome of the game will not be affected if we prohibit moving pebbles from one vertex to another, that is, if we allow the players to play with each p_i exactly once, say, in the i-th round. We denote this variant of $\text{EHR}_r^r(G, H)$ by $\text{EHR}_r(G, H)$ and will mean it whenever the term *Ehrenfeucht game* is used with no specification.

LEMMA 3.2. *Suppose that in the 3-pebble Ehrenfeucht game on* (G, H) *some two vertices* $x, y \in V(G)$ *at distance* n *were selected so that their counterparts* $x', y' \in V(H)$ *are at a strictly larger distance (possibly infinity). Then Spoiler can win in at most* $\lceil \log n \rceil$ *extra moves.*

PROOF. Spoiler sets $u_1 = x$, $u_2 = y$, $v_1 = x'$, $v_2 = y'$, and places a pebble on the middle vertex u in a shortest path from u_1 to u_2 (or either of the two middle vertices if $d(u_1, u_2)$ is odd). Let $v \in V(H)$ be selected by Duplicator in response to u. By the triangle inequality, we have $d(u, u_m) < d(v, v_m)$ for $m = 1$ or $m = 2$. For such m Spoiler resets $u_1 = u$, $u_2 = u_m$, $v_1 = v$, $v_2 = v_m$ and applies the same strategy once again. In this way Spoiler ensures that $d(u_1, u_2) < d(v_1, v_2)$ in each round. Eventually, unless Duplicator loses earlier, $d(u_1, u_2) = 1$ while $d(v_1, v_2) > 1$, that is, Duplicator fails to preserve adjacency.

To estimate the number of moves made, notice that initially $d(u_1, u_2) = n$ and for each subsequent u_1, u_2 this distance becomes at most $f(d(u_1, u_2))$, where $f(\alpha) = (\alpha + 1)/2$. Therefore the number of moves does not exceed the minimum i such that $f^{(i)}(n) < 2$. As $(f^{(i)})^{-1}(\beta) = 2^i\beta - 2^i + 1$, the latter inequality is equivalent to $2^i \geq n$, which proves the bound. $\square$

There is a rather clear connection between Spoiler's strategy designed in the proof of Lemma 3.2 and first-order formula $\Delta'_n(x, y)$ in (5). We will see that, in some strong sense, $\text{EHR}_r(G, H)$ corresponds to first-order logic, while $\text{EHR}_r^k(G, H)$ corresponds to its k-variable fragment. In fact, every logic has its own corresponding game.

In the *k-alternation* variant of $\text{EHR}_r(G, H)$ Spoiler is allowed to switch from one graph to another at most k times during the game, i.e., in at most k rounds he can choose the graph other than that in the preceding round.

In the *counting version* of the game $\text{EHR}_r^k(G, H)$ Spoiler can make a *counting move* consisting of two acts. First, he specifies a set of vertices A in one of the graphs. Duplicator has to respond with a set of vertices B in the other graph so that $|B| = |A|$ (if this is impossible, she immediately loses). Second, Spoiler places a pebble p_i on a vertex $b \in B$. In response Duplicator has to place the other copy of p_i on a vertex $a \in A$. It is clear that, any round with $|A| = 1$ is virtually the same as a round of the standard game.

There is a general analogy between strategies allowing Spoiler to win a game on G and H and first-order sentences distinguishing these graphs: the former can be converted into the latter and vice versa so that the duration of a game will be in correspondence to the quantifier depth and the number of pebbles will be in correspondence to the number of variables.

THEOREM 3.3 (The Ehrenfeucht theorem and its variations). *Let G and H be non-isomorphic graphs.*

1. (Ehrenfeucht [**25**], Fraïssé [**30**][4]) *$D(G,H)$ equals the minimum r such that Spoiler has a winning strategy in* $\mathrm{EHR}_r(G,H)$.
2. (Pezzoli [**60**]) *$D_k(G,H)$ equals the minimum r such that Spoiler has a winning strategy in the k-alternation game* $\mathrm{EHR}_r(G,H)$.
3. (Immerman [**44**], Poizat [**66**]) *$W(G,H)$ equals the minimum k such that Spoiler has a winning strategy in* $\mathrm{EHR}_r^k(G,H)$ *for some r.*
4. (Immerman [**44**], Poizat [**66**]) *$D^k(G,H)$ equals the minimum r such that Spoiler has a winning strategy in* $\mathrm{EHR}_r^k(G,H)$.
5. (Immerman and Lander [**47**]) *$W_\#(G,H)$ equals the minimum k such that Spoiler has a winning strategy in the counting version of* $\mathrm{EHR}_r^k(G,H)$ *for some r. Furthermore, if $k \geq W_\#(G,H)$, then $D_\#^k(G,H)$ equals the minimum r such that Spoiler has a winning strategy in the counting version of* $\mathrm{EHR}_r^k(G,H)$.

We refer the reader to [**45**, Theorem 6.10] for the proof of Parts 3–5. Part 1 follows from Part 4 in view of the facts that $D(G,H) = \min_k D^k(G,H)$ and that any sentence Φ can be equivalently rewritten with the same quantifier depth $D(\Phi)$ and with use of at most $D(\Phi)$ variables.

In view of Lemma 2.5, the Ehrenfeucht theorem provides us with a powerful tool for estimating the logical depth and width of graphs. Consider, for instance, a path P_n. Example 3.1 and Lemma 3.2 are immediately translated into the upper bound $D^3(P_n) < \log n + 3$. On the other hand, a lower bound $D(P_n) \geq \log n - 2$ follows from the existence of a winning strategy for Duplicator in $\mathrm{EHR}_r(P_n, P_{n+1})$ whenever $r \leq \lfloor \log n \rfloor - 1$ (all details can be found in [**72**, Theorem 2.1.3]).

4. The Weisfeiler-Lehman algorithm

Graph Isomorphism is the problem of recognizing if two given graphs are isomorphic. The best known algorithm (Babai, Luks, and Zemlyachenko [**9**]) takes time $2^{O(\sqrt{n \log n})}$, where n denotes the number of vertices in the input graphs. Particular classes of graphs for which Graph Isomorphism is solvable more efficiently are therefore of considerable interest. Somewhat surprisingly, a number of important tractable cases are solvable by a combinatorially simple, uniform approach, namely the multidimensional Weisfeiler-Lehman algorithm. The efficiency of this method depends much on the logical complexity of input graphs.

For the history of this approach to the graph isomorphism problem we refer the reader to [**5, 15**]. We will abbreviate *k-dimensional Weisfeiler-Lehman algorithm* by *k-dim WL*. The 1-dim WL is commonly known as *canonical labeling* or *color refinement algorithm*. It proceeds in rounds; in each round a coloring of the vertices

[4] It was Ehrenfeucht who formally introduced the game. Prior to Ehrenfeucht, Fraïssé obtained virtually the same result using an equivalent language of partial isomorphisms.

of input graphs G and H is defined, which refines the coloring of the previous round. The initial coloring C^0 is uniform, say, $C^0(u) = 1$ for all vertices $u \in V(G) \cup V(H)$. In the $(i+1)$st round, the color $C^{i+1}(u)$ is defined to be a pair consisting of the preceding color $C^{i-1}(u)$ and the multiset of colors $C^{i-1}(w)$ for all w adjacent to u. For example, $C^1(u) = C^1(v)$ iff u and v have the same degree. To keep the color encoding short, after each round the colors are renamed (we never need more than $2n$ color names[5]). As the coloring is refined in each round, it stabilizes after at most $2n$ rounds, that is, no further refinement occurs. The algorithm stops once this happens. If the multiset of colors of the vertices of G is distinct from the multiset of colors of the vertices of H, the algorithms reports that the graphs are not isomorphic; otherwise, it declares them to be isomorphic. Disappointingly, the output is not always correct. The algorithm may report false positives, for example, if both input graphs are regular with the same vertex degree.

Following the same idea, the k-dimensional version iteratively refines a coloring of $V(G)^k \cup V(H)^k$. The initial coloring of a k-tuple $\bar{u}$ is the isomorphism type of the subgraph induced by the vertices in $\bar{u}$ (viewed as a labeled graph where each vertex is labeled by the positions in the tuple where it occurs). Loosely speaking, the refinement step takes into account the colors of all neighbors of $\bar{u}$ in the Hamming metric. Color stabilization is surely reached in $r < 2n^k$ rounds and, thus, the algorithm terminates in polynomial time for fixed k.

Let us give a careful description of the k-dim WL for $k \ge 2$. Given an ordered k-tuple of vertices $\bar{u} = (u_1, \ldots, u_k) \in V(G)^k$, we define the *isomorphism type* of $\bar{u}$ to be the pair

$$\mathrm{tp}(\bar{u}) = \Big(\big\{ (i,j) \in [k]^2 : u_i = u_j \big\}, \big\{ (i,j) \in [k]^2 : \{u_i, u_j\} \in E(G) \big\} \Big), \tag{9}$$

where $[k]$ denotes the set $\{1, \ldots, k\}$. If $w \in V(G)$ and $i \le k$, we let $\bar{u}^{i,w}$ denote the result of substituting w in place of u_i in $\bar{u}$.

The *r-round k-dim WL* takes as an input two graphs G and H and purports to decide if $G \cong H$. The algorithm performs the following operations with the set $V(G)^k \cup V(H)^k$.

Initial coloring. The algorithm assigns each $\bar{u} \in V(G)^k \cup V(H)^k$ color $C^{k,0}(\bar{u}) = \mathrm{tp}(\bar{u})$ (in a suitable encoding).

Color refinement step. In the i-th round each $\bar{u} \in V(G)^k$ is assigned color

$$C^{k,i}(\bar{u}) = \Big(C^{k,i-1}(\bar{u}), \{\!\!\{ \big(C^{k,i-1}(\bar{u}^{1,w}), \ldots, C^{k,i-1}(\bar{u}^{k,w}) \big) : w \in V(G) \}\!\!\} \Big)$$

and similarly with each $\bar{u} \in V(H)^k$.

Here $\{\!\!\{ \ldots \}\!\!\}$ denotes a multiset. In a weaker *count-free* version of the algorithm, this notation will be interpreted as a set. Let

$$C^{k,r}(G) = \{\!\!\{ C^{k,r}(\bar{u}) : \bar{u} \in V(G)^k \}\!\!\} .$$

Computing an output. The algorithm reports that $G \not\cong H$ if

$$C^{k,r}(G) \ne C^{k,r}(H) \tag{10}$$

and that $G \cong H$ otherwise.

In the above description we skipped an important implementation detail. In order to prevent increasing the length of $C^{k,i}(\bar{u})$ at the exponential rate, we arrange

[5]We do not need even more than n because appearance of the $(n+1)$th color indicates non-isomorphism.

colors of all k-tuples of $V(G)^k \cup V(H)^k$ in the lexicographic order and replace each color with its number before every refinement step.

Furthermore, let

$$diag\, C^{k,r}(G) = \{\!\!\{\, C^{k,r}(u^k) : u \in V(G) \}\!\!\}\,,$$

where u^k denotes the k-tuple $(u, \ldots, u)$.

LEMMA 4.1. *In both the standard and the count-free versions of the k-dim WL, inequality*

$$diag\, C^{k,r}(G) \neq diag\, C^{k,r}(H) \tag{11}$$

implies (10), which in its turn implies

$$diag\, C^{k,r+k-1}(G) \neq diag\, C^{k,r+k-1}(H). \tag{12}$$

PROOF. Consider the standard version; the analysis of the count-free case is similar (and even simpler). By the *equality type* of a k-tuple $\bar{u}$ we mean the first component of (9). Note that k-tuples with different equality types never have the same color. Therefore, $C^{k,r}(G)$ and $C^{k,r}(H)$ are different iff they are different on some class of k-tuples with the same equality type. This proves the first implication.

On the other hand, suppose that (10) holds. Let E be an equality type on which $C^{k,r}(G)$ and $C^{k,r}(H)$ differ. Note that each $\bar{u}$ in E contributes color $C^{k,r}(\bar{u})$ (a certain number of times) to color $C^{k,r+k-1}(a^k)$. Moreover, the sum of the contributions over all vertices a is the same for every $\bar{u} \in E$. It follows that, if a color has different multiplicities in $C^{k,r}(G)$ and $C^{k,r}(H)$, its "traces" occur different number of times in $diag\, C^{k,r+k-1}(G)$ and $diag\, C^{k,r+k-1}(H)$, and hence these multisets are distinct. $\square$

As it is easily seen, if ϕ is an isomorphism from G to H, then for all k, i, and $\bar{u} \in V(G)^k$ we have $C^{k,i}(\bar{u}) = C^{k,i}(\phi(\bar{u}))$. This shows that for isomorphic input graphs the output is always correct. If input graphs are non-isomorphic and the dimension k is not big enough, the algorithm can erroneously report isomorphism. A criterion for the optimal choice of the dimension is obtained by Cai, Fürer, and Immerman [**15**], who discovered a connection between the Weisfeiler-Lehman algorithm and the logical complexity of graphs via the Ehrenfeucht game (for the color refinement algorithm this was done by Immerman and Lander [**47**]). The success of the standard version of the algorithm depends on distinguishability of the input graphs in the logic with counting quantifiers, while the count-free version is in the same way related to the standard first-order logic.

Referring to the k-dim WL below, we will **always** assume $k \geq 1$ for the standard version of the algorithm and $k \geq 2$ for its count-free version (we can exclude the case of $k = 1$, whose analysis differs by some details, as the count-free 1-dim WL is of no interest: note that it is unable to distinguish between two graphs of order n without isolated vertices).

Given numbers r, l, and $k \leq l$, graphs G, H, and k-tuples $\bar{u} \in V(G)^k$, $\bar{v} \in V(H)^k$, we use notation $\mathrm{EHR}_r^l(G, \bar{u}, H, \bar{v})$ to denote the r-round l-pebble Ehrenfeucht game on G and H with initial configuration $(\bar{u}, \bar{v})$, that is, the game starts on the board with k already pebbled pairs (u_i, v_i). If the initial configurations is not a partial isomorphism, Duplicator loses $\mathrm{EHR}_r^l(G, \bar{u}, H, \bar{v})$ whatever $r \geq 0$. The following lemma is a key element of our analysis.

LEMMA 4.2 (Cai, Fürer, and Immerman [**15**]). *Let* $\bar{u} \in V(G)^k$ *and* $\bar{v} \in V(H)^k$.

1. *Equality*

$$C^{k,r}(\bar{u}) = C^{k,r}(\bar{v}) \tag{13}$$

holds for (the standard version of) the k*-dim WL iff Duplicator has a winning strategy in the counting version of* $\text{EHR}_r^{k+1}(G, \bar{u}, H, \bar{v})$.

2. *Equality (13) holds for the count-free version of the* k*-dim WL iff Duplicator has a winning strategy in (the standard version of)* $\text{EHR}_r^{k+1}(G, \bar{u}, H, \bar{v})$.

PROOF. We prove only Part 2 (Part 1 is proved in detail in [**15**, Theorem 5.2]). We proceed by induction on r. The base case $r = 0$ is straightforward by the definitions of the initial coloring and the game. Assume that the proposition is true for $r - 1$ rounds.

Let x_i and y_i denote the vertices in G and H respectively marked by the i-th pebble pair. Assume (13) and consider the Ehrenfeucht game on G, H with initial configuration $(x_1, \ldots, x_k) = \bar{u}$ and $(y_1, \ldots, y_k) = \bar{v}$. First of all, this configuration is non-losing for Duplicator since (13) implies that $\text{tp}(\bar{u}) = \text{tp}(\bar{v})$. Further, Duplicator can survive in the first round. Indeed, assume that Spoiler in this round selects a vertex a in one of the graphs, say in G. Then Duplicator selects a vertex b in the other graph H so that $C^{k,r-1}(\bar{u}^{i,a}) = C^{k,r-1}(\bar{v}^{i,b})$ for all $i \le k$. In particular, $\text{tp}(\bar{u}^{i,a}) = \text{tp}(\bar{v}^{i,b})$ for all $i \le k$. Along with $\text{tp}(\bar{u}) = \text{tp}(\bar{v})$, this implies that $\text{tp}(\bar{u}, a) = \text{tp}(\bar{v}, b)$. Assume now that in the second round Spoiler removes j-th pebble, $j \le k$. Then Duplicator's task in the rest of the game is essentially to win $\text{EHR}_{r-1}^{k+1}(G, \bar{u}^{j,a}, H, \bar{v}^{j,b})$. Since $C^{k,r-1}(\bar{u}^{j,a}) = C^{k,r-1}(\bar{v}^{j,a})$, Duplicator succeeds by the induction assumption.

Assume now that (13) is false. It follows that $C^{k,r-1}(\bar{u}) \ne C^{k,r-1}(\bar{v})$ (then Spoiler has a winning strategy by the induction assumption) or there is a vertex a in one of the graphs, say in G, such that for every b in the other graph H we have $C^{k,r-1}(\bar{v}^{j,a}) \ne C^{k,r-1}(\bar{u}^{j,b})$ for some $j = j(b)$. In the latter case Spoiler in his first move places the $(k+1)$-th pebble on a. Let b be the vertex selected in response by Duplicator. In the second move Spoiler will remove the $j(b)$-th pebble, which implies that the players essentially play $\text{EHR}_{r-1}^{k+1}(G, \bar{u}^{j,a}, H, \bar{v}^{j,b})$ from now on. By the induction assumption, Spoiler wins. □

LEMMA 4.3. *Equality* $\textit{diag}\, C^{k,r}(G) = \textit{diag}\, C^{k,r}(H)$ *is true for the standard (resp. count-free) version of the* k*-dim WL iff Duplicator has a winning strategy in the counting (resp. standard) version of* $\text{EHR}_{r+1}^{k+1}(G, H)$.

PROOF. We consider the standard version of the algorithm; the proof for the count-free version is very similar. If the multisets $\textit{diag}\, C^{k,r}(G)$ and $\textit{diag}\, C^{k,r}(H)$ are not equal, Spoiler has a winning strategy in the counting game $\text{EHR}_{r+1}^{k+1}(G, H)$. In the first round he makes a counting move that forces pebbling $a \in V(G)$ and $b \in V(H)$ so that $C^{k,r}(a^k) \ne C^{k,r}(b^k)$. The remainder of the game is equivalent to the counting game $\text{EHR}_r^{k+1}(G, a^k, H, b^k)$, where Spoiler has a winning strategy by Lemma 4.2.

If the multisets $\textit{diag}\, C^{k,r}(G)$ and $\textit{diag}\, C^{k,r}(H)$ are equal, Duplicator is able to play the first round so that $C^{k,r}(a^k) = C^{k,r}(b^k)$ for the pebbled vertices a and b. She wins the remaining game again by Lemma 4.2. □

We say that the r-round k-dim WL *works correctly for a graph G* if its output is correct on all input pairs (G, H) (here H may have any order, not necessary the same as G).

THEOREM 4.4. *The r-round k-dim WL works correctly for G if*

$$k \geq W_{\#}(G) - 1 \quad \text{and} \quad r \geq D_{\#}^{k+1}(G) - 1$$

and only if

$$k \geq W_{\#}(G) - 1 \quad \text{and} \quad r \geq D_{\#}^{k+1}(G) - k.$$

The same holds true for the count-free r-round k-dim WL and the standard logic (without counting).

PROOF. If $G \cong H$, the output is correct in any case. Suppose that $G \ncong H$. By Lemma 4.1, inequality (11) is a sufficient condition for the output being correct while (12) is a necessary condition for this. The theorem now follows from Lemma 4.3, the Ehrenfeucht theorem (Theorem 3.3.4,5), and Lemma 2.5.1 along with its counting version. □

By Theorem 4.4, $k \geq W_{\#}(G) - 1$ is both a sufficient and a necessary condition for a successful work of the k-dim WL on all inputs (G, H). As we already discussed, the number of rounds can be taken $r = O(n^k)$. Therefore, Graph Isomorphism is solvable in polynomial time for any class of graphs C with $W_{\#}(G) = O(1)$ for all $G \in C$. This applies to any class of graphs embeddable into a fixed surface and any class of graphs with bounded treewidth (see Section 5.1).

Sometimes the Weisfeiler-Lehman algorithm gives us even better result, namely the solvability of the isomorphism problem by a parallel algorithm in polylogarithmic time. The concept of polylogarithmic parallel time is captured by the complexity class NC and its refinements: $\mathrm{NC} = \bigcup_i \mathrm{NC}^i$ and $\mathrm{NC}^i \subseteq \mathrm{AC}^i \subseteq \mathrm{TC}^i \subseteq \mathrm{NC}^{i+1}$, where NC^i consists of functions computable by circuits of polynomial size and depth $O(\log^i n)$, AC^i is an analog for circuits with unbounded fan-in, and TC^i is an extension of AC^i allowing threshold gates. As it is well known [**49**], AC^i consists of exactly those functions computable by a CRCW PRAM with polynomially many processors in time $O(\log^i n)$. Grohe and Verbitsky [**42**] point out that the r-round k-dim WL (resp. its count-free version) is implementable in TC^1 (resp. AC^1) as long as $k = O(1)$ and $r = O(\log n)$. If combined with Theorem 4.4, this gives us the following result.

THEOREM 4.5. *Let $k \geq 2$ be a constant.*

1. *Let C be a class of graphs G with $D_{\#}^k(G) = O(\log n)$. Then Graph Isomorphism for C is solvable in* TC^1.
2. *Let C be a class of graphs G with $D^k(G) = O(\log n)$. Then Graph Isomorphism for C is solvable in* AC^1.

We will see applications of Theorem 4.5 in Section 5.1.

Suppose that $k \geq W(G) - 1$ and that we do not know a priori any bounds for $D^{k+1}(G)$. How large has r to be taken in order to ensure that the r-round k-dim WL works correctly for G? An answer is given by an important concept of *color stabilization* that was already discussed in the beginning of this section. We will regard $C^{k,r}$ as a partition of $V(G)^k \cup V(H)^k$. Let R be the minimum number for which $C^{k,R} = C^{k,R-1}$. Of course, it is enough to check the condition (10) for $r = R$; it cannot change for bigger r. Since each $C^{k,r}$ is a refinement of $C^{k,r-1}$, we have

$R \le v(G)^k + v(H)^k$. In fact, we are able to prove a bit more delicate claim: The Weisfeiler-Lehman algorithm can be terminated as soon as $C^{k,r}$ stabilizes at least within $V(G)^k$.

To make this more precise, we introduce some notation. Denote the restriction of the partition $C^{k,r}$ to $V(G)^k$ by $C_G^{k,r}$. Let $Stab^k(G)$ be the smallest number s such that $C_G^{k,s} = C_G^{k,s-1}$. Note that $Stab^k(G)$ is an individual combinatorial parameter of a graph G, not depending on H (we may think that the k-dim WL is run on a single graph G, which is actually a quite meaningful *canonization mode* of the algorithm).

We now state practical termination rules for the k-dim WL.

Rule 1: Once $C^{k,r}(G) \ne C^{k,r}(H)$, terminate and report non-isomorphism.

Rule 2: Once $r = Stab^k(G)$ and $C^{k,r}(G) = C^{k,r}(H)$, terminate and report isomorphism.

Let us argue that these rules are sound for both versions of the algorithm. Suppose that Rule 2 is invoked. Thus $C_G^{k,r} = C_G^{k,r-1}$ and $C^{k,r}(G) = C^{k,r}(H)$. By the latter equality we also have $C^{k,r-1}(G) = C^{k,r-1}(H)$. It follows that in the r-th round the algorithm achieves a proper color refinement on neither G nor H. Thus, the partition $C^{k,r}$ has been stabilized on $V(G)^k \cup V(H)^k$ and the soundness of Rule 2 follows.

THEOREM 4.6.

1. *The r-round k-dim WL recognizes non-isomorphism of G and H if*
$$k \ge W_{\#}(G,H) - 1 \quad \text{and} \quad r \ge Stab^k(G).$$
2. *The r-round k-dim WL works correctly for G if*
$$k \ge W_{\#}(G) - 1 \quad \text{and} \quad r \ge Stab^k(G).$$
3. *Both claims hold true for the count-free version of the algorithm and the standard logic (with no counting).*

We have seen that good bounds for the logical complexity of graphs imply efficiency of the Weisfeiler-Lehman algorithm on these graphs. Now we will get a couple of noteworthy facts on the logical complexity as a consequence of our analysis of the algorithm.

THEOREM 4.7. *Let G be a graph of order n.*

1. *If G is distinguishable from another graph H in the ℓ-variable logic, then $D^\ell(G,H) \le n^{\ell-1} + \ell - 2$.*
2. *If G is definable in the ℓ-variable logic, then $D^\ell(G) \le n^{\ell-1} + \ell - 2$.*

PROOF. Let $k = \ell - 1$. Comparing the sufficient conditions for the correctness of the r-round k-dim WL given by Theorem 4.6 and the necessary conditions given by Theorem 4.4, we have $D^{k+1}(G,H) \le Stab^k(G) + k$ provided $k \ge W(G,H) - 1$ and $D^{k+1}(G) \le Stab^k(G) + k$ provided $k \ge W(G) - 1$. For the former claim we need also the fact, actually established in the proof of Theorem 4.4, that the count-free r-round k-dim WL is able to recognize non-isomorphism of G and H only if $k \ge W(G,H) - 1$ and $r \ge D^{k+1}(G,H) - k$. It remains to notice that $Stab^k(G) \le n^k - 1$. □

A somewhat weaker bound $D^\ell(G) \le n^\ell + \ell + 1$ follows from the work of Dawar, Lindell, and Weinstein [**20**, Corollary 4].

5. Worst case bounds

5.1. Classes of graphs. Here we overview known bounds for the logical depth and width for natural classes of graphs. Several interesting definability effects can be observed even when we focus on so simple graphs as trees. This class is considered at the beginning of this section (and will be further discussed in Sections 6 and 7). We will see that many results about trees admit generalization to graphs with bounded treewidth. We further consider planar graphs. Then we briefly discuss more general cases of graphs embeddable into a fixed surface and graphs with an excluded minor, as well as a few sporadic results on other classes.

5.1.1. *Trees.* The following result is based on Edmonds' algorithm, that dates back to the sixties (see, e.g., [**16**]), and its logical interpretation is due to Immerman and Lander [**47**].

THEOREM 5.1.

1. *The color refinement algorithm succeeds in recognizing isomorphism of trees. Consequently,* $W_{\#}(T, T') \le 2$ *for every two non-isomorphic trees* T *and* T'.
2. $W_{\#}(T) \le 2$ *for every tree* T.

PROOF. *1.* As in Section 4, let C^r denote the coloring appearing after the r-th refinement. Let $N_r(v)$ denote the set of all vertices at the distance at most r from a vertex v. It is not hard to see that, if v is an arbitrary vertex in a tree T, then the subtree spanned by $N_r(v)$ is, up to isomorphism, reconstructible from $C^r(v)$. Let v and v' be arbitrary vertices in trees T and T'. If $T \not\cong T'$, we have $C^r(v) \neq C^r(v')$ at latest for r one greater than the smaller of the eccentricities of v and v'. Therefore, the color refinement algorithm distinguishes between any two non-isomorphic trees. The second statement of Part 1 follows by Lemma 4.2.1 and Theorem 3.3.5.

2. To obtain the desired definability result, we use the equality $W_{\#}(T) = \max_{H \not\cong T} W_{\#}(T, H)$, which is an analog of Lemma 2.5.2 (with a much simpler proof as graphs of different orders are distinguishable with a single counting quantification). Thus, it suffices to prove that $W_{\#}(T, H) \le 2$ whenever $H \not\cong T$. Suppose that H is not a tree for otherwise we are done by Part 1. Also, as it was just mentioned, we can suppose that both T and H have n vertices.

Assume first that H has a connected component T' which is a tree. Note that $T' \not\cong T$ because T' has less than n vertices. Let $v \in V(T)$ and $v' \in V(T')$. Run the color refinement algorithm on input (T, H). As in the proof of Part 1 we have $C^n(v) \neq C^n(v')$ because the coloring C^n on T' is the same as if the algorithm was run on T' instead of H. Therefore, T and H are distinguishable with 2 variables in the counting logic.

If none of the connected components of H is a tree, then H has at least n edges. Since T has exactly $n - 1$ edges, H and T have distinct multisets of vertex degrees and, hence, are distinguishable by a sentence with 2 counting quantifiers. □

The proof of Theorem 5.1 gives us only a linear upper bound $D^2_{\#}(T) = O(n)$ for a tree of order n. We can get a speed-up if we allow more variables.

THEOREM 5.2. *For every tree* T *on* n *vertices we have* $D^3_{\#}(T) < 3 \log n$.

PROOF. By an analog of Lemma 2.5.1 for the counting logic and Theorem 3.3.5, we have to show that Spoiler is able to win the counting game $\mathrm{EHR}^3_r(T, T')$

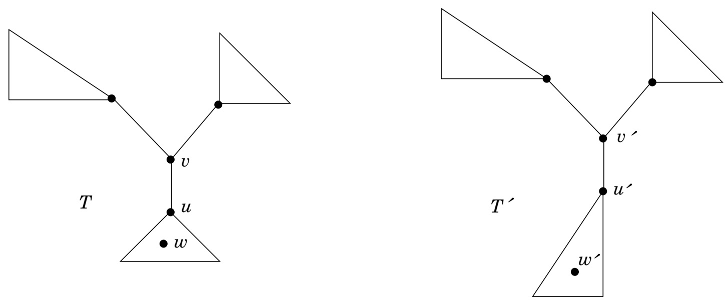


FIGURE 1. A separator strategy of Spoiler.

with some $r < 3 \log n$ for any graph T' non-isomorphic to T. Suppose that T' has the same order n. If T' is disconnected, Spoiler wins (even without counting moves) by Lemma 3.2. If T' is connected and has a cycle, then T and T' have distinct multisets of vertex degrees. Therefore, we will suppose that T' is a tree too.

Every tree T has a single-vertex *separator*, that is, a vertex v such that no branch of $T - v$ has more than $n/2$ vertices; see, e.g., Ore [**59**, Chapter 4.2]. The idea of Spoiler's strategy is to pebble such a vertex and to force further play on some non-isomorphic branches of T and T', where the same strategy can be applied recursively.

Thus, in the first round Spoiler pebbles a separator v in T and Duplicator responds with a vertex v' somewhere in T'. The component of $T - v$ containing a neighbor u of v will be denoted by T_{vu} and considered a rooted tree with the root at u. A similar notation will apply also to T'. In the second round Spoiler makes a counting move and ensures that $u \in N(v)$ and $u' \in N(v')$ are pebbled so that the rooted trees T_{vu} and $T'_{v'u'}$ are non-isomorphic, see Fig. 1. The next goal of Spoiler is to force pebbling adjacent vertices v_1 and u_1 in T_{vu} and adjacent vertices v'_1 and u'_1 in $T'_{v'u'}$ so that $T_{v_1u_1} \not\cong T'_{v'_1u'_1}$, $V(T_{v_1u_1}) \subset V(T_{vu})$, and $v(T_{v_1u_1}) \le v(T_{vu})/2$. Once this is done, the same will be repeated recursively.

To make the transition from T_{vu} to $T_{v_1u_1}$, Spoiler follows three rules.

Rule 1. If T_{vu} has a branch T_{ux} for some $x \in N(u) \setminus \{v\}$ such that $v(T_{ux}) \le v(T_{vu})/2$ and the number of branches isomorphic to T_{ux} is different for T_{vu} and $T'_{v'u'}$, then Spoiler makes a counting move and forces pebbling such x and $x' \in N(u') \setminus \{v'\}$ so that $T_{ux} \not\cong T'_{u'x'}$. The latter two branches will serve as $T_{v_1u_1}$ and $T'_{v'_1u'_1}$. If no such branch is available, Spoiler pebbles a separator w of T_{vu}. Note that Duplicator is forced to respond with a vertex w' in $T'_{v'u'}$. Otherwise we would have $dist(w, u) = dist(w, v) - 1$ while $dist(w', u') = dist(w', v') + 1$. Therefore, some distances among the three pebbled vertices would be different in T and in T' and Spoiler could win in less than $\log v(T_{vu}) + 1$ moves by Lemma 3.2.

Rule 2. If T differs from T' by some branch T_{wx} (having a different number of occurrences in $T' - w'$) that does not contain u, Spoiler makes a counting move with the pebble released from v and forces pebbling such x in T and some x' in T' so that $T_{wx} \not\cong T'_{w'x'}$. These branches will serve as $T_{v_1u_1}$ and $T'_{v'_1u'_1}$. (It is possible that $T'_{w'x'}$ contains u' or T_{wx} contains u but then the distances among u, w, x are not all equal to the distances among u', w', x' and Spoiler quickly wins.)

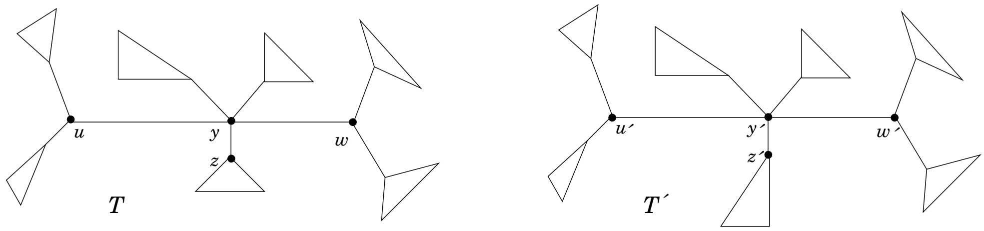


FIGURE 2. Rule 3 invoked.

Rule 3. Denote the branch of $T_{vu} - w$ containing u by $T_{w,u}$ (and similarly for T'). If Rule 2 is not applicable, then $T_{w,u}$ and $T'_{w',u'}$ are non-isomorphic (where an isomorphism would need to respect two pairs of designated vertices, namely u and u' as well as the neighbors of w and w'). Assume the harder subcase that $dist(u, w) = dist(u', w')$. When Spoiler pebbles a vertex y on the path from u to w by moving the pebble from v, Duplicator is forced to pebble the corresponding vertex y' on the path from u' to w'. It is easy to see that a vertex $y \neq w$ can be chosen so that $T - y$ and $T' - y'$ differ by branches containing neither u and u' nor w and w'. Let Spoiler pebble such y as close to w as possible. Note that $y \neq u$ because otherwise Rule 1 was applicable. Now Spoiler can make a counting move with the pebble released from w to force pebbling z and z' for which

$$T_{yz} \not\cong T'_{y'z'}, \tag{14}$$

see Fig. 2. This complies with the goal of finding new $T_{v_1u_1}$ and $T'_{v'_1u'_1}$ because

$$v(T_{yz}) < v(T_{w,u}) \leq v(T_{vu})/2. \tag{15}$$

In fact, Duplicator could try to prevent the fulfillment of (14) by forcing a choice of z' such that $T'_{y'z'}$ would contain either u' or w'. In the former case Spoiler could win by using differences between the distances among u, y, z and among u', y', z'. In the latter case (14) would anyway be true because T_{yz} and $T'_{y'z'}$ would have different orders. Indeed, since Rule 2 was not applicable, the choice of y ensures that the branches of $T - y$ and $T' - y'$ containing w and w' are isomorphic. Thus, we would have $v(T'_{y'z'}) = v(T_{y,w}) \geq v(T_{vu})/2$ (where $T_{y,w}$ denotes the branch of $T - y$ containing w) while $v(T_{yz})$ is strictly smaller by (15).

Given T_{vu}, Spoiler finds a new distinguishing branch $T_{v_1u_1}$ in 3 rounds in the worst case. Also, 2 rounds suffice to win the game once the current subtree T_{vu} has at most 4 vertices. The number of transitions from the initial branch of order at most $\lceil n/2 \rceil$ to one with at most 4 vertices is bounded by $\log\lceil n/2 \rceil - 2$ because $v(T_{vu})$ becomes twice smaller each time. Routine calculations (and Lemma 3.2) imply the desired bound on the length of the game. $\square$

The definability of trees in a finite-variable counting logic within logarithmic quantifier depth can also be derived from a work by Etessami and Immerman [**27**], which also implies that counting quantifiers are here not needed as long as the maximum vertex degree is bounded by a constant.

Curiously, Theorem 5.2 sheds some new light on the history of isomorphism testing for trees. The first record of this history was made by Edmonds, who showed that the problem is solvable in linear time (see Theorem 5.1). Ruzzo [**71**] found an AC^1 algorithm under the condition that the vertex degrees of input trees are at most logarithmic in the number of vertices. Miller and Reif [**55**] established an AC^1 upper bound unconditionally. They wrote [**55**, page 1128]: "No polylogarithmic parallel algorithm was previously known for isomorphism of unbounded-degree trees." However, the 2-dimensional Weisfeiler-Lehman algorithm has been discussed in the literature at least since 1968 (e.g., [**77**]) and, as we now see by combining Theorem 5.2 with Theorem 4.5.1, this algorithm does the job for arbitrary trees in NC^2, i.e., in parallel time $O(\log^2 n)$!

To complete this historical overview, we have to mention a result by Lindell [**53**] who showed that isomorphism of trees is recognizable in logarithmic space. Though Lindell's result is best possible (see Jenner et al. [**48**]), the solvability of the problem by so simple and natural procedure as the Weisfeiler-Lehman algorithm still remains a noteworthy fact.

Note that $D^2_\#(P_n) = \frac{n}{2} - O(1)$ (it is not hard to see that the color refinement algorithm requires at least $\frac{n}{2} - O(1)$ rounds to distinguish between P_n and the disjoint union of P_{n-3} and C_3). Thus, Theorem 5.2 shows a jump from linear to logarithmic quantifier depth under increase the number of variables just by 1. Such width-depth trade-offs were observed and studied by Fürer [**31**].

Theorem 5.1 says that 2 variables and counting quantifiers suffice to define any tree. Moreover, we could well manage without counting quantifiers but then we would need to have $\Delta(T) + 1$ variables. A simple example of a star, where $W(K_{1,m}) = m+1$, shows that a smaller number is not enough. The following bound is a variant of a result by Immerman and Kozen [**46**], who consider definability of trees represented by an asymmetric child-parent relation between vertices.

THEOREM 5.3. *$W(T) \le \Delta(T) + 1$ for any tree T with the exceptions of $T \in \{P_1, P_2\}$.*

The logical depth of a tree can be bounded in terms of the maximum degree and the order.

THEOREM 5.4 (Bohman et al. [**12**]).

1. *For every tree T of order n with maximum vertex degree $\Delta(T) \ge 9$ we have*
$$D(T) \le \left(\frac{\Delta(T)}{2\log(\Delta(T)/2)} + 3\right)\log n + \frac{3\Delta(T)}{2} + O(1).$$
2. *Let $D(n,d)$ be the maximum of $D(T)$ over all trees with n vertices and maximum degree at most $d = d(n)$. If both d and $\log n/\log d$ tend to infinity, then*
$$D(n,d) = \left(\frac{1}{2} + o(1)\right) d\,\frac{\log n}{\log d}.$$

The upper bounds on $D(T)$ comes from Spoiler's strategy similar to that of the proof of Theorem 5.2, that is, Spoiler pebbles a separator v of the given tree T and then tries to restrict the game to one of the components of $T - v$. Informally speaking, the worst case scenario for Spoiler is when $T - v$ has d components of order about n/d of two different isomorphism types, each occurring half of the time. Then Spoiler may need around $d/2$ extra moves to restrict game to a component of $T - v$ (if the components of the counterpart $T' - v'$ are of these two types but with different multiplicities). Thus, roughly, Spoiler "reduces" the order by factor d using $d/2$ moves, which gives the heuristic for the bound of Theorem 5.4.2. The optimality of this bound is given by a recursive construction of a tree T (and another tree $T' \not\cong T$), where at each recursion step we glue together about $d/2$ trees of two different isomorphism types at a common root.

5.1.2. *Graphs of bounded treewidth.* Informally speaking, the *treewidth* of a graph tells us to which extent the graph is representable as a tree-like structure. This concept appeared in the Robertson-Seymour theory and, aside of its theoretical importance, found a lot of applications in design of algorithms on graphs. We do not go into any detail here, referring instead to the books [**22**] and [**23**] that may serve as introductions to, respectively, the structural theory of graphs and the algorithmic applications.

It happens quite often that techniques applicable to trees can be extended to graphs whose treewidth is bounded by a constant. In particular, this is true for the definability parameters.

THEOREM 5.5.

1. (Grohe and Mariño [**40**]) *If a graph G has treewidth k, then $W_{\#}(G) \le k+2$.*
2. (Grohe and Verbitsky [**42**]) *If a graph G on n vertices has treewidth k, then*

$$D_{\#}^{4k+4}(G) < 2(k+1)\log n + 8k + 9.$$

Consequently, isomorphism of graphs whose treewidth does not exceed k is recognizable by the $(4k+3)$-dimensional Weisfeiler-Lehman algorithm in TC^1.

The last claim in the theorem follows a general paradigm provided by Theorem 4.5.1: A low quantifier depth implies solvability of the isomorphism problem in NC. Prior to [**42**], for graphs with bounded treewidth only polynomial-time isomorphism test of Bodlaender [**11**] was known. Very recently Das, Torán, and Wagner [**18**] put the problem in the complexity class LOGCFL.

Like Theorem 5.2, the proof of Theorem 5.5.2 is based on separator techniques. In general, a set $X \subset V(G)$ will be called a *separator* for graph G if any component of $G \setminus X$ has at most $n/2$ vertices. It is well known [**70**] that all graphs of treewidth k have separators of size $k+1$.

5.1.3. *Planar graphs.* The separator techniques in the study of logical complexity of graphs were introduced by Cai, Fürer, and Immerman [**15**], who derived a bound $W_{\#}(G) = O(\sqrt{n})$ for planar graphs from the known fact [**54**] that every planar graph of order n has a separator of size $O(\sqrt{n})$. In fact, this result is a particular case of Theorem 5.5.1 because planar graphs have treewidth bounded by $5\sqrt{5}\,n$; see [**3**, Proposition 4.5]. Later Grohe [**35**] proved that $W_{\#}(G)$ for all planar G is actually bounded by a constant.

Without counting quantifiers we cannot have any nontrivial upper bound for the logical depth in terms of the order of a graph as long as a class under consideration contains all trees. However, some natural classes of planar graphs admit such bounds. A plane drawing of a graph is called *outerplanar* if all the vertices lie on the boundary of the outer face. *Outerplanar graphs* are those planar graphs having an outerplanar drawing. The treewidth of any outerplanar graph is at most 2. As it is well known (see, e.g., [**43**]), any outerplanar graph is representable as a tree of its biconnected components. Note also that an outerplanar graph is biconnected iff it has a Hamiltonian cycle and that such a graph can be geometrically viewed as a dissection of a convex polygon.

THEOREM 5.6 (Verbitsky [**75, 76**]).

1. *If G is a biconnected outerplanar graph of order n, then*

$$D(G) < 22 \log n + 9.$$

2. *For a 3-connected planar graph G of order n we have*

$$D^{15}(G) < 11 \log n + 45.$$

Part 2 shows another case when Theorem 4.5 is applicable. It gives an AC^1 isomorphism test for 3-connected planar graphs and, by a known reduction of Miller and Reif [**55**], for the whole class of planar graphs. This complexity bound for the planar graph isomorphism is not new; it follows from the AC^1 isomorphism test for embeddings designed in [**55**] and the AC^1 embedding algorithm in [**68**]. As a possible advantage of the Weisfeiler-Lehman approach, note that it is combinatorially much simpler and more direct. In particular, we do not need any embedding procedure here. The best possible complexity bound for the planar graph isomorphism is recently obtained by Datta et al. [**19**] who design a logarithmic-space algorithm for this problem.

Theorem 5.6.1 is proved in [**75**] and is based on the existence of a 2-vertex separator in any outerplanar graph. The possibility to avoid counting quantifiers relies on certain rigidity of biconnected outerplanar graphs. The latter is related to the following geometric fact: Any such graph has a unique, up to homeomorphism, outerplanar drawing.

The case of 3-connected planar graphs is much more complicated because the smallest separators in such graphs can have about $\sqrt{n}$ vertices (such examples can be obtained by adding a few edges to the grid graph $P_m \times P_m$). The proof of Theorem 5.6.2 in [**76**] exploits a strong rigidity property of 3-connected planar graphs: By the Whitney theorem (see, e.g., [**56**]), they have a unique, up to homeomorphism, embedding into the sphere. An embedding can be represented as a purely combinatorial structure, called a *rotation system* (see [**56**]), to which one can extend the concepts of definability, isomorphism, the Ehrenfeucht game etc. Defining rotation systems is a simpler business because they admit a kind of coordinatization and hence an analog of the halving strategy from Lemma 3.2 is available for Spoiler. The most essential ingredient of the proof of Theorem 5.6.2 is a strategy for Spoiler in the Ehrenfeucht game on graphs allowing him to simulate the Ehrenfeucht game on the corresponding rotation systems.

5.1.4. *Graphs with an excluded minor.* No graph with treewidth h has K_{h+2} as a minor. The class of graphs embeddable into a closed 2-dimensional surface S is closed under minors and, as follows from the Robertson-Seymour Graph Minor Theorem, no graph from this class contains a minor of K_h for some $h = h(S)$. Extending his earlier work on graphs embeddable into a fixed surface [**36**], Grohe [**38**] recently announced a proof that, if a graph G does not contain K_h as a minor, then $W_{\#}(G)$ is bounded by a constant $c = c(h)$. The case of $h = 5$ is treated in detail in [**37**].

Alon, Seymour, and Thomas [**3**] proved that, if a graph G of order n does not contain a K_h as a minor, then it has a separator of size at most $h^{3/2}\sqrt{n}$. Using this result, for all connected graphs with this property one can prove [**75**] that $D(G) = O(h^{3/2}\sqrt{n}) + O(\Delta(G)\log n)$.

5.1.5. *Other classes of graphs.* A graph is *strongly regular* if all its vertices have equal degrees and, for some λ and μ, each pair of adjacent vertices has exactly λ common neighbors and each pair of non-adjacent vertices has exactly μ common neighbors. Non-isomorphic graphs with the same order, degree, and parameters λ and μ are standard examples of a failure of the 2-dim WL algorithm. Babai studies the isomorphism problem for this class in [**6**]. His individualization-and-refinement technique translates into a bound $W_{\#}(G) \le 2\sqrt{n}\log n$ for all strongly regular graphs of a sufficiently large order n with the exception for the disjoint unions of complete graphs and their complements (for which we have $D_{\#}(G) \le 3$). Further improvements are obtained by Spielman [**74**].

Evdokimov, Ponomarenko, and Tinhofer [**28**] undertake an analysis of the 3-dimensional WL algorithm on the classes of cographs, interval graphs, and even directed path graphs (the latter class extends the class of interval graphs and contains also all ptolemaic graphs, in particular, trees). It follows from [**28**] that $W_{\#}(G) \le 4$ for all G in any of these classes. The boundedness of $W_{\#}(G)$ for interval graphs follows also from the paper of Laubner [**51**], who uses purely logical methods (while Evdokimov et al. develop an algebraic approach that, in fact, originates from the seminal work by Weisfeiler and Lehman).

Grohe [**39**] proves that $W_{\#}(G) = O(1)$ for all chordal line graphs. On the other hand, he shows that there are chordal graphs with $W_{\#}(G) = \Omega(n)$ and the same holds true for line graphs. The latter result is obtained by a reduction to the graphs with $W_{\#}(G) = \Omega(n)$ constructed by Cai, Fürer, and Immerman [**15**] (cf. Theorem 5.7 below). Note that the Cai-Fürer-Immerman graphs are regular of degree 3, where the regularity can be traded for the bipartiteness after a slight modification.

5.2. General case.

5.2.1. *Identification problem.* Recall that

$$W_{\#}(G,H) \le W(G,H) \le D(G,H) \text{ and } W_{\#}(G,H) \le D_{\#}(G,H) \le D(G,H).$$

If we are motivated by the graph isomorphism problem, it is quite natural to focus on these parameters under the assumption that G and H have the same order (even without saying that $D_{\#}(G,H) = 1$ otherwise). Distinguishing a graph G from all non-isomorphic H of *the same order* is sometimes called *identification problem.* In particular, we would like to determine or estimate the maximum of $D(G,H)$ (resp. $D_{\#}(G,H)$) as a function of $n = v(G) = v(H)$. Equivalently, what is the minimum $r = r(n)$ such that Spoiler has a winning strategy in $\textsc{Ehr}_r(G,H)$ for all non-isomorphic G and H of order n?

By taking disjoint unions of complete and empty graphs, it is easy to find G and H with $D(G,H) \ge (n+1)/2$. Bounding $D_{\#}(G,H)$ from below is much more subtle issue. Using a nice nontrivial argument, Cai, Fürer, and Immerman [**15**] came up with a linear lower bound.

THEOREM 5.7 (Cai, Fürer, and Immerman [**15**]). *For infinitely many n there are non-isomorphic graphs G and H both of order n such that $W_{\#}(G,H) \ge cn$, where c is a positive constant.*

The calculation of Pikhurko et al. [**64**, Section 7.5] shows that one can take $c = 0.00465$.

Let us turn to upper bounds. Suppose that $G \not\cong H$ and $v(G) = v(H) = n$. Before reading further, the reader might try to improve the trivial bound $D(G, H) \le n$ at least somewhat. It may be seen as a curious observation that $D(G, H) \le n-1$ follows from the Harary version of the Ulam Reconstruction Conjecture, open for a long time, claiming that non-isomorphic graphs of equal orders have different sets of vertex-deleted subgraphs.

One solution of this exercise, giving $D(G, H) < n - \frac{1}{4}\log n$, is to apply the Erdős-Szekeres bound on Ramsey numbers. It implies that every graph G of large order n contains a homogeneous set of more than $\frac{1}{2}\log n + \frac{1}{4}\log\log n$ vertices. Spoiler pebbles the complement of such a set S in G. Suppose that the unpebbled set is independent (otherwise we can play on the complementary graphs). If Duplicator is lucky, she manages to pebble the complement to an independent set S' in H so that $G \setminus S \cong H \setminus S'$. Identifying the pebbled parts, Spoiler compares the number of vertices in S and in S' with the same neighborhood. These numbers cannot be identical for G and H and, by $v(G) = v(H)$, Spoiler can demonstrate this using at most $(|S| + 1)/2$ further moves in one of the graphs.

After this warm-up, we can state an almost optimal bound.

THEOREM 5.8 (Pikhurko, Veith, and Verbitsky [**63**]). *For every two non-isomorphic graphs G and H of the same order n we have $D(G, H) \le (n + 3)/2$.*

5.2.2. *General bounds for the logical depth and width.* In the case of the counting logic, Theorem 5.7 provides us with infinitely many graphs G for which $D_{\#}(G) \ge W_{\#}(G) > 0.00465\, n$. As usually, n denotes the order of a graph. An upper bound easily follows from Theorem 5.8: we have $D_{\#}(G) \le 0.5\, n + 1.5$ for all G. Though this bound does not use the power of counting quantifiers at all, we are not aware of any better bound.

Consider the standard first-order logic (without counting). At the first sight, everything is clear here. Indeed, the general upper bound $W(G) \le D(G) \le n + 1$ is attained, even for the width, by the complete graph K_n and by the empty graph $\overline{K_n}$. However, these are the only two extremal graphs. In other words, $D(G) \le n$ for all G with exception of $G \in \{K_n, \overline{K_n}\}$. As K_n and $\overline{K_n}$ are the most symmetric graphs, this observation suggests two problems. The first one is to prove a better bound for a class of graphs with restrictions on the automorphism group. The second is to obtain, for as small as possible $l = l(n)$, an explicit or algorithmic description of all order-n graphs whose logical depth (resp. width) exceeds l. We start with the first problem.

DEFINITION 5.9. Let u, v, and s be three vertices and $s \notin \{u, v\}$. We say that s *separates* u and v if s is adjacent to exactly one of the two vertices. Furthermore, we call u and v *twins* if no s separates u and v (or, equivalently, if the transposition of u and v is an automorphism of the graph). A graph is called *twin-free* if it has no twins.

THEOREM 5.10 (Pikhurko, Veith, and Verbitsky [**63**]). *If G is twin-free, then $D_1(G) \le (n + 5)/2$.*

Theorem 5.10 cannot be improved to a sublinear bound. Indeed, consider mP_4, the disjoint union of m copies of P_4. As it is easily seen, mP_4 is twin-free and $D(mP_4) \ge D(mP_4, (m + 1)P_4) > m$ (the reader is welcome to play $\mathrm{EHR}_m(mP_4, (m+1)P_4)$ on Duplicator's side). No sublinear improvement is possible even for connected graphs: the graphs constructed by Cai, Fürer, and Immerman in Theorem 5.7 are twin-free.

We prove Theorem 5.10 based on Lemma 2.5.1 and the Ehrenfeucht Theorem (Theorem 3.3.1). That is, we design a strategy allowing Spoiler to win $\mathrm{EHR}_r(G, H)$ for any $H \not\cong G$, where $r = \lfloor (n+5)/2 \rfloor$. As an important additional feature of the strategy, Spoiler will alternate between the graphs only once. By Theorem 3.3.2, this shows that our bound holds even for the logic with only one quantifier alternation (as it is indicated by the subscript in Theorem 5.10).

DEFINITION 5.11. Let $X \subset V(G)$. Given two vertices $u, v \in V(G) \setminus X$, we call them X-*similar* and write $u \equiv_X v$ if u and v are inseparable by any vertex in X, i.e., if $N(u) \cap X = N(v) \cap X$.

Now, let $y \notin X$. We say that X *sifts out* y if for every $y' \notin X$ the relation $y \equiv_X y'$ implies $y' = y$ (in other words, the vertex y is uniquely identified by its adjacencies to X). Let $\mathcal{S}(X)$ consist of all $x \in X$ and all y sifted out by X. We call X a *sieve*[6] if $\mathcal{S}(X) = V(G)$. Furthermore, X is called a *weak sieve* if $\mathcal{S}(\mathcal{S}(X)) = V(G)$.

Consider the Ehrenfeucht game on non-isomorphic G and H and assume that X is a sieve in G. Let Spoiler pebble all vertices of X. We leave to the reader to verify that Spoiler can win in at most 2 more moves. We now describe a more advanced *Weak Sieve Strategy*.

LEMMA 5.12. *If X is a weak sieve in G, then Spoiler is able, for any $H \not\cong G$, to win* $\mathrm{EHR}_r(G, H)$ *with $r \le |X| + 3$. Moreover, he does not need to jump from one graph to the other more than once during the game.*

PROOF. First, Spoiler selects all of X. Let $X' \subset V(H)$ be the Duplicator's reply. Assume that Duplicator has not lost yet. For the notational simplicity let us identify X and X' so that $V(G) \cap V(H) = X = X'$ and the player's moves coincide on X. Let $Y = \mathcal{S}(X)$ in G and $Y' = \mathcal{S}(X')$ in H.

It is not hard to see that Spoiler wins in at most two extra moves unless the following holds. For any $y \in Y \setminus X$ there is a $y' \in Y' \setminus X$ (and vice versa) such that $N(y) \cap X = N(y') \cap X$. Moreover, this bijective correspondence between Y and Y' establishes an isomorphism between $G[Y]$ and $G'[Y']$.

Suppose that this is the case and identify Y with Y'. Let $Z = V \setminus Y$ and $Z' = V' \setminus Y$. Let $z \in Z$ and define $W'_z = \{\, z' \in Z' : N(z') \cap Y = N(z) \cap Y\}$.

If $W'_z = \emptyset$, Spoiler wins in at most two moves. First, he selects z. Let Duplicator reply with z'. Assume that $z' \in Z'$ for otherwise she has already lost. As the neighborhoods of z, z' in Y differ, Spoiler can demonstrate this by picking a vertex of Y. If $|W'_z| \ge 2$, then Spoiler selects any two vertices in W'_z and wins with at most one more move, as required.

Hence, we can assume that for any z we have $W'_z = \{f(z)\}$ for some $f(z) \in Z'$. Since each vertex in Z is sifted out by Y, the function f is injective. If $f(Z) \neq Z'$, Spoiler easily wins in two moves. Suppose, therefore, that $f : Z \to Z'$ is a bijection. As $G \not\cong H$, the mapping f does not preserve the adjacency relation between some $y, z \in Z$. Now, Spoiler selects both y and z. Duplicator cannot respond with $f(y)$ and $f(z)$; by the definition of f Spoiler can win in one extra move. □

Theorem 5.10 immediately follows from Lemma 5.12 and the next lemma.

[6] Babai [**6**] uses sieves under the name *distinguishing sets*.

LEMMA 5.13. *Any twin-free graph G on n vertices has a weak sieve X with $|X| \le (n-1)/2$.*

PROOF. Given $X \subset V(G)$, let $\mathcal{C}(X)$ denote the partition of $\overline{X} = V(G) \setminus X$ into $\equiv_X$-equivalence classes. Starting from $X = \emptyset$, we repeat the following procedure. As long as there exists $u \in \overline{X}$ such that $|\mathcal{C}(X \cup \{u\})| > |\mathcal{C}(X)|$, we move u to X. As soon as there is no such u, we arrive at X which is $\mathcal{C}$-*maximal*, that is, $|\mathcal{C}(X \cup \{u\})| \le |\mathcal{C}(X)|$ for any $u \in \overline{X}$. Note that $|\mathcal{C}(X)| \ge |X| + 1$ because this inequality is true at the beginning and is preserved in each construction step. Using also the inequality $|X| + |\mathcal{C}(X)| \le n$, we conclude that $|X| \le (n-1)/2$.

We now prove that the X is a weak sieve. Suppose, to the contrary, that u and v are distinct $\mathcal{S}(X)$-similar vertices in $Z = V(G) \setminus \mathcal{S}(X)$. Since G has no twins, these vertices are separated by some s. We cannot have $s \in \mathcal{S}(X)$ by the definition of $\mathcal{S}(X)$-similarity. Thus $s \in Z$. Let C_1 be the class in $\mathcal{C}(X)$ including $\{u, v\}$ and C_2 be the class in $\mathcal{C}(X)$ containing s. Since $s \notin \mathcal{S}(X) \setminus X$, the class C_2 has at least one more element in addition to s. If $C_1 \ne C_2$, moving s to X splits up C_1 and does not eliminate C_2. If $C_1 = C_2$, moving s to X splits up this class and splits up or does not affect the others. In either case $|\mathcal{C}(X)|$ increases, giving a contradiction. □

The proof of Theorem 5.10 is complete. This theorem was significantly extended in [**63**] giving some progress on the second research problem stated above. In particular, it was shown that one can efficiently check whether or not $D(G) \le (n+5)/2$ for the input graph G of order n and, if this is not true, then one can efficiently compute the exact value of $D(G)$. Also, the same holds for $W(G)$.

This result is interesting in view of the fact that algorithmic computability of the logical depth and width of a graph, even with no efficiency requirements, is unclear. A reason for this is that the question if a given first-order sentence defines some graph is known to be undecidable [**61**].

The upper bound of $\frac{1}{2}n + O(1)$ can be improved if we impose a restriction on the maximum vertex degree.

THEOREM 5.14 (Pikhurko, Veith, and Verbitsky [**63**]). *Let $d \ge 2$. Let G be a graph of order n with no isolated vertex and no isolated edge. If $\Delta(G) \le d$, then*

$$D_1(G) < c_d\, n + d^2 + d + 4$$

for a constant $c_d = \frac{1}{2} - \frac{1}{4} d^{-2d-5}$.

Theorem 5.14 aims at showing a constant c_d strictly less than $1/2$ rather than at attempting to find the optimum c_d. In the case of $d = 2$, which is simple and included just for uniformity, an optimal bound is $D_1(G) \le n/3 + O(1)$. Without the assumption that G has no isolated vertex and edge, the theorem does not hold for any fixed $c_d < 1/2$. A counterexample is provided by the disjoint union of isolated edges. Even under the stronger assumption that G is connected, Theorem 5.14 still does not admit any sublinear improvement: the Cai-Fürer-Immerman graphs in Theorem 5.7 are connected and have maximum degree 3.

6. Average case bounds

In Section 5 we investigated the maximum values of logical parameters over graphs of order n. Now we want to know its typical values. A natural setting for this problem is given by the Erdős-Rényi model of a random graph $G_{n,p}$. The latter is a random graph on n vertices where every two vertices are connected by an edge with probability p independently of the other pairs. A particularly important case is $G_{n,1/2}$, when we have the uniform distribution on all graphs on a fixed set of n vertices. Whenever we say that for a random graph of order n something happens *with high probability* (abbreviated as *whp*), we mean a probability approaching 1 as $n \to \infty$.

6.1. Bounds for almost all graphs.

6.1.1. *Logic with counting.* We begin with a simple but useful observation about the color refinement algorithm described at the beginning of Section 4: If the coloring of a graph stabilizes with all color classes becoming singletons, it can be considered a canonical vertex ordering. It turns out that this happens for almost all graphs. This result can be used to estimate the logical complexity of almost all graphs, in particular, to show that almost surely $W_{\#}(G_{n,1/2}) = 2$ (Immerman and Lander [**47**]).

THEOREM 6.1.

1. (Babai, Erdős, and Selkow [**7**]) *2 color refinements split a random graph $G_{n,1/2}$ into color classes which are singletons with probability more than $1 - 1/\sqrt[7]{n}$, for all large enough n. Consequently, $D^2_{\#}(G_{n,1/2}) \le 4$ with this probability.*
2. (Babai and Kučera [**8**]) *3 color refinements split a random graph $G_{n,1/2}$ into color classes which are singletons with probability more than $1 - 1/2^{cn}$, for a constant $c > 0$ and all large enough n. Consequently, $D^2_{\#}(G_{n,1/2}) \le 5$ with this probability.*

The logical conclusions made in Theorem 6.1 are based on the necessity part of Theorem 4.4. It suffices to notice that, once the color refinement splits the vertex set of an input graph G into singletons, one extra round of the algorithm suffices to distinguish G from any non-isomorphic graph.

Next, we are going to show that the upper bound of Theorem 6.1.1 is best possible. Let C^r_G denote the coloring of the vertex set of a graph G produced by the color refinement procedure in r rounds.

LEMMA 6.2. *Whp for $G = G_{n,1/2}$ there exists a non-isomorphic graph H on $V = V(G)$ such that $C^2_G(x) = C^2_H(x)$ for every vertex $x \in V$.*

PROOF. Let G be a typical graph of a sufficiently large order n. In particular, we assume that G satisfies Theorem 6.1.1 and that $|\deg x - n/2| \le m$ for every vertex x of G, where we can take e.g. $m = \sqrt{(n \log n)/2}$ by a simple application of Chernoff's bound (see also [**13**, Corollary 3.4]). By the Pigeonhole Principle, there is a set U of $u = \lceil n/(2m+1) \rceil$ vertices all having the same degree.

Another property of the random graph G that we assume is that every set $X \subset V$ of size u contains distinct vertices w, x, y, z with $wx, yz \in E(G)$ and $xy, zw \notin E(G)$. Indeed, let us fix a u-set $X \subset V$ and estimate the probability that it violates this property. One can find at least $\binom{u}{2}/5$ edge-disjoint 4-cycles inside the complete graph on X. (For example, picking up cycles one by one arbitrarily, we get enough of them by the well-known fact that a C_4-free graph on u vertices has $O(u^{3/2})$ edges, see, e.g., [**2**, Chapter 25.5].) For each 4-cycle on vertices x_1, x_2, x_3, x_4 in this order, at least one of the relations $x_1x_2, x_3x_4 \in E(G)$ and $x_2x_3, x_4x_1 \notin E(G)$ should be false, this having probability 15/16. By the edge-disjointness, these events for different selected cycles are mutually independent. Hence X violates the desired property with probability is at most $(15/16)^{\binom{u}{2}/5} = o(\binom{n}{u}^{-1})$. Since there are $\binom{n}{u}$ candidates for a bad set X, the probability that it exists is $o(1)$, giving the required.

Hence, the equidegree set U contains vertices w, x, y, z with $wx, yz \in E(G)$ and $xy, zw \notin E(G)$. Let H be obtained from G by removing edges wx, yz and adding edges xy, zw. This operation preserves the degree of every vertex as well as the multiset of degrees of its neighbors, that is, $C^2_G(v) = C^2_H(v)$ for every vertex $v \in V$.

Suppose that G and H are isomorphic. Any isomorphism f must preserve the C^2-colors. Since C^2-classes are all singletons, f has to be the identity map on $V(G) = V(H)$. But then the adjacency between, e.g., w and x is not preserved, a contradiction. The lemma is proved. □

Given a typical $G = G_{n,1/2}$, let H be a graph satisfying Lemma 6.2. Thus, the 2-round color refinement fails to distinguish between G and H. By the sufficiency part of Theorem 4.4, we have $D^2_\#(G) > 3$. As an alternative proof, the reader can design a winning strategy for Duplicator in the counting game $\textsc{Ehr}^2_3(G, H)$. Together with Theorem 6.1.1, this bound gives us the exact value $D^2_\#(G_{n,1/2})$.

THEOREM 6.3. *Whp* $D^2_\#(G_{n,1/2}) = 4$.

We always have $D_\#(G) \le D^2_\#(G)$ and, on the other hand, $D_\#(G) \le 2$ implies $D^2_\#(G) \le 2$ because any definition with quantifier depth 2 can be rewritten with using only 2 variables. It follows from Theorem 6.3 that $3 \le D_\#(G_{n,1/2}) \le 4$ whp. Unfortunately, we could not decide whether the typical value of $D_\#(G_{n,1/2})$ is 3 or 4, which seems to be an interesting question.

6.1.2. *Logic without counting.*

THEOREM 6.4 (Kim et al. [**50**]). *Fix an arbitrarily slowly increasing function* $\omega = \omega(n)$. *Then we have whp that*

$$\log n - 2\log\log n + \log\log \mathrm{e} + 1 - o(1) \le W(G_{n,1/2}) \le \\ \le D_1(G_{n,1/2}) \le \log n - \log\log n + \omega.$$

We first prove the lower bound.

DEFINITION 6.5. For an integer $k \ge 1$, we say that a graph G has the *k-extension property* if, for every two disjoint $X, Y \subset V(G)$ with $|X \cup Y| \le k$, there is a vertex $z \notin X \cup Y$ adjacent to all $x \in X$ and non-adjacent to all $y \in Y$.

LEMMA 6.6. *If both G and H have k-extension property, then* $W(G, H) \ge k+2$.

PROOF. By Theorem 3.3.3 it suffices to design a strategy allowing Duplicator to survive in $\textsc{Ehr}^{k+1}(G, H)$ arbitrarily long. Suppose that Spoiler puts pebble p on a new position v in one of the graphs, say, G. Let X (resp. Y) denote the set of pebbled vertices in H whose counter-parts in G are adjacent (resp. non-adjacent) to v. Duplicator moves the other copy of p to a vertex z with the given adjacencies to $X \cup Y$ whose existence is guaranteed by the k-extension property. □

LEMMA 6.7. *Let $\epsilon > 0$ be a real constant. Then the k-extension property holds for $G_{n,1/2}$ whp for any* $k \le \log n - 2\log\log n + \log\log \mathrm{e} - \epsilon$.

PROOF. Let n be large. Any particular X and Y with $|X \cup Y| = k$ falsify the k-extension property with probability $(1 - 2^{-k})^{n-k}$. Since the number of such pairs is $\binom{n}{k}2^k$, a random graph $G_{n,1/2}$ does not have the k-extension property with probability at most

$$\binom{n}{k}2^k(1-2^{-k})^{n-k} \le n^k(1-2^{-k})^n \le \exp\left\{k\ln n - n2^{-k}\right\}.$$

The former inequality is true only if $k \ge 4$ but this makes no problem because the k-extension property implies itself for all smaller values of parameter k. Since the function $f(x) = x\ln n - n2^{-x}$ is monotone, the k-extension property fails with the probability bounded from above by

$$\exp\left\{f(\log n - 2\log\log n + \log\log \mathrm{e} - \epsilon)\right\} = \\ = \exp\left\{(\ln 2)\left(-2^{\epsilon} + 1 + o(1)\right)\log^2 n\right\} = o(1),$$

as it was claimed. □

Fix $\epsilon > 0$. Let n be sufficiently large and set $k = \lfloor \log n - 2\log\log n + \log\log \mathrm{e} - \epsilon \rfloor$. By Lemma 6.7, $G = G_{n,1/2}$ has the k-extension property whp. Let H be a graph which also possesses the k-extension property and is non-isomorphic to G. The existence of such a graph follows also from Lemma 6.7: Given G, let $H = G_{n,1/2}$ be another, independent copy of a random graph. It should be only noticed that $H \cong G$ with probability at most $n!2^{-\binom{n}{2}} = o(1)$. By Lemma 6.6, we have

$$W(G) \ge W(G,H) \ge k+2 > \log n - 2\log\log n + \log\log \mathrm{e} + 1 - \epsilon,$$

thereby proving the lower bound of Theorem 6.4.

To prove the upper bound, we employ the *Weak Sieve Strategy* that was designed in Section 5.2.2. Lemma 5.12 allows us to estimate the parameter $D_1(G)$ by the size of a weak sieve existing in G. The upper bound of Theorem 6.4 follows from Lemmas 6.8 below, that gives us a good enough bound for the size of a weak sieve in a random graph. The paper [**50**] states a slightly weaker upper bound than that in Theorem 6.4 (namely, $\omega = C\log\log\log n$ there). The current more precise estimate is due to Joel Spencer (unpublished).

LEMMA 6.8. *Fix an arbitrarily slowly increasing function $\omega = \omega(n)$. Then whp $G_{n,1/2}$ has a weak sieve of size at most* $\log n - \log\ln n + \omega$.

PROOF. We will consider a random graph $G_{n,1/2}$ on an n-vertex set V. Fix $X \subset V$ with $|X| = \log n - s$, where $s = \log\ln n - \omega$. We generate $G_{n,1/2}$ in two stages.

Stage 1: reveal the edges between X and $V \setminus X$ (needless to say, each such edge appears with probability $1/2$ independently of the others). Our goal at this stage is to show that $\mathcal{S}(X)$ is large whp.

A fixed $y \in V \setminus X$ is sifted out by X with probability

$$(1-2^{-|X|})^{n-|X|-1} = \exp\left\{-2^s(1+o(1))\right\} = n^{-2^{-\omega}(1+o(1))} = n^{-o(1)}.$$

By linearity of expectation

$$\mathbb{E}\left[\,|\mathcal{S}(X)\setminus X|\,\right] = (n-|X|)n^{-o(1)} = n^{1-o(1)}.$$

We can now apply the martingale techniques to show that whp $|\mathcal{S}(X)\setminus X|$ is concentrated near its mean value. More precisely, we need the following estimate:

$$\mathbb{P}\left[\,|\mathcal{S}(X)\setminus X| < \mathbb{E}\left[\,|\mathcal{S}(X)\setminus X|\,\right] - 2\lambda\sqrt{n-|X|}\,\right] < \mathrm{e}^{-\lambda^2/2} \tag{16}$$

for any $\lambda > 0$, where $\mathbb{P}[\,A\,]$ denotes the probability of an event A.

To prove it, consider the probability space consisting of all functions $g : V \setminus X \to 2^X$. Define a random variable L on this space by setting $L(g)$ to be equal to the number of values in 2^X taken on by g exactly once. Note that, if g and g' differ only at one point, then $|L(g) - L(g')| \le 2$. Construct an appropriate martingale as explained in the Alon-Spencer book [**4**, Chapter 7.4]. Namely, let $V \setminus X = \{y_1, \ldots, y_m\}$ and define a sequence of auxiliary random variables $X_0, X_1, \ldots, X_m$ by $X_i = \mathbb{E}\left[\frac{1}{2}L(g) \mid g(y_j) = h(y_j) \text{ for all } j \le i\right]$. By Azuma's inequality (see [**4**, Theorems 7.2.1 and 7.4.2]), for all $\lambda > 0$ we have

$$\mathbb{P}\left[\, L(g) < \mathbb{E}[\,L(g)\,] - 2\lambda\sqrt{m}\,\right] < \mathrm{e}^{-\lambda^2/2},$$

which is exactly what is claimed by (16).

By (16) we have whp that

$$|\mathcal{S}(X) \setminus X| \ge n^{1-o(1)}.$$

Conditioning on $\mathcal{S}(X)$ satisfying this bound, we go to the next stage of generating $G_{n,1/2}$.

Stage 2: reveal the edges inside $V \setminus X$. It is enough to show that $V \setminus \mathcal{S}(X) \subset \mathcal{S}(\mathcal{S}(X) \setminus X)$ whp. If the last claim is false, then there are $z, z' \in V \setminus \mathcal{S}(X)$ having the same adjacencies to $\mathcal{S}(X) \setminus X$. This happens with probability no more than

$$\binom{n}{2} 2^{-|\mathcal{S}(X)\setminus X|} < n^2 2^{-n^{1-o(1)}} = o(1).$$

The proof is complete. □

Theorem 6.4 shows rather close lower and upper bounds for the logical width and depth of a random graph $G_{n,1/2}$. Surprisingly, even this can be improved.

THEOREM 6.9 (Kim et al. [**50**]). *For infinitely many n we have whp*

$$D_2(G_{n,1/2}) \le \log n - 2\log\log n + \log\log \mathrm{e} + 6 + o(1).$$

This upper bound is at most by $5+o(1)$ larger than the lower bound of Theorem 6.4. It follows that, for infinitely many n, the parameters $D_i(G)$ with $i \ge 2$, $D(G)$, and $W(G)$ are all concentrated on at most 6 possible values (while some extra work, see [**50**, Section 4.3], gives a 5-point concentration).

6.1.3. *Bounds for trees.*

THEOREM 6.10 (Bohman et al. [**12**]). *Let T_n denote a tree on the vertex set $\{1, 2, \ldots, n\}$ selected uniformly at random among all n^{n-2} such trees. Whp we have $W(T_n) = (1 + o(1))\frac{\log n}{\log\log n}$ and $D(T_n) = (1 + o(1))\frac{\log n}{\log\log n}$.*

The lower bound for $W(T_n)$ immediately follows from the following property of a random tree: whp T_n has a vertex adjacent to $(1 + o(1))\frac{\log n}{\log\log n}$ leaves. Note that the upper bound for $D(T_n)$ does not follow directly from Theorem 5.4 because whp $\Delta(T_n) = (1 + o(1))\frac{\log n}{\log\log n}$ (see Moon [**57**]).

6.2. An application: The convergency rate in the zero-one law. We will write $G \models \Phi$ to say that a sentence Φ is true on a graph G. Let $p_n(\Phi) = \mathbb{P}[\,G_{n,1/2} \models \Phi\,]$. The *0-1 law* established by Glebskii et al. [**32**] and, independently, by Fagin [**29**] says that, for each Φ, $p_n(\Phi)$ approaches 0 or 1 as $n \to \infty$. Denote the limit by $p(\Phi)$.

Define the *convergency rate function* for the 0-1 law by

$$R(k,n)=\max_{\Phi}\{\,|p_n(\Phi)-p(\Phi)|:\,D(\Phi)\le k\}\,.$$

Note that the maximization here can be restricted to a finite set by Theorem 2.4.1. Therefore, the standard version of the 0-1 law implies that $R(k,n)\to 0$ as $n\to\infty$ for any fixed k. Naor, Nussboim, and Tromer [**58**] showed that $R(\log n-2\log\log n,n)\to 0$. Another result in [**58**] states that one can choose $p(n)=1/2+o(1)$ and $k(n)=(2+o(1))\log n$ such that the probability that $G_{n,p}$ has a $k(n)$-clique is bounded away from 0 and 1. Thus for this probability $p(n)$ the 0-1 law does not hold with respect to formulas of depth $k(n)$.

The following theorem sharpens slightly the first part of the above result and improves on the second part in two aspects: we do not need to change the probability $p=1/2$ and we get an almost best possible upper bound.

Theorem 6.11. *Let $g(n)=\log n-2\log\log n+\log\log\mathrm{e}+c$ with constant c.*

1. *If $c<1$, then $R(g(n),n)\to 0$ as $n\to\infty$.*
2. *If $c>6$, then $R(g(n),n)$ does not tend to 0 as $n\to\infty$. More strongly, for every $\gamma\in[0,1]$ there is a sequence of formulas $\Phi_{n_1},\Phi_{n_2},\dots$ (where $n_i<n_{i+1}$) with $D(\Phi_{n_i})\le g(n_i)$ such that $p_{n_i}(\Phi_{n_i})\to\gamma$ as $i\to\infty$.*

Part 2 follows from Theorem 6.9. The latter implies that (for infinitely many n) actually *any* property $\mathcal{P}$ of graphs on n vertices can be "approximated" by a first-order sentence of depth at most $g(n)$. Indeed, take the conjunction of defining formulas over all graphs in $\mathcal{P}$ of order n and depth at most $g(n)$. The omitted graphs constitute negligible proportion of all graphs by Theorem 6.9.

We now prove Part 1. Like the proof in [**58**] we use the extension property, but we argue in a slightly different way.

Proof of Part 1. Let E_k denote a first-order statement of quantifier depth k expressing the $(k-1)$-extension property. Lemma 6.7 provides us with an infinitesimal $\alpha(n)$ such that

$$1-p_n(E_k)\le\alpha(n)\ \text{ as long as }\ k\le g(n). \tag{17}$$

We will consider $g(x)$ on the range $x\ge 2$. This function is decreasing for $x\le\mathrm{e}^2$ and increasing for $x\ge\mathrm{e}^2$. Since $g(2)<3$, for any $k_0\ge 3$ there is some n_0 such that the conditions $g(n)\ge k_0$ and $n\ge n_0$ are equivalent. We fix a value $k_0\ge 3$ so that $1-\alpha(n)\ge\sqrt{3}/2$ whenever $g(n)\ge k_0$. Note that

$$p_n(E_k)\ge\frac{\sqrt{3}}{2}\ \text{ whenever }\ g(n)\ge k\ge k_0. \tag{18}$$

The result readily follows from the following fact.

Claim A. If $k_0\le k\le g(n)$, then for every first-order statement Φ with $D(\Phi)=k$ we have

$$|p_n(\Phi)-p(\Phi)|\le 2\,\alpha(n). \tag{19}$$

We will prove first a more modest bound.

Claim B. If $k_0\le k\le g(n)$, then for every first-order statement Φ with $D(\Phi)=k$ we have

$$|p_n(\Phi)-p(\Phi)|\le 1/2. \tag{20}$$

Proof of Claim B. Consider a pair of integers k and n such that $k_0 \le k \le g(n)$. Let Φ be a first-order statement with $D(\Phi) = k$. Without loss of generality, suppose that $p(\Phi) = 0$. If $p_N(\Phi) > 1/2$ for some N, let N denote the largest such number. For $M = N + 1$ we have $p_M(\Phi) \le 1/2$. Let G_N and G_M be independent random graphs with, respectively, N and M vertices. Note that

$$\mathbb{P}[\,D(G_N, G_M) \le k\,] \ge \mathbb{P}[\,G_N \models \Phi \;\&\; G_M \not\models \Phi\,] > \frac{1}{4}.$$

On the other hand, Lemma 6.6 implies that

$$\mathbb{P}[\,D(G_N, G_M) > k\,] \ge \mathbb{P}[\,G_N \models E_k \;\&\; G_M \models E_k\,] = p_N(E_k)p_M(E_k).$$

It follows that $p_N(E_k)p_M(E_k) < 3/4$ and, therefore, $p_N(E_k) < \sqrt{3}/2$ or $p_M(E_k) < \sqrt{3}/2$. Comparing this with (18), we conclude that $g(N) < k$ or $g(M) < k$, which implies that $n > N$. The desired bound (20) follows now by the definition of N. $\triangleleft$

Proof of Claim A. Consider a pair of integers k and n such that $k_0 \le k \le g(n)$. Let Φ be a first-order statement with $D(\Phi) = k$. Let G' and G'' be two independent copies of $G_{n,1/2}$. By Lemma 6.6,

$$\mathbb{P}[\,D(G', G'') > k\,] \ge \mathbb{P}[\,G' \models E_k \;\&\; G'' \models E_k\,] = p_n(E_k)^2.$$

On the other hand,

$$\begin{aligned}\mathbb{P}[\,D(G', G'') > k\,] &\le \mathbb{P}[\,G' \text{ and } G'' \text{ are not distinguished by } \Phi\,]\\ &= p_n(\Phi)^2 + (1 - p_n(\Phi))^2.\end{aligned}$$

Combining the two bounds, we obtain

$$2\,p_n(\Phi)(1 - p_n(\Phi)) \le 1 - p_n(E_k)^2.$$

Using Claim B, we immediately infer from here that

$$|p_n(\Phi) - p(\Phi)| \le 1 - p_n(E_k)^2 \le 2(1 - p_n(E_k)).$$

The desired bound (19) follows now from (17). $\triangleleft$ $\square$

6.3. The evolution of a random graph. We now take a dynamical view on a random graph $G_{n,p}$ by letting the edge probability p vary. With p varying from 0 to 1, $G_{n,p}$ evolves from empty to complete. We want to trace the changes of its logical complexity during the evolution. Since the definability parameters do not change when we pass to the complement of a graph, we can restrict ourselves to case $p \le 1/2$.

When p is a constant, one can estimate $D(G)$ within additive error $O(\log\log n)$.

THEOREM 6.12 (Kim et al. [**50**]). *If $0 < p \le 1/2$ is constant, then whp*

$$\log_{1/p} n - c_1 \ln\ln n - O(1) \le W(G_{n,p}) \le D(G_{n,p}) \le \log_{1/p} n + c_2 \ln\ln n,$$

where $c_1 = 2\ln^{-1}(1/p)$ and $c_2 = (2 + o(1))\,(-p\ln p - (1-p)\ln(1-p))^{-1} - c_1$.

SKETCH OF PROOF. Similarly to Theorem 6.4, the lower bound is based on the k-extension property. However, the proof of the upper bound is quite different. In particular, we have hardly any control on the alternation number in this result. The argument is rather complicated so we give only a brief sketch, concentrating more on its logical rather than probabilistic component.

Let $G = G_{n,p}$ be typical and $G' \not\cong G$ be arbitrary. Let $V = V(G)$ and $V' = V(G')$. For a sequence X of vertices, let $V_X = \{y \in V : \forall x \in X\; xy \in E(G)\}$ and $G_X = G[V_X]$. Let the analogous notation (with primes) apply to G'. If there is $x \in V$ such that for every $x' \in V'$ we have $G_x \not\cong G'_{x'}$, then Spoiler selects x. Whatever Duplicator's reply $x' \in V'$ is, Spoiler reduces the game to non-isomorphic graph G_x and $G'_{x'}$. We expect that $|V_x| = (p+o(1))n$ and G_x is also 'typical'. Thus Spoiler used one move to reduce the order of the random graph by a factor of p, which should lead to the upper bound $D(G) \le (1+o(1)) \log_{1/p} n$.

Suppose now that there are $x \in V$ and distinct $y', z' \in V'$ such that $G_x \cong G'_{y'} \cong G'_{z'}$. Spoiler selects $y' \in V'$. Assume that Duplicator replies with $y = x$, for otherwise $G_y \not\cong G'_{y'}$ and Spoiler proceeds as above. Now Spoiler selects z'; let $z \in V$ be the Duplicator's reply. We can assume that $G_{y,z} \cong G'_{y',z'}$, for otherwise Spoiler applies the inductive strategy to the $(G_{y,z}, G_{y',z'})$-game, where the order of the random graph is reduced by factor $(1+o(1))\, p^2$. Let $U = V_{y,z}$ and $U' = V'_{y',z'}$. A first moment calculation shows that there is vertex $v \in V_y \setminus U$ such that no vertex of $V_z \setminus U$ has the same neighborhood in U as v. Let Spoiler select v and let $v' \in V'_{y'} \setminus U'$ be the Duplicator's reply. Two copies $G'_{y'}$ and $G'_{z'}$ of a 'typical' graph G_x have a large vertex intersection. Another first moment calculation shows that whp there is only one way to achieve this, namely that the (unique) isomorphism $f : V'_{y'} \to V'_{z'}$ between $G'_{y'}$ and $G'_{z'}$ is in fact the identity on U'. But then $f(v')$ has the same adjacencies to U' as v'. Spoiler selects $f(v')$ and wins the game in at most one extra move.

Finally, up to a symmetry it remains to consider the case that there is a bijection $g : V \to V'$ such that for any $x \in V$ we have $G_x \cong G'_{g(x)}$.

As $G \not\cong G'$, there are $y, z \in V$ such that g does not preserve the adjacency between y and z. Spoiler selects y. We can assume that Duplicator replies with $y' = g(y)$ for otherwise Spoiler reduces the game to G_y. Now, Spoiler selects z to which Duplicator is forced to reply with $z' \neq g(z)$. Let $w = g^{-1}(z')$. Assume that $G_{y,z} \cong G_{y',z'}$ for otherwise Spoiler applies the inductive strategy to these graphs. But then $G_{y,z}$ is an induced subgraph of $G_w \cong G'_{z'}$, a property that we do not expect to see in a random graph.

In order to convert this rough idea into a rigorous proof one has to show that whp as long as the subgraphs $G_{x_1,x_2,\ldots}$ that can appear in the game are sufficiently large, they have all required properties. Also, one has to design Spoiler's strategy to deals small subgraphs of $G_{n,p}$ at the end of the game. All details can be found in [**50**, Section 3]. $\square$

It is interesting to investigate the behavior, e.g., of $D(G_{n,p})$ when $p = p(n)$ tends to zero. In particular, it is open whether, for every constant $\delta \in (0,1)$ and $n^{-\delta} \le p(n) \le 1/2$ we have whp $D(G_{n,p}) = O(\log n)$.

Some restriction on $p(n)$ from below is necessary here. Indeed, let G be an arbitrary non-empty graph (i.e., G has at least one edge) and let G' be obtained from G by adding one more isolated vertex. It is easy to see that $W(G, G') > d_0(G)$ and $D(G, G') > d_0(G) + 1$, where $d_0(G)$ denotes the number of isolated vertices of G. It follows that

$$W(G) \ge d_0(G) + 1 \quad\text{and}\quad D(G) \ge d_0(G) + 2. \tag{21}$$

It is well known (see, e.g., [**13**]) that

$$d_0(G_{n,p}) = (\mathrm{e}^{-pn} + o(1))\, n \tag{22}$$

whp as long as $p = O(n^{-1})$. In particular, we have $W(G_{n,p}) = (1-o(1))n$ whenever $p = o(n^{-1})$.

In some cases, the lower bounds (21) are sharp.

LEMMA 6.13. *Let $c_F(G)$ denote the number of connected components in a graph G isomorphic to a graph F. Suppose that a non-empty graph G satisfies*

$$c_F(G) + v(F) \leq d_0(G) + 1, \quad \text{for every component } F \text{ of } G. \tag{23}$$

Then $W(G) + 1 = D(G) = D_1(G) = d_0(G) + 2$.

PROOF. Let us show that $D_1(G, H) \leq d_0(G)+2$ for any $H \not\cong G$. Let F be such that $c_F(H) \neq c_F(G)$. For definiteness suppose that $c_F(H) > c_F(G)$. Spoiler marks $c_F(G) + 1$ components of H which are isomorphic to F by pebbling one vertex in each of them. Duplicator is forced either to mark one of the F-components of G twice (by pebbling two vertices, say, u and v in it) or to mark a component F' of G which is not isomorphic to F. In the former case Spoiler wins by pebbling a path from u to v. In the latter case Spoiler pebbles completely the F-component of H corresponding to F'. Duplicator is forced to pebble a connected part F'' of F'. If she has not lost yet, then $F'' \cong F$ and hence F'' is a proper subgraph of F'. Spoiler wins by pebbling another vertex in F' which is adjacent to a vertex in F''. Altogether at most $d_0(G) + 2$ moves are made.

It remains to prove the upper bound on the width. The last move may require using the $(d_0(G)+2)$-th pebble. However, for this purpose Spoiler can reuse a pebble placed earlier in a component different from F'. This trick is unavailable only if $c_F(G) = 1$ and $c_F(H) = 0$ or if $c_F(G) = 0$ and $c_F(H) = 1$. In both cases Spoiler can win in at most $v(F)+1$ rounds (and at most one alternation). Moreover, if this number is at least $d_0(G) + 2$, then $c_F(G) = 0$ and G has no component with $v(F)$ or more vertices by (23). In this case, Spoiler can win in at most $v(F)$ moves. $\square$

THEOREM 6.14 (Kim et al. [**50**], Bohman et al. [**12**]). *If $p = c/n$ with $c = c(n) \geq 0$ being an arbitrary bounded function of n, then $D(G_{n,p}) = (\mathrm{e}^{-c} + o(1))n$ whp.*

SKETCH OF PROOF. It is well known that, observing the evolution process in the scale $p = c/n$, at the point $c = 1$ we encounter the phase transition. If $c < 1-\epsilon$, whp all components of $G_{n,p}$ have $O(\log n)$ vertices each; if $c > 1 + \epsilon$, there appears a unique exception, the so-called *giant* component with a linear number of vertices.

One can check that, for any $c < \alpha - \epsilon$, Condition (23) holds whp (even for the giant component if it exists), where $\alpha = 1.1918...$ is a root of some explicit equation, see [**50**, Theorem 19]. Then, by Lemma 6.13 and Equality (22), $W(G_{n,p})$ and $D_1(G_{n,p})$ (and all parameters in between) are $(\mathrm{e}^{-c} + o(1))\, n$. When c is larger than $\alpha + \epsilon$, then whp the giant component of $G_{n,p}$ violates (23): Its order exceeds the number of isolated vertices. This case is handled in [**12**] as follows.

Denote the giant component of $G = G_{n,p}$ by M. Given $H \not\cong G$, we have to design a strategy allowing Spoiler to fast enough win the Ehrenfeucht game on G and H. The strategy in the proof of Lemma 6.13 does not work only if $c_M(H) = 0$ or $c_M(H) \geq 2$. We adapt it for these cases so that Spoiler, instead of selecting all vertices of M, plays an optimal strategy for M using at most $D(M) + \log n + 1$ moves (instead of $v(M) + 1$ moves as earlier).

First, we can assume that no component of H has diameter n or more. Otherwise Spoiler pebbles u and v at distance n in H. For Duplicator's responses u' and v' in G we have either $dist(u', v') < n$ or $dist(u', v') = \infty$. Hence Spoiler wins in less than $\log n + 1$ moves.

Second, we can assume that Duplicator always respects the connectivity relation (two vertices are in the relation if they are connectable by a path). Indeed, suppose that u and v belong to the same connected component F in one of the graphs while their counter-parts u' and v' are in different components of the other graph. Then Spoiler wins in less than $\log \mathit{diam}(F) + 1$ moves.

Under this assumption, Spoiler easily forces that, starting from the 2nd round, the play goes on components of G and H, of which exactly one is isomorphic to M. One of the main results of [**12**] states that whp

$$D(M) = O\left(\frac{\ln n}{\ln \ln n}\right), \tag{24}$$

which implies that Spoiler is able to win quickly and proves the theorem.

The upper bound (24) is obtained roughly as follows. By iteratively removing vertices of degree 1 from the giant component M, one obtains the *core* C of M (that is, C is a maximum subgraph with minimum degree at least 2). The *kernel* K of G is the *serial reduction* of C, that is, we iterate the following to obtain K: If there is a vertex x of degree 2, then we remove x but add edge $\{y, z\}$, where y and z are the two neighbors of x. The kernel may have loops and multiple edges and has to be modeled as a *colored graph*. The original graph G can be encoded by specifying its kernel K and the structure of rooted trees that correspond to each vertex or edge of K, the latter being viewed as a total coloring of K. It happens that whp every vertex x of K can be *identified* by a small-depth formula Φ_x with one free variable (that is $K, x \models \Phi_x$ while $K, y \not\models \Phi_x$ for every other vertex $y \in V(K)$) in the first-order language of colored graphs. Thus one can define K succinctly by stating that for every $x \in V(K)$ there is a unique vertex satisfying Φ_x, that every vertex satisfies Φ_x for some $x \in V(K)$, and by listing the adjacencies between vertices identified by Φ_x and Φ_y for every $x, y \in V(K)$. The core C can now be defined by specifying the length of the path corresponding to each edge of K, while the giant component M can be defined by specifying the random rooted trees hanging on the vertices of C using Theorem 6.10 (which relies in part on Theorem 5.4). □

The bound (24) is optimal up to a constant factor. This follows from the fact that whp the giant component M has a vertex v adjacent to at least $(1 - \epsilon) \log n / \log \log n$ leaves. (Indeed, consider the graph $M' \not\cong M$ that is obtained from M by attaching an extra leaf at v.) We believe that the lower bound is sharp, that is, whp $D(M) = (1 + o(1)) \log n / \log \log n$, but we were not able to settle this question.

Finally, we consider edge probabilities $p = n^{-\alpha}$ with rational $\alpha \in (0, 1)$. Such p occur as threshold functions for (non-)appearance of particular graphs as induced subgraphs in $G_{n,p}$. What is relevant to our subject is that such p show an irregular behavior of $G_{n,p}$ with respect to first-order properties.

Since the treatment of the general case of rational α would require a considerable amount of technical work, the paper [**50**] focuses on a sample value $\alpha = 1/4$, when $D(G_{n,p})$ falls down and becomes so small as it is essentially possible (cf. Section 7).

THEOREM 6.15 (Kim et al. [**50**]). *If $p = n^{-1/4}$, then whp*

$$\log^* n - \log^* \log^* n - 1 \le D(G_{n,p}) \le D_3(G_{n,p}) \le \log^* n + O(1).$$

SKETCH OF PROOF. The upper bound is based on the following ideas. Let the predicate $C(x_1, x_2, x_3, x_4)$ state that these 4 distinct vertices have no common neighbor. Its probability is $(1-p)^{n-4} = \mathrm{e}^{-1} + o(1)$ and its values over different 4-tuples are rather weakly correlated. Thus, if for a set A and a vertex $v \notin A$, we define $H_v(A)$ be the 3-uniform hypergraph on A with $x_1, x_2, x_3 \in A$ being a hyperedge if and only if $C(v, x_1, x_2, x_3)$ holds, then $H_v(A)$ behaves somewhat like a random hypergraph. As it is shown in [**50**, Lemma 21], one can find 4 vertices such that their common neighborhood A is relatively large (namely, $|A| = \lfloor \ln^{0.3} n \rfloor$) and yet there are vertices a, m such that hypergraphs $H_a(A)$ and $H_m(A)$ encode in some way the multiplication and addition tables for an initial interval of integers. Also, any integer can be succinctly defined in first-order logic with arithmetic operations. Roughly speaking, in order to define an integer j, one can write it in binary $j = b_k \dots b_1$ and specify for every $i \le k$ the i-th bit b_i; crucially, the same binary expansion trick can be used recursively to specify the index i, and so on. This allows us to identify vertices A with very small depth. Next, we consider the set B of vertices of G that have exactly 4 neighbors in A and are uniquely determined by this. Again, the vertices of B are easy to identify (just list the 4 neighbors in A). Finally, if A was chosen carefully, then each vertex w of G is uniquely identified by the hypergraph $H_w(B)$. (The reason that we need an intermediate set B is that the number of possible 3-uniform hypergraphs $H_w(A)$ is at most $2^{\binom{|A|}{3}} < n - |A|$, that is, too small.) Of course, many technical difficulties arise when one tries to realize this approach.

The lower bound in Theorem 6.15 is very general. We use only the simple fact that any particular unlabeled graph with m edges is the value of $G_{n,p}$, where $p = n^{-1/4}$, with probability at most

$$n! p^m (1-p)^{\binom{n}{2} - m} \le n!(1-p)^{\binom{n}{2}} \le \exp\left(-(1/2 - o(1)) n^{7/4}\right).$$

Let $F(k)$ be the number of non-isomorphic graphs definable with depth at most k. Then $\mathbb{P}[\, D(G) \le k \,] \le F(k) \exp\left\{-(1/2 - o(1)) n^{7/4}\right\}$. By Theorem 2.3, $F(k) \le \mathit{Tower}(k + 2 + \log^* k)$. If $k = \log^* n - \log^* \log^* n - 2$, we have $F(k) \le 2^n$ and hence $\mathbb{P}[\, D(G) \le k \,] = o(1)$. $\square$

The above idea (arithmetization of certain vertex sets in graphs) has been previously used by Spencer [**72**, Section 8] to obtain non-convergence and non-separability results on the example of $G_{n,p}$ with $p = n^{-1/3}$.

So far we have considered the evolution of the logical complexity of a random graph in the standard logic with no counting. We conclude this section with an extension of Theorem 6.1.

THEOREM 6.16 (Czajka and Pandurangan [**17**]). *Let $p(n)$ be any function of n such that $\frac{\omega(n) \log^4 n}{n \log\log n} \le p(n) \le 1/2$ where $\omega(n) \to \infty$ as $n \to \infty$. Then 2 color refinements split a random graph $G_{n,p}$ into color classes which are singletons with probability that is higher than $1 - n^{-c}$ for each constant $c > 0$ and all large enough n. Consequently, $D^2_{\#}(G_{n,p}) \le 4$ with this probability.*

Note that, in the case of $p = 1/2$, this result improves the probability bound in Part 1 of Theorem 6.1, while the probability bound in Part 2 is still better.

By elaborating on the argument of Lemma 6.2, we are able to supply Theorem 6.16 with the matching lower bound (that is, $D^2_{\#}(G_{n,p}) \ge 4$ whp) within the range $4\sqrt{\log n/n} \le p \le 1/2$.

7. Best-case bounds: Succinct definitions

As in the preceding sections, we consider the logical depth of graphs with a given number of vertices n. We know that the maximum value $D(G) = n + 1$ is attained by the complete and empty graphs (and only by them) and that the typical values lie around $\log n$ (see Theorem 6.4). Now we are going to look at the minimum. We already have a good starting point: By Theorem 6.15, there are graphs with

$$D_3(G) \le \log^* n + O(1).$$

In order to get such examples, we have to generate a random graph with the edge probability $n^{-1/4}$. In Section 7.1 we give three explicit constructions achieving the same bound. In Section 7.2 we introduce the *succinctness function* $q(n) = \min\{\, D(G) : v(G) = n\}$ and give an account of what is known about it. Section 7.3 is devoted to the question of how succinctly we can define graphs if we are not allowed to make quantifier alternations. In Section 7.4 the bounds on the succinctness function are applied to proving separations results for logical parameters of graphs, in particular, for $D(G)$ and $L(G)$.

7.1. Three constructions.

7.1.1. *First method: Padding.* We describe a "padding" operation that was invented by Joel Spencer (unpublished). It converts any graph G to an exponentially larger graph G^* with the logical depth larger just by 1. G^* includes G as an induced subgraph. In addition, for every subset X of $V = V(G)$, the graph G^* contains a vertex v_X. Denote the set of these vertices by V'. There is no edge inside V' but there are some edges between V and V'. Specifically, $v \in V$ is adjacent to v_X iff $v \in X$. In particular, $v_\emptyset$ is isolated and $N(v_V) = V$.

Vertex v_V will play a special role in our first-order definition of G^*. First of all, we will say that there is a vertex c (assuming $c = v_V$) whose neighborhood spans in G^* a subgraph isomorphic to G. This can be done by relativizing a formula Φ_G defining G to $N(c)$. That is, each universal quantification $\forall x(\Psi)$ in Φ_G has to be modified to

$$\forall_{x\in N(c)}(\Psi) \stackrel{\text{def}}{=} \forall x(x \sim c \to \Psi)$$

and each existential quantification to

$$\exists_{x\in N(c)}(\Psi) \stackrel{\text{def}}{=} \exists x(x \sim c \wedge \Psi).$$

Denote the relativized version of Φ_G by $\Phi_G|_{N(c)}$. Note that relativization does not change the quantifier depth. A sentence defining G^* can now look as follows:

$$\Phi_{G^*} \stackrel{\text{def}}{=} \exists c\Big(\Phi_G|_{N(c)} \wedge \forall_{x\notin N(c)}(N(x) \subset N(c)) \wedge \forall_{x_1\notin N(c)}\forall_{x_2\notin N(c)}(N(x_1) \ne N(x_2)) \\ \wedge \forall_{x_1\notin N(c)}\forall_{y\in N(x_1)}\exists_{x_2\notin N(c)}(N(x_2) = N(x_1) \setminus \{y\})\Big),$$

where we use harmless shorthands for simple first-order expressions.

It is easy to see that $D(\Phi_{G^*}) = \max\{D(\Phi_G), 4\} + 1$ and that, if Φ_G is a $\exists^*\forall^*\exists^*\forall^*$-*formula* (that is, every chain of nested quantifiers is a string of this form), then Φ_{G^*} stays in this class as well. Consider now a sequence of graphs G_k where

$G_1 = P_1$, the single-vertex graph, and $G_{k+1} = (G_k)^*$. Since $v(G^*) = v(G) + 2^{v(G)}$, we have $v(G_k) \ge Tower(k-1)$. It follows that $D_3(G_k) \le \log^* v(G_k) + 3$.

7.1.2. *Second method: Unite and conquer.* Suppose that we have a set C of n-vertex graphs, each of logical depth at most d. Our goal is to construct a much larger set C^* of graphs with a much larger number of vertices n^* and logical depth bounded by $d+3$. An additional technical condition is that all the graphs have diameter 2. We know from Theorem 6.4 that almost all graphs on n vertices have logical depth less than $\log n$ and it is well known that they have diameter 2. Choosing a sufficiently large n, we can start with C being the class of all such graphs. Since almost all graphs are asymmetric, we have $|C| = (1-o(1))2^{\binom{n}{2}}$. Just for the notational simplicity, we prefer that $|C|$ is even.

For each $S \subset C$ such that $|S| = |C|/2$, the set C^* contains graph $G_S = \overline{\bigsqcup_{G\in S} G}$, that is, we take the vertex disjoint union of all graphs in S and complement it. For convenience, we bound the logical depth of the complement $\overline{G_S} = \bigsqcup_{G\in C} G$ rather than that of G. Given an arbitrary $H \not\cong \overline{G_S}$, we analyze the Ehrenfeucht game on the two graphs.

If H has a connected component of diameter at least 3, Spoiler pebbles vertices u and v in H at the distance exactly 3 from one another. For Duplicator's responses u' and v' in $\overline{G_S}$, either $dist(u',v') \le 2$ or $dist(u',v') = \infty$. In any case, Spoiler wins within the next 2 moves. Suppose from now on that all components of H have diameter at most 2. This condition allows us to assume that Duplicator respects the connectivity relation for otherwise Spoiler wins with one extra move (which will be added to the total count of rounds).

If one of the graphs, $\overline{G_S}$ or H, has a connected component A non-isomorphic to any component of the other graph, Spoiler pebbles a vertex in A. Let B be the component of the other graph where Duplicator responds. Starting from the second round, Spoiler plays the Ehrenfeucht game on non-isomorphic graphs A and B and wins in at most d moves.

If such a component does not exist, $\overline{G_S}$ must have a component A with at least two isomorphic copies in H. Then in the first two rounds Spoiler pebbles vertices in these two. Duplicator is forced at least once to respond in a component B of $\overline{G_S}$ non-isomorphic to A, which is an already familiar configuration.

Thus, $D(G_S)$ can be at most 3 larger than the maximum logical depth of graphs in C. At the same time G_S has the much larger number of vertices, namely $n^* = n|C|/2$. It follows that $D(G) < \log\log v(G)$ for any G in C^*.

Note that any graph in C^* is the complement of a disconnected graph and hence has diameter 2. This allows us to iterate the construction. Say, for any $G \in (C^*)^*$ we get $D(G) < \log\log\log v(G)$ and so on. If we fix the initial class C, the iteration procedure gives us graphs with $D(G) < 3\log^* v(G) + O(1)$. This bound is worse than in the preceding section but the extra factor of 3 can be eliminated if Spoiler plays more smartly (see [**62**]).

7.1.3. *Third method: Asymmetric trees.* The two previous examples were artificially constructed with the aim to ensure low quantifier depth. Now we present a natural class of graphs admitting succinct definability.

The *radius* of a graph G is defined by $r(G) = \min_{v\in V(G)} e(v)$, where $e(v)$ denotes the eccentricity of a vertex v. A vertex v is *central* if $e(v) = r(G)$. Any tree has either one or two central vertices (see, e.g., [**59**, Chapter 4.2]).

LEMMA 7.1. *Let T be an asymmetric tree with $r(T) \ge 6$. Then $D(T) \le r(T)+2$.*

PROOF. We will design a strategy for Spoiler in the Ehrenfeucht game on T and a non-isomorphic graph T'. The reader that took the effort to reconstruct the proof of Theorem 5.3 will now definitely benefit.

We can assume that T and T' have equal diameters (in particular, T' is connected) for else Spoiler wins in less than $\log r(T) + 4$ moves by Lemma 3.2. If T' is a non-tree, let Spoiler pebble a vertex v' on a cycle in T'. By this move Spoiler forces the game on $T \setminus v$ and $T' \setminus v'$, where v is Duplicator's response in T. If v is a leaf, Spoiler wins in two moves. Otherwise $T \setminus v$ is disconnected, while $diam(T' \setminus v') \le 3\, diam(T') \le 6\, r(T)$. Lemma 3.2 applies again and Spoiler wins in less than $\log r(T) + 6$ moves. Assume, therefore, that T' is a tree too.

Call a tree *diverging* if every vertex w splits it into pairwise non-isomorphic branches, where each branch is considered rooted at the respective neighbor of w (an isomorphism of rooted trees has to match their roots). Any asymmetric tree is obviously diverging. On the other hand, if a tree is diverging, it is either asymmetric or has a single nontrivial automorphism and the latter transposes two central vertices.

Suppose that T' is diverging. In the first round Spoiler pebbles a central vertex v of T and Duplicator responds with a vertex v' in T'. As it is easily seen, at least one of $T \setminus v$ and $T' \setminus v'$ has a branch B non-isomorphic to any branch in the other tree. Spoiler restricts further play to B by pebbling its root. Continuing in this fashion, that is, each time finding a matchless subbranch, Spoiler forces pebbling two paths in T and T' emanating from v and v' respectively. Spoiler wins at latest when the path in T reaches a leaf.

So suppose that T' is not diverging. Let v' be a central vertex of T' and u' be a vertex at the maximum possible distance from v' with the property that $T' \setminus u'$ has two isomorphic branches B' and B''. Spoiler pebbles the path from v' to u' and the two neighbors of u' in B' and B''. From this point Spoiler can play as before because B' and B'' are diverging and only one of them can be isomorphic to the corresponding branch pebbled by Duplicator in T. $\square$

Lemma 7.1 shows that asymmetric trees are definable with quantifier depth not much larger than their radius. On the other hand, asymmetric trees can grow in breadth, having a huge number of vertices. More precisely, there are asymmetric trees with $v(T) \ge Tower(r(T)-1)$. Indeed, let r_k denote the number of asymmetric rooted trees of height at most k. A simple recurrence $r_k = 2^{r_{k-1}}$, where $r_0 = 1$, shows that $r_k = Tower(k)$. Let $k \ge 3$ and T_k be the (unrooted) tree of radius k with a single central vertex c such that the set of branches growing from c consists of all r_{k-1} pairwise non-isomorphic asymmetric rooted trees of height less than k. (The reader will now surely recognize another instance of the unite-and-conquer method!) Since T_k has even diameter, the central vertex c is fixed under all automorphisms. It easily follows that T_k is asymmetric. This graph will be referred to as the *universal asymmetric tree of radius* k. Note that $v(T_k) \ge r_{k-1}+1 > Tower(k-1)$. Combining it with Lemma 7.1, we obtain $D(T_k) \le k + 2 \le \log^* v(T_k) + 2$.

With a little extra work, trees with low logical depth can be constructed on any given number of vertices. It turns out that the log-star bound is essentially the best what can be achieved for trees.

THEOREM 7.2 (Pikhurko, Spencer, and Verbitsky [**61**]). *For every n there is a tree T on n vertices with $D(T) \le \log^* n + 4$. On the other hand, for all trees T on n vertices we have $D(T) \ge \log^* n - \log^* \log^* n - 4$.*

We will see in the next section that the lower bound of Theorem 7.2 cannot be extended to the class of all graphs.

Universal asymmetric trees have been proved to be a useful technical tool in complexity theory and finite model theory since a long time, see the references in Dawar et al. [**21**]. Lemma 3.4(e) in the latter paper readily implies a succinctness result for the logical *length*.

THEOREM 7.3 (Dawar et al. [**21**]). *For the universal asymmetric tree of radius k we have $L(T_k) = O((\log^* v(T_k))^4)$.*

The theorem shows that, for infinitely many n, there is a tree T on n vertices with $L(T) = O((\log^* n)^4)$. Unlike Theorem 7.2, this result cannot be extended to all n because there are infinitely many n such that all graphs on n vertices have logical length $\Omega\left(\frac{\log n}{\log\log n}\right)$ (see (31) in the proof of Theorem 7.13).

7.2. The succinctness function. Define the *succinctness function* by

$$q(n) = \min\{\, D(G) : v(G) = n\} .$$

Since only finitely many graphs are definable with a fixed quantifier depth (see Theorem 2.3), we have $q(n) \to \infty$ as $n \to \infty$. The examples collected in Section 7.1 show that $q(n)$ increases rather slowly. Let $q_a(n)$ denote the version of $q(n)$ for definitions with at most a quantifier alternations. The padding construction from Section 7.1.1 gives us

$$q_3(n) \le \log^* n + 3 \tag{25}$$

for infinitely many n and, by Theorem 6.15, this bound holds actually for all n, perhaps with a worst additive constant.

Is the log-star bound best possible? The answer is surprising enough: in some strong sense it is but, at the same time, it is very far from being tight. First, let us elaborate on the latter claim.

A *prenex formula* is a formula with all its quantifiers being in front. In this case there is a single sequence of nested quantifiers and the quantifier rank is just the number of quantifiers occurring in a formula. The superscript *prenex* will mean that we allow defining sentences only in prenex form. Thus, $q_a^{prenex}(n)$ is equal to the minimum quantifier depth of a prenex formula with at most a quantifier alternations that defines a graph on n vertices. We obviously have $D(G) \le D_a(G) \le D_a^{prenex}(G)$. Recall that $L_a(G)$ denotes the minimum length of a sentence defining G with at most a quantifier alternations. Since a quantifier-free formulas with k variables is equivalent to a disjunctive normal form over $2\binom{k}{2}$ relations between the variables, we obtain also relation

$$L_a(G) = O(h(D_a^{prenex}(G))) \quad \text{where} \quad h(k) = k^2 2^{k^2}. \tag{26}$$

Recall that a *total recursive function* is an everywhere defined recursive function.

THEOREM 7.4 (Pikhurko, Spencer, and Verbitsky [**61**]). *There is no total recursive function f such that $f(q_3^{prenex}(n)) \ge n$ for all n.*

The theorem implies a superrecursive gap between $v(G)$ and $D_3(G)$ or even $L_3(G)$. In particular, the values of $q_3(n)$ are infinitely often inconceivably smaller even than the values of $\log^* n$. More generally, if a total recursive function $l(n)$ is monotone nondecreasing and tends to infinity, then

$$q(n) < l(n) \quad \text{for infinitely many} \quad n, \tag{27}$$

which actually means that the succinctness function admits no reasonable lower bound.

The proof of Theorem 7.4 is based on simulation of a Turing machine M by a prenex formula Φ_M in which a computation of M determines a graph satisfying Φ_M and vice versa. Such techniques were developed in the classical research on Hilbert's *Entscheidungsproblem* by Turing, Trakhtenbrot, Büchi and other researchers (see [**14**] for survey and references). An important feature of our simulation is that it works if we restrict the class of structures to graphs. As a by-product, we obtain another proof of Lavrov's version of the Trakhtenbrot theorem [**52**] (see also [**26**, Theorem 3.3.3]) saying that the first-order theory of finite graphs is undecidable. The proof actually shows the undecidability of the $\forall^*\exists^p\forall^s\exists^t$-fragment of this theory for some p, s, and t.

We now have to explain why bound (25), though not sharp, is best possible in some sense. Let us define the *smoothed succinctness function* $q^*(n)$ to be the least monotone nondecreasing integer function bounding $q(n)$ from above, that is,

$$q^*(n) = \max_{m \le n} q(m). \tag{28}$$

The following theorem shows that $q^*(n) = (1 + o(1)) \log^* n$ and, therefore, the log-star function is a nearly optimal *monotone* upper bound for the succinctness function $q(n)$.

THEOREM 7.5 (Pikhurko, Spencer, and Verbitsky [**61**]).

$$\log^* n - \log^* \log^* n - 2 \le q^*(n) \le \log^* n + 4.$$

Though the lower bound contains a nonconstant lower order term, it can hardly be distinguished from a constant: for example, $\log^* \log^* n = 3$ for $n = 10^{80}$, which is a rough estimate of the number of elementary particles in the observable universe.

PROOF. Theorem 7.2 implies that $q(n) \le \log^* n + 4$ for all n. Since this bound is monotone, it is a bound on $q^*(n)$ as well. The lower bound for $q^*(n)$ can be derived from Theorem 2.3. According to it, at most $\mathit{Tower}(k + \log^* k + 2)$ graphs are definable with quantifier depth k. Given $n > \mathit{Tower}(3)$, let k be such that $\mathit{Tower}(k + 2 + \log^* k) < n \le \mathit{Tower}(k + 3 + \log^*(k + 1))$. It follows that $k > \log^* n - \log^* \log^* n - 4$. By the Pigeonhole Principle, there will be some $m \le n$ for which no graph of order precisely m is defined with quantifier depth at most k. We conclude that $q^*(n) \ge q(m) > k$ and hence $q^*(n) \ge \log^* n - \log^* \log^* n - 2$. □

We defined $q^*(n)$ to be the "closest" to $q(n)$ monotone function. Notice that $q(n)$ itself lacks the monotonicity, deviating from $q^*(n)$ infinitely often (set $l(n)$ to be the lower bound in Theorem 7.5 and apply (27)).

7.3. Definitions with no quantifier alternation. It is interesting to observe how the succinctness function changes when we put restrictions on the logic. Note that all what we have stated about the succinctness function for first-order logic actually holds true for its fragment with 3 quantifier alternations. Now we consider the first-order logic with no quantifier alternation, consisting of purely existential and purely universal formulas and their monotone Boolean combinations (of course, all negations are supposed to stay in front of relation symbols). It is easy to see that any sentence with no quantifier alternation is equivalent to a sentence in the *Bernays-Schönfinkel class*. The latter consists of prenex formulas in which the existential quantifiers all precede the universal quantifiers, as in

$$\Phi \stackrel{\text{def}}{=} \exists x_1 \ldots \exists x_k \forall y_1 \ldots \forall y_l \Psi(\bar{x}, \bar{y}), \tag{29}$$

where Ψ is quantifier-free. This fragment of first-order logic is provably weak.

To substantiate this claim, consider the *finite satisfiability problem*: Given a first-order sentence Φ about graphs, one has to decide whether or not there is a finite graph satisfying Φ. More generally, let $\mathrm{Spectrum}(\Phi)$ consist of all those n such that there is a graph on n vertices satisfying Φ. Thus, the problem is to decide whether $\mathrm{Spectrum}(\Phi)$ is nonempty.

Lavrov [**52**] proved that this problem is unsolvable even for sentences without equality (for directed graphs this is a classical result on Hilbert's *Entscheidungsproblem*, known as the Trakhtenbrot-Vaught theorem, see [**14**]). However, if we consider only sentences in the Bernays-Schönfinkel class, the finite satisfiability problem becomes decidable. This directly follows from the following simple observation showing that a nonempty spectrum always contains a certain small number.

LEMMA 7.6. *Suppose that a first-order sentence Φ is of the form (29). If Φ is satisfiable, then* $\mathrm{Spectrum}(\Phi)$ *contains k or a smaller number.*

PROOF. Assume that Φ is true on a graph G with more than k vertices and let $U \subset V(G)$ be the set of vertices $x_1, \ldots, x_k$ whose existence is claimed by Φ. Note that the induced subgraph $G[U]$ satisfies Φ as well. $\square$

The solvability of the finite satisfiability problem for the Bernays-Schönfinkel class was observed by Ramsey in [**69**]. Ramsey showed that the spectrum of a Bernays-Schönfinkel formula can be completely determined. This follows from the following result where his famous combinatorial theorem appeared as a technical tool. Recall that a set is *cofinite* if it has finite complement.

THEOREM 7.7 (Ramsey [**69**]). *Any sentence about graphs Φ in the Bernays-Schönfinkel class has either finite or cofinite spectrum. More specifically, if Φ is of the form (29), then either* $\mathrm{Spectrum}(\Phi)$ *contains no number equal to or greater than $2^k 4^l$ or it contains all numbers starting from $k + l$.*

PROOF. Assume that Φ is true on a graph G with at least $2^k 4^l$ vertices and let $U \subset V(G)$ consist of vertices $x_1, \ldots, x_k$ whose existence is claimed by Φ. Recall that *Ramsey number $R(l)$* is equal to the minimum R such that every graph with R or more vertices contains a homogeneous set of l vertices. As it is well known, $R(l) < 4^l$. By the Pigeonhole Principle, $V(G) \setminus U$ contains a subset W of $R(l)$ vertices with the same neighborhood within U. Let X be a homogeneous set of l vertices in $G[W]$. Note that $G[U \cup X]$ satisfies Φ and that X is a set of l twins in this graph. Cloning the twins, we can obtain a graph that satisfies Φ and has any number of vertices larger than $k + l$. $\square$

After this small historical excursion, let us turn back to the definability with no quantifier alternation. First of all, note that even without quantifier alternation all graphs remain definable (see (4)) and, hence, the parameter $D_0(G)$ is well defined.

THEOREM 7.8 (Pikhurko, Spencer, and Verbitsky [**61**]). *$D_0(G)$ is a computable parameter of a graph.*

PROOF. Given $m \geq 0$, one can algorithmically construct a finite set U_m consisting of 0-alternating sentences of quantifier depth m so that every 0-alternating sentence of quantifier depth m has an equivalent in U_m. To decide if $D_0(G) \leq m$, for each sentence $\Upsilon \in U_m$ satisfied by G we have to check if Υ can be satisfied by another graph G'. We first reduce Υ to an equivalent statement Ψ in the Bernays-Schönfinkel class. Suppose that Ψ has k existential quantifiers. It suffices to test all $G' \not\cong G$ with at most $k+1$ vertices. Indeed, if Φ is true on a graph with more than $k+1$ vertices then, by the argument used to prove Lemma 7.6, Φ is as well true on its induced subgraphs with $k+1$ and k vertices (one of which is not isomorphic to G). □

We cannot prove anything similar for $D(G)$ or even $D_1(G)$. The proof of Theorem 7.8 is essentially based on the decidability of whether or not a 0-alternating sentence is defining for some graph. However, in general this problem is undecidable (see [**61**]).

For the logic with no quantifier alternation, the succinctness function has much more regular behavior.

THEOREM 7.9 (Spencer, Pikhurko, and Verbitsky [**62**]).

$$\log^* n - \log^* \log^* n - 2 \leq q_0(n) \leq \log^* n + 22.$$

The lower bound has to be contrasted to Theorem 7.4. It gives us a kind of a quantitative confirmation of the fact that the 0-alternation fragment of first-order logic is strictly less powerful. The upper bound improves upon the alternation number in (25) attaining the optimum. The proof of this bound is based on the unite-and-conquer construction in Section 7.1.2, where more subtle analysis is needed in order to achieve the zero alternation number. All the details can be found in [**62**].

PROOF OF THEOREM 7.9 (LOWER BOUND). Given n, denote $k = q_0(n)$ and fix a graph G on n vertices such that $D_0(G) = k$. The same relation between $L_a(G)$ and $D_a(G)$ as in Theorem 2.2 is proved in [**62**]. By this result, G is definable by a 0-alternating sentence Υ of length less than $\mathit{Tower}(k+\log^* k+2)$. Convert Υ to an equivalent sentence Φ in the Bernays-Schönfinkel class and note that $D(\Phi) \leq L(\Upsilon)$. By Lemma 7.6, Φ must be true on some graph with at most $D(\Phi)$ vertices. Since Φ is true only on G, we have

$$n \leq D(\Phi) \leq L(\Upsilon) < \mathit{Tower}(k + \log^* k + 2).$$

This implies that

$$\log^* n \leq k + \log^* k + 2. \tag{30}$$

Suppose on the contrary to our claim that $k \leq \log^* n - \log^* \log^* n - 3$. Then $\log^* k \leq \log^* \log^* n$ and (30) implies that

$$\log^* n \leq (\log^* n - \log^* \log^* n - 3) + \log^* \log^* n + 2,$$

which is a contradiction, proving the claimed bound. □

Using the lower bound of Theorem 7.9 and the absence of any recursive linkage between $q_3(n)$ and n, we are able to show a superrecursive gap between two parameters in the logical depth hierarchy

$$D(G) \leq D_3(G) \leq D_2(G) \leq D_1(G) \leq D_0(G).$$

THEOREM 7.10 (Pikhurko, Spencer, and Verbitsky [**61**]). *There is no total recursive function f such that $D_0(G) \le f(D_3(G))$ for all graphs G.*

PROOF. Assume that such an f exists. Let G_n be a graph for which $D_3(G_n) = q_3(n)$. Then

$$f(q_3(n)) = f(D_3(G_n)) \ge D_0(G_n) \ge q_0(n) \ge \log^* n - \log^* \log^* n - 2.$$

This implies that $\mathit{Tower}(2f(q_3(n))) \ge n$, contradictory to Theorem 7.4. $\square$

We have seen weighty evidences that the 0-alternating sentences are strictly less expressive than the sentences of the same quantifier depth with quantifier alternations. It is quite surprising that, nevertheless, sometimes we can prove for $D_0(G)$ upper bounds which are just a little worse than the best known bounds for $D(G)$. The following results should be compared with Theorems 5.4, 5.8, and 6.4.

THEOREM 7.11.

1. (Bohman et al. [**12**]) *Let $D_0(n,d)$ denote the maximum of $D_0(T)$ over all trees with n vertices and maximum degree at most $d = d(n)$. If both d and $\log n/\log d$ tend to infinity, then $D_0(n,d) \le (1+o(1))\frac{d \log n}{\log d}$.*
2. (Pikhurko, Veith, and Verbitsky [**63**]) $D_0(G,H) \le \frac{n+5}{2}$ *for all non-isomorphic graphs G and H with the same number of vertices n.*
3. (Kim et al. [**50**]) $D_0(G_{n,1/2}) \le (2+o(1))\log n$ *with high probability.*

We conclude this subsection with a demonstration of somewhat surprising strength of the Bernays-Schönfinkel class. We say that a sentence Φ *identifies* a graph G if it distinguishes G from any non-isomorphic graph of the same order. Let $BS(G)$ denote the minimum quantifier depth of Φ in the Bernays-Schönfinkel class identifying G. We already discussed the identification problem in Section 5.2.1. Note, however, a striking difference. While in Section 5.2.1 we could make the conjunction of all sentences Φ_H distinguishing G from another graph H of the same order, now we have to distinguish G from all such H by a single prenex sentence!

THEOREM 7.12 (Pikhurko and Verbitsky [**65**]).

1. *For any graph G of order n, we have $BS(G) \le \frac{3}{4}n + \frac{3}{2}$.*
2. *With high probability we have $BS(G_{n,1/2}) \le (2+o(1))\log n$. Moreover, the latter bound holds true even if the number of universal quantifiers in an identifying formula is restricted to 2.*

7.4. Applications: Inevitability of the tower function. Succinctly definable graphs can be used to show that the tower function is sometimes unavoidable in relations between logical parameters of graphs. We first observe that the relationship between the logical depth and the logical length in Theorem 2.2 is "nearly" tight.

THEOREM 7.13 (Pikhurko, Spencer, and Verbitsky [**61**]).[7] *There are infinitely many pairwise non-isomorphic graphs G with $L(G) \ge \mathit{Tower}(D(G) - 7)$.*

PROOF. The proof is given by a simple counting argument. A first-order sentence Φ defining a graph G determines a natural binary encoding of G (up to isomorphism) of length $O(L(\Phi)\log L(\Phi))$. It follows that at most $m = 2^{O(k \log k)}$

[7]In [**61**] we stated a better bound $L(G) \ge \mathit{Tower}(D(G) - 6) - O(1)$, which was proved for the variant of $L(G)$ where variable x_i contributes $\log i$, rather than just 1, to the formula length.

graphs can have logical length less than k. By the Pigeonhole Principle, there is $n \le m+1$ such that $L(G) \ge k$ for all G on n vertices. For all these graphs we have

$$L(G) = \Omega\left(\frac{\log n}{\log\log n}\right), \tag{31}$$

which exceeds $\log\log n$ if k is chosen sufficiently large. By Theorem 7.2, there is a graph G_n on n vertices with

$$D(G_n) < \log^* n + 5. \tag{32}$$

Combining (32) and (31), we obtain the desired separation of $L(G_n)$ from $D(G_n)$. Increasing the parameter k, we can have infinitely many such examples. □

One of the consequences of Theorem 7.4 is that prenex formulas are sometimes unexpectedly efficient in defining a graph. We are now able to show that, nevertheless, they generally cannot be competitive against defining formulas with no restriction on structure. More specifically, we have simple relations

$$D(G) \le D^{prenex}(G) < L(G) \le L^{prenex}(G). \tag{33}$$

Combining the second inequality with Theorem 2.2, we obtain

$$D^{prenex}(G) < \mathit{Tower}(D(G) + \log^* D(G) + 2)$$

and we can now see that this relationship between $D^{prenex}(G)$ and $D(G)$ is not so far from being optimal.

Corollary 7.14. *There are infinitely many pairwise non-isomorphic graphs G with $D^{prenex}(G) \ge \mathit{Tower}(D(G) - 8)$.*

The proof of Theorem 7.13 gives us actually a better bound, though somewhat cumbersome, namely $L(G) \ge T/(c\log T)$ with $T = \mathit{Tower}(D(G) - 6)$ and c a constant. Corollary 7.14 follows from here simply by noticing that parameters $D^{prenex}(G)$ and $L(G)$ are exponentially close. The latter fact follows from (33) and a version of (26), namely

$$L(G) = O(h(D^{prenex}(G))) \quad \text{where} \quad h(x) = x^2 2^{x^2}.$$

In conclusion we note that the tower function is essential also in the upper bound for the number of graphs definable with quantifier depth k given by Theorem 2.3.

Corollary 7.15. *There are at least $(1-o(1))\,\mathit{Tower}(k-2)$ first-order sentences of quantifier depth k defining pairwise non-isomorphic graphs and, hence, being pairwise inequivalent.*

Proof. In Section 7.1.3 we noticed that there are exactly $r_h = \mathit{Tower}(h)$ asymmetric rooted trees of height at most h. Basically this follows from the fact that such a tree is completely characterized by the set of its branches from the root, each being an asymmetric rooted tree of height at most $h-1$ (the root is not a part of any branch). Thus, $r_h - r_{h-1}$ asymmetric rooted trees have height exactly h. Note that $(r_{h-1} - r_{h-2})r_{h-1}$ of them have exactly one branch of height $h-1$. Therefore, there are at least $r_h - r_{h-1} - (r_{h-1} - r_{h-2})r_{h-1} = (1 - o(1))\,\mathit{Tower}(h)$ asymmetric rooted trees whose underlying trees (with roots dismissed) have diameter $2h$ and, hence, are asymmetric too. By Lemma 7.1, each of these trees is definable with quantifier depth $h + 2$. □

A lower bound of $Tower(k-2)$ for the number of pairwise inequivalent sentences of quantifier depth k is shown by Spencer [**72**, Theorem 2.2.2].

8. Open problems

Many questions remain open, some of which are included in the main text of the survey alongside the known related results. For reader's convenience we collect a few open problems here that we consider most interesting.

Tomasz Łuczak (Conference on Random Structures and Algorithms, Poznań, 2003) asked if $D(G)$, or $W(G)$, is a computable function of the input graph G.

While the factor of 1/2 in Theorem 5.8 is best possible, we do not know if it can be improved for logic with counting. Surprisingly, we could not resolve even the following question. Is there $\epsilon > 0$ such that for every graph G of sufficiently large order n we have $W_{\#}(G) \le (\frac{1}{2} - \epsilon)n$?

Recall that no sublinear bound is generally possible here because Cai, Fürer, and Immerman [**15**] constructed graphs with linear width in the counting logic; see Theorem 5.7. Automorphisms of these graphs play an essential role in establishing this lower bound. It would be very interesting to estimate $W_{\#}(G)$ from above for asymmetric G. Again, we have only the bound $D_{\#}(G) \le (n+3)/2$ as a straightforward corollary of Theorem 5.8, where no restriction on the automorphism group is supposed.

Another research direction, with applications to the graph isomorphism problem, is identification of natural classes of graphs with $W_{\#}(G)$ bounded by a constant; see Sections 5.1.4 and 5.1.5. Such a bound is known for interval graphs [**28, 51**], and it is interesting if it can be extended to the class of circular-arc graphs. The approach suggested in [**51**] is based on the fact that any maximal clique in an interval graph is definable as the common neighborhood of some two vertices. This prevents any straightforward extension to the class of circular-arc graphs, where the number of maximal cliques can be exponential. The (un)boundedness of $W_{\#}(G)$ is an interesting open question also for disk graphs, yet another extension of the class of interval graphs (Martin Grohe, 2010).

A result of Dawar, Lindell, and Weinstein [**20**] (see also Theorem 4.7) implies an upper bound for $D_{\#}(G)$ in terms of $W_{\#}(G)$ and the order n of G, where $W_{\#}(G)$ disappointedly occurs at the exponent. Can this bound be improved? At the moment we cannot even exclude that $D_{\#}(G) = O(W_{\#}(G) \log n)$. If the latter bound was true for $D^k_{\#}(G)$ with $k = O(W_{\#}(G))$, this would have important consequences for isomorphism testing by Theorem 4.5.

Where do we need the power of counting quantifiers? To keep far away from the trivial example of a complete or empty graph, suppose that a graph G is asymmetric. Is it true or not that $W(G) = O(W_{\#}(G) \log n)$? A random graph shows that this bound would be best possible.

We are still far from having a complete evolutionary picture of the logical complexity for a random graph. Let $\delta \in (0, 1)$ be fixed and p be an arbitrary function of n with $n^{-\delta} \le p \le \frac{1}{2}$. Is it true that whp $D(G_{n,p}) = O(\log n)$?

The local behavior of the succinctness function $q(n)$, that was defined in Section 7.2, is unclear. While it is trivial that $q(n+1) \le q(n) + 1$, we do not know, for example, if $q(n+1) \ge q(n) - C$ for some constant C and all n.

In accordance with our notation system, let $q^k(n)$ denote the succinctness function for the k-variable logic. By slightly modifying the proof of Lemma 7.1, one can show that $q^3(n) \le (1+o(1)) \log^* n$ for all n. Since the satisfiability problem for the 3-variable logic is undecidable (see, e.g., [**34**]), it is not excluded that an analog of Theorem 7.4 can be established for $q^3(n)$.

Given a fixed k, how far apart from one another can the values of $D(G)$ and $D^k(G) < \infty$ be?

Theorem 7.10 says that there is no recursive link between $D_3(G)$ and $D_0(G)$. Can one show a superrecursive gap between $D_a(G)$ and $D_b(G)$ for some $b > a > 0$ or, at least, between $D(G)$ and $D_1(G)$?

Though the case of trees was thoroughly investigated throughout the survey, this class of graphs deserves further attention. One may expect that many logical questions for trees are easier. Note in this respect that the first-order theory of finite trees is decidable due to Rabin [**67**]. Nevertheless, we do not know, for example, whether or not the logical depth $D(T)$ of a tree T is a computable parameter (while it is not hard to show that the logical width $W(T)$ is computable in logarithmic space).

Disappointingly, we were able to collect only a few results on the logical length for this survey. ¿From the fact that there are $2^{(1/2+o(1))\, n^2}$ non-isomorphic graphs of order n, it is easy to derive that whp $L(G_{n,1/2}) = \Omega\left(\frac{n^2}{\log n}\right)$. The obvious general upper bound is $O(n^2)$. This leaves open the question what the logical length of a typical graph is. Also, it would be very interesting to find explicit examples of graphs with large $L(G)$. Pseudo-random graphs can be natural candidates. For example, it is well known (Blass, Exoo, and Harary [**10**]) that Paley graphs share the first-order properties of a truly random graph.

Furthermore, we can define the succinctness function with respect to the logical length by $s(n) = \min\{\, L(G) : v(G) = n\}$. Let $s_a(n)$ be the version of $s(n)$ for the a-alternation logic. From Theorem 7.4 and the relation (26), it follows that $s(n)$, and even $s_3(n)$, can be incomprehensibly smaller than n: for any total recursive function f we must have $f(s_3(n)) < n$ infinitely often. On the other hand, the estimate (31) in the proof of Theorem 7.13 implies that $s(n) = \Omega\left(\frac{\log n}{\log\log n}\right)$ for infinitely many n. Moreover, the same argument shows that $s^*(n) = \Omega\left(\frac{\log n}{\log\log n}\right)$ for *all* n, where $s^*(n)$ denotes the smoothed version of $s(n)$ similarly to (28). How tight is the bound of $\Omega\left(\frac{\log n}{\log\log n}\right)$ in these statements? Another interesting problem is the behavior of the function $s_0(n)$ (recall that for $q_0(n)$ we know the exact asymptotics owing to Theorem 7.9). Note in conclusion that techniques for estimating the length of a first-order formula are worked out, e.g., by Adler and Immerman [**1**], Dawar et al. [**21**], Grohe and Schweikardt [**41**].

Acknowledgment. We are grateful to Joel Spencer for the fruitful collaboration on the subject of this survey and for allowing us to use his unpublished ideas in the proof of Lemma 6.8 and in Section 7.1.1. We also thank Martin Grohe for his detailed comments, in particular, for bringing up a succinctness result of Dawar et al. [**21**] (Theorem 7.3 here) to our attention.

References

[1] M. Adler, N. Immerman. An $n!$ lower bound on formula size. *ACM Transactions on Computational Logic* 4:296–314 (2003).
[2] M. Aigner, G. Ziegler. Proofs from THE BOOK. Springer, 4th ed. (2010).

[3] N. Alon, P. Seymour, R. Thomas. A separator theorem for nonplanar graphs. *J. Am. Math. Soc.* 3:801–808 (1990).
[4] N. Alon, J. Spencer. *The probabilistic method.* Wiley, 3rd ed. (2008).
[5] L. Babai. Automorphism groups, isomorphism, reconstruction. Chapter 27 of the *Handbook of Combinatorics,* pages 1447–1540. Elsevier Publ. (1995).
[6] L. Babai. On the complexity of canonical labeling of strongly regular graphs. *SIAM J. Comput.* 9:212–216 (1980).
[7] L. Babai, P. Erdős, S. M. Selkow. Random graph isomorphism. *SIAM J. Comput.* 9:628–635 (1980).
[8] L. Babai, L. Kučera. Canonical labeling of graphs in linear average time. In: *Proc. of the 20th IEEE Symp. Found. Computer Sci.* 39–46 (1979).
[9] L. Babai and E. M. Luks. Canonical labeling of graphs. In: *Proc. of the 15th ACM Symp. on Theory of Computing* 171–183 (1983).
[10] A. Blass, G. Exoo, F. Harary. Paley graphs satisfy all first-order adjacency axioms. *J. Graph Theory* 5:435–439 (1981).
[11] H. L. Bodlaender. Polynomial algorithms for Graph Isomorphism and Chromatic Index on partial k-trees. *J. Algorithms* 11:631–643 (1990).
[12] T. Bohman, A. Frieze, T. Luczak, O. Pikhurko, C. Smyth, J. Spencer, O. Verbitsky. first-order definability of trees and sparse random graphs. *Combinatorics, Probability and Computing* 16:375-400 (2007).
[13] B. Bollobás. *Random graphs.* Cambridge Univ. Press, 2nd ed. (2001).
[14] E. Börger, E. Grädel, Y. Gurevich. *The classical decision problem.* Springer (1997).
[15] J.-Y. Cai, M. Fürer, N. Immerman. An optimal lower bound on the number of variables for graph identification. *Combinatorica* 12:389–410 (1992).
[16] C. J. Colbourn, K. S. Booth. Linear time automorphism algorithms for trees, interval graphs, and planar graphs. *SIAM J. Comp.* 10:203–225 (1981).
[17] T. Czajka, G. Pandurangan. Improved random graph isomorphism. *J. Discr. Algorithms* 6:85–92 (2008).
[18] B. Das, J. Torán, F. Wagner. Restricted space algorithms for isomorphism on bounded treewidth graphs. *Proc. of the 27th Symp. on Theoretical Aspects of Computer Science*, the Leibniz International Proceedings in Informatics series, 227–238 (2010).
[19] S. Datta, N. Limaye, P. Nimbhorkar, T. Thierauf, F. Wagner. Planar graph isomorphism is in Log-Space. In: *Proc. of the 24th Conf. on Computational Complexity*, 203–214 (2009).
[20] A. Dawar, S. Lindell, S. Weinstein, Infinitary logic and inductive definability over finite structures. *Information and Computation* 119:160–175 (1995).
[21] A. Dawar, M. Grohe, S. Kreutzer, N. Schweikardt. Model theory makes formulas large. In: *Proc. of the 34th Int. Colloquium on Automata, Languages and Programming.* Lecture Notes in Computer Science, Vol. 4596, 913–924 (2007).
[22] R. Diestel. *Graph theory.* Springer, 2nd ed. (2000).
[23] R. G. Downey, M. R. Fellows. *Parameterized complexity.* Springer (1998).
[24] H.-D. Ebbinghaus, J. Flum. *Finite model theory.* Springer, 2nd ed. (1999).
[25] A. Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fundam. Math.* 49:129–141 (1961).
[26] Y. Ershov, I. Lavrov, A. Taimanov, M. Taitslin. Elementary theories (In Russian). *Uspekhi Matematicheskikh Nauk* 20:37–108 (1965). English translation in *Russian Math. Surveys* 20:35–105 (1965).
[27] K. Etessami, N. Immerman. Tree canonization and transitive closure. *Information and Computation* 157:2–24 (2000).
[28] S. Evdokimov, I.N. Ponomarenko, G. Tinhofer. Forestal algebras and algebraic forests (on a new class of weakly compact graphs). *Discrete Mathematics* 225:149–172 (2000).
[29] R. Fagin. Probabilities on finite models. *J. Symb. Logic* 41:50–58 (1976).
[30] R. Fraïssé. Sur quelques classifications des systems de relations. *Publ. Sci. Univ. Alger* 1:35–182 (1954).
[31] M. Fürer. Weisfeiler-Lehman refinement requires at least a linear number of iterations. In: *Proc. of the 28th Int. Colloquium on Automata, Languages, and Programming.* Lecture Notes in Computer Science, Vol. 2076, 322–333 (2001).
[32] Y. Glebskii, D. Kogan, M. Liogonkii and V. Talanov. Range and fraction of satisfiability of formulas in the restricted predicate calculus. *Kibernetika*, Kyiv, 2:17–28 (1969).

[33] E. Grädel. Finite model theory and descriptive complexity. In: *Finite Model Theory and Its Applications*, pages 125–230. Springer (2007).

[34] M. Grohe. Finite variable logics in descriptive complexity theory. *The Bulletin of Symbolic Logic* 4:345–398 (1998).

[35] M. Grohe. Fixed-point logics on planar graphs. In: *Proc. of the 13th IEEE Symp. on Logic in Computer Science*, 6–15 (1998).

[36] M. Grohe. Isomorphism testing for embeddable graphs through definability. In: *Proc. of the 32nd ACM Symp. on Theory of Computing,* 63–72 (2000).

[37] M. Grohe. Definable tree decompositions. In: *Proc. of the 23rd IEEE Symp. on Logic in Computer Science,* 406–417 (2008).

[38] M. Grohe. Fixed-point definability and Polynomial Time on graphs with excluded minors. In: *Proc. of the 25th Symp. on Logic in Computer Science*, 179–188 (2010).

[39] M. Grohe. Fixed-point definability and polynomial time on chordal graphs and line graphs. In: *Fields of Logic and Computation*, Lecture Notes in Computer Science, Vol. 6300, 328–353 (2010).

[40] M. Grohe, J. Mariño. Definability and descriptive complexity on databases of bounded tree-width. In: *Proc. of the 7th Int. Conf. on Database Theory*, Lecture Notes in Computer Science, Vol. 1540, 70–82 (1999).

[41] M. Grohe, N. Schweikardt. The succinctness of first-order logic on linear orders, *Logical Methods in Computer Science* 1(1:6):1–25 (2005).

[42] M. Grohe, O. Verbitsky. Testing graph isomorphism in parallel by playing a game. In: *Proc. of the 33rd Int. Colloquium on Automata, Languages, and Programming.* Lecture Notes in Computer Science, Vol. 4051, 3–14 (2006).

[43] F. Harary. *Graph theory.* Addison-Wesley, Reading MA (1969).

[44] N. Immerman. Upper and lower bounds for first-order expressibility. *J. Comput. Syst. Sci.* 25:76–98 (1982).

[45] N. Immerman. *Descriptive complexity.* Springer (1999).

[46] N. Immerman, D. Kozen. Definability with bounded number of bound variables. *Information and Computation* 83:121–139 (1989).

[47] N. Immerman, E. Lander. Describing graphs: a first-order approach to graph canonization. In: *Complexity theory retrospective*, pages 59–81. Springer (1990).

[48] B. Jenner, J. Köbler, P. McKenzie, J. Torán. Completeness Results for Graph Isomorphism. *J. Comp. Syst. Sci.* 66:549–566 (2003).

[49] R. M. Karp, V. Ramachandran. Parallel algorithms for shared-memory machines. In: *Algorithms and complexity. Handbook of theoretical computer science.* Vol. A, pages 869–941. Elsevier (1990).

[50] J.-H. Kim, O. Pikhurko, J. Spencer, O. Verbitsky. How complex are random graphs in first-order logic? *Random Structures and Algorithms* 26:119–145 (2005).

[51] B. Laubner. Capturing polynomial time on interval graphs. In: *Proc. of the 25th Symp. on Logic in Computer Science*, 199–208 (2010).

[52] I. Lavrov. Effective inseparability of the sets of identically true and finitely refutable formulae for certain elementary theories (In Russian). *Algebra i Logika* 2:5–18 (1963).

[53] S. Lindell. A logspace algorithm for tree canonization. In: *Proc. of the 24th ACM Symp. on Theory of Computing* 400–404 (1992).

[54] R. J. Lipton, R. E. Tarjan. A separator theorem for planar graphs. *SIAM J. Appl. Math.* 36:177-189 (1979).

[55] G. L. Miller, J. H. Reif. Parallel tree contraction. Part 2: further applications. *SIAM J. Comp.* 20:1128–1147 (1991).

[56] B. Mohar, C. Thomassen. *Graphs on surfaces.* The John Hopkins University Press (2001).

[57] J. W. Moon. On the maximum degree in a random tree. *Michigan Math. J.*, 15:429–432, 1968.

[58] M. Naor, A. Nussboim, E. Tromer. Efficiently constructible huge graphs that preserve first order properties of random graphs. In: *Theory of Cryptography.* Lecture Notes in Computer Science, Vol. 3378, 66–85 (2005).

[59] O. Ore. *Theory of graphs.* American Mathematical Society (1962).

[60] E. Pezzoli. Computational complexity of Ehrenfeucht-Fraïssé games on finite structures. In: *Proc. of the 12th Conf. on Computer Science Logic 1998.* Lecture Notes in Computer Science, Vol. 1584, 159–170 (1999).

[61] O. Pikhurko, J. Spencer, O. Verbitsky. Succinct definitions in first-order graph theory. *Annals of Pure and Applied Logic* 139:74–109 (2006).

[62] O. Pikhurko, J. Spencer, O. Verbitsky. Decomposable graphs and definitions with no quantifier alternation. *European J. Comb.* 28:2264-2283 (2007).

[63] O. Pikhurko, H. Veith, O. Verbitsky. The first-order definability of graphs: upper bounds for quantifier depth. *Discrete Applied Mathematics* 154:2511–2529 (2006).

[64] O. Pikhurko, H. Veith, O. Verbitsky. The first-order definability of graphs: upper bounds for quantifier rank. E-print: *http://arxiv.org/abs/math.CO/0311041* (2003).

[65] O. Pikhurko, O. Verbitsky. Descriptive complexity of finite structures: saving the quantifier rank. *J. Symb. Logic* 70:419–450 (2005).

[66] B. Poizat. Deux ou trois choses que je sais de L_n. *J. Symb. Logic* 47:641–658 (1982).

[67] M. O. Rabin. Decidability of second order theories and automata on infinite trees. *Trans. of the AMS* 141:1–35 (1965).

[68] V. Ramachandran, J. Reif. Planarity testing in parallel. *J. Comput. Syst. Sci.* 49:517–561 (1994).

[69] F. Ramsey. On a problem of formal logic. *Proc. of the London Math. Soc. 2-nd series,* 30:264–286 (1930).

[70] N. Robertson, P.D. Seymour. Graph minors II. Algorithmic aspects of tree-width. *J. Algorithms* 7:309–322 (1986).

[71] W. L. Ruzzo. On uniform circuit complexity. *J. Comput. Syst. Sci.* 21:365–383 (1981).

[72] J. Spencer. *The strange logic of random graphs.* Springer (2001).

[73] J. Spencer, K. St. John. The complexity of random ordered structures. *Annals of Pure and Applied Logic* 152:174–179 (2008)

[74] D.A. Spielman. Faster isomorphism testing of strongly regular graphs. In: *Proc. of the 28th ACM Symp. on Theory of Computing* 576–584 (1996).

[75] O. Verbitsky. The first-order definability of graphs with separators via the Ehrenfeucht game. *Theoretical Computer Science* 343:158–176 (2005).

[76] O. Verbitsky. Planar graphs: logical complexity and parallel isomorphism tests. In: *Proc. of the 24th Symp. on Theoretical Aspects of Computer Science*, Lecture Notes in Computer Science, Vol. 4393, 682–693 (2007).

[77] B. Yu. Weisfeiler, A. A. Lehman. A reduction of a graph to a canonical form and an algebra arising during this reduction. *Nauchno-Technicheskaya Informatsia, Seriya 2*, 9:12–16 (1968). In Russian.

Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA.

Institute for Applied Problems of Mechanics and Mathematics, 79060 Lviv, Ukraine.

This page intentionally left blank

Contemporary Mathematics
Volume **558**, 2011

Methods for Algorithmic Meta Theorems

Martin Grohe and Stephan Kreutzer

Abstract. Algorithmic meta-theorems state that certain families of algorithmic problems, usually defined in terms of logic, can be solved efficiently. This is a survey of algorithmic meta-theorems, highlighting the general methods available to prove such theorems rather than specific results.

1. Introduction

Faced with the seeming intractability of many common algorithmic problems, much work has been devoted to studying restricted classes of admissible inputs on which tractability results can be retained. A particularly rich source of structural properties which guarantee the existence of efficient algorithms for many problems on graphs comes from structural graph theory, especially graph minor theory. It has been found that most generally hard problems become tractable on graph classes of bounded tree-width and many remain tractable on planar graphs or graph classes excluding a fixed minor.

Besides many specific results giving algorithms for individual problems, of particular interest are results that establish tractability of a large class of problems on specific classes of instances. These results come in various flavours. Here we are mostly interested in results that take a descriptive approach, i.e. results that use a logic to describe algorithmic problems and then provide general tractability results for all problems definable in that logic on specific classes of inputs. Results of this form are usually referred to as *algorithmic meta-theorems*. The first explicit algorithmic meta-theorem was proved by Courcelle [**3**] establishing tractability of decision problems definable in monadic second-order logic (even with quantification over edge sets) on graph classes of bounded tree-width, followed by similar results for monadic second-order logic with only quantification over vertex sets on graph classes of bounded clique-width [**4**], for first-order logic on graph classes of bounded degree [**45**], on planar graphs and more generally graph classes of bounded local tree-width [**23**], on graph classes excluding a fixed minor [**20**], on graph classes locally excluding a minor [**6**] and graph classes of bounded local expansion [**13**].

The natural counterpart to any algorithmic meta-theorem establishing tractability for all problems definable in a given logic L on specific classes of structures are corresponding lower bounds, i.e. results establishing intractability results for L

1991 *Mathematics Subject Classification.* Primary 68Q19; Secondary 68Q25.

with respect to structural graph parameters. Ideally, one would aim for results of the form: all problems definable in L are tractable if a graph class $\mathcal{C}$ has a specific property $\mathcal{P}$, such as bounded tree-width, but if $\mathcal{C}$ does not have this property, then there are L definable properties that are hard.

Early results on lower bounds have either focused on graph classes with very strong closure properties such as being closed under minors [**37**], or on very specific graph classes such as the class of all cliques [**4**]. Recently, however, much more general lower bounds have been established giving much tighter bounds on the tractability of monadic second-order logic [**30, 34**] and first-order logic [**31**] with respect to structural parameters.

In this paper we give a survey of the most important methods used to obtain algorithmic meta-theorems. All these methods have a model-theoretic flavor. In the first part of the paper we focus on upper bounds, i.e. tractability results. Somewhat different to the existing surveys on algorithmic meta-theorems [**26, 31**], this survey is organised along the core methods used to establish the results, rather than the specific classes of graphs they refer to. We put an emphasis on the most recent results not yet covered in the earlier surveys. In particular, we devote the longest section of this article to the recent linear time algorithm for deciding first-order definable properties on graphs of bounded expansion [**13**], which was established by a completely new technique that we call the "Colouring Technique" here. Moreover, for the first time we also survey lower bounds for algorithmic meta theorems, most of which are very recent as well.

2. Preliminaries

We assume familiarity with basic concepts of logic and graph theory and refer to the textbooks [**14, 27, 10**] for background. Our notation is standard; in the following we review a few important points.

If M, N are two sets, we define $M \dot{\cup} N$ as the *disjoint union* of M and N, obtained by taking the union of M and a copy N' of N disjoint from M. We write $\mathbb{Z}$ for the set of integers and $\mathbb{N}$ for the set of non-negative integers.

All graphs in this article are finite, undirected and simple, i.e., have no multiple edges and no self loops. We denote the vertex set of a graph G by $V(G)$ and its edge set by $E(G)$. We usually denote an edge between vertices v and w as vw, i.e., without parenthesis. We use standard graph theoretic notions like *sub-graphs*, *paths* and *cycles*, *trees* and *forests*, *connectedness* and *connected components*, the *degree* of a vertex, etc, without further explanation. Occasionally, we also need to work with *directed graphs* (for short: *digraphs*). Here we also use standard terminology.

A graph G is a *minor* of a graph H (we write $G \preceq H$) if G is isomorphic to a graph obtained from a subgraph of H by contracting edges. (*Contracting* an edge means deleting the edge and identifying its endvertices.) A graph H is an *excluded minor* for a class $\mathcal{C}$ of graphs if H is not a minor of any graph in $\mathcal{C}$.

All structures in this article are finite and without functions. Hence a *signature* is a finite set of relation symbols and constant symbols. Each relation symbol $R \in \sigma$ is equipped with its *arity* $ar(R) \in \mathbb{N}$. Let σ be a signature. A σ-structure A is a tuple consisting of a finite set $V(A)$ of elements, the *universe*, for each relation symbol $R \in \sigma$ of arity r an r-ary relation $R(A) \subseteq V(A)^r$, and for each constant symbol $c \in \sigma$ a constant $c(A) \in V(A)$. (Hence if σ contains constant symbols then $V(A)$ must be nonempty.) A signature is *relational* if it contains no constant symbols, and a structure is *relational* if its signature is. A signature is *binary*, if the arity of all relation symbols in it is at most 2. The *order* $|A|$ of a σ-structure A is $|V(A)|$ and its *size* $||A||$ is $|\sigma| + |V(G)| + \sum_{R\in\sigma} |R(G)|$.[1] For σ-structures A, B, we write $A \cong B$ to denote that A and B are isomorphic.

We may view graphs as $\{E\}$-structures, where E is a binary relation symbol. The *Gaifman-graph* $G(A)$ of a σ-structure A is the graph with vertex set $V(A)$ and edge set $\{bc : \text{there is an } R \in \sigma \text{ and a tuple } \overline{a} \in R(A) \text{ such that } b, c \in \overline{a}\}$. Here and elsewhere, for a tuple $\overline{a} = (a_1, \ldots, a_k)$ and an element b we write $b \in \overline{a}$ instead of $b \in \{a_1, \ldots, a_k\}$.

We denote the class of all structures by $\mathcal{S}$ and the class of all graphs by $\mathcal{G}$. If $\mathcal{C}$ is a class of graphs, we let $\mathcal{S}(\mathcal{C})$ be the class of all structures with Gaifman graph in $\mathcal{C}$. If $\mathcal{C}$ is a class of structures and σ a signature, we let $\mathcal{C}(\sigma)$ be the class of all σ-structures in $\mathcal{C}$.

Let σ be a relational signature and $A, B \in \mathcal{S}(\sigma)$. Then A is a *substructure* of B (we write $A \subseteq B$) if $V(A) \subseteq V(B)$ and $R(A) \subseteq R(B)$ for all $R \in \sigma$. If $A \subseteq B$ and $R(A) = R(B) \cap V(A)^{ar(R)}$ for all $R \in \sigma$ then A is an *induced* substructure of B. For a set $W \subseteq V(B)$, we let $B[W]$ be the induced substructure of B with universe W, and we let $B \setminus W := B[V(B) \setminus W]$.

Formulas of *first-order logic* FO are built from variables ranging over elements of the universe of a structure, atomic formulas $R(t_1, \ldots, t_k)$ and $t_1 = t_2$, where the t_i are *terms*, i.e., variables or constant symbols, the usual Boolean connectives $\wedge, \vee, \rightarrow, \neg$, and existential and universal quantification $\exists x$, $\forall x$, where x is a variable.

In *monadic second order logic* MSO we also have "set variables" ranging over sets of elements of the universe, new atomic formulas $X(t)$, where X is a set variable and t a term, and quantification over set variables. In the context of algorithmic meta-theorems, an extension MSO_2 of MSO is often considered. MSO_2 is a logic only defined on graphs, and in addition to variables ranging over sets of vertices it has also variables ranging over sets of edges of a graph. The generalisation of MSO_2 to arbitrary structures is known as *guarded second-order logic* GSO. It has variables ranging over relations of arbitrary arities, but for relations of arity greater than one only allows guarded quantification $\exists X \subseteq R$ and $\forall X \subseteq R$, where R is a relation symbol.

We write $\varphi(x_1, \ldots, x_k)$ to denote that the free variables of a formula (of some logic) are among $x_1, \ldots, x_k$, and for a structure A and elements $a_1, \ldots, a_k \in V(A)$, we write $A \models \varphi[a_1, \ldots, a_k]$ to denote that A satisfies φ if x_i is interpreted by a_i. Furthermore, we let

$$\varphi(A) = \{(a_1, \ldots, a_k) \mid A \models \varphi[a_1, \ldots, a_k]\}.$$

If φ is a *sentence*, i.e., a formula without free variables, we just write $A \models \varphi$ to denote that A satisfies φ.

3. Algorithmic Meta-Theorems and Model-Checking Problems

As described in the introduction, the algorithmic meta-theorems we are interested in here have the following general form:

Algorithmic Meta Theorem (Nonuniform Version). *Let* L *be a logic (typically* FO *or* MSO*),* $\mathcal{C}$ *a class of structures (most often a class of graphs), and* $\mathcal{T}$ *a*

---

[1] Up to constant factors, $||A||$ corresponds to the size of a representation of A in an appropriate model of computation, random access machines with a uniform cost measure (cf. [**19**]).

class of functions on the natural numbers (typically the class of all linear functions or the class of all polynomial functions). Then for all L*-definable properties* π *of structures in* $\mathcal{C}$*, there is a function* $t \in \mathcal{T}$ *and an algorithm that tests if a given structure* $A \in \mathcal{C}$ *has property* π *in time* $t(||A||)$*.*

Of course we may also restrict the algorithm's consumption of memory space or other resources, but most known meta-theorems are concerned with running time. (One notable exception is [**15**].) We note that in MSO we can define NP-complete properties of graphs, for example 3-colourability. Hence unless P = NP, there are MSO-definable properties of graphs for which no polynomial time algorithm exists. All FO-definable properties of graphs have a polynomial time algorithm, but the exponent of the running time of the algorithm will usually depend on the formula defining the property. There are generally believed assumptions from parameterized complexity theory (see [**12, 21**]) which imply that for every constant c there are FO-definable properties of graphs that cannot be decided by an $O(n^c)$-algorithm, where n is the size of the input graph.

Stated as above, meta-theorems are non-uniform in the sense that there is no direct connection between a property π and the corresponding algorithm. It would certainly be desirable to be able to construct the algorithm from an L-definition of the property. Fortunately, we usually obtain such uniform versions of our meta-theorems, which can be phrased in the following form:

Algorithmic Meta Theorem (Uniform Version). *Let* L *be a logic,* $\mathcal{C}$ *a class of structures, and* $\mathcal{T}$ *a class of functions on the natural numbers. Then there is an algorithm that, given an* L*-sentence* φ *and a structure* $A \in \mathcal{C}$*, decides whether* A *satisfies* φ*. Moreover, for every* L*-sentence* φ *there is a function* $t_\varphi \in \mathcal{T}$ *such that the running time of the algorithm on input* φ, A *is bounded by* $t_\varphi(||A||)$*.*

Hence in this uniform version, our meta-theorems are just statements about the complexity of model-checking problems. The *model-checking problem for the logic* L *on the class* $\mathcal{C}$ *of structures* is the following decision problem:

> Given an L-sentence φ and a structure $A \in \mathcal{C}$, decide if A satisfies φ.

We denote this problem by $\mathrm{MC}(\mathrm{L}, \mathcal{C})$. It is well-known that both $\mathrm{MC}(\mathrm{FO}, \mathcal{G})$ and $\mathrm{MC}(\mathrm{MSO}, \mathcal{G})$ are PSPACE-complete [**49**]. Hence we cannot hope to obtain polynomial time algorithms. We say that $\mathrm{MC}(\mathrm{L}, \mathcal{C})$ is *fixed-parameter tractable* (for short: *fpt*) if it can be decided by an algorithm running in time

$$f(k) \cdot n^c \tag{3.1}$$

for some function f and some constant c. Here k denotes the length of the input formula φ and n the size of the input structure A. We say that $\mathrm{MC}(\mathrm{L}, \mathcal{C})$ is fpt by *linear time parameterized algorithms* if we can let $c = 1$ in (3.1). Now for $\mathcal{T}$ being the class of all linear functions, we can concisely phrase our algorithmic meta-theorem as follows:

Algorithmic Meta Theorem (Uniform Version for Linear Time). *Let* L *be a logic and* $\mathcal{C}$ *a class of structures. Then* $\mathrm{MC}(\mathrm{L}, \mathcal{C})$ *is fpt by linear time parameterized algorithms.*

This is the form in which we usually state our meta-theorems. Even though we usually think of the L-sentence φ as being fixed, we may wonder what the dependence of the running time of our fixed-parameter tractable model-checking algorithm on φ is, i.e., what we can say about the function f in (3.1). For all known meta-theorems f is easily seen to be computable. However, f usually grows very quickly. Even for very simple classes $\mathcal{C}$ (such as the class of all trees, which is contained in all classes that appear in the meta-theorems surveyed in this article), it has been shown [**24**] that, under generally believed complexity theoretic assumptions, all fpt algorithms for $\mathrm{MC}(\mathrm{FO}, \mathcal{C})$, and hence for $\mathrm{MC}(\mathrm{MSO}, \mathcal{C})$, have a non-elementary running time (i.e., f grows faster than any stack of exponentials of fixed-height).

Part I: Upper Bounds

In this first part of the paper we consider the most successful methods for establishing algorithmic meta-theorems.

4. The Automata Theoretic Method

The automata theoretic approach to algorithmic meta-theorems can best be explained with a familiar algorithmic problem, *regular expression pattern matching*. The goal is to decide whether a text matches a regular expression (or equivalently, has a sub-string matching a regular expression). One efficient way of doing this is to translate the regular expression into a deterministic finite automaton and then run the automaton on the text. The translation of the regular expression into the automaton may cause an exponential blow-up in size, but running the automaton on the text can be done in time linear in the length of the text, so this leads to an algorithm running in time of $O(2^k + n)$, where k is the length of the regular expression and n the length of the text. In practise, we usually match a short regular expression against a long text. Thus k is much smaller than n, and this algorithm, despite its exponential running time, may well be the best choice.

We can use the same method for MSO model-checking on words, suitably encoded as relational structures. By the Büchi-Elgot-Trakhtenbrot Theorem [**2, 16, 48**], we can translate every MSO-formula φ to a finite automaton A_φ that accepts precisely the words satisfying the formula. Hence to test if a word W satisfies φ we just need to run A_φ on W and see if it accepts. This leads to a linear time fpt algorithm for MSO model-checking on the class of words. The same method even works for trees. Since we are going to use this result later, let us state it more formally: with every finite alphabet Σ we associate a signature τ_Σ consisting of two binary relation symbols E_L and E_R and a unary relation symbol P_a for every $a \in \Sigma$. A *binary Σ-tree* is a τ_Σ-structure whose underlying graph is a tree in which E_L is the "left-child relation" and E_R is the "right-child relation" and in which every vertex belongs to exactly one P_a. Let $\mathcal{B}_\Sigma$ denote the class of binary Σ-trees.

THEOREM 4.1 ([**11, 46**]). *For every finite alphabet Σ, the model-checking problem* $\mathrm{MC}(\mathrm{MSO}, \mathcal{B}_\Sigma)$ *is fpt by a linear-time parameterized algorithm.*

5. The Reduction Method

In this section, we will show how *logical reductions* can be used to transfer algorithmic meta-theorems between classes of structures. We start by reviewing *syntactic interpretations* or *transductions*, a well known tool from model theory. Recall that for a formula $\varphi(\overline{x})$ and a structure A, we denote by $\varphi(A)$ the set of all tuples $\overline{a}$ such that $A \models \varphi[\overline{a}]$.

Definition 5.1. *Let σ, τ be signatures and let $\mathrm{L} \in \{\mathrm{MSO}, \mathrm{FO}\}$. A (one-dimensional)* L-transduction *from τ to σ is a sequence*

$$\Theta := \big(\varphi_{valid}, \varphi_{univ}(x), (\varphi_R(\overline{x}))_{R\in\sigma}, (\varphi_c(x))_{c\in\sigma}\big)$$

of $\mathrm{L}[\tau]$ formulas where for all relation symbols $R \in \sigma$, the number of free variables in φ_R is equal to the arity of R. Furthermore, for all τ-structures A such that $A \models \varphi_{valid}$ and all constant symbols $c \in \sigma$ there is exactly one element $a \in V(A)$ satisfying φ_c.

If A is a τ-structure such that $A \models \varphi_{valid}$ we define $\Theta(A)$ as the σ-structure B with universe $V(B) := \varphi_{univ}(A)$, $R(B) := \varphi_R(A)$ for each $R \in \sigma$ and $c(B) := a$, where a is the uniquely defined element with $\{a\} = \varphi_c(A)$.

Finally, if $\mathcal{C}$ is a class of τ-structures we let $\Theta(\mathcal{C}) := \{\Theta(A) : A \in \mathcal{C}, A \models \varphi_{valid}\}$.

Every L-transduction from τ to σ naturally defines a translation of L-formulas from $\varphi \in \mathrm{L}[\sigma]$ to $\varphi^* := \Theta(\varphi) \in \mathrm{L}[\tau]$. Here, φ^* is obtained from φ by recursively replacing

- first-order quantifiers $\exists x\varphi$ by $\exists x(\varphi_{univ}(x) \wedge \varphi^*)$ and quantifiers $\forall x\varphi$ by $\forall x(\varphi_{univ}(x) \rightarrow \varphi^*)$,
- second-order quantifiers $\exists X\varphi$ and $\forall X\varphi$ by $\exists X\big(\forall y(Xy \rightarrow \varphi_{univ}(y)) \wedge \varphi^*\big)$ and $\forall X\big(\forall y(Xy \rightarrow \varphi_{univ}(y)) \rightarrow \varphi^*\big)$ respectively and
- atoms $R(\overline{x})$ by $\varphi_R(\overline{x})$ and
- every atom $P(\overline{u})$, where $\overline{u}$ is a sequence of terms containing a constant symbol $c \in \sigma$ and $P \in \sigma \cup \{=\}$ by the sub-formula $\exists c'(\varphi_c(c') \wedge P(\overline{u}'))$, where $\overline{u}'$ is obtained from $\overline{u}$ by replacing c by c' and c' is a new variable not occurring elsewhere in $\overline{u}$.

The following lemma is easily proved (see [**27**]).

Lemma 5.2. *Let Θ be an* MSO*-transduction from τ to σ. Then for all* $\mathrm{MSO}[\sigma]$*-formulas and all τ-structures $A \models \varphi_{valid}$*

$$A \models \Theta(\varphi) \iff \Theta(A) \models \varphi.$$

The analogous statement holds for FO*-transductions.*

Transductions map τ-structures to σ-structures and also map formulas back the other way. This is not quite enough for our purposes as we need a mapping that takes both formulas and structures over σ to formulas and structures over τ.

Definition 5.3. *Let σ, τ be signatures and $\mathrm{L} \in \{\mathrm{MSO}, \mathrm{FO}\}$. Let $\mathcal{C}$ be a class of σ-structures and $\mathcal{D}$ be a class of τ-structures. An* L-reduction from $\mathcal{C}$ to $\mathcal{D}$ *is a pair $(\Theta, \mathcal{A})$, where Θ is an* L*-transduction from τ to σ and $\mathcal{A}$ is a polynomial time algorithm that, given a σ-structure $A \in \mathcal{C}$, computes a τ-structure $\mathcal{A}(A) \in \mathcal{D}$ with $\Theta(\mathcal{A}(A)) \cong A$.*

The next lemma is now immediate.

Lemma 5.4. *Let $(\Theta, \mathcal{A})$ be an* L*-reduction from a class $\mathcal{C}$ of σ-structures to a class $\mathcal{D}$. If* $\mathrm{MC}(\mathrm{L}, \mathcal{D})$ *is fpt then* $\mathrm{MC}(\mathrm{L}, \mathcal{C})$ *is fpt.*

Note that if the algorithm $\mathcal{A}$ in the reduction is a linear time algorithm then we can also transfer fixed-parameter tractability by linear time parameterized algorithms from $\mathrm{MC}(\mathrm{L}, \mathcal{D})$ to $\mathrm{MC}(\mathrm{L}, \mathcal{C})$. Of course, we can also apply the lemma to prove hardness of $\mathrm{MC}(\mathrm{L}, \mathcal{D})$ in cases where $\mathrm{MC}(\mathrm{L}, \mathcal{C})$ is known to be hard. In our context, L-reductions play a similar role as polynomial time many-one reductions in complexity theory. We start with two simple applications of the reduction technique.

EXAMPLE 5.5. The first example is a reduction from graphs to their complements. Let $\Theta := (\varphi_{valid}, \varphi_{univ}, \varphi_E)$ be a transduction from σ_{graph} to σ_{graph} defined by $\varphi_{valid} = \varphi_{univ} := \mathit{true}$ and $\varphi_E(x,y) := \neg E(x,y)$. Then, for all graphs G, $\Theta(G)$ is the complement of G, i.e. the graph $\overline{G} := (V(G), \overline{E(G)})$.

Now if $\mathcal{A}$ is the algorithm which on input G outputs $\overline{G}$, then $(\Theta, \mathcal{A})$ is a first-order reduction reducing a pair (G, φ), where G is a graph and $\varphi \in \mathrm{FO}$, to $(\overline{G}, \Theta(\varphi))$ such that $G \models \varphi$ if, and only if, $\overline{G} \models \Theta(\varphi)$.

Hence, if $\mathcal{C}$ is the class of complements of binary trees then this reduction together with Theorem 4.1 implies that $\mathrm{MC}(\mathrm{FO}, \mathcal{C})$ is fpt. The same is true for MSO.

Obviously, the same reduction shows that if $\mathcal{D}$ is any class of graphs for which we will show $\mathrm{MC}(\mathrm{FO}, \mathcal{D})$ to be fpt in the remainder of this paper, then $\mathrm{MC}(\mathrm{FO}, \mathcal{C})$ is also fpt, where $\mathcal{C}$ is the class of graphs whose complements are in $\mathcal{D}$. $\dashv$

EXAMPLE 5.6. In this example, we transfer the fixed-parameter tractability of MSO model-checking from binary to arbitrary Σ-trees. Let us fix a finite alphabet Σ, and let σ_Σ be the signature consisting of the binary relation symbols E and a unary relation symbol P_a for every $a \in \Sigma$. A Σ-*tree* is a τ_Σ-structure whose underlying graph is a directed tree with edges directed away from the root in which every vertex belongs to exactly one P_a.

With every Σ-tree T we associate a binary Σ-tree B_T as follows: we order the children of each node of T arbitrarily. Now each node has a "first child" and a "last child", and each child of a node except the last has a "next sibling". We let B_T be the binary Σ-tree with $V(B_T) = V(T)$ such that $E_L(B_T)$ is the "first child" relation of T and $E_R(B_T)$ is the "next sibling" relation. Clearly, there is a linear time algorithm $\mathcal{A}$ that computes B_T from T. (More precisely, the algorithm computes "some" B_T constructed in the way described, because the tree B_T depends on the choice of the ordering of children.)

Now we define an MSO-transduction $\Theta = (\varphi_{valid}, \varphi_{univ}, \varphi_E, (\varphi_{P_a})_{a\in\Sigma})$ from τ_Σ to σ_Σ such that for every Σ-tree T we have $\Theta(B_T) \cong T$. We let $\varphi_{valid} = \varphi_{univ}(x) := \mathit{true}$ and $\varphi_{P_a}(x) := P_a(x)$ for all $a \in \Sigma$. Moreover, we let

$$\varphi_E(x,y) := \forall X\Big[\Big(\forall z(E_L(x,z) \to X(z)) \\ \wedge \forall z \forall z'\big((X(z) \wedge E_R(z,z')) \to X(z')\big)\Big) \to X(y)\Big].$$

We leave it to the reader to verify that in B_T this formula indeed defines the edge relation of T. $\dashv$

The rest of this section is devoted to a more elaborate application of the reduction technique: we will show that monadic second-order logic is fpt on all classes of structures of bounded tree-depth, a concept introduced in [**40**]. We first need some preparation.

In the following, we shall work with directed trees and forests. As before, all edges are directed away from the root(s). The *height* of a vertex v in a forest is its distance from the root of its tree. The *height* of the forest is the maximum of the heights of its vertices. A vertex v is an *ancestor* of a vertex w if there is a path from v to w. Furthermore, v is a *descendant* of w if w is an ancestor of v.

We define the *closure* of a forest F to be the (undirected) graph $clos(F)$ with vertex set $V(clos(F)) := V(F)$ and edge set

$$E(clos(F)) := \{vw \mid v \text{ is an ancestor of } w\}.$$

DEFINITION 5.7 ([**40**]). *A graph G has* tree-depth d *if it is a sub-graph of the closure of a rooted forest F of height d. We call F a* tree-depth decomposition *of G.*

It was proved in [**40**] that there is an algorithm which, given a graph G of tree-depth at most d, computes a tree-depth decomposition in time $f(d) \cdot |G|$, for some computable function d. For our purposes, an algorithm for computing approximate tree-depth decompositions will be enough. One such algorithm simply computes a *depth-first search forest* of its input graph. In [**40**] it was shown that a simple path of length 2^d has tree-depth exactly d. As the tree-depth of a sub-graph $H \subseteq G$ is at most the tree-depth of G, no graph of tree-depth d can contain a path of length greater than 2^d. This implies the following lemma.

LEMMA 5.8. *Let G be a graph of tree-depth d and let F be a forest obtained from a depth-first search (DFS) in G. Let F' be the closure of F. Then $G \subseteq F'$ and the depth of F is at most 2^d.*

We will now define an FO-reduction from the class $\mathcal{D}_d$ of graphs of tree-depth at most d to the class of Σ_d-trees, for a suitable alphabet Σ_d. For simplicity, we only present the reduction for connected graphs. We fix d and let $k = 2^d$ and $\Sigma_d = \{0,1\}^{\leq k}$, the set of all $\{0,1\}$-strings of length at most k.

We first associate a Σ_d-tree T_G with every connected graph $G \in \mathcal{D}_d$. Let T be a DFS-tree of G. By Lemma 5.8, the height of T is at most k, and $G \subseteq clos(T)$. We define a Σ_d-labelling of the vertices of T to obtain the desired Σ_d-tree T_G as follows: let $v \in V(T)$ be a vertex of height h, and let $u_0, \dots, u_{h-1}, v$ be the vertices on the path from the root to v in T. Then v is labelled by $i_0 \dots i_{h-1} \in \{0,1\}^h \subseteq \Sigma_d$, where $i_j = 1 \leftrightarrow u_j v \in E(G)$. It is not difficult to show that T_G can be computed from G by a linear time algorithm $\mathcal{A}$.

We leave it to the reader to define a first-order transduction Θ from σ_{Σ_k} to $\{E\}$ such that for every connected graph $G \in \mathcal{D}_d$ we have $\Theta(T_G) = G$. This yields the desired reduction.

By Example 5.6 and Theorem 4.1 we obtain the next corollary.

COROLLARY 5.9. $\mathrm{MC}(\mathrm{MSO}, \mathcal{D}_d)$ *is fpt by a linear time parameterized algorithm.*

6. The Composition Method

The next method we consider is based on *composition theorems* for first-order logic and monadic second-order logic, which allow to infer the formulas satisfied by a structure composed of simpler pieces from these pieces. The best known such composition theorems are due to Feferman and Vaught [**18**].

Let L be either first-order or monadic second-order logic. Recall that the *quantifier rank* of an L-formula is the maximum number of nested quantifiers in the formula (counting both first-order and second-order quantifiers). Let A be a structure, $\overline{a} = (a_1, \dots, a_k) \in V(A)^k$ and $q \in \mathbb{N}$. Then the *q-type of $\overline{a}$ in A* is the set $\mathrm{tp}^A_{\mathrm{L},q}(\overline{a})$ of all L-formulas $\varphi(\overline{x})$ of quantifier-rank at most q such that $A \models \varphi[\overline{a}]$.

As such, the type of a tuple is an infinite class of formulas. However, we can syntactically normalise first-order and second-order formulas so that every formula can effectively be transformed into an equivalent normalised formula of the same quantifier-rank and furthermore for every quantifier-rank there are only finitely many pairwise non-equivalent normalised formulas. Hence, we can represent types by finite sets of normalised formulas. We will do so tacitly whenever we work with types in this paper. The following basic composition lemma can easily be proved using Ehrenfeucht-Fraïssé games (see [**36**] for a proof).

LEMMA 6.1. *Let A, B be σ-structures and $\overline{a} \in V(A)^k, \overline{b} \in V(B)^\ell$, and $\overline{c} \in V(A \cap B)^m$ such that all elements of $V(A \cap B)$ appear in $\overline{c}$. Let $q \in \mathbb{N}$, and let $\mathrm{L} \in \{\mathrm{FO}, \mathrm{MSO}\}$.*

Then $\mathrm{tp}^{A\cup B}_{\mathrm{L},q}(\overline{a}\overline{b}\overline{c})$ is uniquely determined by $\mathrm{tp}^{A}_{\mathrm{L},q}(\overline{a}\overline{c})$ and $\mathrm{tp}^{B}_{\mathrm{L},q}(\overline{b}\overline{c})$. Furthermore, there is an algorithm that computes $\mathrm{tp}^{A\cup B}_{\mathrm{L},q}(\overline{a}\overline{b}\overline{c})$ from $\mathrm{tp}^{A}_{\mathrm{L},q}(\overline{a}\overline{c})$ and $\mathrm{tp}^{B}_{\mathrm{L},q}(\overline{b}\overline{c})$.

As an application and an illustration of the method, we sketch a proof of Courcelle's well-known meta-theorem for monadic second-order logic on graphs of bounded tree width. A *tree decomposition* of a graph G is a pair (T, β) where T is a tree and β a mapping that assigns a subset $\beta(t) \subseteq V(G)$ with every $t \in V(T)$, subject to the following conditions:

(1) For every vertex $v \in V(G)$ the set $\{t \in V(T) \mid v \in \beta(t)\}$ is nonempty and connected in T.
(2) For every edge $vw \in E(G)$ there is a $t \in V(T)$ such that $v, w \in \beta(t)$.

The *width* of a tree decomposition (T, β) is $\max\{|\beta(t)| : t \in V(T)\} - 1$, and the *tree width* of a graph G is the minimum of the widths of all tree decompositions of G. Intuitively, tree width may be viewed as a measure for the similarity of a graph with a tree. Bodlaender [**1**] proved that there is an algorithm that, given a graph G of tree width w, computes a tree decomposition of G of width w in time $2^{O(w^3)}|G|$. It is an easy exercise to show that every graph G of tree-depth at most h also has tree-width at most h.

THEOREM 6.2 ([**3**]). *Let $\mathcal{C}$ be a class of graphs of bounded tree-width. Then $\mathrm{MC}(\mathrm{MSO}, \mathcal{C})$ is fpt by linear time parameterized algorithms.*

Proof sketch. Let $G \in \mathcal{C}$ and $\varphi \in \mathrm{MSO}$ be given. Let w be the tree width of G (which is bounded by some constant), and let q be the quantifier rank of φ. We first use Bodlaender's algorithm to compute a tree decomposition (T, β) of G of width w. We fix a root r of T arbitrarily. For every $t \in V(T)$, we let T_t be the sub-tree of T rooted at t, and we let G_t be the induced subgraph of G with vertex set $\bigcup_{u \in V(T_t)} \beta(u)$. Moreover, we let $\overline{b}_t$ be a $(w+1)$-tuple of vertices that contains precisely the vertices in $\beta(t)$. (Without loss of generality we assume $\beta(t)$ to be nonempty.)

Now, beginning from the leaves, we inductively compute for each $t \in V(T)$ the type $\mathrm{tp}^{G_t}_{\mathrm{MSO},q}(\overline{b}_t)$. We can do this by brute force if t is a leaf and hence $|G_t| \leq w+1$, and we use Lemma 6.1 if t is an inner node.

Finally, we check whether $\varphi \in \mathrm{tp}^{G_r}_{\mathrm{MSO},q}(\overline{b}_r) = \mathrm{tp}^{G}_{\mathrm{MSO},q}(\overline{b}_r)$. $\square$

Courcelle's theorem can easily be generalised from graphs to arbitrary structures, and it can be extended from monadic to guarded second-order logic (and thus to MSO_2 on graphs). An alternative proof of Courcelle's Theorem is based on Theorem 4.1 and the reduction method of Section 5.

By a similar application of the composition method it can be proved that $MC(MSO, \mathcal{C})$ is fpt for all classes $\mathcal{C}$ of graphs of bounded clique width (see [**4**]). Further applications of the composition method can be found in [**5, 36**].

Finally, let us mention an analogue of Courcelle's theorem for logarithmic space, recently proved in [**15**]: for every class $\mathcal{C}$ of graphs of bounded tree width, there is an algorithm for $MC(MSO_2, \mathcal{C})$ that uses space $O(f(k) \cdot \log n)$, where f is a computable function and k, n denote the size of the input formula and structure, respectively, of the model-checking problem.

7. Locality based arguments

In Section 5 we have seen how logical reductions can be used to transfer tractability results from a one class of structures to another. In this section we will look at a tool that will allow us to transfer tractability results from a class $\mathcal{C}$ of a structures to the class of all structures that locally look like a structure from $\mathcal{C}$.

We start with a simple example to explain the basic idea. Recall that a *homomorphism* from a graph H to a graph G is a function $\pi : V(H) \to V(G)$ such that whenever $uv \in E(H)$ then $\pi(u)\pi(v) \in E(G)$. The *graph homomorphism problem* asks, given two graphs H and G, whether there is a homomorphism from H to G. The homomorphism problem can trivially be solved in time $O(|G|^{|H|} \cdot |H|^2)$. The question is if we can solve it in time $f(|H|)|G|^c$, for some computable function f and constant c. In general, this is not possible, but it becomes possible if the graph G is "locally simple".

To explain the idea, suppose first that H is a connected graph. Then if there is a homomorphism π from H to G then the distance between any two vertices in the image $\pi(H)$ is at most $|H| - 1$. To exploit this observation, we define the *k-neighbourhood* of a vertex v in a graph G to be the subgraph of G induced by the set of all vertices of distance at most k from v. Then to test if there is a homomorphism from a connected k-vertex graph H to a graph G, we test for all $v \in V(G)$ whether the $(k-1)$-neighbourhood of v contains a homomorphic image of H. Of course in general, this does not help much, but it does help if the $(k-1)$-neighbourhoods in G are structurally simpler than the whole graph G. For example, if the girth of G (that is, the length of the shortest cycle) is at least k, then the $(k-1)$-neighbourhood of every vertex is a tree, and instead of testing whether there is a homomorphism from H to arbitrary graph we only need to test whether there is a homomorphism from H to a family of trees (of depth at most $k-1$), and this is much easier than the general homomorphism problem. Or if the maximum degree of G is d, then the order of the $(k-1)$-neighbourhood of every vertex in G is less than $(d+1)^k$, and instead of testing whether there is a homomorphism from H to graph of arbitrary size we only need to test whether there is a homomorphism from H to a family of graphs of size less than $(d+1)^k$. To apply the same method if H is not connected, we just check for each connected component of H separately if there is a homomorphism to G.

We can apply the same idea to the model-checking problem for first-order logic, because by Gaifman's Locality Theorem, first-order logic is *local* in the following sense: if σ is a relational signature and A is a σ-structure, we define the distance $d^A(a, b)$ between any two vertices $a, b \in V(A)$ to be the length of the shortest path from a to b in the Gaifman-graph $G(A)$ of A.[2] We define the *r-neighbourhood* $N_r^A(a)$ of a vertex $a \in V(A)$ to be the induced substructure of A with universe $\{b \mid d^A(a,b) \leq r\}$. A first-order formula $\varphi(x)$ is *r-local* if for every structure A and all $a \in V(A)$

$$A \models \varphi[a] \qquad \text{iff} \qquad N_r^A(a) \models \varphi[a].$$

Hence, truth of an r-local formula at an element a only depends on its r-neighbourhood. A *basic local sentence* is a first-order sentence of the form

$$\exists x_1 \dots \exists x_k \Big(\bigwedge_{1 \leq i < j \leq k} \operatorname{dist}(x_i, x_j) > 2r \wedge \bigwedge_{i=1}^{k} \vartheta(x_i) \Big) \tag{7.1}$$

where $\vartheta(x)$ is r-local. Here $\operatorname{dist}(x,y) > 2r$ is a first-order formula stating that the distance between x and y is greater than $2r$.

Theorem 7.1 (Gaifman's Locality Theorem[**25**]). *Every first-order sentence is equivalent to a Boolean combination of basic local sentences. Furthermore, there is an algorithm that, given a first-order formula as input, computes an equivalent Boolean combination of basic local sentences.*

We can exploit Gaifman's Theorem to efficiently solve the model-checking problem for first-order logic in structures that are "locally simple" as follows: Given a structure A and a first-order sentence φ, we first compute a Boolean combination φ' of basic local sentences that is equivalent to φ. Then we check for each of the basic local sentences appearing in φ' whether they hold in A. We can easily combine the results to check whether the Boolean combination φ' holds. To check whether a basic local sentence of the form (7.1) holds in A, we first compute the set $T(\vartheta)$ of all $a \in V(A)$ such that $N_r^A(a) \models \varphi[a]$, or equivalently, $A \models \varphi[a]$. For this, we only need to look at the r-neighbourhoods of the elements of a, and as we assumed A to be "locally simple", we can do this efficiently. It remains to check whether $T(\vartheta)$ contains k vertices of pairwise distance greater than $2r$. It turns out that this can be reduced to a "local" problem as well, and as A is "locally simple", it can be solved efficiently.

This idea yields the following lemma, which captures the core of the locality method. We say that first-order model-checking is *locally fpt* on a class $\mathcal{C}$ of structures if there is an algorithm that, given a structure $A \in \mathcal{C}$, an element $a \in V(A)$, a sentence $\varphi \in \mathrm{FO}$, and an $r \in \mathbb{N}$, decides whether $N_r^A(a) \models \varphi$ in time $f(r, |\varphi|) \cdot |A|^{O(1)}$, for some computable function f.

Lemma 7.2 ([**23, 6**]). *Let $\mathcal{C}$ be a class of structures on which first-order model-checking is locally fpt. Then* $\mathrm{MC}(\mathrm{FO}, \mathcal{C})$ *is fpt.*

Maybe surprisingly, there are many natural classes of graphs on which first-order model-checking is locally fpt, among them planar graphs and graphs of bounded degree. The most important of these are the classes of *bounded local tree width.* A class $\mathcal{C}$ of graphs has this property if for every $r \in \mathbb{N}$ there is a $k \in \mathbb{N}$ such that for every $G \in \mathcal{C}$ and every $v \in V(G)$ we have $\operatorname{tw}(N_r^G(v)) \leq k$. Examples of classes of graphs of bounded local tree width are all classes of graphs that can be embedded in a fixed surface, all classes of bounded degree, and (trivially) all classes of bounded tree width. Bounded local tree width was first considered

[2]See Section 2 for a definition of Gaifman-graphs.

by Eppstein [**17**] in an algorithmic context (under the name "diameter tree width property").

In [**6**] Lemma 7.2 is applied in a context that goes beyond bounded local tree width to show that first-order model-checking is fpt on all classes of graphs locally excluding a minor.

8. Colouring and Quantifier-Elimination

In Section 7 we have seen a method for establishing tractability results based on structural properties of r-neighbourhoods in graphs. Another way of presenting the locality method is that we cover the graph by local neighbourhoods which have a simpler structure than the whole graph. More generally we could use other forms of covers, i.e. cover the graph by arbitrary induced sub-graphs whose structure is simple enough to allow tractable model-checking. The main difficulty is to infer truth of a formula in the whole graph from the truth of (possibly a set of) formulas in the individual sub-graphs used in the cover. In the case of neighbourhood covers, Gaifman's locality theorem provided the crucial step in the construction which allowed us to reduce the model-checking problem in the whole graph to model-checking in individual r-neighbourhoods.

In this section we present a similar method. Again the idea is that we cover the graph by induced sub-graphs. However this time r-neighbourhoods will not necessarily be contained in a single sub-graph. This will make combining model-checking results in individual sub-graphs to the complete graph much more complicated.

The method we present is based on vertex colourings of graphs. Basically, we will colour a graph with a certain number c of colours, where c will depend on the formula we want to check, such that for some $k < c$, the union of any k colours induces a sub-graph of simple structure.

This technique was first developed by DeVos et al. [**9**] for graph classes excluding a fixed minor. They showed that if $\mathcal{C}$ excludes a minor then there is a constant d such that any $G \in \mathcal{C}$ can be 2 coloured so that any colour class induces a graph of tree-width at most d. See also [**7, 8**] for generalisations and algorithmic versions of this result and various applications.

The technique was later generalised by Nešetřil and Ossona de Mendez to graph classes of bounded expansion [**38**] and to nowhere dense classes of graphs [**41**]. In this section we will show how this can be used to establish tractability results for first-order model-checking on graph classes of bounded expansion.

To formally define classes of bounded expansion we first need some preparation. Recall that a graph H is a *minor* of G if it can be obtained from a sub-graph $G' \subseteq G$ by contracting edges. An equivalent, sometimes more intuitive, characterisation of the minor relation can be obtained using the concept of *images.* An *image map* of H into G is a map μ mapping each $v \in V(H)$ to a tree $\mu(v) \subseteq G$ and each edge $e \in E(H)$ to an edge $\mu(e) \in E(G)$ such that if $u \neq v \in V(H)$ then $\mu(v) \cap \mu(u) = \varnothing$ and if $uv \in E(H)$ then $\mu(uv) = u'v'$ for some $u' \in V(\mu(u))$ and $v' \in V(\mu(v))$. The union $\bigcup_{v \in V(H)} \mu(v) \cup \bigcup_{e \in E(H)} \mu(e) \subseteq G$ is called the *image* of H in G. It is not difficult to see that $H \preccurlyeq G$ if, and only if, there is an image of H in G.

The *radius* of a graph is G is the least r such that there is a vertex $v \in V(G)$ with $G = N_r^G(v)$. For $r \geq 0$, a graph H is an *minor at depth* r of a graph G, denoted $H \preccurlyeq_r G$, if H has an image map μ in G where for all $v \in V(H)$, $\mu(v)$ is a tree of radius at most r.

Definition 8.1 (bounded expansion). *Let G be a graph. The* greatest reduced average density *of G with rank r is*

$$\nabla_r(G) := \max\Big\{\frac{|E(H)|}{|V(H)|} : H \preccurlyeq_r G\Big\}.$$

A class $\mathcal{D}$ of graphs has bounded expansion *if there is a computable*[3] *function $f : \mathbb{N} \to \mathbb{N}$ such that $\nabla_r(G) \leq f(r)$ for all $G \in \mathcal{D}$ and $r \geq 0$. Finally, a class $\mathcal{C}$ of σ-structures has* bounded expansion *if $\{G(A) : A \in \mathcal{C}\}$ has bounded expansion, where $G(A)$ denotes the Gaifman-graph of the structure A (see Section 2).*

As every graph of average degree at least $c \cdot k\sqrt{\log k}$, for some constant c, contains a k-clique as a minor [**28, 29, 47**], it follows that every class of graphs excluding a minor also has bounded expansion.

The next definition formally defines the concept of colourings such that any constant number of colour classes together induce a sub-graph of small tree-depth.

Definition 8.2. *Let σ be a signature. Let $\mathcal{C}$ be a class of σ-structures of bounded expansion and let $A \in \mathcal{C}$.*

(1) Let $\gamma : V(A) \to \Gamma$ be a vertex colouring of A. If $\overline{C} \in \Gamma^s$ is a tuple of colours, we write $A_{\overline{C}}$ for the sub-structure of A induced by the union $\{v \in V(A) : \gamma(v) \in \overline{C}\}$ of the colour classes in $\overline{C}$.
(2) For $k \geq 0$, a vertex-colouring $\gamma : V(A) \to \Gamma$ of A is a td-k-colouring *if $A_{\overline{C}}$ has tree-depth at most k, for all $\overline{C} \in \Gamma^k$.*

It was shown in [**38, 39**] that for graph classes of bounded expansion, td-k-colourings using a constant number of colours exist and can be computed efficiently.

Theorem 8.3 ([**38, 39**]). *If $\mathcal{C}$ is a class of σ-structures of bounded expansion then there are computable functions $f, N_{\mathcal{C}} : \mathbb{N} \to \mathbb{N}$ and an algorithm which, given $A \in \mathcal{C}$ and k, computes a td-k-colouring of A with at most $N_{\mathcal{C}}(k)$ colours in time $f(k) \cdot |A|$.*

To demonstrate the application of td-k-colourings for model-checking, we prove the following result that will be used later.

Theorem 8.4 ([**42**]). *Let σ be a signature and let $\mathcal{C}$ be a class of σ-structures of bounded expansion. There is a computable function $f : \mathbb{N} \to \mathbb{N}$ such that given an existential first-order formula $\varphi \in \mathrm{FO}[\sigma]$ and a σ-structure $A \in \mathcal{C}$, $A \models \varphi$ can be decided in time $f(|\varphi|) \cdot |A|$.*

Proof. W.l.o.g. we assume that φ is in prenex normal form, i.e. of the form

$$\varphi := \exists x_1 \dots \exists x_q \vartheta,$$

with ϑ quantifier-free. Let q be the number of quantifiers in φ.

Using Theorem 8.3 we first compute a td-q-colouring $\gamma : V(A) \to \Gamma$ of A, where Γ is a set of $N_{\mathcal{C}}(q)$ colours, in time $f_1(q) \cdot |A|$, for some computable function f_1.

Clearly, $A \models \varphi$ if, and only if, there are vertices $a_1, \dots, a_q \in V(A)$ such that $A \models \vartheta[\overline{a}]$ and therefore $A_{\gamma(a_1),\dots,\gamma(a_q)} \models \varphi$. Hence, $A \models \varphi$ if, and only if, there is a tuple $\overline{C} \in \Gamma^q$ such that $A_{\overline{C}} \models \varphi$. Therefore, to check whether $A \models \varphi$ we go through all tuples $\overline{C} \in \Gamma^q$ and decide whether $A_{\overline{C}} \models \varphi$. As all $A_{\overline{C}}$ have tree-depth

[3] The original definition in [**38**] does not require f to be computable. This would imply that some of the following fpt-algorithms are non-uniform.

at most q, by Corollary 5.9, this check can be performed in time $f_2(q) \cdot |A|$, where f_2 is a computable function.

Hence, the complete algorithm runs in time $\big(f_1(|\varphi|) + N_{\mathcal{C}}(|\varphi|)^q \cdot f_2(|\varphi|)\big) \cdot |A|$. □

We are now ready to state the main result of this section. Our presentation follows [**33**].

THEOREM 8.5 ([**13**]). *Let σ be a relational signature and let $\mathcal{C}$ be a class of σ-structures of bounded expansion. Then* MC(FO, $\mathcal{C}$) *is fpt by linear time parameterized algorithms.*

We first give a high-level description of the proof. Given a structure $A \in \mathcal{C}$ and a formula φ with at most q quantifiers, we will define an equivalence relation of finite index on q-tuples of elements such that if $\overline{a}, \overline{b}$ fall into the same equivalence class then they satisfy the same formulas of quantifier-rank at most q. The equivalence class of a tuple $\overline{a}$ is called its *full type*. Suppose that for each tuple $\overline{b}$ of length at most q we can compute its equivalence class. Then, to decide whether $A \models \varphi$, we can use the naive evaluation algorithm, i.e. for each quantifier we test all possibilities. However, as equivalent tuples satisfy the same formulas, we only need to check one witness for each equivalence class and therefore we can implement the evaluation algorithm in constant time, depending only on the size of the formula. For this to work we need to compute the full types by a linear time fpt algorithm.

We proceed in stages. As a first step we define for each k-tuple $\overline{C}$ of colours the *local type of quantifier-rank q* of a tuple $\overline{a}$ of elements in $A_{\overline{C}}$. If two tuples have the same local type in some $A_{\overline{C}}$ then they satisfy the same first-order formulas in $A_{\overline{C}}$ up to quantifier-rank q.

The second step is the definition of the *global type* of a tuple $\overline{a}$, which is simply the collection of all local types of $\overline{a}$ in the individual sub-graphs $A_{\overline{C}}$, for all $\overline{C}$ of length at most k (see Definition 8.13). We will show that global types can be defined by existential first-order formulas.

Finally, we will use the global types as the basis for the definition of full types. A full type of a tuple $\overline{a}$ describes the complete quantifier-rank q first-order type of $\overline{a}$ and therefore determines which formulas of quantifier-rank at most q are true at $\overline{a}$. The main difficulty is to decide which full types are realised in the structure. For this we will show that each full type can be described by an existential first-order formula.

The existential formulas describing full types in a structure A will not be over the structure A itself, but over an expansion of A by the edges of tree-depth decompositions. We first introduce these expansions and then define the various types we are using.

Recall (from Section 6) that the first-order q-type $\mathrm{tp}^{A}_{\mathrm{FO},q}(a)$ of an element $a \in V(A)$ in a structure A is the class of all first-order formulas $\varphi(x) \in \mathrm{FO}[\sigma]$ of quantifier-rank at most q such that $A \models \varphi[v]$. As we are only dealing with first-order logic in this section, we drop the index FO from now on.

Notation. For the rest of this section we fix a signature σ and a class $\mathcal{C}$ of σ-structures of bounded expansion. Let $k, q \geq 0$ and let $c := (3k + q + 4) \cdot 2^{k+q+1}$. Let $A \in \mathcal{C}$ be a structure and $\gamma : V(A) \to \Gamma$ be a td-$(k + q)$-colouring of A, where Γ is a set of $N_{\mathcal{C}}(k + q)$ colours. As before, for each tuple $\overline{C} \in \Gamma^{k+q}$, let $A_{\overline{C}}$ be the sub-structure of A induced by the elements $\{v \in V(A) : \gamma(v) \in \overline{C}\}$. For each $\overline{C}$ we fix a depth-first search (DFS) forest $F_{\overline{C}}$ of $G(A_{\overline{C}})$ and add a self-loop to every root of a tree in $F_{\overline{C}}$.

Finally, we agree that for the rest of this section all formulas are "normalised" (see beginning of Section 6). In particular, this implies that we can test effectively whether a formula belongs to a given type. To increase readability, the formulas stated explicity in this section will not be normalised. However, they can easily be brought into normalised form as the normalisation process for first-order formulas is effective. □

Recall that, by Lemma 5.8, the closure of a DFS-forest $F_{\overline{C}}$ is a tree-depth decomposition of $A[\overline{C}]$ and as $A_{\overline{C}}$ has tree-depth at most $k+q$ the lemma implies that the height of $F_{\overline{C}}$ is at most 2^{k+q}.

Definition 8.6. *(1) For $\overline{C} \in \Gamma^{k+q}$ we define $(A_{\overline{C}}, F_{\overline{C}})$ as the $\sigma\dot\cup\{F_{\overline{C}}\}$-expansion of $A_{\overline{C}}$ with $F_{\overline{C}}((A_{\overline{C}}, F_{\overline{C}})) := E(F_{\overline{C}})$.*

(2) We define $\tau(\Gamma, \sigma, k+q) := \sigma \dot\cup \{F_{\overline{C}}, T_{\overline{C},t} : \overline{C} \in \Gamma^{k+q}$ and $t(x)$ is a finite set of formulas $\varphi(x) \in \mathrm{FO}[\sigma\dot\cup\{F_{\overline{C}}\}]$ of quantifier-rank at most $c\}$.

(3) The $\tau(\Gamma, \sigma, k+q)$-structure $A(\gamma)$ is defined as the $\tau(\Gamma, \sigma, k+q)$-expansion of A with $F_{\overline{C}}(A(\gamma)) := E(F_{\overline{C}})$, and

$$T_{\overline{C},t}(A(\gamma)) := \{v \in V(A) : t = \mathrm{tp}_c^{(A_{\overline{C}}, F_{\overline{C}})}(v)\}.$$

Note that $A(\gamma)$ depends on the particular choice of $F_{\overline{C}}$ and is therefore not unique. But the precise choice will never matter and all results remain true independent of a particular choice of DFS-forest.

Essentially, to obtain $A(\gamma)$ we fix a tree-depth decomposition for each sub-structure induced by $k+q$ colours and add the edges of the decomposition to A, giving them a different edge colour $F_{\overline{C}}$ for any tuple $\overline{C} \in \Gamma^{k+q}$. Furthermore, for each $v \in V(A)$ and each sub-structure $(A_{\overline{C}}, F_{\overline{C}})$ induced by $k+q$ colours $\overline{C}$ which contains v we label v by its q-type in $(A_{\overline{C}}, F_{\overline{C}})$. The reason we work with DFS-forests rather than general tree-depth decompositions is that we can add the edges of a DFS-forest to A without introducing new edges in the Gaifman-graph. Hence, if $\mathcal{C}$ is a class of σ-structures of bounded expansion then the class $\{A(\gamma) : A \in \mathcal{C}$ and γ a td-$(k+q)$-colouring of $A\}$ also has bounded expansion for all choices of γ and DFS-forests.

8.1. Local types. We show next that the formulas true at a given tuple $\overline{a}$ in a sub-structure $(A_{\overline{C}}, F_{\overline{C}})$ only depend on the formulas true at each individual element a_i and the relative position of the a_i within the tree-depth decomposition. By adding the edges of the tree-depth decomposition to the structure $A(\gamma)$, this relative position becomes first-order definable in $A(\gamma)$, a fact that will be used later in our model-checking algorithm.

Definition 8.7. *Let $\overline{C} \in \Gamma^{k+q}$ be a tuple of colours and let $x, y \in V(A_{\overline{C}})$ be two vertices contained in the same tree in $F_{\overline{C}}$.*

- *The* least common ancestor $\mathrm{lca}_{\overline{C}}(x,y)$ *of x and y in $F_{\overline{C}}$ is the element of $F_{\overline{C}}$ of maximal height that is an ancestor of both x and y.*
- *We define $\mathrm{lch}_{\overline{C}}(x,y)$ to be the height of $\mathrm{lca}_{\overline{C}}(x,y)$ in $F_{\overline{C}}$ and define $\mathrm{lch}_{\overline{C}}(x,y) := \infty$ if x and y are not in the same component of $F_{\overline{C}}$.*

The following simple lemma shows that $\mathrm{lch}_{\overline{C}}$ and $\mathrm{lca}_{\overline{C}}$ are first-order definable in $(A_{\overline{C}}, F_{\overline{C}})$ for all $\overline{C} \in \Gamma^{k+q}$.

LEMMA 8.8. *For all $r \leq 2^{k+q}$ there is a first-order formula $\mathrm{lch}_r^{\overline{C}}(x, y) \in \mathrm{FO}[F_{\overline{C}}]$ of quantifier-rank at most $2^{k+q} + 1$ such that for all $a, b \in V(A_{\overline{C}})$ we have*

$$A(\gamma) \models \mathrm{lch}_r^{\overline{C}}(a, b) \iff (A_{\overline{C}}, F_{\overline{C}}) \models \mathrm{lch}_r^{\overline{C}}(a, b) \iff \mathrm{lch}_{\overline{C}}(a, b) = r$$

The next lemma says that truth of a formula $\varphi(\overline{x})$ of quantifier-rank at most q at a tuple $\overline{a} := (a_1, \ldots, a_k)$ only depends on the relative position of the a_i and the formulas of quantifier-rank at most $(k+q)2^{k+q+1}$ true at each a_i.

LEMMA 8.9. *Let $\overline{C} \in \Gamma^{k+q}$ and let $\varphi(x_1, \ldots, x_k) \in \mathrm{FO}[\sigma \cup \{F_{\overline{C}}\}]$ be a formula of quantifier-rank at most q.*

If $u_1, \ldots, u_k, v_1, \ldots, v_k \in V(A_{\overline{C}})$ are such that for all $1 \leq i \leq k$ and $1 \leq i \leq j \leq 2^{k+q}$,

$$\mathrm{tp}_{(k+q)2^{k+q+1}}^{(A_{\overline{C}}, F_{\overline{C}})}(v_j) = \mathrm{tp}_{(k+q)2^{k+q+1}}^{(A_{\overline{C}}, F_{\overline{C}})}(u_j) \text{ and } \mathrm{lch}(v_i, v_j) = \mathrm{lch}(u_i, u_j)$$

then $(A_{\overline{C}}, F_{\overline{C}}) \models \varphi(v_1, \ldots, v_k)$ if, and only if, $(A_{\overline{C}}, F_{\overline{C}}) \models \varphi(u_1, \ldots, u_k)$.

We are now ready to define the first equivalence relation on tuples of vertices, the *local type* of a tuple.

DEFINITION 8.10 (Local Types). *(1) For all $\overline{C} \in \Gamma^{k+q}$ we define the set $\mathrm{Loc}(\overline{C}, \sigma, k, q)$ of* local types *as the set of all tuples*

$$\big(t_1, \ldots, t_k, (r_{i,j})_{1 \leq i < j \leq k}\big),$$

where t_i is a finite set of formulas $\varphi(x) \in \mathrm{FO}[\sigma \dot\cup \{F_{\overline{C}}\}]$ of quantifier-rank at most c and $r_{i,j} \leq 2^{k+q}$ for all i, j.

(2) For $\overline{C} \in \Gamma^{k+q}$ and $\overline{a} := a_1, \ldots, a_k \in V(A_{\overline{C}})$ we define the local type *$loc_q(\overline{a}; \overline{C}) \in \mathrm{Loc}(\overline{C}, k, q)$ as*

$$\big(\mathrm{tp}_c^{(A_{\overline{C}}, F_{\overline{C}})}(a_1), \ldots, \mathrm{tp}_c^{(A_{\overline{C}}, F_{\overline{C}})}(a_k), (\mathrm{lch}_{\overline{C}}(a_i, a_j))_{1 \leq i < j \leq k}\big).$$

We will prove next that the local type $loc_q(\overline{a}; \overline{C}) := (t_1, \ldots, t_k, (r_{i,j})_{1 \leq i < j \leq k})$ of a tuple $\overline{a}$ of vertices completely describes the formulas of quantifier-depth at most q which are true at $\overline{a}$ within the sub-structure $(A_{\overline{C}}, F_{\overline{C}})$. Note that we require the t_i to be quantifier-rank c-types of a_i even though we are only interested in formulas $\varphi(x_1, \ldots, x_k)$ of quantifier-rank q. The reason for this will become clear in the following lemma.

LEMMA 8.11. *Let $l := (t_1, \ldots, t_k, (r_{i,j})_{1 \leq i < j \leq k}) \in \mathrm{Loc}(\overline{C}, \sigma, k, q)$ be a local type. Then for all formulas $\varphi(\overline{x})$ with quantifier-rank at most q and all k-tuples $\overline{a} \in V(G)^k$ with $loc_q(\overline{a}, \overline{C}) = l$, $(A_{\overline{C}}, F_{\overline{C}}) \models \varphi[\overline{a}]$ if, and only if, if t_1 contains the formula*

$$\varphi^*(x_1) := \exists x_2 \ldots \exists x_k \begin{array}{l} \bigwedge_{1 \leq i < j \leq k} \mathrm{lch}_{r_{i,j}}^{\overline{C}}(x_i, x_j) \wedge \\ \bigwedge_{1 \leq i \leq k} \bigvee_{t \in \mathrm{Loc}(\overline{C}, \sigma, k, q) : t_i \cap \mathcal{T}_{(k+q) \cdot 2^{k+q+1}} \subseteq t} T_{\overline{C}, t}(x_i) \wedge \\ \varphi(x_1, \ldots, x_k), \end{array}$$

where $\mathcal{T}_{(k+q) \cdot 2^{k+q+1}}$ is the finite set of all formulas in $\mathrm{FO}[\sigma \dot\cup \{F_{\overline{C}}\}]$ of quantifier-rank at most $(k+q) \cdot 2^{k+q+1}$.

Proof. Recall that the height of $F_{\overline{C}}$ is at most 2^{k+q}. Hence, by Lemma 8.8, the quantifier-rank of the formula φ^* is at most c.

Suppose $(A_{\overline{C}}, F_{\overline{C}}) \models \varphi[\overline{a}]$. Choosing $a_1, \ldots, a_k$ as witnesses for $x_1, \ldots, x_k$ it is obvious that $(A_{\overline{C}}, F_{\overline{C}}) \models \Big(\bigwedge_{1\leq i<j\leq k} \mathrm{lch}_{r_{i,j}}(x_i, x_j) \wedge \bigwedge_{1\leq i\leq k} T_{\overline{C},t_i}(x_i)\Big)[\overline{a}]$. Hence, $\varphi^*(x_1)$ is contained in t_1.

Conversely, suppose that $\varphi^*(x_1)$ is contained in t_1 and hence $(A_{\overline{C}}, F_{\overline{C}}) \models \varphi^*[a_1]$. Hence, there are $b_2, \ldots, b_k \in V(A_{\overline{C}})$ such that $\mathrm{lch}_{\overline{C}}(b_i, b_j) = r_{i,j}$, for all $1 \leq i < j \leq k$, where we set $b_1 := a_1$ to simplify notation, and further $\mathrm{tp}^{(A_{\overline{C}}, F_{\overline{C}})}_{(k+q)\cdot 2^{k+q+1}}(b_i) = \mathrm{tp}^{(A_{\overline{C}}, F_{\overline{C}})}_{(k+q)\cdot 2^{k+q+1}}(a_i)$, for all $1 \leq i \leq k$. Hence, by Lemma 8.9, $\overline{a}$ and $\overline{b}$ satisfy the same formulas of quantifier-rank at most q in $(A_{\overline{C}}, F_{\overline{C}})$ and therefore $(A_{\overline{C}}, F_{\overline{C}}) \models \varphi[\overline{a}]$. □

Recall that we are only working with normalised formulas in this section. However, the normalisation process for first-order formulas is effective and hence a normalised version of the formula φ^* can be computed effectively from the formula φ. Hence, the lemma implies that whether a tuple $\overline{a}$ with local type l satisfies a formula $\varphi(\overline{x})$ within some $(A_{\overline{C}}, F_{\overline{C}})$ can be read off directly from the local type l independent of the actual tuple $\overline{a}$. This motivates the following definition.

DEFINITION 8.12. *A local type l defined as $l := (t_1, \ldots, t_k, (r_{i,j})_{1\leq i<j\leq k}) \in$* $\mathrm{Loc}(\overline{C}, \sigma, k, q)$ entails *a formula $\varphi(x_1, \ldots, x_k)$ of quantifier-rank at most q, denoted $l \models \varphi$, if t_1 contains the formula $\varphi^*(x_1)$ defined in Lemma 8.11.*

8.2. Global types. As the second step towards defining the full type of a tuple $\overline{a}$ we now define the *global type* of $\overline{a}$, which is the collection of their local types over all combinations of colours.

DEFINITION 8.13.

1. *We define* $\mathrm{Glob}(\Gamma, \sigma, k, q) := \{(l_{\overline{C}})_{\overline{C}\in\Gamma^{k+q}} : l_{\overline{C}} \in \mathrm{Loc}(\overline{C}, \sigma, k, q)\}$.
2. *For $\overline{a} \in V(G)$ we define the* global type *of $\overline{a}$ as*

$$glob_q(\overline{a}, \Gamma) := \big(loc_q(\overline{a}, \overline{C})\big)_{\overline{C}\in\Gamma^{k+q}} \in \mathrm{Glob}(\Gamma, \sigma, k, q).$$

We now extend Lemma 8.11 to tuples having the same global type in G. However, this only applies to existential formulas and can be shown to be false for formulas with quantifier alternation.

LEMMA 8.14. *If $\overline{a} := a_1, \ldots, a_k, \overline{b} := b_1, \ldots, b_k \in V(G)$ are tuples such that $glob_q(\overline{a}) = glob_q(\overline{b})$, then $\overline{a}$ and $\overline{b}$ satisfy in A the same existential formulas $\varphi \in$* $\mathrm{FO}[\sigma]$ *with at most q quantifiers.*

More precisely, $A \models \varphi[\overline{a}]$ if, and only if, $glob_q(\overline{a})$ contains a local type l which entails φ.

Proof. Let $\varphi(x_1, \ldots, x_k) \in \mathrm{FO}[\sigma]$ be an existential first-order formula with at most q quantifiers. W.l.o.g. we assume that φ is in prenex normal form, i.e. $\varphi := \exists z_1 \ldots z_q \vartheta(\overline{x}, \overline{z})$, where ϑ is quantifier-free.

Suppose $A \models \varphi[\overline{a}]$. Let $u_1, \ldots, u_q$ be witnesses for the existential quantifiers in φ, i.e. $A \models \vartheta[\overline{a}, \overline{u}]$, and let $\overline{C} := (\gamma(a_1), \ldots, \gamma(a_k), \gamma(u_1), \ldots, \gamma(u_q))$. Then, $A_{\overline{C}} \models \varphi[\overline{a}]$ and therefore $loc_q(\overline{a}, \overline{C})$ entails φ. As $\overline{b}$ has the same global type as $\overline{a}$ it also has the same local types, i.e. $loc_q(\overline{b}, \overline{C}) = loc_q[\overline{a}, \overline{C}]$ and therefore $A_{\overline{C}} \models \varphi(\overline{b})$. As the argument is symmetric, this concludes the proof. □

Again we define entailment between types and formulas.

DEFINITION 8.15. *Let* $l := \big(l_{\overline{C}}\big)_{\overline{C}\in\Gamma^{k+q}} \in \mathrm{Glob}(\Gamma,\sigma,k,q)$ *and let* $\varphi(\overline{x}) \in \mathrm{FO}[\sigma]$ *be an existential formula with at most* q *quantifiers and* k *free variables* $x_1,\dots,x_k$. *The type* l entails φ, *denoted* $l \models \varphi$, *if there is* $\overline{C} \in \Gamma^{k+q}$ *such that* $l_{\overline{C}}$ *entails* φ.

LEMMA 8.16. *For each* $l \in \mathrm{Glob}(\Gamma,\sigma,k,q)$ *there is an existential first-order formula* $\varphi_l(\overline{x})$ *such that for all* $\overline{a} \in V(A)^k$,

$$glob_q(\overline{a},\Gamma) = l \quad \text{if, and only if,} \quad A(\gamma) \models \varphi_l[\overline{a}].$$

Furthermore, the formula depends only on Γ,σ,k *and* q *but not on a specific colouring or structure.*

Proof. Suppose $l := (l_{\overline{C}})_{\overline{C}\in\Gamma^{k+q}}$, where $l_{\overline{C}} := (t_1,\dots,t_k,(r_{i,j})_{1\le i<j\le k})$ are local types. For each $l_{\overline{C}}$ define

$$\varphi_{l_{\overline{C}}}(\overline{x}) := \bigwedge_{i=1}^{k} x_i \in P_{\overline{C},t_i} \wedge \bigwedge_{1\le i<j\le k} \mathrm{lch}^{\overline{C}}_{r_{i,j}}(x_i,x_j).$$

Then, $A(\gamma) \models \varphi_{l_{\overline{C}}}[\overline{a}]$ if, and only if, $loc_q(\overline{a};\overline{C}) = l_{\overline{C}}$. Hence.

$$\varphi_l(\overline{x}) := \bigwedge_{\overline{C}\in\Gamma^{k+q}} \varphi_{l_{\overline{C}}}(\overline{x})$$

says that the global type of $\overline{x}$ is l. ☐

8.3. Full Types. Finally, we give the definition of full types, the main equivalence relation between tuples used in our algorithm. We will define the full type $\mathrm{ft}_i^q(\overline{a})$ of an i-tuple $\overline{a} \in V(A)^i$ such that if $\overline{a}$ and $\overline{b}$ have the same full type they satisfy the same formulas of quantifier-rank at most $q-i$.

DEFINITION 8.17. *For* $0 \le i \le q$ *we define the set* $\mathfrak{F}_i^q$ *of full types of* i*-tuples and the full type* $\mathrm{ft}_i^q(\overline{a})$ *of* $\overline{a} := a_1\dots a_i \in V(A)^i$ *inductively as follows.*

(1) *For* $i = q$ *we set* $\mathfrak{F}_q^q := \mathrm{Glob}_{\mathcal{C}}(\Gamma,\sigma,q,0)$ *and for* $a_1,\dots,a_q \in V(A)$ *we define* $\mathrm{ft}_q^q(\overline{a}) := glob_0(\overline{a},\Gamma)$.

(2) *For* $i < q$ *we define* $\mathfrak{F}_i^q := \{\Phi : \Phi \subseteq \mathfrak{F}_{i+1}^q\}$ *and for* $\overline{a} := a_1,\dots,a_i$

$$\mathrm{ft}_i^q(\overline{a}) := \{\mathrm{ft}_{i+1}^q(\overline{a},a_{i+1}) : a_{i+1} \in V(A)\}.$$

A full type $t \in \mathfrak{F}_i^q$ *is* realised in A and γ *if there is* $\overline{a} \in V(A)^i$ *such that* $t = \mathrm{ft}_i^q(\overline{a})$. *We define* $\mathfrak{R}_i^q(A,\gamma) \subseteq \mathfrak{F}_i^q$ *as the set of types realised in* A *and* γ.

Note that the cardinality of $\mathfrak{F}_i^q$ only depends on q and $\mathcal{C}$. A straight forward Ehrenfeucht-Fraïssé-game argument establishes the following lemma.

LEMMA 8.18. *If* $\overline{a},\overline{b} \in V(G)^i$ *are such that* $\mathrm{ft}_i^q(\overline{a}) = \mathrm{ft}_i^q(\overline{b})$ *then* $\overline{a}$ *and* $\overline{b}$ *satisfy the same formulas of quantifier-rank at most* $q-i$ *in* A.

We show next that the full type of a tuple can be described by an existential first-order formula. As a consequence, we can check in linear time whether a full type is realised. For this, we first establish two lemmas which show that we can express Boolean combinations of existential formulas in a structure $A \in \mathcal{C}$ by an existential formula. However, this formula will not be over A but over an expansion $A(\gamma)$ for a suitable td-l-colouring $\gamma : V(A) \to \Gamma$, for some l. The next lemmas therefore no longer refer to the structure $A \in \mathcal{C}$ and colouring γ fixed at the beginning and we therefore state them in full generality.

LEMMA 8.19. *Let $k, q \geq 0$. Let $\mathcal{D}$ be a class of σ'-structures of bounded expansion and let $\varphi(\overline{x}) \in \mathrm{FO}[\sigma']$ be an existential formula with q quantifiers and k free variables $\overline{x} := x_1, \dots, x_k$. Let Γ be a set of $N_{\mathcal{D}}(k+q)$ colours.*

There is an existential formula $\overline{\varphi}(\overline{x}) \in \mathrm{FO}[\tau(\Gamma, \sigma', q+k)]$ such that for all $A \in \mathcal{D}$ and all td-$(q+k)$-colourings $\gamma : V(A) \to \Gamma$ and all $\overline{a} \in V(A)^k$

$$A \not\models \varphi[\overline{a}] \quad \textit{if, and only if,} \quad A(\gamma) \models \overline{\varphi}[\overline{a}].$$

Proof. W.l.o.g. we can assume that φ is in prenex normal form, i.e. of the form $\varphi := \exists \overline{y} \vartheta$, where ϑ is quantifier-free.

By Lemma 8.14, $A \models \varphi[\overline{a}]$ if, and only if, $glob_q(\overline{a}, \Gamma) \models \varphi$. It follows that $\overline{a}$ does not satisfy φ in A if $glob_q(\overline{a}, \Gamma) \not\models \varphi$.

As, by Lemma 8.16, global types l can be expressed by existential formulas φ_l, we can express that $A \not\models \varphi(\overline{a})$ by the existential $\mathrm{FO}[\tau(\Gamma, \sigma', q+k)]$-formula

$$\overline{\varphi}(\overline{x}) := \bigvee_{l \in \mathrm{Glob}(\Gamma, \sigma', k, q), l \not\models \varphi(\overline{x})} \varphi_l(\overline{x}).$$

□

LEMMA 8.20. *Let $k, q \geq 0$. Let $\mathcal{D}$ be a class of σ'-structures of bounded expansion and let $\varphi_1(\overline{x}), \dots, \varphi_n(\overline{x}) \in \mathrm{FO}[\sigma']$ be existential formulas with k free variables each.*

Then there is a $q \geq 0$ and a set Γ of $N_{\mathcal{C}}(k+q)$ colours and for each $I \subseteq \{1, \dots, n\}$ an existential formula $\varphi_I \in \mathrm{FO}[\tau(\Gamma, \sigma', k+q)]]$ such that for all $A \in \mathcal{D}$ and all td-$(q+k)$-colourings $\gamma : V(A) \to \Gamma$ and all $\overline{a} \in V(A)^k$,

$$A \models \psi_I[\overline{a}] \quad \textit{if, and only if, } A(\gamma) \models \varphi_I[\overline{a}],$$

where $\psi_I := \bigwedge_{i \in I} \exists x \varphi_i \wedge \bigwedge_{i \notin I} \neg \exists x \varphi_i(x)$.

Proof. To define an existential formula φ_I equivalent to ψ_I we have to replace the $\neg \exists \varphi_i(x)$ parts by existential statements. Let q' be the maximum number of quantifiers in any φ_i, $1 \leq i \leq n$ and let $q := q' + 1$. Let Γ be a set of $N_{\mathcal{C}}(k+q)$ colours.

As $\exists x \varphi_i$ is an existential formula, Lemma 8.19 implies that there is an existential $\mathrm{FO}[\tau(\Gamma, \sigma', k+q)]$-formula $\overline{\varphi}$ such that for all td-$(k+q)$-colourings $\gamma : V(A) \to \Gamma$, $A \not\models \exists x \varphi_i(\overline{a})$ if, and only if, $A(\gamma) \models \overline{\varphi}_i(\overline{a})$.

Hence, for all $I \subseteq \{1, \dots, n\}$

$$A \models \psi(\overline{a}) \quad \text{if, and only if,} \quad A(\gamma) \models \bigwedge_{i \in I} \exists x \varphi_i \wedge \bigwedge_{i \notin I} \overline{\varphi}_i$$

□

The previous two lemmas immediately imply the following.

LEMMA 8.21. *Let $\mathcal{D}$ be a class of σ'-structures of bounded expansion. Let $q \geq 0$. Let $A \in \mathcal{D}$ and $\gamma : V(A) \to \Gamma$ be a td-q-colouring.*

There is $r := r(q, \sigma', \mathcal{D}) \in \mathbb{N}$ such that for all $1 \leq i \leq q$ and all $l \in \mathfrak{F}_i^q$ there are existential first-order formulas $\varphi_l(x_1, \dots, x_i), \varphi_l(x_1, \dots, x_i), \varphi_l^e, \varphi_l^{\neg e} \in \mathrm{FO}[\tau(\Gamma', \sigma, r)]$, where Γ' is a set of $N_{\mathcal{C}}(r)$ colours disjoint from Γ, such that for every $A \in \mathcal{D}$,

$\overline{a} \in V(A)^i$ and td-r-colouring $\gamma' : V(A) \to \Gamma'$

$$\begin{array}{rll} A(\gamma') \models \varphi_l[\overline{a}] & \textit{if, and only if,} & \mathrm{ft}_i^q(\overline{a}) = l. \\ A(\gamma') \models \varphi_l^e & \textit{if, and only if,} & \textit{the type } l \textit{ is realised in } A(\gamma) \\ A(\gamma') \models \varphi_l^\neg[\overline{a}] & \textit{if, and only if,} & \mathrm{ft}_i^q(\overline{a}) \neq l. \\ A(\gamma') \models \varphi_l^{\neg e} & \textit{if, and only if,} & \textit{the type } l \textit{ is not realised in } A(\gamma). \end{array}$$

Proof. For $l \in \mathfrak{F}_q^q$ the existence of φ_l was proved in Lemma 8.16. Then $\varphi_l^e := \exists\overline{x}\varphi_l$. Furthermore, $\varphi_l^\neg$ and $\varphi_l^{\neg e}$ can be obtained from φ_l, φ_l^e by Lemma 8.19.

For the induction step, let $l \in \mathfrak{F}_i^q$ for some $i < q$. Then $l \subseteq \mathfrak{F}_{i+1}^q$ is a set of types $t \in \mathfrak{F}_{i+1}^q$ which, by induction hypothesis, can all be defined by existential formulas. Hence, $\varphi_l' := \bigwedge_{t\in l} \varphi_t \wedge \bigwedge_{t\notin l} \neg\varphi_t$ defines l and, by Lemma 8.20, can equivalently be written as an existential formula. $\varphi_l^e, \varphi_l^\neg, \varphi_l^{\neg e}$ can be defined as before.

Note that each step increases the signature and number of colours so that we finally obtain r and τ as required. □

As a consequence of the previous lemma we get that the set of types realised in a given structure can be computed in parameterized-linear time.

Corollary 8.22. *Let $\mathcal{C}$ be a class of bounded expansion. There is a computable function $f : \mathbb{N} \to \mathbb{N}$ such that on input $A \in \mathcal{C}$, $q \geq 0$ and a td-q-colouring $\gamma : V(A) \to \Gamma$ the set $\mathfrak{R}_i^q$ can be computed in time $f(q) \cdot |G|$, for all $1 \leq i \leq q$.*

8.4. Model-checking in classes of bounded expansion. We are now going to describe our model-checking algorithm for classes of bounded expansion. Let $\mathcal{C}$ be a class of σ-structures of bounded expansion and $A \in \mathcal{C}$. Let $\varphi \in \mathrm{FO}[\sigma]$ be a formula with at most q quantifiers. W.l.o.g. we assume that φ is in prenex normal form and of the form $\varphi := \exists x_1 Q_2 x_2 \dots Q_q x_q \vartheta(x_1, \dots, x_q)$ with ϑ quantifier free and $Q_i \in \{\exists, \forall\}$. For $i \geq 1$ we define $\varphi_i(x_1, \dots, x_i) := Q_{i+1}x_{i+1} \dots Q_q x_q \vartheta$.

We can now check whether $A \models \varphi$ as follows. First, we compute a td-q-colouring $\gamma : V(A) \to \Gamma$ of A. By Corollary 8.22 there is a computable function $f : \mathbb{N} \to \mathbb{N}$ such that the sets $\mathfrak{R}_i^q$, for $i \leq q$, can be computed in time $f(q) \cdot |A|$.

For each $t \in \mathfrak{R}_0^q$ we can now simply test whether $t \models \varphi$ and return true if such a type exists.

It is easily seen that the algorithm is correct. Furthermore, its running time only depends on the size of φ and the size of $\bigcup_{1\leq i\leq q} \mathfrak{R}_{q-i}^i$ which again depends only on φ and σ.

Hence, by Corollary 8.22 and Theorem 8.3 the algorithm runs in time $f(|\varphi|)\cdot|G|$, for some computable function $f : \mathbb{N} \to \mathbb{N}$. This concludes the proof of Theorem 8.5.

Part II: Lower Bounds

In the previous part we have presented a range of tools for establishing tractability results for logics on specific classes of graphs. In this section we consider the natural counterparts to these results, namely lower bounds establishing limits beyond which the tractability results cannot be extended. Ideally, we aim for logics L such as FO, MSO_1 or MSO_2 for a structural property P_{L} such that model-checking for L is tractable on a class of structures if, and only if, it has the property P_{L}. As the general model-checking problem for $\mathrm{FO}, \mathrm{MSO}_1, \mathrm{MSO}_2$ is PSPACE-complete, any proof that model-checking for any of the logics is not fpt on a class $\mathcal{C}$ would separate

PTIME from PSPACE. We can therefore only hope to find such a property subject to assumptions from complexity theory and possibly subject to further restrictions.

In this section we first review recent lower bounds for monadic second-order logic with edge set quantification (MSO_2) and then comment briefly on lower bounds for FO and MSO_1.

9. Lower Bounds for MSO with Edge Set Quantification

In this section we review the known lower bounds for monadic second-order logic MSO_2. To make the results as strong as possible, we will concentrate on simple undirected graphs.

Recall that by Courcelle's theorem (Theorem 6.2), MSO_2 model-checking is fixed-parameter tractable on any class of structures of bounded tree-width. The aim of this section is to establish intractability results for classes of graphs of unbounded tree-width. As explained above, the lower bounds reported below are conditional on some complexity theoretical assumptions. Consequently, the results usually are proved by a reduction from some NP-hard problems.

At the core of all results reported below is the observation that the run of a Turing machine M on some input $w \in \{0,1\}^*$ can be simulated by an MSO_2 formula on a suitable sub-graph of a large enough grid. Here, the $(n \times m)$-grid is the graph $G_{n,m}$ with vertex set $\{(i,j) : 1 \leq i \leq n, 1 \leq j \leq m\}$ and edge set $\{\big((i,j),(i',j')\big) : |i-i'| + |j-j'| = 1\}$. Essentially, the grid provides the drawing board on which the time space diagram of a run of M on w can be guessed using set quantification. This yields the following result which is part of the folklore (see [**32**] for an exposition).

THEOREM 9.1. *Let* $\mathcal{G}^* := \{H \subseteq G_{n\times n} : n > 0\}$ *be the class of sub-graphs of grids. If* PTIME $\neq$ NP *then* $\text{MC}(\text{MSO}_1, \mathcal{G}^*)$ *is not fpt.*

We can use the result to obtain the following lower bound for MSO_1 on graph classes closed under taking minors, first obtained by Makowsky and Mariño.

THEOREM 9.2 ([**37**]). *Let* $\mathcal{C}$ *be a class of graphs closed under taking minors. If* $\mathcal{C}$ *has unbounded tree-width then* $\text{MC}(\text{MSO}_1, \mathcal{C})$ *is not fpt unless* PTIME = NP. *The same is true if* $\mathcal{C}$ *is only closed under topological minors.*

The result follows from Theorem 9.1 and the following structural result about graph classes with large tree-width established by Robertson and Seymour [**44**].

THEOREM 9.3 (Excluded Grid Theorem [**44**]). *There is a computable function* $f : \mathbb{N} \to \mathbb{N}$ *such that for all* $k \geq 0$, *every graph of tree-width at least* $f(k)$ *contains a* $(k \times k)$*-grid as a minor.*

It follows that if $\mathcal{C}$ is closed under minors and has unbounded tree-width, then the Excluded Grid Theorem implies that $\mathcal{G}^* \subseteq \mathcal{C}$. Intractability of MSO_2 on $\mathcal{C}$ therefore follows from Theorem 9.1. The generalisation to topological minors can be proved along the lines using *walls* instead of grids.

However, another consequence of the excluded grid theorem is that any (topological) minor closed class $\mathcal{C}$ of graphs of unbounded tree-width has very large tree-width, as it contains all grids and therefore graphs whose tree-width is roughly the square root of their order. Hence, there is a very large gap between the classes of graphs of bounded tree-width to which Courcelle's tractability results apply and the lower bound provided by Theorem 9.2.

To close this gap we will establish lower bounds for classes $\mathcal{C}$ of graphs of unbounded tree-width. Towards this aim we first need to measure the degree of unboundedness of the tree-width of classes $\mathcal{C}$ of graphs. We will do so by relating the tree-width of a graph in $\mathcal{C}$ to its order.

DEFINITION 9.4. *Let σ be a binary signature. Let $f : \mathbb{N} \to \mathbb{N}$ be a function and $p(n)$ be a polynomial.*

The tree-width of a class $\mathcal{C}$ of σ-structures is (f, p)-unbounded*, if for all $n \geq 0$*

(1) there is a graph $G_n \in \mathcal{C}$ of tree-width $\mathrm{tw}(G_n)$ between n and $p(n)$ such that $\mathrm{tw}(G_n) > f(|G|)$ and

(2) given n, G_n can be constructed in time $2^{n^{\varepsilon}}$, for some $\varepsilon < 1$.

The degree of $p(n)$ is called the gap degree. *The tree-width of $\mathcal{C}$ is* poly-logarithmically unbounded *if there are polynomials $p_i(n)$, $i \geq 0$, so that $\mathcal{C}$ is $(\log^i, p_i)$-unbounded for all i.*

The next theorem shows that essentially MSO_2 model-checking is fixed-parameter intractable on any class of graphs closed under sub-graphs with logarithmic tree-width. A similar result for classes of coloured graphs (but not closed under sub-graphs) was obtained in [**30**].

THEOREM 9.5. [**35, 34**] *Let $\mathcal{C}$ be a class of graphs closed under sub-graphs, i.e. $G \in \mathcal{C}$ and $H \subseteq G$ implies $H \in \mathcal{C}$.*

(1) If the tree-width of $\mathcal{C}$ is $(\log^{28\gamma} n, p(n))$-unbounded, where p is a polynomial and $\gamma > 1$ is larger than the gap-degree of $\mathcal{C}$, then $\mathrm{MC}(\mathrm{MSO}_2, \mathcal{C})$ is not fpt unless SAT *can be solved in sub-exponential time $2^{o(n)}$.*

(2) If the tree-width of $\mathcal{C}$ is poly-logarithmically unbounded then $\mathrm{MC}(\mathrm{MSO}_2, \mathcal{C})$ is not fpt unless all problems in the polynomial-time hierarchy can be solved in sub-exponential time.

At its very core, the proof of the previous result also relies on a definition of large grids in graphs $G \in \mathcal{C}$. However, as the tree-with of graphs in $\mathcal{C}$ is only logarithmic in their order, the excluded grid theorem only yields grids of double logarithmic size which is not good enough. Instead the proof uses a new replacement structure for grids, called *grid-like minors* developed by Reed and Wood [**43**]. These structures do not exist in the graphs $G \in \mathcal{C}$ itself but only in certain intersection graphs of paths in G which makes their definition in MSO much more complicated. See [**34**] for details.

The previous results narrow the gap to Courcelle's theorem significantly. But clearly there still is a gap, between classes of bounded tree-width and those of super-logarithmic tree-width. In [**37**], Makowsky and Mariño exhibit a class of graphs of logarithmic tree-width which is closed under sub-graphs and on which MSO_2 model-checking becomes tractable. So there is no hope to improve the results in the previous theorem to classes with sub-logarithmic tree-width.

All previous results refer to classes which are closed under sub-graphs (or allow colourings which in some sense amounts to the same thing). We have seen that MSO_1 is fixed-parameter tractable even on classes of bounded clique-width. As clique-width is not closed under sub-graphs, one might wonder if even MSO_2 could be tractable on such classes. The question was answered in the negative by Courcelle et al. in [**4**] who showed that MSO_2 model-checking is not even tractable on the class of cliques, unless EXPTIME = NEXPTIME. The model checking problem on the class of cliques might be considered as being slightly artificial. It is worth noticing, therefore, that the observation that MSO_2 is not tractable on classes of bounded clique-width has subsequently been observed also in purely algorithmic form [**22**] on graph classes of bounded clique-width. In particular, they show that problems such as HAMILTONIAN PATH, which are MSO_2 but not MSO_1 definable, are W[1]-hard when parameterized by the clique-width.

10. Further results on lower bounds

We close this part by commenting on lower bounds for first-order logic. It was shown in [**31**] that if a class $\mathcal{C}$ of graphs is closed under sub-graphs and not nowhere dense, then it has intractable first-order model-checking (subject to some technical condition). A class of graphs is *nowhere dense* if for every $r \geq 0$ there is a graph H_r such that $H_r \not\preccurlyeq_r G$ for all $G \in \mathcal{C}$. Nowhere dense classes of graphs are slightly more general than classes of bounded expansion considered in Section 8. Hence, there is again a gap between the lower and upper bound for first-order logic.

Finally, very little is known about lower bounds for MSO_1. Again, if $\mathcal{C}$ has unbounded tree-width and is closed under minors or topological minors then it has intractable model-checking (unless $P = \mathrm{NP}$). To obtain similar results as Theorem 9.5, we would first have to find an analogue of grid-like minors but to date not even a good candidate is known. Hence, we first need to understand obstructions for rank- and clique-width much better before any lower bounds can be shown.

References

[1] Hans L. Bodlaender, *A linear-time algorithm for finding tree-decompositions of small tree-width*, SIAM Journal on Computing **25** (1996), 1305 – 1317.

[2] Julius R. Büchi, *Weak second-order arithmetic and finite automata*, Zeitschrift für Mathematische Logik und Grundlagen der Mathematik **6** (1960), 66–92.

[3] Bruno Courcelle, *Graph rewriting: An algebraic and logic approach*, Handbook of Theoretical Computer Science (J. van Leeuwen, ed.), vol. 2, Elsevier, 1990, pp. 194 – 242.

[4] Bruno Courcelle, Johann Makowsky, and Udi Rotics, *Linear time solvable optimization problems on graphs of bounded clique-width*, Theory of Computing Systems **33** (2000), no. 2, 125–150.

[5] Bruno Courcelle and Johann A. Makowsky, *Fusion in relational structures and the verification of monadic second-order properties*, Mathematical Structures in Computer Science **12** (2002), no. 2, 203–235.

[6] Anuj Dawar, Martin Grohe, and Stephan Kreutzer, *Locally excluding a minor*, Logic in Computer Science (LICS), 2007, pp. 270–279.

[7] Erik D. Demaine, Mohammad Taghi Hajiaghayi, and Ken ichi Kawarabayashi, *Algorithmic graph minor theory: Decomposition, approximation, and coloring*, 46th Annual Symposium on Foundations of Computer Science (FOCS), 2005, pp. 637–646.

[8] ———, *Decomposition, approximation, and coloring of odd-minor-free graphs*, SODA, 2010, pp. 329–344.

[9] Matt DeVos, Guoli Ding, Bogdan Oporowski, DP. Sanders, Bruce Reed, Paul Seymour, and Dirk Vertigan, *Excluding any graph as a minor allows a low tree-width 2-coloring*, Journal of Combinatorial Theory, Series B **91** (2004), 25 – 41.

[10] Reinhard Diestel, *Graph theory*, 3rd ed., Springer-Verlag, 2005.

[11] John Doner, *Tree acceptors and some of their applications*, Journal of Computer and System Sciences **4** (1970), 406–451.

[12] Rodney G. Downey and Michael R. Fellows, *Parameterized complexity*, Springer, 1998.

[13] Zdeněk Dvořák, Daniel Král, and Robin Thomas, *Deciding first-order properties for sparse graphs*, 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2010.

[14] Heinz-Dieter Ebbinghaus, Jörg Flum, and Wolfgang Thomas, *Mathematical logic*, 2nd ed., Springer, 1994.
[15] Michael Elberfeld, Andreas Jakoby, and Till Tantau, *Logspace versions of the theorems of Bodlaender and Courcelle*, Proceedings of the 51th Annual IEEE Symposium on Foundations of Computer Science, 2010, pp. 143–152.
[16] Calvin C. Elgot, *Decision problems of finite automata design and related arithmetics*, Transactions of the American Mathematical Society **98** (1961), 21–51.
[17] David Eppstein, *Subgraph isomorphism in planar graphs and related problems*, J. of Graph Algorithms and Applications **3** (1999), no. 3, 1–27.
[18] Solomon Feferman and Robert L. Vaught, *The first order properties of products of algebraic systems*, Fundamenta Mathematicae **47** (1959), 57–103.
[19] Jörg Flum, Markus Frick, and Martin Grohe, *Query evaluation via tree-decompositions*, J. ACM **49** (2002), no. 6, 716–752.
[20] Jörg Flum and Martin Grohe, *Fixed-parameter tractability, definability, and model checking*, SIAM Journal on Computing **31** (2001), 113 – 145.
[21] Jörg Flum and Martin Grohe, *Parameterized complexity theory*, Springer, 2006, ISBN 3-54-029952-1.
[22] Fedor V. Fomin, Petr A. Golovach, Daniel Lokshtanov, and Saket Saurabh, *Intractability of clique-width parameterizations*, SIAM Journal of Computing **39** (2010), no. 5, 1941–1956.
[23] Markus Frick and Martin Grohe, *Deciding first-order properties of locally tree-decomposable structures*, Journal of the ACM **48** (2001), 1148 – 1206.
[24] ———, *The complexity of first-order and monadic second-order logic revisited*, Ann. Pure Appl. Logic **130** (2004), no. 1-3, 3–31.
[25] Haim Gaifman, *On local and non-local properties*, Herbrand Symposium, Logic Colloquium '81 (J. Stern, ed.), North Holland, 1982, pp. 105 – 135.
[26] Martin Grohe, *Logic, graphs, and algorithms*, Logic and Automata – History and Perspectives (E.Grädel T.Wilke J.Flum, ed.), Amsterdam University Press, 2007.
[27] Wilfrid Hodges, *A shorter model theory*, Cambridge University Press, 1997.
[28] Alexandr V. Kostochka, *The minimum Hadwiger number for graphs with a given mean degree of vertices*, Metody Diskret. Analiz. **38** (1982), 37–58, in Russian.
[29] ———, *Lower bound of the Hadwiger number of graphs by their average degree*, Combinatorica **4** (1984), no. 4, 307–316.
[30] Stephan Kreutzer, *On the parameterised intractability of monadic second-order logic*, Proc. of Computer Science Logic (CSL), 2009.
[31] ———, *Algorithmic meta-theorems*, Finite and Algorithmic Model Theory (Javier Esparza, Christian Michaux, and Charles Steinhorn, eds.), London Mathematical Society Lecture Note Series, Cambridge University Press, 2011, a preliminary version is available at Electronic Colloquium on Computational Complexity (ECCC), TR09-147, http://www.eccc.uni-trier.de/report/2009/147, pp. 177–270.
[32] ———, *On the parameterized intractability of monadic second-order logic*, CSL 2009 Special Issue in Logical Methods in Computer Science (LMCS) (2011).
[33] Stephan Kreutzer and Anuj Dawar, *Parameterized complexity of first-order logic*, Electronic Colloquium on Computational Complexity (ECCC) **16** (2009), 131.
[34] Stephan Kreutzer and Siamak Tazari, *Lower bounds for the complexity of monadic second-order logic*, Logic in Computer Science (LICS), 2010.
[35] ———, *On brambles, grid-like minors, and parameterized intractability of monadic second-order logic*, Symposium on Discrete Algorithms (SODA), 2010.
[36] Johann A. Makowsky, *Algorithmic aspects of the Feferman-Vaught theorem*, Annals of Pure and Applied Logic **126** (2004), no. 1–3, 159 – 213.
[37] Johann A. Makowsky and Julian Mariño, *Tree-width and the monadic quantifier hierarchy*, Theor. Comput. Sci. **1** (2003), no. 303, 157–170.
[38] Jaroslav Nesetril and Patrice Ossona de Mendez, *Grad and classes with bounded expansion i. decompositions*, Eur. J. Comb. **29** (2008), no. 3, 760–776.
[39] ———, *Grad and classes with bounded expansion ii. algorithmic aspects*, Eur. J. Comb. **29** (2008), no. 3, 777–791.
[40] Jaroslav Nešetřil and Patrice Ossona de Mendez, *Tree depth, subgraph coloring and homomorphisms*, European Journal of Combinatorics (2005).

[41] Jaroslav Nešetřil and Patrice Ossona de Mendez, *On nowhere dense graphs*, European Journal of Combinatorics (2008), submitted.

[42] ______, *First order properties on nowhere dense structures*, Journal of Symbolic Logic **75** (2010), no. 3, 868–887.

[43] Bruce Reed and Damian Wood, *Polynomial treewidth forces a large grid-like minor*, unpublished. Available at arXiv:0809.0724v3 [math.CO], 2008.

[44] Neil Robertson and Paul D. Seymour, *Graph minors V. Excluding a planar graph*, Journal of Combinatorial Theory, Series B **41** (1986), no. 1, 92–114.

[45] Detlef Seese, *Linear time computable problems and first-order descriptions*, Mathematical Structures in Computer Science **5** (1996), 505–526.

[46] J. W. Thatcher and J. B. Wright, *Generalised finite automata theory with an application to a decision problem of second-order logic*, Mathematical Systems Theory **2** (1968), 57–81.

[47] Andrew Thomason, *An extremal function for contractions of graphs*, Math. Proc. Cambridge Philos. Soc. **95** (1984), no. 2, 261–265.

[48] Boris Trakhtenbrot, *Finite automata and the logic of monadic predicates*, Doklady Akademii Nauk SSSR **140** (1961), 326–329.

[49] Moshe Vardi, *On the complexity of relational query languages*, Proc. of the 14th Symposium on Theory of Computing (STOC), 1982, pp. 137–146.

Humboldt-Universität zu Berlin, Germany
E-mail address: `grohe@informatik.hu-berlin.de`

Oxford University Computing Laboratory, Oxford, U.K. and Chair for Logic and Semantics, Technical University Berlin, Germany
E-mail address: `stephan.kreutzer@tu-berlin.de`

This page intentionally left blank

Contemporary Mathematics
Volume **558**, 2011

On Counting Generalized Colorings

Tomer Kotek, Johann A. Makowsky, and Boris Zilber

Abstract. The notion of graph polynomials definable in Monadic Second Order Logic, **MSOL**, was introduced by B. Courcelle, J.A. Makowsky and U. Rotics in 2001. It was shown later that the Tutte polynomial and generalizations of it, as well as the matching polynomial, the cover polynomial and the various interlace polynomials fall into this category.

In this article we present a uniform model theoretic framework for studying graph polynomials. In particular we study an infinite class of graph polynomials based on counting functions of generalized colorings definable in full second order logic **SOL**.

2010 *Mathematics Subject Classification.* Primary: 05C15, 05C31, 03C35 .

Partially supported by the Fein Foundation and the Graduate School of the Technion - Israel Institute of Technology.

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

Partially supported by MODNET: Marie Curie Research Training Network MRTN-CT-2004-512234 .

1. Introduction

1.1. Graph invariants and graph polynomials. A *graph invariant* is a function from the class of (finite) graphs $\mathcal{G}$ into some domain $\mathcal{D}$ such that isomorphic graphs have the same image. Usually such invariants are uniformly defined in some formalism. If $\mathcal{D}$ is the two-element Boolean algebra we speak of *graph properties*. Examples are the properties of being connected, planar, Eulerian, Hamiltonian, etc. If $\mathcal{D}$ consists of the natural numbers, we speak of *numeric graph invariants*. Examples are the number of connected components, the size of the largest clique or independent set, the diameter, the chromatic number, etc. But $\mathcal{D}$ could also be a polynomial ring $\mathbb{Z}[\overline{X}]$ over $\mathbb{Z}$ with a set of indeterminates $\overline{X}$. Here examples are the characteristic polynomial, the chromatic polynomial, the Tutte polynomial, etc.

There are many graph invariants discussed in the literature, which are polynomials in $\mathbb{Z}[\overline{X}]$, but there are hardly any papers discussing classes of graph polynomials as an object of study in its generality. An outline of such a study was presented in [**49, 50**].

The results reported in this article are part of an ongoing research project which aims to develop a general theory of graph polynomials[1].

1.2. Graph polynomials as generating functions. We denote by **SOL** Second Order Logic and by **MSOL** Monadic Second Order Logic, where quantification of relations is restricted to unary relations. In [**46, 47, 48**] J.A. Makowsky introduced the **MSOL**-definable and the **SOL**-definable graph polynomials, the class of graph polynomials where the range of summation is definable in (monadic) second order logic. He has verified that all the examples of graph polynomials discussed in the literature, with the exception of the weighted graph polynomial of [**53**], are **SOL**-polynomials over some expansions (by adding order relations) of the graph, cf. also [**49**]. Actually the most prominent graph polynomials in the literature, such as the characteristic polynomial, the chromatic polynomial, the matching polynomial, the Tutte polynomial and the interlace polynomial, are **MSOL**-definable. In some cases this is straightforward, but in some other cases it follows from intricate theorems. However, since the publication of our conference version of this paper, [**42**], we have found new **SOL**-graph polynomials, which are provably not **MSOL**-definable, [**31**]. **SOL**-definable graph polynomials are polynomials by definition. They are generalizations of generating functions counting subgraphs with prescribed properties definable in **SOL**. Without a definability condition imposed, these polynomials will be called *subset expansion polynomials*. The first graph polynomials of this type which appeared in the literature were the generating matching polynomial and the independence polynomial, [**34**]. They are indeed **MSOL**-definable.

Traditionally, generating functions use standard monomials, i.e., products of powers of the indeterminates, and their coefficients have some specific combinatorial interpretations. The **SOL**-definable polynomials in [**47**] were defined like this. However, some graph polynomials use different representations, where the powers of the occurences of indeterminates raised to the power X^i are replaced by falling factorial $X_{(i)} = X\cdot(X-1)\cdot\ldots\cdot(X-i+1)$, the function $\binom{X}{i}$, or other combinatorially motivated polynomials. To accommodate these cases we use here, as in [**42**], wider

[1]See `http://www.cs.technion.ac.il/ janos/RESEARCH/gp-homepage.html`

classes of **SOL**-definable polynomials, the *standard (falling factorial, and Newton)* **SOL**-*polynomials.*

1.3. Graph polynomials arising from generalized colorings. There are other ways graph polynomials occur naturally, namely as parametrized graph invariants $P(G, k_1, \ldots k_\alpha)$ with one or several parameters $k_1, \ldots, k_\alpha$ ranging over non-negative integers. The best known examples are the chromatic polynomial and the Tutte polynomial. As a matter of fact, the chromatic polynomial was the first graph polynomial to appear in the literature, [**11**]. Our main focus in this paper, expanding on [**51, 42**], is the study of generalized colorings, where we have several simultaneous colorings using several color sets, subject to conditions formulated by a formula θ in **SOL**. It will become clear in Section 2 that θ has to be subject to certain semantic restrictions such as: invariance under permutation of the colors, the existence of a bound on the colors used, and independence from the colors not used.

The associated counting function $\chi_\theta(G, k_1, \ldots, k_\alpha)$ counts the number of generalized colorings satisfying θ as a function of $(k_1, \ldots, k_\alpha)$.

The starting point of our investigation is the following fact:

PROPOSITION A. *Let $\overline{k} = (k_1, \ldots, k_\alpha)$ be the cardinalities of the various color sets. For ϕ subject to the conditions above, the counting function $\chi_\phi(G, \overline{k})$ is a polynomial in $\overline{k}$.*

This will lead us to observe that many known graph theoretic counting functions are in fact graph polynomials, which previously were not recognized as such.

We shall compare the counting functions $\chi_\theta(G, \overline{k})$ of generalized colorings with **SOL**-polynomials. A natural question which now arises is under what conditions an **SOL**-polynomial can be viewed as counting generalized colorings, and vice versa. The purpose of this paper is to answer this question. We shall present a framework in which we can prove the following:

THEOREM B. *Every* **SOL**-*polynomial (standard or* FF*) is a counting function of a generalized coloring of ordered graphs definable in* **SOL**.

THEOREM C. *Every counting function of a generalized coloring of ordered graphs definable in* **SOL** *is an* **SOL**-*polynomial, where the choice of standard,* FF*or Newton* **SOL**-*polynomial depends on the exact definition of generalized coloring.*

1.4. A model theoretic view. In B. Zilber's study of the structure of models of totally categorical first order theories, [**59, 60, 61**], the growth function of the size of finite approximations of models generated by finite sets of indiscernibles plays an important rôle. It is proved there, for theories which are $\aleph_0$-categorical and ω-stable, that this growth function is a polynomial. The totally categorical theories are a special case of this. It turns out that our counting functions of generalized colorings can be seen in this framework. Although this connection between graph polynomials and model theory may be of little interest to the finite combinatorists, it will allow us to associate with a graph G an infinite structure $\mathfrak{M}(G)$, which can be viewed as the most general graph invariant.

1.5. Outline of the paper. We assume the reader is familiar with the basics of graph theory as, say, presented in [**23, 12**]. We also assume the reader is familiar the basics of model theory and finite model theory as, say, presented in [**26, 25, 43, 39, 54**].

Section 2 is a prelude to our general discussion. In it we discuss the chromatic polynomial and the bivariate matching polynomial and motivate our general approach. In Section 3 we introduce our notion of counting functions of generalized colorings definable in **SOL**. We prove they are polynomials in the number of colors and show examples of graph polynomials from the literature which fall under this class of graph polynomials. In Section 4 we give precise definition of the various versions of **SOL**-definable polynomials and prove Proposition A. In Section 5 we discuss the various choices of monomials as the basis of the **SOL**-definable polynomials. In Section 6 we state and prove Theorems B and C precisely. In Section 7 we discuss under which conditions graph polynomials are or are not **MSOL**-definable. In Section 8 we discuss how counting functions of generalized colorings fit into the framework of the model theory of totally categorical theories, cf. [**61**]. Section 9 presents conclusions and open problems.

An earlier version of this article was posted as [**51**] and a short version was published as [**42**].

2. Prelude: two typical graph polynomials

Before we introduce our general definitions, we discuss two typical graph polynomials studied in the literature, the classical *chromatic polynomial* $\chi(G;X)$ and the *bivariate matching polynomial* $M(G;X,Y)$. Both have a very rich literature. For an exhaustive monograph on the chromatic polynomial, cf. [**24**]. For the matching polynomial the reader may consult [**45, 32**].

We denote by $[n]$ the set $\{1,\ldots,n\}$ and by $\mathcal{G}$ the set of undirected (labeled) graphs of the form $G=(V,E)$ where $V=[n]$ for some $n\in\mathbb{N}$. Let $G\in\mathcal{G}$. A *vertex-k-coloring of G* is a function $f:[n]\to[k]$. The coloring f is a *proper vertex-k-coloring of G* if additionally it satisfies that, whenever $(u,v)\in E$, then $f(u)\neq f(v)$. $\chi(G;k)$ denotes the number of proper vertex-k-colorings of G. For a fixed graph G this defines a function $\chi_G:\mathbb{N}\to\mathbb{N}$ which can be proved to be a polynomial in k with integer coefficients. Therefore it can be interpreted as a polynomial in $\mathbb{Z}[X]$ or even $\mathbb{R}[X]$. We denote this polynomial by $\chi_G(X)$. We shall discuss several proofs of this below. We define its coefficients by

$$\chi(G,X)=\sum_i c_i X^i=\sum_i b_i X_{(i)} \tag{2.1}$$

where $X_{(i)}=X\cdot(X-1)\cdot\ldots\cdot(X-i+1)$, the falling factorial.

Let $M\subseteq E$ be a set of edges. M is a *matching* if it consists of isolated edges. We denote by $cov(M)$ the set of vertices $v\in V$ such that there is an $e=(u,v)\in M$. We note that $|cov(M)|=2|M|$. Consider the graph parameter $m_i(G)$ which counts the number of matchings of G which consist of i (isolated) edges. The *bivariate matching polynomial* is defined as

$$M(G;X,Y)=\sum_i m_i(G)X^iY^{|V|-2i} \tag{2.2}$$

For a fixed graph G this defines a function $M_G:\mathbb{N}^2\to\mathbb{N}$. Here M_G is a polynomial by definition. Two matching polynomials are obtained from $M(G;X,Y)$ as substitution instances: the *generating matching polynomial*

$$g(G, X) = M(G; X, 1) \tag{2.3}$$

and the *defect matching polynomial* or *acyclic polynomial*

$$m(G, X) = M(G; -1, X). \tag{2.4}$$

We note that $\chi_G(X)$ and $M(G; X, Y)$ both really denote a family of polynomials indexed by graphs from $\mathcal{G}$. These families are, furthermore, uniformly defined based on some of the properties of the graph G. We are interested in various formalisms in which such definitions can be given.

2.1. Recursive definitions of $\chi(G; X)$ and $M(G; X, Y)$. The first proof that $\chi_G(X)$ is a polynomial used the observation that $\chi_G(X)$ has a recursive definition using the order of the edges, which can be taken as the order induced by the lexical ordering on $[n]^2$. However, the object defined does not depend on the particular order of the edges. For details, cf. [**10, 12**]. We shall also give a similar definition of $M(G; X, Y)$. The essence of the proof is as follows:

For $e = (v_1, v_2)$, we put

(i) $G - e = (V, E')$ with $E' = E - \{e\}$.
The operation of passing from G to $G - e$ is called *edge deletion*.

(ii) $G/e = (V', E')$ with $V' = V - \{v_2\}$
and $E' = (E \cap (V')^2) \cup \{(v_1, v) : (v_2, v) \in E\}$.
The operation of passing from G to G/e is called *edge contraction*.

(iii) $G \dagger e = (V', E')$ with $V' = V - \{v_1, v_2\}$ and $E' = E \cap (V')^2$.
The operation of passing from G to $G \dagger e$ is called *edge extraction*.

REMARK 2.1. *If we want to ensure that in the resulting graph the universe $V' = [n-1]$ or $V' = [n-2]$ then a suitable relabeling is needed. We omitted this from the definitions to keep the notation simple.*

It is easy to verify that these operations commute.

LEMMA 2.2. *Let e, f be two edges of G. Then we have*

(i) $(G - e) - f = (G - f) - e$,
$(G/e) - f = (G - f)/e$,
$(G - e)/f = (G/f) - e$ *and*
$(G/e)/f = (G/f)/e$.

(ii) *If e, f are disjoint edges, then we also have*
$(G \dagger e) \dagger f = (G \dagger f) \dagger e$,
$(G - e) \dagger f = (G \dagger f) - e$, *and*
$(G \dagger e)/f = (G/f) \dagger e$.

Next one notes the following recurrence relations, cf. [**24, 32**]:

LEMMA 2.3. *Let $G = (V, E)$ be a graph and $e \in E$ an edge.*

$$\chi(G; X) = \chi(G - e; X) - \chi(G/e; X) \tag{2.5}$$

$$M(G; X, Y) = M(G - e; X, Y) + X \cdot M(G \dagger e; X, Y) \tag{2.6}$$

Furthermore, if $G = G_1 \sqcup G_2$ is the disjoint union of G_1 and G_2, then we have multiplicativity, i.e.

$$\chi(G_1 \sqcup G_2; X) = \chi(G_1; X) \cdot \chi(G_2; X) \tag{2.7}$$

$$M(G_1 \sqcup G_2; X, Y) = M(G_1; X, Y) \cdot M(G_2; X, Y) \tag{2.8}$$

Let $E_n = ([n], \emptyset)$. To compute the polynomials recursively we note that

$$\chi(E_n; X) = X^n \tag{2.9}$$

$$M(E_n, X, Y) = Y^n \tag{2.10}$$

Let $E = (e_0, e_1, \ldots, e_m)$ be the enumeration of the edges in lexicographic order. One can compute $\chi_G(X)$ and $M(G; X, Y)$ by eliminating edges in this order. It also turns out, using Lemma 2.2, that the result is *independent* of the ordering of the edges.

Other graph polynomials from the literature which satisfy similar recursive definitions are the Tutte polynomial and its many variations and substitution instances, [**12, 13**, Chapter X], the Cover polynomial for directed graphs, [**20**], and the various Interlace polynomials, [**2, 5, 6**]. A systematic study of polynomials which are defined recursively using edge and vertex eliminations may be found in I. Averbouch's thesis [**8**].

2.2. Generating functions and explicit descriptions. The bivariate matching polynomial was defined by 2.2 as a *generating function*. This can be rewritten as

$$\begin{aligned} M(G; X, Y) &= \sum_i m_i(G) X^i Y^{|V|-2i} = \sum_{M \subseteq E} X^{|M|} Y^{|V|-|cov(M)|} \\ &= \sum_{M \subseteq E} \left(\sum_{C = V - cov(M)} X^{|M|} Y^{|C|} \right) \end{aligned} \tag{2.11}$$

where the summation is over all matchings $M \subseteq E$. The properties "M is a matching" and "$C = V - cov(M)$" can be expressed by formulas in Second Order Logic **SOL**. We call this an **SOL**-*polynomial* presentation for $M(G; X, Y)$. A formal definition will be given in Section 4.

In [**24**, Theorem 1.4.1] an explicit description of $\chi(G; X)$ is given: Let $a(G, m)$ be the number of partitions of V into m independent sets of vertices. Then

$$\chi(G; X) = \sum_m a(G, m) \cdot X_{(m)} = \sum_m b_m X_{(m)} \tag{2.12}$$

In other words, the coefficients b_m from Equation 2.1 have a combinatorial interpretation.

This again can be written as

$$\chi(G; X) = \sum_{P : indpart(P, A_P, V)} X_{(card(A_P))} \tag{2.13}$$

where $indpart(P, A_P, V)$ says that "P is an equivalence relation on V" and "each equivalence class induces an independent set" and "A_P consists of the first elements (with respect to the order on $V = [n]$) of each equivalence class". This can be expressed in **SOL**, and therefore is an **SOL**-*subset expansion* for $\chi(G; X)$ which uses an order relation on the vertices V of $G = (V, E)$. However, the choice of the particular order does not matter, the definition is *order invariant.*

Another explicit description for $\chi(G; X)$ is given in [**24**, Theorem 2.2.1]. It can be obtained from a two-variable dichromatic polynomial[2] $Z_G(X, Y)$ defined by

$$Z_G(X,Y) = \sum_{S:S\subseteq E} \left(\prod_{v:fcomp(v,S)} X \cdot \prod_{e:e\in S} Y \right) = \sum_{S:S\subseteq E} \left(X^{k(S)} \cdot \prod_{e:e\in S} Y \right)$$

where $fcomp(v, S)$ is the property "v is the first vertex in the order of V of some connected component of the spanning subgraph $< S : V >$ on V induced by S", and $k(S)$ is the number of connected components of $< S : V >$. Again this is *order invariant.* Now it is well known, [**55**], that

$$\chi(G; X) = Z_G(X, -1) \tag{2.14}$$

Hence, $\chi(G; X)$ is a *substitution instance of an order invariant* **SOL**-*subset expansion* of the graph G with an order on the vertices.

The bivariate matching polynomial has a presentation as an **SOL**-polynomial in the pure language of graphs, in other words, which does not use an order at all. It is natural to ask, whether such a presentation as an **SOL**-polynomial can also be found for the chromatic polynomial $\chi(G; X)$? We now show that this is not possible.

PROPOSITION 2.4. *$\chi(G; X)$ has no presentation as an* **SOL**-*polynomial in the pure language of graphs.*

PROOF. To see this, assume that

$$\chi(G; X) = \sum_{A\subseteq V^p:\phi_1(A)} X^{|A|} \tag{2.15}$$

or

$$\chi(G; X) = \sum_{A\subseteq V^p:\phi_2(A)} X_{(|A|)} \tag{2.16}$$

where A ranges over all subsets satisfying an **SOL**-property $\phi_1(A)$ and $\phi_2(A)$, respectively.

We set $X = 2$ and look at the graphs C_n, the 2-regular connected graphs on n vertices. This can be written as an **SOL**-formula $Cycle(G)$. Clearly, $\chi(C_n, 2) = 0$ if n is odd, and $\chi(C_n, 2) = 2$ if n is even. We first deal with Equation (2.15). So we have

$$\chi(C_{2n}; 2) = \sum_{A\subseteq V^p:\phi_1(A)} 2^{|A|} = 2 \tag{2.17}$$

and

$$\chi(C_{2n+1}; 2) = \sum_{A\subseteq V^p:\phi_1(A)} 2^{|A|} = 0 \tag{2.18}$$

It follows that $Cycle(G)$ and $\phi_1(A)$ imply that A is a singleton and uniquely defined on G. But this is a contradiction, because C_n has non-trivial automorphisms for $n \geq 3$.

[2] $Z_G(X, Y)$ is related to the Potts model in statistical mechanics and is related to the Tutte polynomial by a prefactor, cf. [**12**].

For Equation (2.16) the argument is similar, observing that $2^m = 2_{(m)}$ for $m \leq 2$ and $2_{(m)} = 0$ for $m \geq 3$. $\square$

2.3. Recursive definitions vs presentations as SOL-polynomials. It is a recurrent theme in the literature about graph polynomials to look both for recursive definitions and for presentations as **SOL**-polynomials. In the classical literature these presentations are called subset expansions if the summation formula is true for all subsets, and spanning tree expansions if the formula requires that the spanning subgraph (V, A) is a tree. Sometimes both cases are called subset expansions. Good examples of polynomials which have both recursive definitions and subset expansions are the Tutte polynomial and its many variations and substitution instances, [**12, 13**, Chapter X], the Cover polynomial for directed graphs, [**20**], the various Interlace polynomials, [**2, 5, 6**], and all the polynomials defined recursively by vertex and edge eliminations studied in [**57, 7, 8**]. In all these cases the subset expansions turn out to be **SOL**-polynomials. We introduced the notion of "having a presentation as an **SOL**-polynomial" as a generalization which encompasses all the cases which we have encountered in the literature.

In [**30**], a general theorem is formulated and proved which states that, under rather general definitions, every recursively defined graph polynomial can be presented as an **SOL**-polynomial. The converse is likely not to be true. The recursive definitions here are more general than vertex and edge eliminations, and are based on local operations. The framework does cover all the above mentioned cases.

2.4. The bivariate matching polynomial counts colorings. Recall that a proper k-coloring of a graph $G = (V, E)$ is a function $f : V \to [k]$ such that $f^{-1}(i)$ induces an independent set of G. Alternatively, we can define it as a relation $r \subseteq V \times [k]$ satisfying the **SOL**-property $\varphi(r)$ saying "r is a total function such that $\{v \in V : \exists j\ r(v, j)\}$ induces an independent set".

A natural setting in which to interpret the formula $\varphi(r)$ is that of the two-sorted structure of the form $\mathfrak{M}_k = \langle V, [k]; E, r\rangle$, where V and $[k]$ are two universes, E is the edge relation of the graph, and $r \subseteq V \times [k]$ is a relation.

Now $\chi(G, k)$ can be written as

$$\chi(G,k) = \mid \{r \subseteq V \times [k] : \mathfrak{M}_k \models \varphi(r)\} \mid \tag{2.19}$$

In other words, $\chi(G, k)$ counts the number of φ-colorings of G.

We want to define $M(G; X, Y)$ in a similar way. We first do it for

$$g(G, X) = M(G; X, 1) = \sum_M X^{|M|} \tag{2.20}$$

where M ranges over all matchings of $G = (V, E)$. To do this we replace $\varphi(r)$ by $\varphi_1(r)$ which says that "$r \subseteq E \times [k]$ is a partial function the domain of which is a matching of G". In other words r is a partial edge-coloring such that for each $i \in [k]$ the set $\{e \in E : (e, i) \in r\}$ is an independent set of edges. For each matching $M \subseteq E$ there are $k^{|M|}$ many functions with domain M. Hence

$$g(G, k) = \mid \{r \subseteq E \times [k] : \mathfrak{M}_k \models \varphi_1(r)\} \mid = \sum_M X^{|M|} \tag{2.21}$$

This shows that $g(G; X)$ counts the number of φ_1-colorings of G.

We can obtain a similar presentation for $g^*(G,k) = \sum_M X_{(|M|)}$ by writing

$$g^*(G,k) = |\ \{r \subseteq E \times [k] : \mathfrak{M}_k \models \varphi_2(r)\}\ | = \sum_M X_{(|M|)} \tag{2.22}$$

where $\varphi_2(r)$ is the formula "$\varphi_1(r)$ and r is injective". This shows that $g^*(G;X)$ counts the number of φ_2-colorings of G.

To interpret the bivariate polynomial $M(G;X,Y)$ as counting colorings we use two sorts of colors $[k_1]$ and $[k_2]$, the three-sorted structure $\mathfrak{M}_k = \langle V, [k_1], [k_2]; E, r_1, r_2\rangle$, with two coloring relations $r_1 \subseteq E \times [k_1]$ and $r_2 \subseteq V \times [k_2]$ and a formula $\varphi_3(r_1, r_2)$ which says that
"$r_1 \subseteq E \times [k_1]$ is a partial function the domain M of which is a matching of G".
and "$r_2 \subseteq V \times [k_1]$ is a partial function with domain $V - cov(M)$".

2.5. Generalized chromatic polynomials. As stated earlier, the interpretation of $M(G;X,Y)$ as counting colorings will be generalized and formally defined in Section 3. Here we want to informally prepare our general definition.

Any relation $r \subseteq V^\alpha \times [k]$ on $\mathfrak{M}_k = \langle V, [k]; E, r\rangle$ satisfying a formula $\varphi(r)$ of **SOL** can be viewed as a coloring relation over the graph $G = (V,E)$. We denote by

$$\chi_\varphi(G,k) = |\ \{r \subseteq V^\alpha \times [k] : \mathfrak{M}_k \models \varphi(r)\}\ | \tag{2.23}$$

the number of φ-colorings of G. What interests us here is when $\chi_\varphi(G,k)$ is a polynomial in k. It turns out that this is true under rather general conditions, as stated in Proposition A, in the introduction. The conditions are

(i) the coloring r is invariant under permutations of the colors, i.e., if π is a permutation of the colors, then r satisfies $\varphi(r)$ iff the composition r' of r with π satisfies $\varphi(r')$.
(ii) the number of colors used by each r satisfying $\varphi(r)$ is bounded by the size of V, and
(iii) the property $\varphi(r)$ is independent of the colors not used.

This is obviously the case for $\chi(G;X)$ and $M(G;X,Y)$ and, in general, is easily verified.

2.6. Previously unnoticed graph polynomials. The literature contains many papers on generalized colorings, and their authors are interested either in questions of extremal graph theory or in the complexity of deciding the existence of these colorings. Counting the number of generalized colorings is rarely studied. However, it turns out that counting the number of colorings with k colors very often gives rise to previously unnoticed graph polynomials. We list here a few examples, which we think deserve further investigations.

F. Harary introduced the notion of *P-colorings*, [**36, 38, 15, 16**]. Here P is any graph property. Given a graph $G = (V,E)$, a function $f : V \to [k]$ is a *P-k-coloring* if for all $i \in [k]$ in the range of f, the set $f^{-1}(i)$ induces a graph in P. If P is definable in **SOL**, there is a formula $\varphi(f)$ defining the P-colorings.

Examples of P colorings include:

(i) The proper k-colorings with P being the edgeless graphs.
(ii) The *convex colorings* [**52**], with P being the connected graphs. Convex colorings have applications in computational biology.
(iii) The G-free colorings studied in [**14, 1**], where P consists of G-free graphs.

(iv) The partitions of graphs into cographs, [**29**]. The family of cographs is the smallest class of graphs that includes K_1 and is closed under complementation and disjoint union. They can be characterized as the P_4-free graphs.

(v) the *mcc_t-colorings* defined in [**3**] and further studied in [**44**]. Here $t \in \mathbb{N}$ and $P = P_t$ consist of all graphs the connected components of which are of size at most t. For $t = 1$ these are just the proper colorings.

This list is far from being complete and the reader will easily find more examples of P-colorings.

It is easy to verify that P-colorings satisfy the three informal conditions required in Proposition A. The precise version of Proposition A is stated in Section 3 as Proposition 3.10.

COROLLARY 2.5. *Let P be a graph property and let $\chi_P(G, k)$ be the number of P-colorings with k colors. Then $\chi_P(G, k)$ is a polynomial in k.*

Note that the definability condition is not needed here. We only need that P is closed under isomorphisms.

REMARK 2.6. *For fixed t, $\chi_{P_t}(G, k)$ is a polynomial in k. However, it is not a bivariate polynomial in t and k.*

Harary's definition can be generalized in various ways, using edge colorings, rather than vertex colorings, or by requiring that the union of any s color classes induces a graph in P. An example of the latter with $s = 2$ and P the class of forests is the set of acyclic colorings introduced in [**33**] and further studied in [**4**]. Also these generalizations satisfy the three conditions required in Proposition A, hence they give rise to previously unnoticed graph polynomials.

Not all coloring properties in the literature can be formulated as variants of P-colorings.

(i) *Injective colorings* are vertex colorings such that for any three vertices $u, v, w \in V$ such that (u, v) and (u, w) are edges in E then $f(v) \neq f(w)$, [**35**].

(ii) A coloring f of a graph $G = (V, E)$ is *harmonious*, if it is a proper coloring and every pair of colors occurs at most once along an edge. Harmonious colorings were first studied in [**40**] and extensively studied in [**27**].

(iii) There are various notions of *rainbow colorings*, which all are edge colorings and impose some injectivity condition on certain configurations of edges. For example in [**18**], the condition is that any two vertices are joined by a path such that all edges on it have different colors.

It is easy to verify that these colorings are not P-colorings but still satisfy the three conditions required in Proposition A. Hence, counting injective colorings, harmonious colorings and rainbow colorings with k colors again gives rise to new graph polynomials.

F. Harary [**37**] also introduced the notion of a *complete k-coloring* of a graph, and the associated chromatic number, which he named the *achromatic number*. A coloring f of a graph $G = (V, E)$ is *complete* if (i) it is a proper coloring and (ii) every pair of colors from $[k]$ occurs at least once along an edge. For a survey on this topic, cf. [**41**].

If f is a complete k-coloring, f cannot be a complete $(k+1)$-coloring, therefore complete colorings do not satisfy our conditions. Indeed, the number of complete k-colorings of a graph is not a polynomial in k, as it vanishes for k with $\mid E \mid < \binom{k}{2}$.

3. Counting generalized colorings

3.1. φ-colorings. Previously we only considered graph colorings; now we expand our discussion to include τ structures $\mathfrak{M}$, where τ is a finite vocabulary for relational structures. We shall also use the formalism of many-sorted structures. We think of many-sorted structures as having a single big universe (the union of the universes corresponding to the sorts), and with unary predicates whose interpretations are the universes of the sorts.

Let $\mathfrak{M}$ be a τ-structure with universe M. We assume without loss of generality that $M = [n] = \{1, \ldots, n\}$ for $n > 1$. We will assume that all our structures are ordered, i.e. there exists a binary relation symbol $\mathbf{R}_{\leq}$ in τ which is always interpreted as the natural linear ordering of the universe. To simplify notation, we omit the order relation from structures.

Let k be a natural number. We denote by $\mathfrak{M}_{F,k}$ the two-sorted structure

$$\mathfrak{M}_{F,k} = \langle \mathfrak{M}, [k], F \rangle$$

where $F : M \to [k]$ is a function. We think of $\mathfrak{M}_{F,k}$ as the colored structure induced by the function F on $\mathfrak{M}$. The set $[k]$ will be referred to as the color set. Note that the order relation does not extend to the second sort $[k]$.

Let $\mathbf{F}$ be a unary function symbol and let $\tau_{\mathbf{F}} = \tau \cup \{\mathbf{F}\}$, then $\mathfrak{M}_{F,k}$ is a $\tau_{\mathbf{F}}$-structure. On the other hand, every two-sorted $\tau_{\mathbf{F}}$-structure $\mathfrak{A}$ with second sort $[k]$ can be thought of as $\mathfrak{M}_{F,k}$ for a unique tuple $\langle \mathfrak{M}, [k], F \rangle$ consisting of a τ-structure $\mathfrak{M}$ with universe M, the set $[k]$ and a function $F : M \to [k]$.

Let $\mathcal{L}$ be a fragment of **SOL**. We will only be interested in formulas whose interpretations are invariant under the choice of a linear ordering of the universe.

Definition 3.1 ($\mathcal{L}(\tau_{\mathbf{F}})$-Coloring formulas). *We say $\varphi \in \mathcal{L}(\tau_{\mathbf{F}})$ is an $\mathcal{L}(\tau_{\mathbf{F}})$-*coloring formula *if φ does not quantify over the second sort, but instead all first and second order quantifiers are on the first sort only. We say it is an $\mathcal{L}(\tau_{\mathbf{F}})$-*coloring sentence *if φ is a sentence.*

Definition 3.2 (φ-colorings and coloring properties). *Let φ be an $\mathcal{L}(\tau_{\mathbf{F}})$-coloring sentence.*

(i) $\mathfrak{M}_{F,k}$ *is called a* φ-colored τ-structure *and F is called a* φ-coloring of $\mathfrak{M}$ *if $\mathfrak{M}_{F,k}$ is a two-sorted $\tau_{\mathbf{F}}$-structure such that*

$$\mathfrak{M}_{F,k} \models \varphi$$

(ii) *Let $\mathcal{P}_\varphi$ be the class*

$$\mathcal{P}_\varphi = \{\mathfrak{M}_{F,k} : \mathfrak{M}_{F,k} \models \varphi\}$$

of φ-colored τ-structures. Then $\mathcal{P}_\varphi$ is called a coloring property.

Example 3.3 (Proper coloring and variations). *Let τ_{Graphs} be the vocabulary consisting of one binary relation symbol $\mathbf{E}$ as well as the order relation $\mathbf{R}_{\leq}$. Let $G = (V, E, \leq)$ be a τ_{Graphs}-structure, where $V = [n]$.*

(i) *A function* $F : V \to [k]$ *is a* proper coloring*, if it satisfies*

$$\varphi_{proper} = \forall u, v(\mathbf{E}(u,v) \to \mathbf{F}(u) \neq \mathbf{F}(v))$$

The class $\mathcal{P}_{\varphi_{proper}}$ *is the class of tuples* $\langle G, [k], F\rangle$ *of graphs together with a color-set* $[k]$ *and a proper coloring* F*.*

(ii) *A function* $F : V \to [k]$ *is* pseudo-complete *if it satisfies*

$$\forall x, y \exists u, v(\mathbf{E}(u,v) \wedge \mathbf{F}(u) = x \wedge \mathbf{F}(v) = y).$$

We will now see that pseudo-complete colorings do not form a coloring property.

Coloring properties $\mathcal{P}_\varphi$ satisfy the following two properties:

Permutation property: Let $\pi : [k] \to [k]$ be a permutation and let F_π be the function obtained from F by applying π, i.e. $F_\pi(v) = \pi(F(v))$. Then

$$\mathfrak{M}_{F,k} \in \mathcal{P}_\varphi \text{ iff } \mathfrak{M}_{F_\pi,k} \in \mathcal{P}_\varphi.$$

In other words, $\mathcal{P}_\varphi$ is closed under permutation of the color-set $[k]$.

Remark 3.4. *It follows from the Permutation Property, that we can assume that* $Range(F)$ *is of the form* $[k_0]$ *for some* $k_0 \leq k$*.*

Extension property: For every $\mathfrak{M}$, F with $Range(F) = [k_0]$, $k' \geq k_0$

$$\mathfrak{M}_{F,k_0} \in \mathcal{P}_\varphi$$

iff

$$\mathfrak{M}_{F,k'} \in \mathcal{P}_\varphi$$

Namely, the extension property requires that increasing or decreasing the number of colors not in $Range(F)$ does not affect whether it belongs to the property.

Let φ be a $\mathcal{L}(\tau_{\mathbf{F}})$-coloring sentence. The class $\mathcal{P}_\varphi$ satisfies the permutation property because it is definable in **SOL**.

Proposition 3.5. $\mathcal{P}_\varphi$ *satisfies the extension property.*

Proof. We prove by induction a slightly stronger statement:

(*) Let $\mathfrak{A}_1$ be a τ-structure with universe A_1. Let $\mathfrak{A}$ be a two-sorted $\tau_{\mathbf{F}}$-structure $\langle \mathfrak{A}_1, A_2, F\rangle$ and let $Range(F) = F(A_1)$. Then for every formula $\theta \in \mathcal{L}(\tau_{\mathbf{F}})$ with no variable which ranges over the second sort A_2,

$$\mathfrak{A} \models \theta \text{ iff } \langle \mathfrak{A}_1, Range(F), F\rangle \models \theta$$

Basis: Let θ be an atomic formula. Assume first that θ does not contain $\mathbf{F}$. Any relation symbol, function symbol or constant symbol in θ is interpreted in $\mathfrak{A}$ over A_1 only. Since all the variables in θ range over the first sort only, the truth-value of θ does not depend on A_2.

On the other hand, if θ contains $\mathbf{F}$, then θ must be of the form $\mathbf{F}(x_1) = \mathbf{F}(x_2)$, where x_1 and x_2 are first order variables (which range over the first sort). In this case, the truth-value of θ depends only on the elements of A_2 which can be obtained as $\mathbf{F}(a)$ for some $a \in A_1$. I.e., θ depends only on $Range(\mathbf{F})$.

Closure: If θ_1, θ_2 satisfy (*), then clearly so does any Boolean combination of them. We need only deal with universal quantifiers $\forall x, \forall X$, as we get the existential quantifiers as their negation. Let $\theta = \forall z \theta'(z)$ where z is a first order or second order variable. We extend the vocabulary τ with symbol s_z as follows. The symbol s_z is a constant symbol $s_z = c_z$ if z is a first order variable. The symbol s_z is a relation symbol $s_z = R_z$ of arity ρ if z is a second order variable of arity of ρ. We note $\mathfrak{A} \models \theta$ iff for every interpretation a of s_z it holds that $\langle \mathfrak{A}, a \rangle \models \theta'$. By the induction hypothesis, this happens iff for every interpretation a of s_z it holds that $\langle \mathfrak{A}_1, a, Range(F), F \rangle \models \theta'$. The latter occurs iff $\langle \mathfrak{A}_1, Range(F), F \rangle \models \forall z \theta$.

□

REMARK 3.6. *The class $\mathcal{P}_1$ of structures $\mathfrak{M}_{F,k}$ where F is a proper coloring which uses all the colors in the color-set $[k]$ is not a coloring property, since $\mathcal{P}_1$ violates the extension property. Similarly, the class of pseudo-complete colorings is not a coloring property.*

DEFINITION 3.7 (Counting functions of φ-colorings). *Let φ be a $\mathcal{L}(\tau_{\mathbf{F}})$-coloring sentence. Let χ_φ be a function from pairs $\langle \mathfrak{M}, k \rangle$ which consist of a τ-structure $\mathfrak{M}$ and $k \in \mathbb{N}^+$ defined as follows. The function $\chi_\varphi(\mathfrak{M}, k)$ is*

$$| \{ F : \mathfrak{M}_{F,k} \in P_\varphi \} |$$

I.e., $\chi_\varphi(\mathfrak{M}, k)$ counts the number of φ-colorings of $\mathfrak{M}$ with k colors.

For example, treating k as an indeterminate, $\chi_{\varphi_{proper}}(G, k)$ is the chromatic polynomial of the graph G.

Denote by $c_\varphi(\mathfrak{M}, j)$ the number of φ-colorings of $\mathfrak{M}$ with color-set $[j]$ which use all the colors of $[j]$.

PROPOSITION 3.8 (Special case of Proposition A). *Let φ be an $\mathcal{L}(\tau_{\mathbf{F}})$-coloring sentence. For every $\mathfrak{M}$ the number $\chi_\varphi(\mathfrak{M}, k)$ is a Newton polynomial in k of the form*

$$\sum_{j=1}^{|M|} c_\varphi(\mathfrak{M}, j) \binom{k}{j}$$

PROOF. We first observe that any φ-coloring F uses at most $|M|$ of the k colors. By the permutation property, if F is a φ-coloring which uses j colors then any function obtained by permuting the colors is also a φ-coloring. Therefore, given k colors, the number of φ-colorings that use exactly j of the k colors is the product of $c_\varphi(\mathfrak{M}, j)$ and the binomial coefficient $\binom{k}{j}$. So

$$\chi_\varphi(\mathfrak{M}, k) = \sum_{j=0}^{|M|} c_\varphi(\mathfrak{M}, j) \binom{k}{j}$$

The right side here is a polynomial in k, because each of the binomial coefficients is. We also use that for $k < j$ we have $\binom{k}{j} = 0$. □

Remark 3.9.

(i) *Since for a coloring property* φ *the function* $\chi_\varphi(\mathfrak{M}, k)$ *is a polynomial, it is now defined not only for positive integer values of* k*, but rather for every* $k \in \mathbb{R}$*. Still, these instances of* $\chi_\varphi(\mathfrak{M}, k)$ *may have a combinatorial meaning. E.g., the chromatic polynomial* $\chi_G(k)$ *is well known to have meaningful evaluations for the negative integers. In particular,* $\chi_G(-1)$ *is the number of acyclic orientations of the graph* G*, see* [**56**].

(ii) *The restriction to coloring properties in Proposition 3.8 is essential. Denote by* $\chi_{onto}(G, k)$ *the number of functions* $f : V \to [k]$ *which are onto. Clearly, this is not a polynomial in* k *since, for* $k >\mid V \mid$*, it always vanishes, so it should vanish identically, if it were a polynomial.*

(iii) *The proof of Proposition 3.8 does not garantee that the coefficients of the power of* k *are integers. However, the proof of Proposition 3.10 below does garantee it.*

In fact, it holds that $\chi_\varphi(\mathfrak{M}, k)$ is in $\mathbb{Z}[k]$. For two functions $f_1, f_2 : [n] \to [k]$ we write $\sim_{perm}$ if there exists a permutation $\pi : [k] \to [k]$ such that for all $i \in [n]$ we have $\pi(f_1(i) = f_2(i)$. In other words, f_1 and f_2 are equivalent if they can be obtained from one another by applying some permutation of the color set $[k]$. Let $d_\varphi(\mathfrak{M}, j)$ be the number of colorings F with $Range(F) = [j]$.

Proposition 3.10 (Special case of Proposition A). *Let* φ *be an* $\mathcal{L}(\tau_{\mathbf{F}})$*-coloring sentence. For every* $\mathfrak{M}$ *the number* $\chi_\varphi(\mathfrak{M}, k)$ *is an* FF *polynomial in* k*, namely*

$$\sum_{j=1}^{|M|} d_\varphi(\mathfrak{M}, j) \cdot k_{(j)}$$

where $k_{(j)}$ *is the falling factorial,* $k_{(j)} = k \cdot (k-1) \cdots (k-j+1) = \binom{k}{j} \cdot j!$.

Recall that FF polynomials are polynomials where the monomials are falling factorials.

Proof. By proposition 3.8,

$$\chi_\varphi(\mathfrak{M}, k) = \sum_{j=1}^{|M|} c_\varphi(\mathfrak{M}, j) \binom{k}{j}$$

By the Permutation Property, F is a φ-coloring iff the functions which are $\sim_{perm}$ equivalent to F are φ-colorings. Therefore,

$$\chi_\varphi(\mathfrak{M}, k) = \sum_{j=1}^{|M|} d_\varphi(\mathfrak{M}, j) \cdot j! \binom{k}{j} = \sum_{j=1}^{|M|} d_\varphi(\mathfrak{M}, j) \cdot k_{(j)}$$

□

Remark 3.11. *There exists a coloring property* $P_{no\ ext}$ *which does not satisfy the extension property and yet* $\chi_{no\ ext}(\mathfrak{M}, k)$ *is a polynomial. Let* $P_{no\ ext}$ *consist of all structures* $\mathfrak{M}_{F,k}$ *which satisfy the following condition:*

- *Let* γ_1, γ_2 *and* γ_3 *be the least, second least and third least elements in the linear ordering of* $M = [n]$*. The function* $F : M \to [k]$ *satisfies either* $F(\gamma_1) = F(\gamma_2) \neq F(\gamma_3)$ *or* $F(\gamma_1) = F(\gamma_3) \neq F(\gamma_2)$*. If* $F(\gamma_1) = F(\gamma_2)$ *then* F *is onto and if* $F(\gamma_1) = F(\gamma_3)$ *then* F *is not onto.*

Let $F : M \to [k]$ be a function such that $F(\gamma_1) = F(\gamma_2) \neq F(\gamma_3)$ and F is onto. The addition of an unused color to $[k]$ (i.e., looking at F as a function from M to $[k+1]$) implies that F is no longer onto and yet $F(\gamma_1) \neq F(\gamma_3)$. Hence, $\mathfrak{M}_{F,k+1} \notin P_{no\ ext}$, so $P_{no\ ext}$ does not satisfy the extension property. On the other hand, the number of such structures equals the number of functions $F : [n] \to [k]$ for which $F(1) = F(2) \neq F(3)$, so $\chi_{no\ ext}(\mathfrak{M}, k) = k^{n-2}(k-1)$ is a polynomial.

3.2. Multi-colorings. To construct graph polynomials in several variables, we extend in this and the next subsections the definitions in order to deal with several color-sets.

Let $\mathfrak{M}$ be a τ-structure with universe M. Let $\mathfrak{M}_{F,\overline{k}}$ be the $(\alpha+1)$-sorted structure $\langle \mathfrak{M}, [k_1] \ldots, [k_\alpha], F\rangle$ with

$$F : M^m \to [k_1]^{m_1} \times \ldots \times [k_\alpha]^{m_\alpha}$$

We denote by $\tau_{\alpha,\mathbf{F}}$ the corresponding vocabulary.

We extend the definitions of $\mathcal{L}(\tau_{\alpha,\mathbf{F}})$-coloring formulas, φ-colorings and coloring properties naturally to $\mathcal{L}(\tau_{\alpha,\mathbf{F}})$*-multi-coloring formulas, φ-multi-colorings and multi-coloring properties*. Multi-coloring properties $\mathcal{P}_\varphi$ satisfy a version of the permutation and extension properties:

Permutation property: Let $\overline{\pi} = (\pi_1, \ldots, \pi_\alpha)$ be permutations of $[k_1], \ldots, [k_\alpha]$ respectively. Let $F : M^m \to [k_1]^{m_1} \times \ldots \times [k_\alpha]^{m_\alpha}$ and let $F_{\overline{\pi}}$ be the function obtained by applying the permutations $\pi_1, \ldots, \pi_\alpha$ on F. Then

$$\mathfrak{M}_{F,\overline{k}} \in \mathcal{P}_\varphi \text{ iff } \mathfrak{M}_{F_\pi,\overline{k}} \in \mathcal{P}_\varphi$$

Namely, $\mathcal{P}_\varphi$ is closed under permutations of the color-sets.

Extension property: For every $\mathfrak{M}$, $\overline{k} = k_1, \ldots, k_\alpha$, $\overline{k'} = k'_1, \ldots, k'_\alpha$, and F such that $k_1 \leq k'_1, \ldots, k_\alpha \leq k'_\alpha$,
we have

$$\mathfrak{M}_{F,\overline{k}} \in \mathcal{P}_\varphi$$

iff

$$\mathfrak{M}_{F,\overline{k'}} \in \mathcal{P}_\varphi$$

The multi-coloring properties $\mathcal{P}_\varphi$ satisfy the following property as well:

Non-occurrence property: Assume $F : M^m \to [k_1]^{m_1} \times \ldots \times [k_\alpha]^{m_\alpha}$ with $m_i = 0$. [3] Then for every $b \in \mathbb{N}$,

$$\langle \mathfrak{M}, [k_1], \ldots, [k_\alpha], F\rangle \in \mathcal{P}_\varphi$$

iff

$$\langle \mathfrak{M}, [k_1], \ldots, [b], \ldots, [k_\alpha], F\rangle = \mathfrak{M}_{(k_1,\ldots,k_{i-1},b,k_{i+1},\ldots,k_\alpha),F} \in \mathcal{P}_\varphi.$$

The extension property and the non-occurrence property require that increasing and decreasing the number of colors not used by F, respectively adding or removing unused color-sets, does not affect whether $\mathfrak{M}_{F,\overline{k}}$ belongs to the multi-coloring property $\mathcal{P}_\varphi$. The proofs that these properties hold for multi-coloring properties are similar to the one variable case.

We denote by $\chi_\varphi(\mathfrak{M}, k_1, \ldots, k_\alpha)$ the number of φ-multi-colorings with color-sets $[k_1], \ldots, [k_\alpha]$.

[3] For a set S the set S^0 is the singleton set which has as its unique element the empty tuple.

PROPOSITION 3.12 (Special case of Proposition A). *Let $\mathcal{P}_\varphi$ be a multi-coloring property with $\mathcal{L}(\tau_{\alpha,\mathbf{F}})$-coloring formula φ. For every $\mathfrak{M}$, $\chi_\varphi(\mathfrak{M}, k_1, \dots, k_\alpha)$ is an* FF*-polynomial in $k_1, \dots, k_\alpha$ of the form*

$$\sum_{j_1 \le N_M} \sum_{j_2 \le N_M} \cdots \sum_{j_\alpha \le N_M} d_{\varphi(F)}(\mathfrak{M}, \bar{j}) \prod_{1 \le \beta \le \alpha} k_{\beta(j_\beta)}$$

where $\bar{j} = (j_1, \dots, j_\alpha$, $d_{\phi(R)}(\mathfrak{M}, \bar{j})$ is the number of φ-multi colorings F with color-sets $[j_1], \dots, [j_\alpha]$ which uses all the colors in every color-set up to permutations of the color sets, and N_M is a polynomial in $|M|$.

PROOF. Similar to the one variable case. □

EXAMPLE 3.13. *Recall that, in the prelude, $\chi_{mcc(t)}(G, k)$ denoted the number of vertex colorings for which no color induces a graph with a connected component of size larger than t. Let $\chi_{mcc}(G, k, t) = \chi_{mcc(t)}(G, k)$ be the counting function of multi-colorings satisfying the above condition, where t is considered a color-set. These multi-colorings do not satisfy the non-occurrence property. Indeed, we will now see that $\chi_{mcc}(G, k, t)$ is not a polynomial in t.*

Let $F : V \to [k]$ be any function. If $t \ge |V|$ then for every color $c \in [k]$ it holds that $|f^{-1}(c)| < t$ and, in particular, no color induces a graph which has a connected component larger than t. Therefore, for such t, $\chi_{mcc}(G, k, t) = k^{|V|}$. Since $\chi_{mcc}(G, k, t)$ does not always agree with $k^{|V|}$ on small values of t, $\chi_{mcc}(G, k, t)$ cannot be a polynomial in t.

This example shows the motivation for requiring the non-occurrence property of coloring properties.

3.3. Multi-colorings – the general cases.

3.3.1. *Multi-colorings with partial functions.* We extend our definition in two ways. First, we now allow $F \subseteq M^m \times [k_1]^{m_1} \times \ldots \times [k_\alpha]^{m_\alpha}$ to be a partial function. For this purpose $\mathbf{F}$ must now be a relation symbol. Second, we also allow several simultaneous coloring predicates $F_1, \dots, F_s$ and the corresponding number of relation symbols. A coloring property P_φ will therefore consist of structures

$$\mathfrak{M}_{F_1, \dots, F_s, k_1, \dots, k_\alpha} = \langle \mathfrak{M}, [k_1], \dots, [k_\alpha], F_1, \dots, F_s \rangle$$

which satisfy φ and for which each F_i is a (possibly partial) function.

We may call multi-coloring properties and multi-coloring simply also coloring properties and colorings, if the situation is clear from the context. The permutation, extension and non-occurrence properties extend naturally to this case and Proposition A holds as well:

PROPOSITION 3.14 (Proposition A for several partial functions). *Let φ be an $\mathcal{L}(\tau_{\mathbf{F_1}, \dots, \mathbf{F_s}})$-multi-coloring sentence. For every $\mathfrak{M}$ the number $\chi_\varphi(\mathfrak{M}, k)$ of φ-multi-colorings with several partial functions is an* FF*-polynomial in k of the form*

$$\sum_{j_1 \le N_M} \sum_{j_2 \le N_M} \cdots \sum_{j_\alpha \le N_M} d_{\varphi(\overline{F})}(\mathfrak{M}, \bar{j}) \prod_{1 \le \beta \le \alpha} k_{\beta(j_\beta)} \tag{3.1}$$

where $N_M \in \mathbb{N}$. Moreover, N_M is bounded by a polynomial in $|M|$.

3.3.2. *Multi-colorings with bounded relations.* Here we extend the multi-colorings with several partial functions by allowing several relations which are bounded in a certain way.

Definition 3.15 (Bounded relations and multi-coloring properties).

(i) *We say a relation* $R \subseteq M^m \times [k_1]^{m_1} \times \cdots \times [k_\alpha]^{m_\alpha}$ *is* d-bounded if the set of tuples of colors used by R,

$$\{\bar{c} \mid \text{ there exists } \bar{x} \in M^m \text{ such that } (\bar{x}, \bar{c}) \in R\},$$

is of size at most $|M|^d$.

(ii) A $\varphi(\overline{\mathbf{R}})$-multi-coloring property P_φ is the class of τ-structures

$$\mathfrak{M}_{\overline{R},\overline{k}} = \langle \mathfrak{M}, [k_1], \ldots, [k_\alpha], \overline{R} \rangle$$

which satisfy φ. We say P_φ is *bounded* if there exists d such that each R in every structure $\mathfrak{M}_{\overline{R},\overline{k}}$ in P_φ is d-bounded.

Again, the permutation, extension and non-occurrence properties extend naturally and Proposition A holds. However, in this case we do not have a polynomial of the form of Equation (3.1). In particular, in this case we may have polynomials which do not belong to $\mathbb{Z}[\overline{k}]$.

Proposition 3.16 (Proposition A for several d-bounded relations). *Let* $d \in \mathbb{N}$ *and let* φ *be an* $\mathcal{L}(\tau_{\mathbf{R_1},\ldots,\mathbf{R_s}})$*-multi-coloring sentence. For every* $\mathfrak{M}$ *the number* $\chi_\varphi(\mathfrak{M}, k)$ *of* φ*-multi-colorings with several d-bounded relations is a Newton polynomial in* k *of the form*

$$\sum_{j_1 \le N_M} \sum_{j_2 \le N_M} \cdots \sum_{j_\alpha \le N_M} c_{\varphi(\overline{R})}(\mathfrak{M}, \bar{j}) \prod_{1 \le \beta \le \alpha} \binom{k_\beta}{j_\beta}$$

where $N_M \in \mathbb{N}$.

3.4. Closure properties for the general cases.

Proposition 3.17 (Sums and products). *Let* ϕ, ψ *be coloring properties. Then there are* $\theta_1, \theta_2 \in$ **SOL** *such that*

(i) $\chi_{\theta_1(\overline{\mathbf{F}}_3)}(\mathfrak{M}, \overline{k}, 1) = \chi_{\phi(\overline{\mathbf{F}}_1)}(\mathfrak{M}, \overline{k}) + \chi_{\psi(\overline{\mathbf{F}}_2)}(\mathfrak{M}, \overline{k})$

(ii) $\chi_{\theta_2(\overline{\mathbf{F}}_3)}(\mathfrak{M}, \overline{k}) = \chi_{\phi(\overline{\mathbf{F}}_1)}(\mathfrak{M}, \overline{k}) \cdot \chi_{\psi(\overline{\mathbf{F}}_2)}(\mathfrak{M}, \overline{k})$

Proof. Since we are dealing with ordered structures, we can define $\varphi_{min}(\mathbf{F}')$ which requires that F' is a total function $F' : M \to [k']$, where $[k']$ is a new color set. Let

$$\begin{aligned} \theta_1(\overline{\mathbf{F}}_1, \overline{\mathbf{F}}_2, \mathbf{F}') \quad &= \quad \big(\varphi_{min}(\mathbf{F}') \wedge \phi(\overline{\mathbf{F}}_1) \wedge \mathbf{F}_2 = \emptyset\big) \vee \\ & \qquad \big(\mathbf{F}' = \emptyset \wedge \mathbf{F_1} = \emptyset \wedge \psi(\overline{\mathbf{F}}_2)\big) \end{aligned}$$

It holds that $\chi_{\theta_1}(\mathfrak{M}, \overline{k}, k') = (k')^{|M|} \cdot \chi_\phi(\mathfrak{M}, \overline{k}) + \chi_\psi(\mathfrak{M}, \overline{k})$. Taking k' to be 1, we get that $\chi_{\theta_1}(\mathfrak{M}, \overline{k}, 1)$ is the sum.

For the product we take $\chi_{\theta_2}(G, \lambda)$ with

$$\theta_2(\overline{\mathbf{F}}_3) = \big(\phi(\overline{\mathbf{F}}_1) \wedge \psi(\overline{\mathbf{F}}_2)\big)$$

where $\overline{\mathbf{F}}_3 = (\overline{\mathbf{F}}_1, \overline{\mathbf{F}}_2)$. ☐

3.5. The dichromatic polynomial $Z(G;X,Y)$. The dichromatic polynomial is sometimes considered a version of the Tutte polynomial, cf. [**55**]. To illustrate Theorem B we show how to convert $Z(G;X,Y)$ into a counting function of a φ-multi-coloring. We use the dichromatic polynomial in the following form:

$$Z(G;X,Y) = \sum_{A \subseteq E} X^{k(A)} Y^{|A|}$$

where $k(A)$ is the number of connected components of the spanning subgraph (V,A). For this purpose we look at the three-sorted structure

$$G_{(A,F_1,F_2),(k,l)} = \langle \mathfrak{G}, [k], [l], A, F_1, F_2 \rangle$$

with $A \subseteq E$, $F_1 : V \to [k]$ and $F_2 : A \to [l]$ such that $(u,v) \in A$ implies $F_1(u) = F_1(v)$. This is expressed in the formula $dichromatic(A, F_1, F_2)$. As we saw for the matching polynomial, this can be easily converted into a coloring property definable in **SOL**. Now we have

$$\chi_{dichromatic(A,F_1,F_2)}(G;k,l) = \sum_{A \subseteq E} k^{k(A)} l^{|A|}$$

which is the evaluation of $Z(G;X,Y)$ for $X = k, Y = l$.

4. SOL-polynomials and subset expansion

We are now ready to introduce the **SOL**-polynomials, which generalize subset expansions and spanning tree expansions of graph polynomials as encountered in the literature.

4.1. SOL-polynomials. Let $\mathcal{R}$ be a commutative semi-ring, which contains the semi-ring of natural numbers $\mathbb{N}$. For our discussion $\mathcal{R} = \mathbb{Z}$ suffices, but the definitions generalize. Our polynomials have a fixed finite set of variables (indeterminates, if we distinguish them from the variables of **SOL**), $\mathbf{X}$. We denote by $card_{\mathfrak{M},\overline{v}}(\varphi(\overline{v}))$ the number of tuples $\overline{v}$ of elements of the universe that satisfy φ. We again assume τ contains a relation symbol $\mathbf{R}_{\leq}$ which is always interpreted as a linear ordering of the universe.

Let $\mathfrak{M}$ be a τ-structure. We first define the *standard (or geometric)* $\mathbf{SOL}(\tau)$-*monomials* inductively.

Definition 4.1 (standard **SOL**-monomials).

(i) *Let $\phi(\overline{v})$ be a formula in $\mathbf{SOL}(\tau)$, where $\overline{v} = (v_1, \ldots, v_m)$ is a finite sequence of first order variables. Let $r \in \mathbf{X} \cup (\mathbb{Z} - \{0\})$ be either an indeterminate or an integer. Then*

$$r^{card_{\mathfrak{M},\overline{v}}(\phi(\overline{v}))}$$

is a standard $\mathbf{SOL}(\tau)$-monomial (whose value depends on $card_{\mathfrak{M},\overline{v}}(\phi(\overline{v}))$).

(ii) *Finite products of standard $\mathbf{SOL}(\tau)$-monomials are standard $\mathbf{SOL}(\tau)$-monomials.*

Even if r is an integer, and $r^{card_{\mathfrak{M},\overline{v}}(\phi(\overline{v}))}$ does not depend on $\mathfrak{M}$, the monomial stands as it is, and is not evaluated.

The *falling factorial (*FF*)* $\mathbf{SOL}(\tau)$*-monomials* and the *Newton* $\mathbf{SOL}(\tau)$*-monomials* are defined similarly as follows:

Definition 4.2 (FF **SOL**-monomials). *The* FF **SOL**(τ)*-monomials are defined as in Definition 4.1, except we replace the power*

$$r^{card_{\mathfrak{M},\overline{v}}(\phi(\overline{v}))}$$

with the falling factorial

$$r_{(card_{\mathfrak{M},\overline{v}}(\phi(\overline{v})))}$$

Definition 4.3 (Newton **SOL**-monomials). *The Newton* **SOL**(τ)*-monomials are defined as in Definition 4.1, except we replace the power*

$$r^{card_{\mathfrak{M},\overline{v}}(\phi(\overline{v}))}$$

with the binomial coefficient

$$\binom{r}{card_{\mathfrak{M},\overline{v}}(\phi(\overline{v}))}$$

Note the degree of a monomial is polynomially bounded by the cardinality of $\mathfrak{M}$.

Definition 4.4 (**SOL**-polynomials). *The polynomials definable in* **SOL**(τ) *are defined inductively:*

(i) *standard (respectively* FF *respectively Newton)* **SOL**(τ)*-monomials are standard (respectively* FF *respectively Newton)* **SOL**(τ)*-polynomials.*

(ii) *Let ϕ be a $\tau \cup \{\overline{\mathbf{R}}\}$-formula in* **SOL** *where $\overline{\mathbf{R}} = (\mathbf{R}_1, \ldots, \mathbf{R}_m)$ is a finite sequence of relation symbols not in τ. Let t be a standard (respectively* FF *respectively Newton)* **SOL**$(\tau \cup \{\overline{\mathbf{R}}\})$*-polynomial. Then*

$$\sum_{\overline{R}:\langle\mathfrak{M},\overline{R}\rangle\models\phi(\overline{R})} t$$

is a standard (respectively FF *respectively Newton)* **SOL**(τ)*-polynomial.*

For simplicity we refer to **SOL**(τ)-polynomials as **SOL**-polynomials when τ is clear from the context. Among the **SOL**-polynomials we find most of the known graph polynomials from the literature, cf. [**50**]. We will discuss the choice of basis of the **SOL**-polynomials in Section 5.

4.2. Properties of SOL-polynomials.

Lemma 4.5.

(i) *Every indeterminate $x \in \mathbf{X}$ can be written as a standard,* FF *and Newton* **SOL***-monomial.*

(ii) *Every integer c can be written as a standard,* FF *and Newton* **SOL***-monomial.*

Proof. The minimal element f_1 in the linear ordering of the universe is definable in **SOL**. For every $r \in (\mathbb{Z} - \{0\}) \cup \mathbf{X}$, the term r is a standard, FF and Newton **SOL**-monomial since

$$r = r^{card_{\mathfrak{M},v}(v=f_1)} = r_{(card_{\mathfrak{M},v}(v=f_1))} = \binom{r}{card_{\mathfrak{M},v}(v = f_1)}$$

$\square$

LEMMA 4.6 (Normal form).

(i) *Let $P(\mathfrak{M})$ be a standard,* FF *or Newton* **SOL***−polynomial. Then $P(\mathfrak{M})$ can be written in the form*

$$\sum_{R_1:\phi_1(R_1)} \dots \sum_{R_s:\phi_s(R_1,\dots,R_s)} \Phi(\mathfrak{M}, \overline{R}) \tag{4.1}$$

where $\phi_1, \dots, \phi_s \in$ **SOL** *and $\Phi(\mathfrak{M}, \overline{R})$ is a standard,* FF *or Newton* **SOL***-monomial respectively.*

(ii) *Let $\Phi(\mathfrak{M})$ be a standard* **SOL***-monomial. Then $\Phi(\mathfrak{M})$ can be written in the form*

$$r_1^{card_{\mathfrak{M},\overline{v}}(\varphi_1(\overline{v}))} \cdots r_t^{card_{\mathfrak{M},\overline{v}}(\varphi_t(\overline{v}))} \tag{4.2}$$

where $\varphi_t \in$ **SOL** *and $r_1, \dots, r_t \in (\mathbb{Z} - \{0\}) \cup$* **X** *are distinct.*

(iii) *Let $\Phi(\mathfrak{M})$ be an* FF *or Newton* **SOL***-monomial. Then $\Phi(\mathfrak{M})$ can be written in the form of Equation (4.2), except we replace $r_i^{card_{\mathfrak{M},\overline{v}}(\varphi_i(\overline{v}))}$ with $r_{i(card_{\mathfrak{M},\overline{v}}(\varphi_i(\overline{v})))}$ or $\binom{r_i}{card_{\mathfrak{M},\overline{v}}(\varphi_i(\overline{v}))}$ respectively, and the r_i's might not be distinct.*

PROOF. (i) follows directly from the definitions. The proof of (ii) is as follows. From the definitions, it is easy to see that Φ is of the desired form, except $r_1, \dots, r_t$ may not be distinct, i.e. $\Phi(\mathfrak{M})$ can be written as

$$r_1^{card_{\mathfrak{M},\overline{v}}(\varphi_{1,1}(\overline{v}))} \cdots r_1^{card_{\mathfrak{M},\overline{v}}(\varphi_{1,h_1}(\overline{v}))} \cdots \cdot r_t^{card_{\mathfrak{M},\overline{v}}(\varphi_{t,1}(\overline{v}))} \cdots r_t^{card_{\mathfrak{M},\overline{v}}(\varphi_{t,h_t}(\overline{v}))}$$

We will prove that

$$r_i^{card_{\mathfrak{M},\overline{v}}(\varphi_{i,1}(\overline{v}))} \cdots r_i^{card_{\mathfrak{M},\overline{v}}(\varphi_{i,h_i}(\overline{v}))}$$

can be written as $r_i^{card_{\mathfrak{M},\overline{v}}\theta(\overline{v})}$ for some formula θ. Let m be the maximum number of free variables in any of the $\varphi_{i,j}$. Without loss of generality, we may assume that all $\varphi_{i,j}$ have m free variables. We may do so because otherwise we can add free variables $v_q, \dots, v_m$ to those $\varphi_{i,j}$ which have less m free variables. We then change $\varphi_{i,j}$ to require also that each new variable is equal to v_1, thus keeping the same number of tuples satisfying $\varphi_{i,j}$. The first element f_1 and the second element f_2 of the linear ordering of $\mathfrak{M}$ are definable in **SOL**. Let $\theta(v_1, \dots, v_m, u_1, \dots, u_t)$ be the formula such that $(a_1, \dots, a_m, b_1, \dots, b_{h_i})$ satisfies θ iff:

(i) exactly one of the b_j is f_1 and all other b_j's are f_2, and
(ii) if $b_j = f_1$ then $a_1, \dots, a_m$ satisfy $\varphi_{i,j}$.

The formula θ uses the u_j variables to choose the formula $\varphi_{i,j}$. The tuples corresponding to the different $\varphi_{i,j}$ are disjoint which implies that

$$x_i^{card_{\mathfrak{M},\overline{v}}(\theta(\overline{v}))} = x_i^{card_{\mathfrak{M},\overline{v}}(\varphi_{i,1}(\overline{v}))} \cdots x_i^{card_{\mathfrak{M},\overline{v}}(\varphi_{i,h_i}(\overline{v}))}$$

The proof of (ii) follows. Finally, (iii) holds by definition. □

PROPOSITION 4.7. *The pointwise product of two standard,* FF *or Newton* **SOL***-polynomials is again a standard,* FF *or Newton* **SOL***-polynomial respectively.*

PROOF. We prove it for standard **SOL**-polynomials. The proof for FF-polynomials and Newton polynomials is identical. Let $P_1(\mathfrak{M})$ and $P_2(\mathfrak{M})$ be standard **SOL**-polynomials. Every **SOL**-polynomial can be written in the form of Equation (4.1). Without loss of generality, we may assume P_1 and P_2 have the same number of sums (otherwise we add dummy sums of the form $\sum_{U:U=\emptyset}$). We

proceed by induction on the number of summations in P_1 and P_2:
Basis: By definition $P_1(\mathfrak{M}) \cdot P_2(\mathfrak{M})$ is a **SOL**-monomial.
Step: For $i \in \{1, 2\}$, let

$$P_i(\mathfrak{M}) = \sum_{R_i : \phi_i(R_i)} \Phi_i(\langle \mathfrak{M}, R_i \rangle)$$

Then

$$P_1(\mathfrak{M}) \cdot P_2(\mathfrak{M}) = \sum_{R_1, R_2 : \phi_1(R_1) \wedge \phi_2(R_2)} \Phi_1(\langle \mathfrak{M}, R_1 \rangle) \cdot \Phi_2(\langle \mathfrak{M}, R_2 \rangle)$$

By the induction hypothesis, this is a standard **SOL**-polynomial. □

LEMMA 4.8. *Let τ be a vocabulary and let* $\mathbf{S}$ *be a relation symbol not in τ. Let $P(\mathfrak{M})$ be a standard,* FF- *or Newton* $\mathbf{SOL}(\tau)$*-polynomial and let $A \in \{\emptyset, M\}$. Let $P^A(\mathfrak{M}, S)$ be a graph polynomial which satisfies*

$$P^A(\mathfrak{M}, S) = \begin{cases} P(\mathfrak{M}) & S = A \\ 1 & otherwise \end{cases}$$

Then $P^A(\mathfrak{M}, S)$ is a standard, FF- *respectively Newton* $\mathbf{SOL}(\tau \cup \{\mathbf{S}\})$*-polynomial.*

PROOF. We prove the lemma for standard **SOL**-polynomials by induction on the structure of $P(\mathfrak{M})$. The proof for FF-**SOL**-polynomial and Netwon **SOL**-polynomial is similar.
Basis:

(i) Let $P(\mathfrak{M}) = r^{card_{\mathfrak{M},\overline{v}}(\varphi(\overline{v}))}$. Then $P^A(\mathfrak{M}, S) = r^{card_{\mathfrak{M},\overline{v}}(\varphi(\overline{v}) \wedge (S=A))}$ satisfies the conditions.
(ii) $P(\mathfrak{M}) = P_1(\mathfrak{M}) \cdot P_2(\mathfrak{M})$. Then $P^A(\mathfrak{M}) = P_1^A(\mathfrak{M}) \cdot P_2^A(\mathfrak{M})$ satisfies the conditions.

Step: Let

$$P(\mathfrak{M}) = \sum_{\overline{R} : \langle \mathfrak{M}, \overline{R} \rangle \models \phi(\overline{R})} t$$

where t is a $(\tau \cup \{\overline{\mathbf{R}}\})$-polynomial. Let

$$P^A(\mathfrak{M}, S) = \sum_{\overline{R} : \langle \mathfrak{M}, \overline{R} \rangle \models \phi(\overline{R}) \wedge \mu(\overline{R}, S)} t^A$$

where $\mu(\overline{R}, S)$ says that either each relation in $\overline{R}$ is equal to $\emptyset$ or $S = A$ (or both). Then $P^A(\mathfrak{M}, S)$ satisfies the conditions. □

PROPOSITION 4.9. *The pointwise sum of two standard,* FF *or Newton* **SOL***-polynomials is again a standard,* FF *or Newton* **SOL***-polynomial respectively.*

PROOF. We prove it for standard **SOL**-polynomials. The proof for FF-polynomials and Newton polynomials is identical.

Let $P_1(\mathfrak{M})$ and $P_1(\mathfrak{M})$ be standard **SOL**-;polynomials. Let $P_1^{\emptyset}(\mathfrak{M}, S)$ and $P_2^M(\mathfrak{M}, S)$ be the **SOL**-polynomials given in Lemma 4.8. Then the sum of $P_1(\mathfrak{M})$ and $P_2(\mathfrak{M})$ is given by

$$\sum_{S \in \{\emptyset, M\}} P_1^{\emptyset}(\mathfrak{M}, S) \cdot P_2^M(\mathfrak{M}, S)$$

□

PROPOSITION 4.10. *Let*

$$P_1(\mathfrak{M}) = \sum_{\overline{R}:\theta} \prod_{\overline{b}:\psi} \sum_{\overline{a}:\phi} P_2(\langle \mathfrak{M}, \overline{R}, \overline{a}, \overline{b}\rangle)$$

where $P_2(\mathfrak{A})$ *is a standard,* FF *or Newton* **SOL***-monomial and the product and inner summation are on tuples of elements of the universe. It holds that* $P_1(\mathfrak{M})$ *is a standard,* FF *or Newton* **SOL***-polynomial respectively.*

PROOF. We can expand the product

$$\prod_{\overline{b}:\psi} \sum_{\overline{a}:\phi} P_2(\mathfrak{A}) = \sum_{f:\vartheta} \prod_{\overline{a},\overline{b}:\varphi} P_2(\mathfrak{A})$$

where ϑ says the relation f is a function

$$f : \{\overline{b} \mid \langle \mathfrak{A}, \overline{b}\rangle \models \psi\} \to \{\overline{a} \mid \langle \mathfrak{A}, \overline{a}, \overline{b}\rangle \models \phi\},$$

and $\varphi = \big(f(\overline{b}, \overline{a})\big) \wedge \psi \wedge \phi$. So the proposition holds. □

By induction the last proposition holds for functions defined by alternating $\prod$ and $\sum$, as long as all $\sum$ within the scope of a $\prod$ iterate over elements (and not over relations).

4.3. Combinatorial polynomials. As for the case of counting φ-multi-colorings, it is noteworthy that the following combinatorial invariants can be written as standard, FF and Newton **SOL**-polynomials.

Cardinality, I: The cardinality of a definable set

$$card_{\mathfrak{M},\overline{v}}(\varphi(\overline{v})) = \sum_{\overline{v}:\varphi(\overline{v})} 1$$

is an an **SOL**-polynomial.

Cardinality, II: Exponentiation of cardinalities

$$card_{\mathfrak{M},\overline{v}}(\varphi(\overline{v}))^{card_{\mathfrak{M},\overline{v}}(\psi(\overline{v}))} = \prod_{\overline{v}:\psi(\overline{v})} \sum_{\overline{u}:\varphi(\overline{u})} 1$$

is equivalent to a **SOL**-polynomial by proposition 4.10.

Factorials: The factorial of the cardinality of a definable set

$$card_{\mathfrak{M},\overline{v}}(\varphi(\overline{v}))! = \sum_{\pi:\varphi(\overline{v}) \xrightarrow{1-1} \varphi(\overline{v})} 1$$

is an **SOL**-polynomial.

5. Standard vs FF vs Newton SOL-polynomials

We have introduced three notions of **SOL**-polynomials: standard, FF, and Newton **SOL**-polynomials. The sets of monomials $X^i : i \in \mathbb{N}$ (powers of X) and $X_{(i)} : i \in \mathbb{N}$ (falling factorials of X) each form a basis of the polynomial ring $\mathbb{Z}[X]$ as a module over $\mathbb{Z}$. The sets of monomials $\binom{X}{i} : i \in \mathbb{N}$ (binomials of X) form a basis in the polynomial ring $\mathbb{Q}[X]$. Over $\mathbb{Q}$ each of these bases can be transformed into the other using linear transformations. In this section we discuss transformations of one basis into another using substitution by **SOL**-definable polynomials.

5.1. Standard vs FF polynomials. In the statement and proof of Proposition 3.10, the polynomial obtained is of the form

$$\sum_{j=1}^{|M|} d_{\varphi}(\mathfrak{M}, j) \cdot k_{(j)}$$

In the literature on graph polynomials mixed presentations also occur, e.g., the cover polynomial for directed graphs [**20**] is such a case.

We extend the definition of **SOL**-polynomials by allowing both monomials of the form

$$r^{card_{\mathfrak{M},\overline{v}}(\phi(\overline{v}))}$$

and

$$r_{(card_{\mathfrak{M},\overline{v}}(\phi(\overline{v})))}$$

We call the polynomials obtained like this *extended* **SOL**-*polynomials.*

In the following we show that every extended **SOL**-polynomial on ordered structures can be written both as a standard **SOL**-polynomial and as a FF **SOL**-polynomial.

PROPOSITION 5.1. *Let* $\Phi(\mathfrak{M}) = r^{card_{\mathfrak{M},v_1,\ldots,v_y}(\phi(\overline{v}))}$ *be a standard* **SOL**-*monomial with* $r \in \mathbf{X} \cup (\mathbb{Z} - \{0\})$. *There is a* FF **SOL**-*polynomial* Φ' *such that for all structures* $\mathfrak{M}$ *we have*

$$\Phi(\mathfrak{M}) = \Phi'(\mathfrak{M})$$

PROOF. First assume r is a positive integer. The monomial $\Phi(\mathfrak{M})$ counts functions from the set

$$D_{\phi} = \{\overline{a} \mid \mathfrak{M} \models \phi(\overline{a})\}$$

to $[r]$. On the other hand, the monomial $(r)_{card_{\mathfrak{M},\overline{v}}(\phi(v))}$ counts *injective* functions from D_{ϕ} to $[r]$. Let $\Phi'(\mathfrak{M})$ be given by

$$\Phi'(\mathfrak{M}) = \sum_{A \subseteq V^y : \langle \mathfrak{M}, A\rangle \models \psi_1} \; \sum_{R \subseteq V^y : \langle \mathfrak{M}, A, R\rangle \models \psi_2} (r)_{card_{\mathfrak{M},\overline{v}}(\overline{v} \in A)}$$

where

(i) ψ_1 requires that A is a subset of D_{ϕ}, and
(ii) ψ_2 requires that R is an equivalence relation over D_{ϕ} such that for every two distinct tuples $\overline{a} \in A$ and $\overline{b} \in D_{\phi}$, if $\overline{a}$ and $\overline{b}$ belong to the same equivalence class in R, then $\overline{a} < \overline{b}$ with respect to the order on the structure $\mathfrak{M}$. Moreover, for every equivalence class in R there exists some $\overline{a} \in A$ which belongs to it.

Taking any injective function f from A to $[r]$, we may extend it to a function from D_{ϕ} to $[r]$ by assigning every $\overline{b} \in D_{\phi} - A$ with the same value as the $\overline{a} \in A$ for which $(\overline{a}, \overline{b}) \in R$. This extension is determined uniquely by f and R, and forms a bijection between the set of functions $g : D_{\phi} \to [r]$ and the set of triples (A, R, f) such that A and R satisfy ψ_1 and ψ_2 and $f : A \to [r]$ is injective. Hence, $\Phi'(\mathfrak{M}) = \Phi(\mathfrak{M})$. If $r \in \mathbf{X}$, we get that $\Phi'(\mathfrak{M})$ and $\Phi(\mathfrak{M})$ agree on every evaluation of r to a positive integer, and thus, by interpolation, $\Phi'(\mathfrak{M}) = \Phi(\mathfrak{M})$. Therefore, in particular $\Phi'(\mathfrak{M})$ and $\Phi(\mathfrak{M})$ also agree on all non-positive evaluations of r. $\square$

Proposition 5.2. *Let* $\Phi(\mathfrak{M}) = r_{(card_{\mathfrak{M},v_1,\ldots,v_y}(\phi(\overline{v})))}$ *be a* FF **SOL**-*monomial for* $r \in \mathbf{X} \cup (\mathbb{Z} - \{0\})$. *There is a standard* **SOL**-*polynomial* Φ' *such that for all structures* $\mathfrak{M}$ *we have*

$$\Phi(\mathfrak{M}) = \Phi'(\mathfrak{M}).$$

Proof. By definition,

$$(r)_{card_{\mathfrak{M},\overline{v}}(\phi(\overline{v}))} = \prod_{i=0}^{|D_\phi|-1} (r-i)$$

Therefore,

$$(r)_{card_{\mathfrak{M},\overline{v}}(\phi(\overline{v}))} = \prod_{\overline{a}:\phi(\overline{a})} \left(r - \sum_{\overline{b}:\phi(\overline{b})\wedge \overline{b}<\overline{a}} 1 \right)$$

where $\overline{b} < \overline{a}$ means $\overline{b}$ is smaller than $\overline{a}$ in the lexicographic order induced by the order on the elements of the structure $\mathfrak{M}$. By Proposition 4.10 we need only show that

$$r - \sum_{\overline{b}:\phi(\overline{b})\wedge \overline{b}<\overline{a}} 1 \tag{5.1}$$

is a standard **SOL**-polynomial with summation on elements only. The expression in (5.1) is given by

$$\sum_{\overline{b}:\phi(\overline{b})\wedge(\overline{b}\leq\overline{a})} r^{card_{\mathfrak{M},\overline{w}}(\overline{w}=\overline{a}\wedge\overline{b}=\overline{a})} \cdot (-1)^{card_{\mathfrak{M},\overline{w}}(\overline{w}=\overline{a}\wedge\overline{b}\neq\overline{a})}$$

Hence, $(r)_{card_{\mathfrak{M},\overline{v}}(\phi(\overline{v}))}$ is a standard **SOL**-polynomial. ☐

5.2. Newton polynomials. In Proposition 3.8 we used monomials of the form $\binom{X}{i}$. However, writing these as standard polynomials, they have rational but not integer coefficients; hence they are polynomials in $\mathbb{Q}[X]$. Furthermore, the coefficients of standard and FF **SOL**-polynomials are always integers by definition. Therefore, $\binom{X}{i}$ cannot be written as a standard or FF **SOL**-polynomial.

On the other hand, FF **SOL**-monomials can be written as Newton **SOL**-polynomials. To see this we note

$$X_{(|A|)} = \mid A \mid ! \cdot \binom{X}{\mid A \mid} = \sum_{R\subseteq A^2} \binom{X}{\mid A \mid}$$

where R ranges over all permutations of A (as binary relations over A).

6. Equivalence of counting φ-colorings and SOL-polynomials

The following three theorems relate the counting functions of multi-colorings to **SOL** polynomials. Theorems 6.1 and 6.2 show that the class of the counting functions of **SOL**-definable $\varphi(\overline{F})$-multi-colorings with several partial functions $\overline{F}$ and the classes of standard and FF **SOL**-polynomials coincide.

Theorem 6.1. *Let* $\mathcal{P}_\varphi(\overline{F})$ *be an* **SOL**-*definable multi-coloring property. The graph polynomial* $\chi_{\varphi(\overline{F})}(\mathfrak{M};\overline{k})$ *is both a standard and an* FF **SOL**-*polynomial.*

The proof of Theorem 6.1 is given in Subsection 6.1

Theorem 6.2 states that every standard or FF **SOL**-polynomial is an evaluation of some $\chi_\varphi(\mathfrak{M},\overline{k})$.

THEOREM 6.2. *Let $P(\mathfrak{M}; k_1, \dots, k_m)$ be either a standard or an* FF **SOL**-*polynomial. There exists an* **SOL**-*definable multi-coloring property $\mathcal{P}_\varphi$ with $m+l$ color-sets, $[k_1], \dots, [k_{m+l}]$, and $a_1, \dots, a_l \in \mathbb{Z}$ such that*

$$\chi_\varphi(\mathfrak{M}; k_1, \dots, k_m, a_1, \dots, a_l) = P(\mathfrak{M}; k_1, \dots, k_m)$$

where $\chi_\varphi(\mathfrak{M}; k_1, \dots, k_m, k_m, a_1, \dots, a_l)$ is obtained by evaluating the indeterminates $k_{m+1}, \dots, k_{m+l}$ to $a_1, \dots, a_l$ respectively in $\chi_\varphi(\mathfrak{M}; k_1, \dots, k_{m+l})$.

In fact, it will not be difficult to see that it is enough to have $l = 1$ with $a_{m+1} = -1$. The proof of Theorem 6.1 is given in Subsection 6.2

Theorem 6.3 shows that the class of counting functions of **SOL**-definable $\varphi(\overline{R})$-multi-colorings with bounded relations $\overline{R}$ and the class of Newton **SOL**-polynomials coincide.

THEOREM 6.3. *Let P be a function from the class of finite τ-structures to the ring $\mathbb{Q}[\overline{x}]$. The following statements are equivalent:*

(i) *P is a Newton* **SOL**-*polynomial.*
(ii) *P is an evaluation of the counting function $\chi_{\varphi(\overline{R})}(\mathfrak{M}; \overline{k})$ of an* **SOL**-*definable multi-coloring where the relations in $\overline{R}$ are bounded.*

The proof of a theorem similar to Theorem 6.3 is given in the conference version of this paper [**42**]. In Subsection 6.3 we motivate the need to extend partial functions to bounded relations in order to capture the Newton **SOL**-polynomials and sketch this direction of the proof. For the other direction of Theorem 6.3, one augments the proof of Theorem 6.1 given in Subsection 6.1 by using Proposition 3.8 instead of Proposition 3.10.

6.1. Proof of Theorem 6.1. We prove the theorem in the case of φ-colorings. The case of multiple indeterminates and several simultaneous functions is similar. Let $\mathcal{P}_\varphi$ be an **SOL**-definable coloring property. From Proposition 3.10 we know that for every $\mathfrak{M}$ the number of elements given by $\chi_{\varphi(R)}(\mathfrak{M}, k)$ is a polynomial in k of the form

$$\sum_{j=0}^{d \cdot |M|^m} d_\varphi(\mathfrak{M}, j) \cdot k_{(j)}$$

where $d_\varphi(\mathfrak{M}, j)$ is the number of φ-colorings F using all the colors in $[j]$. In other words, if $[k]$ was ordered, $d_\varphi(\mathfrak{M}, j)$ would count the number of φ-colorings F with a fixed set of j colors which are minimal lexicographically among φ-colorings F' obtained from F by permuting the color-set. The total number of colors used is bounded by $N = d \cdot \mid M \mid^m$. Hence we can interpret the set of colors used inside $\mathfrak{M}$ by the set $[M]^{d \cdot m}$. Since $\mathfrak{M}$ has a linear order $\leq_M$, a lexicographic order on $[M]^{d \cdot m}$ is definable in **SOL**.

We replace F by a relation where each occurrence of a color is substituted by a $(d \cdot m)$-tuple, and call this new relation S. We also modify the formula φ to a formula ψ by adding the requirement that all the colors used by S form an initial segment and that S is the smallest in the lexicographic order induced on the colors among its permutations. Let us denote by I_S the initial segment of this lexicographic ordering of the colors used by S. Clearly I_S is definable in **SOL** by a formula ρ.

We have that $\chi_{\varphi(R)}(\mathfrak{M}, k)$ is an FF **SOL**-polynomial given by

$$\sum_{S:\psi(S)} \sum_{I_S:\rho} x_{(card_{\mathfrak{M},\overline{v}}(\overline{v} \in I_S))} \tag{6.1}$$

By Proposition 5.2, $\chi_{\varphi(R)}(\mathfrak{M}, k)$ is a standard **SOL**-polynomial.

6.2. Proof of Theorem 6.2. We prove Theorem 6.2 first for standard **SOL**-monomials only in Lemma 6.4, then for **SOL**-polynomials.

LEMMA 6.4. *Every standard* **SOL**-*monomial* $\Psi(\mathfrak{M})$ *is an evaluation over* $\mathbb{Z}$ *of the counting function of* φ-*multi-colorings.*

PROOF. By Lemma 4.6, $\Psi(\mathfrak{M})$ is of the form

$$r_1^{card_{\mathfrak{M},\overline{v}}(\varphi_1(\overline{v}))} \cdots r_t^{card_{\mathfrak{M},\overline{v}}(\varphi_t(\overline{v}))}$$

$\varphi_1, \ldots, \varphi_t \in$ **SOL** and $r_1, \ldots, r_t \in (\mathbb{Z}-\{0\}) \cup \mathbf{X}$. Without loss of generality, assume $r_1 = x_1, \ldots, r_{t'} = x_{t'} \in \mathbf{X}$ and $r_{t'+1} = c_{t'+1}, \ldots, r_t = c_t \in \mathbb{Z} - \{0\}$. Then

$$x^{card_{\mathfrak{M},\overline{v}}(\varphi)(\overline{v})} = \chi_{\psi(F)}$$

where $\psi(F)$ counts functions F such that if $\overline{a}$ does not satisfy φ then $F(\overline{a}) = f_1$, where f_1 is the minimal element in the linear ordering of $\mathfrak{M}$. Similarly, it holds that $c^{card_{\mathfrak{M},\overline{v}}(\varphi(\overline{v}))} = \chi_{\psi(F,c)}$, i.e. $c^{card_{\mathfrak{M},\overline{v}}(\varphi(\overline{v}))}$ is obtained by evaluating $\overline{x}$ to $\overline{c}$ in $\chi_{\psi(F)}$. By Proposition 3.17, the set of counting functions for φ-multi-colorings is closed under finite product. □

Let $P(\mathfrak{M}, \overline{k})$ be a standard **SOL**-polynomial. As described in Subsection 3.3, we may assume P is given as follows:

$$P(\mathfrak{M}, \overline{k}) = \sum_{\overline{F}:\langle \mathfrak{M},\overline{F}\rangle \models \phi(\overline{F})} t(\langle \mathfrak{M}, \overline{F}\rangle)$$

where $\overline{F}$ is a tuple of functions and $t(\langle \mathfrak{M}, \overline{F}\rangle)$ is an **SOL**-polynomial. By induction on the structure of $t(\langle \mathfrak{M}, \overline{F}\rangle)$ is the evaluation of some counting function of φ-multi-colorings

$$t(\langle \mathfrak{M}, \overline{F}\rangle) = \chi_{\theta(\overline{F'})}(\langle \mathfrak{M}, \overline{F}\rangle, \overline{k}, \overline{a})$$

Then $P(\mathfrak{M}, \overline{k}) = \chi_{\theta(\overline{F},\overline{F'}) \wedge \phi(\overline{F})}(\mathfrak{M}, \overline{k}, \overline{a})$. Using Proposition 5.2, the case of FF **SOL**-polynomials follows.

6.3. From Newton SOL-polynomials to counting bounded relations. Now we prove, by example, direction (i) → (ii) of Theorem 6.3. Note that in this example the coloring relations cannot be replaced by partial functions.

Let

$$N(\mathfrak{M}) = \sum_{A \subseteq M} \binom{x}{card_{\langle \mathfrak{M},A\rangle,v}(v \in A)} = \sum_{A \subseteq M} \binom{x}{|A|} \tag{6.2}$$

We will show how to transform $N(\mathfrak{M})$ into a counting function of multi-colorings. The coloring property needs to consist of structures $\mathfrak{M}_{\overline{R},x}$ with the $\overline{R}$ bounded relations which are not necessarily partial functions.

The term $\binom{x}{|A|}$ counts the number of ways to choose a set of colors of size $|A|$ from $[x]$. Therefore, $N(\mathfrak{M})$ counts relations $R \subseteq A \times [x]$ such that there exists an $I \subseteq [x]$ for which $R = A \times I$ and $|I| = |A|$. It is not difficult to see that this can be expressed in **SOL** by a formula $\varphi_{choose}(R)$. Moreover, every such R is 1-bounded, so the multi-coloring property $P_{\varphi_{choose}}$ is bounded. However, R is not a (partial) function.

Let $N_\theta(\mathfrak{M})$ be the Newton polynomial obtained by adding a definability condition on A under the summation of Equation (6.2),

$$N_\theta(\mathfrak{M}) \sum_{A \subseteq M : \theta} \binom{x}{|A|}$$

Then

$$N_\theta(\mathfrak{M}) = \chi_{\varphi_{choose}(R) \wedge \theta(domain(R))}(\mathfrak{M}, x)$$

The extension to any Newton **SOL**-polynomial is not difficult.

7. MSOL-polynomials

An **SOL**-polynomial $P(\mathfrak{M})$ is an **MSOL**-polynomial if the summations are over unary relations and all the formulas involved are **MSOL**-formulas.

A simple example is the *independence polynomial* $Ind(G, X)$ with

$$Ind(G, X) = \sum_{A \subseteq V} X^{|A|}$$

where $G = (V, E)$ and A ranges over all independent sets of G. The condition of being an independent set can be expressed in **MSOL** where A is a free set variable.

If we look at the graph G as a two-sorted structure $G = (V, E; R)$, where we have a sort for vertices V and a sort for edges E, and an incidence relation R, then the matching polynomial $g(G, X)$ is also an **MSOL**-polynomial. In general, over ordered graphs, many classical graph polynomials, such as the dichromatic polynomial, the Tutte polynomial and the interlace polynomials, can be written as **MSOL**-polynomials. For the case of the various interlace polynomials this needs a proof, cf. [**21**]. In general, computing the coefficients of **SOL**-polynomials, and even **MSOL**-polynomials, can be hard, in fact $\sharp$**P**-hard. However, **MSOL**-polynomials are easy to compute for graph classes of bounded tree-width, cf. [**48, 47, 22**].

DEFINITION 7.1. *A φ-multi-coloring is an* **MSOL**-*multi-coloring if the coloring relations $\overline{F} = (F_1, \ldots, F_k)$ in the formula φ are all unary functions and $\varphi \in$* **MSOL**$(\tau_{\overline{\mathbf{F}}})$.

Inspecting the proof of 6.2 one can verify the following:

PROPOSITION 7.2. *Every* **MSOL**-*polynomial $P(\mathfrak{M})$ is an evaluation of some* **MSOL**-*multi-coloring $\chi_\varphi(\mathfrak{M}, \overline{k})$.*

The converse is, unfortunately, not true. As an example, we look at the harmonious colorings $F : V \to [k]$, which are proper vertex colorings such that each pair of colors occurs at most once along some edge. This can be written as an **MSOL**-formula φ_{harm}. In [**31**, Theorem 10] it is shown that the counting function of harmonious colorings is not an **MSOL**-polynomial. Combining the results of [**28**] and [**31**] one can show that computing its coefficients is **NP**-hard even for trees.

Using the methods developed in [**31**] one can also show other graph polynomials are not **MSOL**-polynomials, e.g., the counting function of rainbow colorings.

8. Enter categoricity

In this section we present an even more general approach to graph polynomials, using advanced first order model theory, in particular the theory of categorical structures. We would like to remind the reader that in this section we require some background in model theory, which goes beyond what was needed in the previous sections. A good background reference is [**39**]. A bit more elementary and still providing necessary background on categoricity is the monograph [**54**].

We first describe a uniform method of attaching to each member G of a family of finite structures $\mathcal{G}$ an infinite structure $M(G)$. The reader can think of $\mathcal{G}$ as the class of finite graphs, but our construction works for arbitrary finite τ-structures.

In the simplest case, the structure $M(G) = M(G, D)$ depends on an infinite set $D = \mathbb{N}$. The structure $M(G, \mathbb{N})$ encodes the family of structures $\langle G, [j], F\rangle = G_{F,j}$ introduced in Section 3, but contains not only the coloring function F, but an infinite set of possible colorings, all first order definable in $M(G)$, and using the infinite set D as colors. The coloring functions (or relations) appear here as elements, and **SOL**-definability reduces to **FOL**-definability.

This approach is extended to definable sets in $M(G, \mathbb{N})$. Correspondingly, if instead of $\mathbb{N}$ we use k-many copies of $\mathbb{N}$ we get generalized multi-colorings. The novelty here is that we allow D to carry more structure, giving rise to a richer class of generalized colorings.

8.1. Background on categoricity. We quote from [**39, 54**]. We assume that all vocabularies are countable or finite. A *theory* $T \subseteq \mathbf{FOL}(\tau)$ is a consistent (satisfiable) set of first order sentences over the vocabulary τ. For a τ-structure $\mathfrak{M}$ we denote by $Th(\mathfrak{M})$ the set of $\mathbf{FOL}(\tau)$-sentences true in $\mathfrak{M}$.

Definition 8.1 (Background). *Let $T \subseteq \mathbf{FOL}(\tau)$ be a theory.*

(i) *T is* complete*, if it is maximal consistent.*
(ii) *T has the* finite model property*, if each finite subset of T has a finite model.*
(iii) *Let κ be a cardinal (initial ordinal). T is κ-categorical if T has an infinite model and any two models of cardinality κ are isomorphic.*
(iv) *An element $a \in M$ is* algebraic *over $C \subseteq M$ if there is τ-formula $\phi(x, \overline{c})$ with one free variable x and parameters $\overline{c}$ from C, such that the set*
$$\{b \in M : \mathfrak{M} \models \phi(b, \overline{c})\}$$
is finite and $\mathfrak{M} \models \phi(a, \overline{c})$.
(v) *In a structure $\mathfrak{M}$ we define the* algebraic closure of a set *$C \subseteq M$, denoted by $acl(C)$, as the set of elements in M which are algebraic over C.*

Facts 1.

(i) *If T is κ-categorical for some infinite κ, and has no finite models, then T is complete (Vaught's Test).*
(ii) *If T is κ-categorical for some uncountable κ, then T is κ'-categorical for all uncountable κ' (Morley's Theorem).*
(iii) *Hence there are two cases which can occur independently in all combinations: T is (or is not) ω-categorical, or T is (or is not) ω_1-categorical. A complete theory which is categorical in all infinite powers is called* totally categorical.

For the more complex notions related to the structure theory of totally categorical theories, such as *C-definable sets, minimal and strongly minimal sets, rank, dimension, ω-stability, etc.*, we refer the reader to the standard texts, e.g. [**9, 39, 54, 17**]. These notions are not used in our technical proofs, but they are mentioned in theorems needed in the proofs. Given a structure $\mathfrak{M}$ the rank of a subset $S \subseteq M$ is denoted by $\mathrm{rk}(S)$.

8.2. The Functor. Let $\mathcal{G} = \mathcal{G}(\tau_0)$ be a class of finite structures for a finite vocabulary τ_0. Let $D_1, \ldots, D_k$ be countable infinite structures for finite vocabularies $\tau_1, \ldots, \tau_k$, respectively.

For every $G \in \mathcal{G}$ we construct the structure $M(G, F, D_1, \ldots, D_k)$ with sorts $G, F, D_1, \ldots, D_k$, and with the vocabulary $\tau = \tau_0 \cup \tau_1 \cdots \cup \tau_k$ and an extra function symbol

$$\Phi : G \times F \to D_1 \times \ldots \times D_k$$

The sort F encodes all the functions from G to $D_1 \times \ldots \times D_k$. We think of these functions as colorings of elements (vertices) of G with a tuple of k colors from the color sets $D_1, \ldots, D_k$. If we wanted to color edges, given as pairs of vertices, or more general, tuples of elements of G, one has to modify our construction correspondingly.

To ensure that the elements of F encode all functions, we require that Φ satisfies the following conditions:

(i) $\exists f \in F$

(ii) $\forall f, f' \in F([\forall g \in G\ \Phi(g, f) = \Phi(g, f')] \to f = f')$

(iii) $\forall f \in F \forall g \in G \forall \overline{d} \in D_1 \times \ldots \times D_k \exists f' \in F$
$\big((\forall g'(g \neq g' \to \Phi(g', f) = \Phi(g', f')) \wedge \Phi(g, f') = \overline{d})\big)$

(i) says that F is not empty, (ii) says that the elements of F are functions, and (iii) says that every one point modification of a function in F is again a function in F. Because G is finite, this ensures that all functions from G to $D_1 \times \ldots \times D_k$ are in F. In other words we have the canonical identification

$$\Phi^\star : F \leftrightarrow (D_1 \times \cdots \times D_k)^G$$

and fixing an enumeration of G we may identify the right-hand-side with the Cartesian power

$$(D_1 \times \cdots \times D_k)^{|G|}$$

We write $f(g)$ instead of $\Phi(g, f)$ and so identify elements $f \in F$ with functions $G \to D_1 \times \cdots \times D_k$.

Remark 8.2. *By the virtue of the construction, given $D_1, \ldots, D_k$, the isomorphism type of $M(G, F, D_1, \ldots, D_k)$ depends only on G. Obviously, G can be recovered from $M(G, F, D_1, \ldots, D_k)$. So, $M(G, F, D_1, \ldots, D_k)$ can be seen as the complete invariant of G. In particular, every definable subset S of F is an invariant of G.*

Proposition 8.3. *$M(G, F, D_1, \ldots, D_k)$ is definable using parameters in the disjoint union $D_1 \sqcup \cdots \sqcup D_k$.*

Proof. Obviously $M(G, F, D_1, \ldots, D_k)$ is definable in the disjoint union of G, F and $D_1, \ldots, D_k$. But as G is finite, one can interpret this sort using $|G|$ constants. $\square$

COROLLARY 8.4.

(i) *Assume that the theory of each* D_i *is* ω*-categorical. Then the theory* $\mathrm{Th}[M(G, D_1, \ldots, D_k)]$ *is* ω*-categorical.*

(ii) *Assume that the theory of each* D_i *is strongly minimal. Then the theory* $\mathrm{Th}[M(G, D_1, \ldots, D_k)]$ *is* ω*-stable with* k *independent dimensions. If* $k = 1$ *then the theory is categorical in uncountable cardinals.*

THEOREM 8.5 (B. Zilber). *Any theory satisfying the conclusions of (i) and (ii) has the finite model property. Moreover any countable model* M *can be represented as a union of an increasing chain of finite substructures* M_i (logically) approximating M*, i.e.,*

$$M = \bigcup_{i=1}^{\infty} M_i$$

PROOF. This follows from Theorem 7 of [**19**], where also more details may be found. □

REMARK 8.6. *The finite model property takes a very simple form for a strongly minimal structure* D*. Namely,* D *has the finite model property if and only if* $\mathrm{acl}\,(X)$ *is finite for any finite* $X \subseteq D$*.*

8.3. Counting functions for definable sets. A very important consequence of the finite model property is the possibility to introduce a stronger *counting function* on definable sets.

We prove here the existence of the counting polynomials in a special case, for the theory $\mathrm{Th}[M(G, D_1, \ldots, D_k)]$. The more general case of ω-categorical ω-stable theories can be found in [**19**, Proposition 5.2.2.]. The more special case of theories categorical in all infinite cardinals has been proved in [**59, 60**] and can be found in [**61**]. The proof under the special assumptions needed in this paper is really elementary and does not require any model-theoretic terminology if one assumes the D_i's to be just sets. It really is a slight generalization of the proof given for Proposition 3.8.

THEOREM 8.7. *Let* $M = M(G, D_1, \ldots, D_k)$*. Assume the finite model property holds in the strongly minimal structures* $D_1, \ldots D_k$*. Then for every finite* $C \subseteq M$ *and any* C*-definable set* $S \subseteq M^\ell$ *there is a polynomial* $p_S \in \mathbb{Q}[x]$ *and there is a number* n_S *such that for every finite* $X \subseteq M$ *with* $C \subseteq X$*,*

(i) *if* $|D_i \cap \mathrm{acl}\,(X)| = x_i \geq n_S$*, we have* $|S \cap \mathrm{acl}\, X| = p_S(x_1, \ldots, x_k)$*;*

(ii) $\mathrm{rk}(S) = \deg(p_S)$*, the degree of the polynomial;*

(iii) *if g(S)=T for some automorphism* g *of* M *then* $p_S = p_T$ *and* $n_S = n_T$*.*

Furthermore, if $C = \emptyset$ *we can take* $n_S = 0$*.*

PROOF. We construct the polynomial for a given S by induction on $\mathrm{rk}(S)$.

W.l.o.g. we may assume that S is an atom over C, that is defined by a principal type over C.

Let $f = \langle f_1, \ldots, f_\ell \rangle \in S$. Recall that each f_i is determined by the values of $f_i(g) \in D_1 \times \cdots \times D_k$, for $g \in G$. Denote $f_{im}(g)$ the mth co-ordinate of $f_i(g)$, an element of D_m.

Suppose $f_{im}(g) \in \mathrm{acl}\,(C)$ for all $i \leq \ell$, $m \leq k$ and $g \in G$. Then $f \in \mathrm{acl}\,(C)$. Since S is an atom, $S \subseteq \mathrm{acl}\,(C)$ and hence

$$|S \cap \mathrm{acl}\,(X)| = |S \cap \mathrm{acl}\,(C)|$$

is a constant, independent of X. So, we are done in this case.

We may now assume that $f_{11}(g_0) \notin \mathrm{acl}\,(C)$. So, we have the partition

$$S = \bigcup_{a \in D_1 \smallsetminus \mathrm{acl}\,(C)} S_a, \quad S_a = \{f \in S : f_{11}(g_0) = a\}$$

Since $D_1 \smallsetminus \mathrm{acl}\,(C)$ is an atom over C (use the strong minimality of D_1) and, of course $G \subseteq \mathrm{acl}\,(\emptyset)$, the subgroup of the automorphism group of M fixing C acts transitively on $D_1 \smallsetminus \mathrm{acl}\,(C)$. Hence all the fibers S_a are conjugated by automorphisms over C and have the same Morley rank. The latter implies by the addition formula for ranks that

$$\mathrm{rk}(S_a) = \mathrm{rk}(S) - 1$$

So, we may apply the induction hypothesis. By (iii) we get that

$$p_{S_a} = p_0, \text{ for all } a \in D_1 \smallsetminus \mathrm{acl}\,(C)$$

for some polynomial p_0. By (ii) $\deg(p_0) = \mathrm{rk}(S_a) = \mathrm{rk}(S) - 1$.

Let $c_0 = |\mathrm{acl}\,(C)|$. So,

$$|(D_1 \smallsetminus \mathrm{acl}\,(C)) \cap \mathrm{acl}\,(X)| = (x_1 - c_0).$$

We further calculate

$$S \cap \mathrm{acl}\,(X) = \bigcup_{a \in (D_1 \smallsetminus \mathrm{acl}\,(C)) \cap \mathrm{acl}\,(X)} S_a \cap \mathrm{acl}\,(X) = (x_1 - c_0) \cdot p_0(x_1, \ldots, x_k)$$

$\square$

8.4. Generalized chromatic polynomials revisited. In the light of Theorem 8.7 let us look first at the generalized colorings of Section 3.

We discuss them for the class $\mathcal{G}(\tau)$ of finite (purely relational) τ-structures. We denote by $\mathbf{SOL}^n(\tau)$ the set of $\mathbf{SOL}(\tau)$-formulas where all second order variables have arity at most n. Let $\phi(\overline{\mathbf{R}}, \mathbf{F}) \in \mathbf{SOL}^n(\tau)$ define a notion of generalized coloring where $\overline{\mathbf{R}}$ is a list of relation parameters, and $\mathbf{F}$ denotes the coloring function. So the generalized chromatic polynomial on a τ-structure $\mathfrak{A}$ is defined as

$$\chi_{\phi(\overline{\mathbf{R}}, \mathbf{F})}(\mathfrak{A}, k) = |\ \{(\overline{R}, F) : \langle \mathfrak{A}, \overline{R}, F, [k] \rangle \models \phi(\overline{R}, F)\}\ |$$

We first expand $\mathfrak{A}$ so that quantification over relations becomes quantification over elements. So for each $\ell \leq n$ we add the set $\wp(A^\ell)$ with the corresponding membership relation $\in_\ell$. We define the $\tau^\star$-structure

$$\mathfrak{A}^\star = \langle \mathfrak{A}, \wp(A^\ell) \in_\ell, \ell \leq n \rangle$$

and apply our functor $M(\mathfrak{A}^\star, \mathbb{N})$ to it with $D_1 = \mathbb{N}$. Let $\tau^\sharp$ be the vocabulary of $M(\mathfrak{A}^\star, \mathbb{N})$.

Now the formula $\phi(\overline{\mathbf{R}}, \mathbf{F}) \in \mathbf{SOL}^n(\tau)$ has a straightforward translation

$$\phi^\sharp(\overline{c}_{\overline{R}}, d_F) \in \mathbf{FOL}(\tau^\sharp)$$

where the function symbol F becomes a variable d_F, the relation symbols $\overline{\mathbf{R}}$ become variables $\overline{c}_{\overline{\mathbf{R}}}$ of the appropriate sorts. Furthermore, it has no additional parameters. It follows that $C = \emptyset$.

Let $X \subseteq \mathbb{N}$ be finite. Due to Theorem 8.7(iii), w.l.o.g., $C = [k]$ for some $k \in \mathbb{N}$. Let $\mathfrak{A}^\star_k$ be the substructure of $M(\mathfrak{A}^\star, \mathbb{N})$ with universe $acl([k])$ and $F_k \subseteq F$ be its part of the sort F. We now easily verify that:

(i) $\mathfrak{A}^\star_{\mathfrak{k}}$ contains all of $\mathfrak{A}^\star$.

(ii) F_k consists exactly of all functions f with range $Rg(f) \subseteq [k]$.
(iii) For $S = \{(\overline{c}, d) \in M(\mathfrak{A}^\star, \mathbb{N}) : M(\mathfrak{A}^\star, \mathbb{N}) \models \phi^\sharp(\overline{c}, d)\}$
we have that $| S \cap \mathrm{acl}([k]) | = p_S(k)$ is a polynomial for every $k \geq n_S = 0$.
(iv) $\chi_{\phi(\overline{\mathbf{R}}, \mathbf{F})}(\mathfrak{A}, k) = | \{(\overline{c}, d) \in \mathfrak{A}^\star_k : M(\mathfrak{A}^\star, \mathbb{N}) \models \phi^\sharp(\overline{c}, d)\} | = p_S(k)$.

This proves Theorem B for the case of generalized chromatic polynomials in one variable.

8.5. Proof of Theorem B. To prove Theorem B in its full generality proceed as before. We observe the following points:

- For multi-colorings we use several copies of $\mathbb{N}$ as strongly minimal sets.
- If the generalized colorings are relations $r \subseteq G^\alpha \times \mathbb{N}^\beta$ the proof still works, provided $M(G, \mathbb{N}, \ldots, \mathbb{N})$ is ω-stable. This is where we use, in our definition of generalized multi-coloring that, for each $\overline{x} \in G^\alpha$, the set $r_{\overline{a}} = \{\overline{b} \in \mathbb{N}^\beta : r(\overline{a}, \overline{b})\}$ is bounded by a fixed finite number d. Without this restriction ω-categoricity is violated.

8.6. The full generality. The general theorem allows for more complicated strongly minimal structures to be used for D_1. A simple example would consist of a countable set of disjoint copies of a fixed finite structure such as a finite field $GF(p^q)$. The colors then would be pairs (n, a) where $n \in \mathbb{N}$ and $a \in GF(p^q)$. We could request that a graph coloring f of a graph $G = (V, E)$ satisfies, say,

$$[((u, v) \in E \wedge f(u) = (n_u, a_u) \wedge f(v) = (n_v, a_v)) \rightarrow$$
$$(n_u \neq n_v \wedge a_u + a_v = 0)]$$

It seems possible that such colorings may be useful in modeling wiring conditions when labeled graphs model network devices.

9. Conclusions

Starting with the classical chromatic polynomial we have introduced multi-colorings of graphs. We have shown that the corresponding counting functions are always polynomials. We have then shown that the class of counting functions of multi-colorings is very rich and covers virtually all examples of graph polynomials which have been studied in the literature. Additionally, it gives rise to counting graph invariants which previously were not recognized to be graph polynomials.

Motivated by the class of **SOL**-definable graph polynomials introduced in [**47**], we introduced variations of **SOL**-definable polynomials using different bases of the polynomial ring: the standard basis, the falling factorial bases and the Newton polynomials. We have then shown that the class of **SOL**-graph polynomials coincides with the class of **SOL**-definable generalized chromatic polynomials. This, along with the extensive scope of the class, suggests that the frameworks presented in this paper are natural for the study of graph polynomials.

Finally, we have constructed functors which map graphs (or other finite relational structures) into $\aleph_0$-categorical ω-stable structures of rank k which in a precise sense encode all **SOL**-graph polynomials in k indeterminates.

Theorems B and C can also be used to analyze the complexity of evaluations of **SOL**-definable polynomials at integer points. They fit nicely into the framework developed by S. Toda in his unpublished thesis and in [**58**].

Acknowledgments. We would like to thank I. Averbouch, A. Blass, S. Burris, B. Courcelle, B. Godlin, E. Hrushovski, S. Shelah and M. Ziegler for valuable discussions and suggestions. We would like to thank A. Blass for allowing us to incorporate the simple proof of Theorem 3.8, which he suggested. Finally we would like to thank the referees for their careful reading and constructive suggestions.

References

[1] D. Achlioptas, *The complexity of G-free colourability*, Discrete Math **165** (1997), 21–30.
[2] M. Aigner and H. van der Holst, *Interlace polynomials*, Linear Algebra and Applications **377** (2004), 11–30.
[3] N. Alon, G. Ding, B. Oporowski, and D. Vertigan, *Partitioning into graphs with only small components*, Journal of Combinatorial Theory, Series B **87** (2003), 231–243.
[4] N. Alon, C. McDiarmid, and B. Reed, *Acyclic coloring of graphs*, Random Structures and Algorithms **2.3** (1991), 277–288.
[5] R. Arratia, B. Bollobás, and G.B. Sorkin, *The interlace polynomial of a graph*, Journal of Combinatorial Theory, Series B **92** (2004), 199–233.
[6] ———, *A two-variable interlace polynomial*, Combinatorica **24.4** (2004), 567–584.
[7] I. Averbouch, B. Godlin, and J.A. Makowsky, *An extension of the bivariate chromatic polynomial*, European Journal of Combinatorics **31.1** (2010), 1–17.
[8] I. Averbouch, *Completeness and universality properties of graph invariants and graph polynomials*, Ph.D. thesis, Technion - Israel Institute of Technology, Haifa, Israel, 2011.
[9] J.T. Baldwin, *Fundamentals of stability theory*, Perspectives in Mathematical Logic, vol. 12, Springer, 1988.
[10] N. Biggs, *Algebraic graph theory, 2nd edition*, Cambridge University Press, 1993.
[11] G.D. Birkhoff, *A determinant formula for the number of ways of coloring a map*, Annals of Mathematics **14** (1912), 42–46.
[12] B. Bollobás, *Modern graph theory*, Springer, 1999.
[13] B. Bollobás and O. Riordan, *A Tutte polynomial for coloured graphs*, Combinatorics, Probability and Computing **8** (1999), 45–94.
[14] J. I. Brown, *The complexity of generalized graph colorings*, Discrete Applied Mathematics **69** (1996), no. 3, 257 – 270.
[15] J. I. Brown and D. G. Corneil, *On generalized graph colorings*, J. Graph Theory **11** (1987), no. 1, 87–99.
[16] ———, *Graph properties and hypergraph colourings*, Discrete Mathematics **98** (1991), no. 2, 81 – 93.
[17] S. Buechler, *Essential stability theory*, Perspectives in Mathematical Logic, vol. 4, Springer, 1996.
[18] G. Chartrand, G. L. Johns, K. A. McKeon, and P. Zhang 0004, *The rainbow connectivity of a graph*, Networks **54** (2009), no. 2, 75–81.
[19] G. Cherlin and E. Hrushovski, *Finite structures with few types*, Annals of Mathematics Studies, vol. 152, Princeton University Press, 2003.
[20] F.R.K. Chung and R.L. Graham, *On the cover polynomial of a digraph*, Journal of Combinatorial Theory, Ser. B **65** (1995), no. 2, 273–290.
[21] B. Courcelle, *A multivariate interlace polynomial and its computation for graphs of bounded clique-width*, The Electronic Journal of Combinatorics **15** (2008), R69.
[22] B. Courcelle, J.A. Makowsky, and U. Rotics, *On the fixed parameter complexity of graph enumeration problems definable in monadic second order logic*, Discrete Applied Mathematics **108** (2001), no. 1-2, 23–52.
[23] R. Diestel, *Graph theory*, Graduate Texts in Mathematics, Springer, 1996.
[24] F.M. Dong, K.M. Koh, and K.L. Teo, *Chromatic Polynomials and Chromaticity of Graphs*, World Scientific, 2005.
[25] H.-D. Ebbinghaus and J. Flum, *Finite model theory*, Perspectives in Mathematical Logic, Springer, 1995.
[26] H.-D. Ebbinghaus, J. Flum, and W. Thomas, *Mathematical logic, 2nd edition*, Undergraduate Texts in Mathematics, Springer-Verlag, 1994.

[27] K. Edwards, *The harmonious chromatic number and the achromatic number*, Survey in Combinatorics (R. A. Bailey, ed.), London Math. Soc. Lecture Note Ser., vol. 241, Cambridge Univ. Press, 1997, pp. 13–47.

[28] K. Edwards and C. McDiarmid, *The complexity of harmonious colouring for trees*, Discrete Appl. Math. **57** (1995), no. 2-3, 133–144.

[29] J. Gimbel and J. Nešetřil, *Partitions of graphs into cographs*, Electronic Notes in Discrete Mathematics **11** (2002), 705 – 721, The Ninth Quadrennial International Conference on Graph Theory, Combinatorics, Algorithms and Applications.

[30] B. Godlin, E. Katz, and J.A. Makowsky, *Graph polynomials: From recursive definitions to subset expansion formulas*, Journal of Logic and Computation **in press** (2011), xx–yy, DOI: 10.1093/logcom/exq006.

[31] B. Godlin, T. Kotek, and J.A. Makowsky, *Evaluation of graph polynomials*, 34th International Workshop on Graph-Theoretic Concepts in Computer Science, WG08, Lecture Notes in Computer Science, vol. 5344, 2008, pp. 183–194.

[32] C.D. Godsil, *Algebraic combinatorics*, Chapman and Hall, 1993.

[33] B. Grunbaum, *Acyclic colorings of planar graphs*, Israel J. Math. **14** (1973), 390–412.

[34] I. Gutman and F. Harary, *Generalizations of the matching polynomial*, Utilitas Mathematicae **24** (1983), 97–106.

[35] G. Hahn, J. Kratochvil, J. Siran, and D. Sotteau, *On the injective chromatic number of graphs*, Discrete mathematics **256.1-2** (2002), 179–192.

[36] F. Harary, *Conditional colorability in graphs*, Graphs and Applications, Proc. First Col. Symp. Graph Theory (Boulder, Colo., 1982), Wiley-Intersci. Publ., Wiley, New York, 1985, pp. 127–136.

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

[38] F. Harary and L.H. Hsu, *Conditional chromatic number and graph operations*, Bull. Inst. Math. Acad. Sinica **19.2** (1991), 125–134.

[39] W. Hodges, *Model theory*, Encyclopedia of Mathematics and its Applications, vol. 42, Cambridge University Press, 1993.

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

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

[42] T. Kotek, J.A. Makowsky, and B. Zilber, *On counting generalized colorings*, Computer Science Logic, CSL'08, Lecture Notes in Computer Science, vol. 5213, 2008, p. 339353.

[43] L. Libkin, *Elements of finite model theory*, Springer, 2004.

[44] N. Linial, J. Matoušek, O. Sheffet, and G. Tardos, *Graph coloring with no large monochromatic components*, arXiv:math/0703362v1, 2007.

[45] L. Lovász and M.D. Plummer, *Matching theory*, Annals of Discrete Mathematics, vol. 29, North Holland, 1986.

[46] J.A. Makowsky, *Colored Tutte polynomials and Kauffman brackets on graphs of bounded tree width*, Proceedings of the 12th Symposium on Discrete Algorithms, SIAM, 2001, pp. 487–495.

[47] ———, *Algorithmic uses of the Feferman-Vaught theorem*, Annals of Pure and Applied Logic **126.1-3** (2004), 159–213.

[48] ———, *Colored Tutte polynomials and Kauffman brackets on graphs of bounded tree width*, Disc. Appl. Math. **145** (2005), no. 2, 276–290.

[49] ———, *From a zoo to a zoology: Descriptive complexity for graph polynomials*, Logical Approaches to Computational Barriers, Second Conference on Computability in Europe, CiE 2006, Swansea, UK, July 2006 (A. Beckmann, U. Berger, B. Löwe, and J.V. Tucker, eds.), Lecture Notes in Computer Science, vol. 3988, Springer, 2006, pp. 330–341.

[50] ———, *From a zoo to a zoology: Towards a general theory of graph polynomials*, Theory of Computing Systems **43** (2008), 542–562.

[51] J.A. Makowsky and B. Zilber, *Polynomial invariants of graphs and totally categorical theories*, MODNET Preprint No. 21,
`http://www.logique.jussieu.fr/modnet/Publications/Preprint%20 server`, 2006.

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

[53] S.D. Noble and D.J.A. Welsh, *A weighted graph polynomial from chromatic invariants of knots*, Ann. Inst. Fourier, Grenoble **49** (1999), 1057–1087.
[54] P. Rothmaler, *Introduction to model theory*, Algebra, Logic and Applications Series, vol. 15, Gordon and Breach Science Publishers, 1995.
[55] A. Sokal, *The multivariate Tutte polynomial (alias Potts model) for graphs and matroids*, Survey in Combinatorics, 2005, London Mathematical Society Lecture Notes, vol. 327, 2005, pp. 173–226.
[56] R. P. Stanley, *Acyclic orientations of graphs*, Discrete Mathematics **5** (1973), 171–178.
[57] P. Tittmann, I. Averbouch, and J.A. Makowsky, *The enumeration of vertex induced subgraphs with respect to the number of components*, European Journal of Combinatorics, to appear in 2011.
[58] S. Toda and O. Watanabe, *Polynomial time 1-Turing reductions from* #PH *to* #P, Theor. Comp. Sc. **100** (1992), 205–221.
[59] B. Zilber, *Strongly minimal countably categorical theories, II*, Siberian Math.J. **25.3** (1984), 71–88.
[60] ———, *Strongly minimal countably categorical theories, III*, Siberian Math.J. **25.4** (1984), 63–77.
[61] ———, *Uncountably categorical theories*, Translations of Mathematical Monographs, vol. 117, American Mathematical Society, 1993.

Faculty of Computer Science, Technion–Israel Institute of Technology, 32000 Haifa, Israel
E-mail address: {tkotek,janos}@cs.technion.ac.il

Mathematical Institute, University of Oxford, Oxford, OX1 3LB, United Kingdom
E-mail address: zilber@maths.ox.ac.uk

This page intentionally left blank

Contemporary Mathematics
Volume **558**, 2011

Counting Homomorphisms and Partition Functions

Martin Grohe and Marc Thurley

1. Introduction

Homomorphisms between relational structures are not only fundamental mathematical objects, but are also of great importance in an applied computational context. Indeed, *constraint satisfaction problems*, a wide class of algorithmic problems that occur in many different areas of computer science such as artificial intelligence or database theory, may be viewed as asking for homomorphisms between two relational structures [**FV98**]. In a logical setting, homomorphisms may be viewed as witnesses for *positive primitive formulas* in a relational language. As we shall see below, homomorphisms, or more precisely the numbers of homomorphisms between two structures, are also related to a fundamental computational problem of statistical physics. Homomorphisms of graphs are generalizations of colorings, and for that reason a homomorphism from a graph G to a graph H is also called an *H-coloring* of G. Note that if H is the complete graph on k-vertices, then an H-coloring of a graph G may be viewed as a proper k-coloring of G in the usual graph theoretic sense that adjacent vertices are not allowed to get the same color.

It is thus no surprise that the computational complexity of various algorithmic problems related to homomorphisms, in particular the *decision problem* of whether a homomorphism between two given structures exists and the *counting problem* of determining the number of such homomorphisms, have been intensely studied. (For the decision problem, see, for example, [**BKN09, Bul06, BKJ05, Gro07, HN90**]. References for the counting problem will be given in Section 3. Other related problems, such as optimization or enumeration problems, have been studied, for example, in [**Aus07, BDGM09, DJKK08, Rag08, SS07**].)

In this article, we are concerned with the complexity of counting homomorphisms from a given structure A to a fixed structure B. Actually, we are mainly interested in a generalization of this problem to be introduced in the next section. We almost exclusively focus on graphs. The first part of the article, consisting of the following two sections, is a short survey of what is known about the problem. In the second part, consisting of the remaining Sections 4-9, we give a proof of a theorem due to Bulatov and the first author of this paper [**BG05**], which classifies

1991 *Mathematics Subject Classification.* 68Q17, 05C30.

The second author was supported in part by Marie Curie Intra-European Fellowship 271959 at the Centre de Recerca Matemàtica, Bellaterra, Spain.

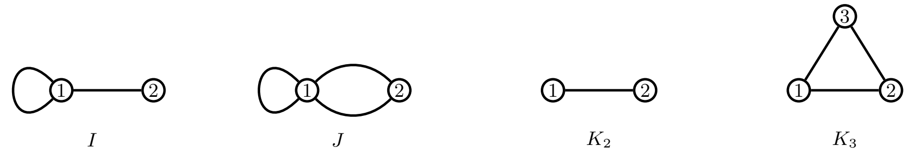


FIGURE 1. The graphs I, J, K_2, K_3

$$A(I) = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \quad A(J) = \begin{pmatrix} 1 & 2 \\ 2 & 0 \end{pmatrix} \quad A(K_2) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad A(K_3) = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

FIGURE 2. The adjacency matrices of the graphs I, K_2, K_3

the complexity of partition functions described by matrices with non-negative entries. The proof we give here is essentially the same as the original one, with a few shortcuts due to [**Thu09**], but it is phrased in a different, more graph theoretical language that may make it more accessible to most readers.

2. From Homomorphisms to Partition Functions

For a fixed graph H we let Z_H be the "homomorphism-counting function" that maps each graph G to the number of homomorphisms from G to H. Several well-known combinatorial graph invariants can be expressed as homomorphism counting functions, as the following examples illustrate:

EXAMPLE 2.1. *Let I be the first graph displayed in Figure 1, and let G be an arbitrary graph. Remember that an* independent set *(or* stable set*) of G is set of pairwise nonadjacent vertices of G. For every set $S \subseteq V(G)$, we define a mapping $h_S : V(G) \to V(I)$ by $h_S(v) = 2$ if $v \in S$ and $h_S(v) = 1$ otherwise. Then h_S is a homomorphism from G to I if and only if S is an independent set. Thus the number $Z_I(G)$ of homomorphisms from G to I is precisely the number of independent sets of G.*

EXAMPLE 2.2. *For every positive integer k, let K_k be the complete graph with vertex set $[k] := \{1, \dots, k\}$ (see Figure 1). Let G be a graph. Recall that a* (proper) k-coloring *of G is a mapping $h : V(G) \to [k]$ such that for all $vw \in E(G)$ it holds that $h(v) \neq h(w)$. Observe that a mapping $h : V(G) \to [k]$ is a k-coloring of G if and only if it is a homomorphism from G to K_k. Hence $Z_{K_k}(G)$ is the number of k-colorings of G.*

Unless mentioned otherwise, graphs in this article are undirected, and they may have loops and parallel edges. Graphs without loops and parallel edges are called *simple*. We always assume the edge set and the vertex set of a graph to be disjoint. The class of all graphs is denoted by $\mathcal{G}$. A *graph invariant* is a function defined on $\mathcal{G}$ that is invariant under isomorphism. The *adjacency matrix* of a graph H is the square matrix $A := A(H)$ with rows and columns indexed by vertices of H, where the entry $A_{v,w}$ at row v and column w is the number of edges from v to w. Figure 2 shows the adjacency matrices of the graphs in Figure 1. For all graphs G, H, we define a *homomorphism* from $G \to H$ to be a mapping $h : V(G) \cup E(G) \to V(H) \cup E(H)$ such that for all $v \in V(G)$ it holds that $h(v) \in V(H)$

and for all edges $e \in E(G)$ with endvertices v, w it holds that $h(e) \in E(H)$ is an edge with endvertices $h(v), h(w)$.[1] The following observation expresses a homomorphism counting function Z_H in terms of the adjacency matrix of H:

OBSERVATION 2.3. *Let H be a graph and $A := A(H)$. Then for every graph G,*

$$Z_H(G) = \sum_{\sigma: V(G) \to V(H)} \prod_{\substack{e \in E(G) \text{ with} \\ \text{endvertices } v,w}} A_{\sigma(v),\sigma(w)}. \tag{2.1}$$

To simplify the notation, we write $\prod_{vw \in E(G)}$ instead of $\prod_{\substack{e \in E(G) \text{ with} \\ \text{endvertices } v,w}}$ in similar expressions in the following. By convention, the empty sum evaluates to 0 and the empty product evaluates to 1. Thus for the empty graph $\emptyset$ we have $Z_H(\emptyset) = 1$ for all H and $Z_\emptyset(G) = 0$ for all $G \neq \emptyset$.

EXAMPLE 2.4. *Consider the second graph J displayed in Figure 1, and let G be an arbitrary graph. $Z_J(G)$ is a weighted sum over all independent sets of G: For all sets $S, T \subseteq V(G)$ we let $e(S,T)$ be the number of edges between S and T. Then*

$$Z_J(G) = \sum_{\substack{S \subseteq V(G) \\ \textit{independent set}}} 2^{e(S, V(G) \setminus S)}.$$

Equation (2.1) immediately suggests the following generalization of the homomorphism counting functions Z_H: For every symmetric $n \times n$ matrix A with entries from some ring $\mathbb{S}$ we let $Z_A : \mathcal{G} \to \mathbb{S}$ be the function that associates the following element of $\mathbb{S}$ with each graph $G = (V, E)$:

$$Z_A(G) := \sum_{\sigma: V \to [n]} \prod_{vw \in E} A_{\sigma(v),\sigma(w)}. \tag{2.2}$$

We call functions Z_A, where A is a symmetric matrix over $\mathbb{S}$, *partition functions* over $\mathbb{S}$. (All rings in this paper are commutative with a unit. $\mathbb{S}$ always denotes a ring.)

With each $n \times n$ matrix A we associate a simple graph $H(A)$ with vertex set $[n]$ and edge set $\{ij \mid A_{i,j} \neq 0\}$. We may view A as assigning nonzero weights to the edges of $G(A)$, and we may view $Z_A(G)$ as a weighted sum of homomorphisms from G to $H(A)$, where the *weight* of a mapping $\sigma : V \to [n]$ is

$$\omega_A(G, \sigma) := \prod_{vw \in E} A_{\sigma(v),\sigma(w)}.$$

Homomorphisms are precisely the mappings with nonzero weight. Inspired by applications in statistical physics (see Section 2.1), we often call the elements of the index set of a matrix, usually $[n]$, *spins*, and we call mappings $\sigma : V \to [n]$ assigning a spin to each vertex of a graph *configurations*.

---

[1]Usually, homomorphisms from G to H are defined to be mappings $g : V(G) \to V(H)$ that preserve adjacency. A mapping $g : V(G) \to V(H)$ is a homomorphism in this sense if and only if it has an extension $h : V(G) \cup E(G) \to V(H) \cup E(H)$ that is a homomorphism as defined above. Thus the two notions are closely related. However, if H has parallel edges, then there a different numbers of homomorphisms for the two notions.

EXAMPLE 2.5. *Recall that a graph G is Eulerian if there is a closed walk in G that traverses each edge exactly once. It is a well-known theorem, which goes back to Euler, that a graph is Eulerian if and only if it is connected and every vertex has even degree. Consider the matrix*

$$U = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

It is not hard to show that for every N-vertex graph G we have $Z_U(G) = 2^N$ if every vertex of G has even degree and $Z_U(G) = 0$ otherwise. Hence on connected graphs, $(1/2^N) \cdot Z_U$ is the characteristic function of Eulerianicity.

EXAMPLE 2.6. *Consider the matrix*

$$B = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Let G be a graph. Then for every $\sigma : V(G) \to [2]$ it holds that

$$\omega_B(G,\sigma) = \begin{cases} 1 & \text{if the induced subgraph } G\big[\sigma^{-1}(2)\big] \text{ has an even number of edges,} \\ -1 & \text{otherwise.} \end{cases}$$

It follows that for every N-vertex graph G,

$$\frac{1}{2} Z_B(G) + 2^{N-1}$$

is the number of induced subgraphs of G with an even number of edges.

EXAMPLE 2.7. *Recall that a* cut *of a graph is a partition of its vertex set into two parts, and the* weight *of a cut is the number of edges from one part to the other. A* maximum cut *is a cut of maximum weight. Consider the matrix*

$$C := \begin{pmatrix} 1 & X \\ X & 1 \end{pmatrix}$$

over the polynomial ring $\mathbb{Z}[X]$. It is not hard to see that for every graph G, the degree of the polynomial $Z_C(G)$ is the weight of a maximum cut of G and the leading coefficient the number of maximum cuts.

Graph polynomials present another important way to uniformly describe families of graph invariants. Examples of graph polynomials are the *chromatic polynomial* and the *flow polynomial*. Both of these are subsumed by the bivariate *Tutte polynomial*, arguably the most important graph polynomial. The following example exhibits a relation between the Tutte polynomial and partition functions. For more on graph polynomials, we refer the reader to the survey [**KMZ11**] in this volume.

EXAMPLE 2.8. *Let $G = (V, E)$ be a graph with N vertices, M edges, and Q connected components. For a subset $F \subseteq E$, by $q(F)$ we denote the number of connected components of the graph (V, F). The* Tutte polynomial *of G is the bivariate polynomial $T(G; X, Y)$ defined by*

$$T(G; X, Y) = \sum_{F \subseteq E} (X-1)^{q(F)-Q} \cdot (Y-1)^{|F|-N+q(F)}.$$

It is characterized by the following contraction-deletion equalities. *For an edge $e \in E$, we let $G \setminus e$ be the graph obtained from G by* deleting *e, and we let G/e be the graph obtained from G by* contracting *e. A* bridge *of G is an edge $e \in E$ such* *that $G \setminus e$ has more connected components than G. A* loop *is an edge that is only incident to one vertex.*

$$T(G;X,Y) = \begin{cases} 1 & \textit{if } E(G)=\emptyset, \\ X \cdot T(G\setminus e;X,Y) & \textit{if } e \in E(G) \textit{ is a bridge}, \\ Y \cdot T(G/e;X,Y) & \textit{if } e \in E(G) \textit{ is a loop}, \\ T(G\setminus e;X,Y) + T(G/e;X,Y) & \textit{otherwise.} \end{cases}$$

Let $r,s \in \mathbb{C}$. It can be shown that the partition function of the $n\times n$ matrix $A(n,r,s)$ with diagonal entries r and off-diagonal entries s satisfies similar contraction-deletion equalities, and this implies that it can be expressed in terms of the Tutte polynomial as follows:

$$Z_{A(n,r,s)}(G) = s^{M-N+Q} \cdot (r-s)^{N-Q} \cdot n^Q \cdot T\left(G; \frac{r+s\cdot(n-1)}{(r-s)}, \frac{r}{s}\right). \tag{2.3}$$

This implies that for all $x,y\in\mathbb{C}$ such that $n := (x-1)\cdot(y-1)$ is a positive integer, it holds that

$$T(G;x,y) = (y-1)^{Q-N} \cdot n^{-Q} \cdot Z_{A(n,y,1)}(G).$$

Example 2.9. *For simplicity, in this example we assume that $G=(V,E)$ is a* simple *graph, that is, a graph without loops and without multiple edges. For every positive integer k, a k*-flow *in G is a mapping $f : V \to \mathbb{Z}_k$ (the group of integers modulo k) such that the following three conditions are satisfied:*

(i) *$f(v,w)=0$ for all $v,w\in V$ with $vw \notin E$;*
(ii) *$f(v,w) = -f(w,v)$ for all $v,w\in V$;*
(iii) $\displaystyle\sum_{\substack{w\in V \text{ with} \\ vw\in E}} f(v,w) = 0$ *for all $v \in V$.*

The k-flow f is nowhere zero *if $f(v,w)\neq 0$ for all $vw\in E$. Let $F(G,k)$ be the number of nowhere zero k-flows of G. It can be shown that*

$$F(G,k) = (-1)^{M-N+Q}\cdot T(G;0,1-k) = k^{-N}\cdot Z_{A(k,k-1,-1)}(G),$$

where T denotes the Tutte polynomial and $A(k,k-1,1)$ the $k\times k$ matrix with diagonal entries $(k-1)$ and off-diagonal entries -1.

2.1. Partition functions in statistical physics. The term "partition function" indicates the fact that the functions we consider here do also have an origin in statistical physics. A major aim of this branch of physics is the prediction of phase transitions in dynamical systems from knowing only the interactions of their microscopic components. In this context, partition functions are the central quantities allowing for such a prediction. As a matter of fact many of these partition functions can be described in the framework we defined above.

Let us see an example for this connection — the partition function of the *Ising model*. Originally introduced by Ising in 1925 [**Isi25**] this model was developed to describe the phase transitions in ferromagnets. For some given graph G, the model associates with each vertex v a spin σ_v which may be either $+1$ or -1. Then the energy of a state σ is given by the *Hamiltonian* defined by

$$H(\sigma) = -J \sum_{uv\in E} \sigma_u\sigma_v \tag{2.4}$$

where $-J\sigma_u\sigma_v$ is the contribution of the energy of each pair of nearest neighbor particles. Let T denote the temperature of the system and k be Boltzmann's constant, define $\beta = (kT)^{-1}$. Then, for a graph $G = (V, E)$ with N vertices, M edges, and Q connected components, we have

$$Z(G,T) \quad = \sum_{\sigma:V\to\{+1,-1\}} e^{-\beta H(\sigma)}$$

We straightforwardly get $Z(G,T) = e^{\beta JM} Z_A(G)$ for the matrix

$$A = A(T) = \begin{pmatrix} e^{2\beta J} & 1 \\ 1 & e^{2\beta J} \end{pmatrix}. \tag{2.5}$$

An extension of this model to systems with more than two spins is the *n-state Potts model*, whose partition function satisfies

$$Z_{\text{Potts}}(G; n, v) \quad = \sum_{\sigma:V\to[n]} \prod_{uv\in E} (1 + v \cdot \delta_{\sigma(u),\sigma(v)}).$$

In fact, for $n = 2$ and $v = e^{2\beta J} - 1$ we see that $Z(G,T) = e^{\beta JM} Z_{\text{Potts}}(G; n, v)$. Moreover, this model can be seen as a specialization of the Tutte Polynomial. Expanding the above sum over connected components of G, we obtain

$$Z_{\text{Potts}}(G; n, v) = \sum_{A\subseteq E} n^{q(A)} v^{|A|}$$

whence it is not difficult to see that, if $(X-1)(Y-1) \in \mathbb{N}$,

$$T(G; X, Y) = (X-1)^{-q(E)}(Y-1)^{-N} Z_{\text{Potts}}(G; (X-1)(Y-1), Y-1).$$

Since $Z_{\text{Potts}}(G; n, v) = Z_{A(n,v+1,1)}(G)$ this relation is actually a special case of equation (2.3).

2.2. Which functions are partition functions? In this section, we shall state (and partially prove) precise algebraic characterizations of partition functions over the real and complex numbers.

A graph invariant $f : \mathcal{G} \to \mathbb{S}$ is *multiplicative* if $f(\emptyset) = 1$ and $f(G \cdot H) = f(G) \cdot f(H)$. Here $\emptyset$ denotes the empty graph and $G \cdot H$ denotes the disjoint union of the graphs G and H. An easy calculation shows:

Observation 2.10. *All partition functions are multiplicative.*

To characterize the class of partition functions over the reals, let $f : \mathcal{G} \to \mathbb{R}$ be a graph invariant. Consider the(infinite) real matrix $M = (M_{G,H})_{G,H\in\mathcal{G}}$ with entries $M_{G,H} := f(G \cdot H)$. It follows from the multiplicativity that if f is a partition function then M has rank 1 and is positive semidefinite. (Here an infinite matrix is *positive semidefinite* if each finite principal submatrix is positive semidefinite.) The criterion for a graph invariant being a partition function is a generalization of this simple criterion. Let $k \geq 0$. A *k-labeled graph* is a graph with k distinguished vertices. Formally, a k-labeled graph is a pair (G, ϕ), where $G \in \mathcal{G}$ and $\phi : [k] \to V(G)$ is injective. The class of all k-labeled graphs is denoted by $\mathcal{G}_k$. For two k-labeled graphs $(G, \phi), (H, \psi) \in \mathcal{G}_k$, we let $(G, \phi) \cdot (H, \psi)$ be the k-labeled graph obtained from the disjoint union of G and H by identifying the labeled vertices $\phi(i)$ and $\psi(i)$ for all $i \in [k]$ and keeping the labels where they are. We extend f to $\mathcal{G}_k$ by letting $f(G, \phi) := f(G)$ and define a matrix

$$M(f,k) = \big(M(f,k)_{(G,\phi),(H,\psi)}\big)_{(G,\phi),(H,\psi)\in\mathcal{G}_k}$$

by letting $M(f,k)_{(G,\phi),(H,\psi)} = f\big((G,\phi)\cdot(H,\psi)\big)$. Note that if we identify 0-labeled graphs with plain graphs, then $M(f,0)$ is just the matrix M defined above. The matrices $M(f,k)$ for $k \geq 0$ are called the *connection matrices* of f. Connection matrices were first introduced by Freedman, Lovász, and Schrijver [**FLS07**] to prove a theorem similar to the following one (Theorem 2.14 below).

THEOREM 2.11 (Schrijver [**Sch09**]). *Let $f : \mathcal{G} \to \mathbb{R}$ be a graph invariant. Then f is a partition function if and only if it is multiplicative and all its connection matrices are positive semidefinite.*

We first give the simple proof of the forward direction. A proof sketch for the backward direction will be given at the end of this subsection. Suppose that $f = Z_A$ for a symmetric matrix $A \in \mathbb{R}^{n\times n}$. For all mappings $\chi := [k] \to [n]$ and k-labeled graphs $(G,\phi) \in \mathcal{G}_k$ we let

$$Z_{A,\chi}(G,\phi) := \sum_{\substack{\sigma:V(G)\to[n]\\ \sigma(\phi(i))=\chi(i) \text{ for all } i\in[k]}} \prod_{vw\in E(G)} A_{\sigma(v),\sigma(w)}.$$

Note that for $(H,\psi) \in \mathcal{G}_k$ we have

$$Z_{A,\chi}\big((G,\phi)\cdot(H,\psi)\big) = Z_{A,\chi}(G,\phi)\cdot Z_{A,\chi}(H,\psi).$$

Hence the matrix $M(Z_{A,\chi},k)$ with entries $M(Z_{A,\chi},k)_{(G,\phi),(H,\psi)} = Z_{A,\chi}\big((G,\phi)\cdot(H,\psi)\big)$ is positive semidefinite. Furthermore,

$$f(G,\phi) = Z_A(G) = \sum_{\chi:[k]\to[n]} Z_{A,\chi}(G,\phi),$$

and thus $M_{f,k}$ is the sum of n^k positive semidefinite matrices, which implies that it is positive semidefinite. Incidentally, the same argument shows that the row rank of the kth connection matrix $M(Z_A,k)$ of a partition function of a matrix $A \in \mathbb{R}^{n\times n}$ is at most n^k.

Schrijver also obtained a characterization of the class of partition functions over the complex numbers, which looks surprisingly different from the one for the real numbers. In the following, let $\mathbb{S} = \mathbb{C}$ or $\mathbb{C}[\mathbf{X}]$ for some tuple $\mathbf{X}$ of variables. For a symmetric $n\times n$ matrix $A \in \mathbb{S}^{n\times n}$ we define the "injective" partition function $Y_A : \mathcal{G} \to \mathbb{S}$ by

$$Y_A(G) := \sum_{\substack{\tau:V(G)\to[n]\\ \text{injective}}} \prod_{vw\in E(G)} A_{\tau(v),\tau(w)}.$$

Note that $Y_A(G) = 0$ for all G with $|V(G)| > n$. For a graph G and a partition P of $V(G)$, we let G/P be the graph whose vertex set consists of the classes of P and whose edge set contains an edge between the class of v and the class of w for every edge $vw \in E(G)$. Note that in general G/P will have many loops and multiple edges. For example, if P has just one class, then G/P will consist of a single vertex with $|E(G)|$ many loops. We denote the set of all partitions of a set V by $\Pi(V)$, and for a graph G we let $\Pi(G) := \Pi\big(V(G)\big)$. For $P,Q \in \Pi(V)$, we write $P \leq Q$ if P refines Q. Then we have

$$Z_A(G) = \sum_{P\in\Pi(G)} Y_A(G/P). \tag{2.6}$$

We can also express Y_A in terms of Z_A. For every finite set V there is a unique function $\mu : \Pi(V) \to \mathbb{Z}$ satisfying the following equation for all $P \in \Pi(V)$:

$$\sum_{\substack{Q\in\Pi(V)\\ Q\le P}} \mu(Q) = \begin{cases} 1 & \text{if } P = T_V, \\ 0 & \text{otherwise.} \end{cases} \tag{2.7}$$

Here T_V denotes the trivial partition $\{\{v\} \mid v \in V\}$. The function μ is a restricted version of the *Möbius function* of the partially ordered set $\Pi(V)$ (for background, see for example [**Aig07**]). Now it is easy to see that

$$Y_A(G) = \sum_{P\in\Pi(G)} \mu(P) \cdot Z_A(G/P). \tag{2.8}$$

We close our short digression on Möbius inversion by noting that for all sets V with $|V| =: k$ we have the following polynomial identity:

$$\sum_{P\in\Pi(V)} \mu(P) \cdot X^{|P|} = X(X-1)\cdots(X-k+1). \tag{2.9}$$

To see this, note that it suffices to prove it for all $X \in \mathbb{N}$. So let $X \in \mathbb{N}$, and let A be the $(X \times X)$-identity matrix. Furthermore, let G be the graph with vertex set V an no edges. Then $Y_A = X(X-1)\cdots(X-k+1)$ and $Z_A(G/P) = X^{|P|}$ for every partition $P \in \Pi(V)$, and (2.9) follows from (2.8).

Now we are ready to state Schrijver's characterization of the partition functions over the complex numbers.

THEOREM 2.12 (Schrijver [**Sch09**]). *Let $f : \mathcal{G} \to \mathbb{C}$ be a graph invariant. Then f is a partition function if and only if it is multiplicative and*

$$\sum_{P\in\Pi(G)} \mu(P) \cdot f(G/P) = 0 \tag{2.10}$$

for all $G \in \mathcal{G}$ with $|V(G)| > |f(K_1)|$.

To understand the condition $|V(G)| > |f(K_1)|$, remember that K_1 denotes the graph with one vertex and no edges and note that for every $n \times n$ matrix A it holds that $Z_A(K_1) = n$.

Proof of Theorem 2.12 (sketch). The forward direction is almost trivial: Suppose that $f = Z_A$ for a symmetric matrix $A \in \mathbb{C}^{n\times n}$. Then f is multiplicative by Observation 2.10, and we have

$$\sum_{P\in\Pi(G)} \mu(P) \cdot f(G/P) = Y_A(G) = 0$$

for all G with $|V(G)| > n = f(K_1)$, where the first equality holds by (2.8) and the second because for G with $|V(G)| > n$ there are no injective functions $\sigma : V(G) \to [n]$.

It is quite surprising that these trivial conditions are sufficient to guarantee that f is a partition function. To see that they are, let $f : \mathcal{G} \to \mathbb{C}$ be a multiplicative graph invariant such that (2.10) holds for all $G \in \mathcal{G}$ with $|V(G)| > f(K_1)$.

We first show that $n := f(K_1)$ is a non-negative integer. Let $k := \lceil |f(K_1)| \rceil$, and let I_k be the graph with vertex set $[k]$ and no edges. (Hence $I_k = \emptyset$ if $k = 0$.)

Suppose that n is not a non-negative integer. Then $k > |n|$, and by (2.10), the multiplicativity of f, and (2.9) we have

$$0 = \sum_{P \in \Pi(I_k)} \mu(P) \cdot f(I_k/P) = \sum_{P \in \Pi([k])} \mu(P) \cdot n^{|P|} = n \cdot \big(n-1\big) \cdots \big(n-k+1\big) \neq 0.$$

This is a contradiction.

A *quantum graph* is a formal linear combination of graphs with coefficients from $\mathbb{C}$, that is, an expression $\sum_{i=1}^{\ell} a_i G_i$, where $\ell \geq 0$ and $a_i \in \mathbb{C}$ and $G_i \in \mathcal{G}$ for all $i \in [\ell]$. The class of all quantum graphs is denoted by $\mathcal{QG}$. The quantum graphs obviously form a vector space over $\mathbb{C}$, and by extending the product "disjoint union" linearly from $\mathcal{G}$ to $\mathcal{QG}$, we turn this vector space into an algebra. We also extend the function f linearly from $\mathcal{G}$ to $\mathcal{QG}$. Observe that f is an algebra homomorphism from $\mathcal{QG}$ to $\mathbb{C}$, because it is multiplicative.

For all $i, j \in [n]$ we let $X_{\{i,j\}}$ be a variable, and we let $\mathbf{X}$ be the tuple of all these variables ordered lexicographically. Furthermore, we let X be the $n \times n$ matrix with $X_{i,j} := X_{j,i} := X_{\{i,j\}}$. We view X as a matrix over the polynomial ring $\mathbb{C}[\mathbf{X}]$. We extend the function $Z_X : \mathcal{G} \to \mathbb{C}[X]$ linearly from $\mathcal{G}$ to $\mathcal{QG}$. Then Z_X is an algebra homomorphism. It is not too hard to show that the image $Z_X(\mathcal{QG})$ consists precisely of all polynomials in $\mathbb{C}[\mathbf{X}]$ that are invariant under all permutations $X_{\{i,j\}} \mapsto X_{\{\pi(i),\pi(j)\}}$ of the variables for permutations π of $[n]$. Using (2.8) and other properties of the Möbius inversion, it can be shown that the kernel of Z_X is contained in the kernel of f. This implies that there is an algebra homomorphism $\hat{f}$ from the image $Z_X(\mathcal{QG})$ to $\mathbb{C}$ such that $f = \hat{f} \circ Z_X$.

Then

$$I := \big\{p \in Z_X(\mathcal{QG}) \mid \hat{f}(p) = 0\big\}$$

is an ideal in the subalgebra $Z_X(\mathcal{QG}) \subseteq \mathbb{C}[\mathbf{X}]$. We claim that the polynomials in I have a common zero. Suppose not. Let I' be the ideal generated by I in $\mathbb{C}[\mathbf{X}]$. Then by Hilbert's Nullstellensatz it holds that $1 \in I'$. Using the fact that the the subalgebra $Z_X(\mathcal{QG})$ consists precisely of all polynomials in $\mathbb{C}[\mathbf{X}]$ that are invariant under all permutations $X_{\{i,j\}} \mapsto X_{\{\pi(i),\pi(j)\}}$ for $\pi \in S_n$, one can show that actually $1 \in I$. But then

$$0 = \hat{f}(1) = \hat{f}(Z_X(K_0)) = f(K_0) = 1,$$

which is a contradiction.

Thus the polynomials in I have a common zero. Let $\mathbf{A} = (A_{\{i,j\}} \mid i, j \in [n])$ be such a zero, and let $A \in \mathbb{C}^{n \times n}$ be the corresponding symmetric matrix. Observe that for each graph G it holds that $Z_X(G) - f(G) \in I$, because $\hat{f}(Z_X(G)) - \hat{f}(f(G)) = f(G) - f(G) = 0$. Hence $Z_A(G) - f(G) = 0$ and thus $f(G) = Z_A(G)$. This completes our proof sketch. □

Even though Theorems 2.11 and 2.12 look quite different, it is not hard to derive Theorem 2.11 from Theorem 2.12. In the remainder of this subsection, we sketch how this is done.

Proof of Theorem 2.11 (sketch). We have already proved the forward direction. For the backward direction, let $f : \mathcal{G} \to \mathbb{R}$ be a multiplicative graph invariant such that all connection matrices of f are positive semidefinite. We first prove that f satisfies condition (2.10): Let $n := f(K_1)$ (we do not know yet that n is an integer, but it will turn out to be), and let $k > n$ be a non-negative integer. Let I_k be the graph with vertex set $[k]$ and no edges, and for every partition P of $[k]$, define

$\phi_P : [k] \to V(I_k/P)$ to be the canonical projection, that is, $\phi_P(i)$ is the class of i in the partition P. Then $(I_k/P, \phi_P) \in \mathcal{G}_k$. As the kth connection matrix $M(f, k)$ is positive semidefinite, we have

$$\sum_{P,Q \in \Pi([k])} \mu(P) \cdot \mu(Q) \cdot f\big((I_k/P, \phi_P) \cdot (I_k/Q, \phi_Q)\big) \geq 0.$$

For partitions $P, Q \in \Pi([k])$, let $P \vee Q$ be the least upper bound of P and Q in the partially ordered set $\Pi([k])$, and note that $(I_k/P, \phi_P) \cdot (I_k/Q, \phi_Q) = (I_k/P \vee Q, \phi_{P \vee Q})$. Hence by the multiplicativity of f we have $f\big((I_k/P, \phi_P) \cdot (I_k/Q, \phi_Q)\big) = n^{|P \vee Q|}$. A calculation similar to the one that leads to (2.9) shows that $\sum_{P,Q \in \Pi[k]} \mu(P) \cdot \mu(Q) \cdot X^{|P \vee Q|} = X \cdot (X-1) \cdots (X-k+1)$. Hence for all non-negative integers $k > n$ we have

(2.11)
$$0 \leq \sum_{P,Q \in \Pi([k])} \mu(P) \cdot \mu(Q) \cdot f\big((I_k/P, \phi_P) \cdot (I_k/Q, \phi_Q)\big) = n \cdot (n-1) \cdots (n-k+1).$$

This is only possible if n is a non-negative integer, in which case equality holds.

Now let G be an arbitrary graph with $k := |V(G)| > n$. Without loss of generality we may assume that $V(G) = [k]$. Let ψ be the identity on $[k]$. Consider the $(|\Pi([k])| + 1) \times (|\Pi([k])| + 1)$-principal submatrix M_0 of $M(f, k)$ with rows and columns indexed by the k-labeled graphs $(I_k/P, \phi_P)$ for all $P \in \Pi([k])$ and (G, ψ). For every $x \in \mathbb{R}$, let $\mathbf{v}_x$ be the column vector with entries $\mu(P)$ for all $P \in \Pi([k])$ followed by x. By the positive semi-definiteness of $M(f, k)$, we have

$$\begin{aligned} 0 &\leq \mathbf{v}_x^\top M_0 \mathbf{v}_x \\ &= \sum_{P,Q \in \Pi([k])} \mu(P) \cdot \mu(Q) \cdot f\big((I_k/P, \phi_P) \cdot (I_k/Q, \phi_Q)\big) \\ &\quad + 2x \cdot \sum_{P \in \Pi([k])} \mu(P) \cdot f\big((I_k/P, \phi_P) \cdot (G, \psi)\big) + x^2 \cdot f\big((G, \psi) \cdot (G, \psi)\big) \\ &= 2x \cdot \sum_{P \in \Pi([k])} \mu(P) \cdot f\big((I_k/P, \phi_P) \cdot (G, \psi)\big) + x^2 \cdot f\big((G, \psi) \cdot (G, \psi)\big) \end{aligned}$$

This is only possible for all $x \in \mathbb{R}$ if $\sum_{P \in \Pi([k])} \mu(P) \cdot f\big((I_k/P, \phi_P) \cdot (G, \psi)\big) = 0$. Note that for every $P \in \Pi([k])$ it holds that $(I_k/P, \phi_P) \cdot (G, \psi) = (G/P, \phi_P)$. Thus

$$\sum_{P \in \Pi(G)} \mu(P) \cdot f(G/P) = 0.$$

If follows from Theorem 2.12 that $f = Z_A$ for some matrix $A \in \mathbb{C}^{n \times n}$.

For every $\ell \geq 0$, let F_ℓ be the graph with two vertices and ℓ parallel edges between these vertices. Exploiting the positive semi definiteness of the principal submatrix of $M(f, 2)$ with rows and columns indexed by the graphs F_ℓ for $0 \leq \ell \leq n(n+1)/2$, it is not hard to show that the matrix A is real. $\square$

2.3. Generalizations.

Vertex weights. We have mentioned that a partition function Z_A can be viewed as mapping a graph G to a weighted sum of homomorphisms to the edge-weighted graph represented by the matrix A. We may also put weights on the vertices of a graph. It will be most convenient to represent vertex weights on a graph H by a diagonal matrix $D = (D_{v,w})_{v,w \in V(H)}$, where $D_{v,v}$ is the weight of vertex v and $D_{v,w} = 0$ if $v \neq w$. Then we define the *weight* of a mapping σ from G to H to be the product of the weights of the images of the edges and the images of the vertices. More abstractly, for every symmetric matrix $A \in \mathbb{S}^{n \times n}$ and diagonal matrix $D \in \mathbb{S}^{n \times n}$ we define a function $Z_{A,D} : \mathcal{G} \to \mathbb{S}$ by

$$Z_{A,D}(G) := \sum_{\sigma : V(G) \to [k]} \prod_{vw \in E(G)} A_{\sigma(v),\sigma(w)} \cdot \prod_{v \in V(G)} D_{\sigma(v),\sigma(v)}.$$

We call $Z_{A,D}$ a *partition functions with vertex weights* over $\mathbb{S}$. ([Freedman, Lovász, and Schrijver [**FLS07**] use the term *homomorphism functions.*) The vertex weights enable us to get rid of the constant factors in some of the earlier examples and give smoother formulations. For example:

EXAMPLE 2.13. *Recall that by Example 2.9, the number $F(G,k)$ of nowhere-zero k-flows of an N-vertex graph G is $k^{-N} \cdot Z_A(G)$ for the $k \times k$ matrix A with diagonal entries $(k-1)$ and off-diagonal entries -1. Let D be the $(k \times k)$-diagonal matrix with entries $D_{ii} := 1/k$ for all $i \in [k]$. Then $F(G,k) = Z_{A,D}(G)$.*

Freedman, Lovász, and Schrijver gave an algebraic characterization of the class of partition functions with non-negative vertex weights over the reals, again using connection matrices. However, they only consider graph invariants defined on the class $\mathcal{G}'$ of graphs without loops (but with parallel edges). For a graph invariant $f : \mathcal{G}' \to \mathbb{R}$, we define the kth connection matrix $M(f,k)$ as for graph invariants defined on $\mathcal{G}$, except that we omit all rows and columns indexed by graphs with loops. We call a matrix $A \in \mathbb{R}^{n \times n}$ *non-negative* if all its entries are non-negative.

THEOREM 2.14 (Freedman, Lovász, and Schrijver [**FLS07**]). *Let $f : \mathcal{G}' \to \mathbb{R}$ be a graph invariant. Then the following two statements are equivalent:*

1. *There are a symmetric matrix $A \in \mathbb{R}^{n \times n}$ and a non-negative diagonal matrix $D \in \mathbb{R}^{n \times n}$ such that $f(G) = Z_{A,D}(G)$ for all $G \in \mathcal{G}'$.*
2. *There is a $q \geq 0$ such that for all $k \geq 0$ the matrix $M(f,k)$ is positive semidefinite and has row rank at most q^k.*

Despite the obvious similarity of this theorem with Theorem 2.11, the known proofs of the two theorems are quite different.

Asymmetric matrices and directed graphs. Of course we can also count homomorphism between directed graphs. We denote the class of all directed graphs by $\mathcal{D}$; as undirected graphs we allow directed graphs to have loops and parallel edges. For every directed graph H, we define the homomorphism counting function $Z_H : \mathcal{D} \to \mathbb{Z}$ by letting $Z_H(D)$ be the number of homomorphisms from D to H. For every square matrix $A \in \mathbb{S}^{n \times n}$ we define the function $Z_A : \mathcal{D} \to \mathbb{S}$ by

$$Z_A(D) := \sum_{\sigma : V(D) \to [n]} \prod_{(v,w) \in E(D)} A_{\sigma(v),\sigma(w)}.$$

Note that if A is a symmetric matrix, then $Z_A(D) = Z_A(G)$, where G is the underlying undirected graph of G.

Hypergraphs and relational structures. Recall that a *hypergraph* is a pair $H = (V, E)$ where V is a finite set and $E \subseteq 2^V$. Elements of V are called *vertices*, elements of E *hyperedges.* A hypergraph is r-uniform, for some $r \geq 1$, if all its hyperedges have cardinality r. Thus a 2-uniform hypergraph is just a simple graph. A *homomorphism* from a hypergraph G to a hypergraph H is a mapping $h : V(G) \to V(H)$ such that $h(e) \in E(H)$ for all $e \in E(H)$. Of course, for $e = \{v_1, \dots, v_r\}$ we let $h(e) = \{h(v_1), \dots, h(v_r)\}$.[2] Hypergraph homomorphism counting functions and partition functions have only been considered for uniform hypergraphs. The class of all r-uniform hypergraphs is denoted by $\mathcal{H}_r$. For all $H \in \mathcal{H}_r$, we define the homomorphism counting function $Z_H : \mathcal{H}_r \to \mathbb{Z}$ in the obvious way. The natural generalization of partition functions to r-uniform hypergraphs is defined by symmetric functions $A : [n]^r \to \mathbb{S}$. We define $Z_A : \mathcal{H}_r \to \mathbb{S}$ by

$$Z_A(G) := \sum_{\sigma : V(G) \to [n]} \prod_{e \in E(G)} A(\sigma(e)),$$

where for $e = \{v_1, \dots, v_r\}$ we let $A(\sigma(e)) := A(\sigma(v_1), \dots, \sigma(v_r))$. This is well-defined because A is symmetric.

Let us finally consider homomorphisms between relational structures. A *(relational) vocabulary* σ is a set of relation symbols; each relation symbol comes with a prescribed finite arity. A *σ-structure* B consist of a set $V(B)$, which we call the *universe* or *vertex set* of B, and for each r-ary relation symbol $S \in \sigma$ an r-ary relation $S(B) \subseteq V(B)^r$. Here we assume all structures to be finite, that is, to have a finite universe and vocabulary. For example, a simple graph may be viewed as an $\{E\}$-structure G, where E is a binary relation symbol and $E(G)$ is irreflexive and symmetric. An r-uniform hypergraph H may be viewed as an $\{E_r\}$-structure, where E_r is an r-ary relation symbol and $E_r(H)$ is symmetric and contains only tuples of pairwise distinct elements. For each vocabulary σ, we denote the class of all σ-structures by $\mathcal{S}_\sigma$. For each σ-structure B we define the homomorphism counting functions $Z_B : \mathcal{S}_\sigma \to \mathbb{Z}$ in the usual way.

To generalize partition functions, we consider *weighted structures.* For a ring $\mathbb{S}$ and a vocabulary σ, an *$\mathbb{S}$-weighted σ-structure* B consists of a finite set $V(B)$ and for each r-ary $R \in \sigma$ a mapping $R(\mathbb{S}) : V(B)^r \to \mathbb{S}$. Then we define $Z_B : \mathcal{S}_\sigma \to \mathbb{S}$ by

$$Z_B(A) := \sum_{\sigma : V(G) \to [n]} \prod_{R \in \sigma} \prod_{\substack{(a_1, \dots, a_r) \in V(A)^r, \\ \text{where } r \text{ is the arity of } R}} R(\mathbb{S})(\sigma(a_1), \dots, \sigma(a_r)).$$

Note that this does not only generalize plain partition functions on graphs, but also partition functions with vertex weights, because we may view a graph with weights on the vertices and edges as a weighted $\{E, P\}$-structure, where E is a binary and P a unary relation symbol.

It is a well-known observation due to Feder and Vardi [**FV98**] that *constraint satisfaction problems* may be viewed as homomorphism problems between relational structures and vice versa. Thus counting homomorphisms between relational structures correspond to counting solutions to constraint satisfaction problems.

[2]There is no need to define hypergraph homomorphisms as mappings from $V(G) \cup E(G)$ to $V(H) \cup E(H)$ because we do not allow parallel hyperedges.

Edge Models. Partition functions and all their generalizations considered so far are weighted sums over mappings defined on the vertex set of the graph or structure. In the context of statistical physics, they are sometimes called *vertex models.* There is also a notion of *edge model.* An edge model is given by a function $F : \mathbb{N}^n \to \mathbb{S}$. Let $G = (V, E)$ be a graph. Given a "configuration" $\tau : E \to [n]$, for every vertex v we let $t(\tau, v) := (t_1, \dots, t_n)$, where t_i is the number of edges e incident with v such that $\tau(e) = i$. We define a function $\tilde{Z}_F : \mathcal{G} \to \mathbb{S}$ by

$$\tilde{Z}_F(G) := \sum_{\tau:E\to[n]} \prod_{v\in V} F\big(t(\tau, v)\big).$$

Szegedy [**Sze07**] gave a characterization of the graph invariants over the reals expressible by edge models that is similar to the characterizations of partition functions (vertex models) given in Theorems 2.11 and 2.14.

EXAMPLE 2.15. *Let $F : \mathbb{N}^2 \to \mathbb{Z}$ be defined by $F(i, j) := 1$ if $j = 1$ and $F(i, j) = 0$ otherwise. Let $G = (V, E)$ be a graph. Observe that for every $\tau : E(G) \to \{1, 2\}$ and every $v \in V$ it holds that $F\big(t(\tau, v)\big) = 1$ if and only if there is exactly one edge in $\tau^{-1}(2)$ that is incident with v. Hence $\prod_{v\in V} F\big(t(\tau, v)\big) = 1$ if $\tau^{-1}(2)$ is a perfect matching of G and $\prod_{v\in V} F\big(t(\tau, v)\big) = 0$ otherwise. It follows that $\tilde{Z}_F(G)$ is the number of perfect matchings of G.*

It can be proved that the function $f = \tilde{Z}_F$ counting perfect matchings is not a partition function, because the connection matrix $M(f, 1)$ is not positive semi-definite [**FLS07**].

Related to the edge models is a class of functions based on Valiant's holographic algorithms [**Val08**]: so-called *holant functions* which have been introduced by Cai, Lu, and Xia [**CLX09**]. A holant function is given by a *signature grid* $\Omega = (G, \mathcal{F}, \pi)$ where $G = (V, E)$ is a graph and, for some $n \in \mathbb{N}$, the set $\mathcal{F}$ contains functions $f : [n]^{a_f} \to \mathbb{S}$, each of some arity a_f. Further π maps vertices $v \in V$ to functions $f_v \in \mathcal{F}$ such that $d(v) = a_{f_v}$. Let, for some vertex v, denote $E(v)$ the set of edges incident to v. The holant function over this signature grid is then defined as

$$\mathrm{Holant}_\Omega = \sum_{\tau:E\to[n]} \prod_{v\in V} f_v(\tau|_{E(v)}).$$

Note that we assume here implicitly that E has some ordering, therefore $f_v(\tau|_{E(v)})$ is well-defined. Edge models are a special case of this framework, since $\tilde{Z}_F(G) = \mathrm{Holant}_\Omega$ for the signature grid which satisfies $f_v(\tau|_{E(v)}) = F(t(\tau, v))$ for all $v \in V$. More generally it can be shown that partition functions on relational structures can be captured by holant functions.

3. Complexity

Partition functions tend to be hard to compute. More precisely, they tend to be hard for the complexity class #P introduced by Valiant in [**Val79**], which may be viewed as the "counting analogue" of NP. A counting problem C (that is, a function with values in the non-negative integers) belongs to #P if and only if there is a nondeterministic polynomial time algorithm A such that for every instance x of C it holds that $C(x)$ is the number of accepting computation paths of A on input x. There is a theory of reducibility and #P-completeness much like the theory of NP-completeness. (The preferred reductions in the complexity theory of counting problems are *Turing reductions*. We exclusively work with Turing reductions in this article.) Most NP-complete decision problems have natural #P-complete counting problems associated with them. For example, the problem of counting the number of independent sets of a graph and the problem of counting the number of 3-colorings of a graph are both #P-complete. There is a counting analogue of a well-known Theorem due to Ladner [**Lad75**] stating that there are counting problems in #P that are neither #P-complete nor in FP (the class of all counting, or more generally functional problems solvable in polynomial time); indeed the counting complexity classes between FP and #P form a dense partial order.

Even though the class of partition functions is very rich, it turns out that partition functions and also their various generalizations discussed in Section 2.3 exhibit a *complexity theoretic dichotomy*: Some partition functions can be computed in polynomial time, most are #P-hard, but there are no partition functions of intermediate complexity. A first dichotomy theorem for homomorphism counting functions of graphs was obtained by Dyer and Greenhill:

THEOREM 3.1 (Dyer and Greenhill [**DG00**]). *Let H be a graph without parallel edges. Then Z_H is computable in polynomial time if each connected component of H is either a complete graph with a loop at every vertex or a complete bipartite graph.*[3] *Otherwise, Z_H is #P-complete.*

When thinking about the complexity of partition functions over the real or complex numbers, we face the problem of which model of computation to use. There are different models which lead to different complexity classes. To avoid such issues, we restrict our attention to algebraic numbers, which can be represented in the standard bit model.[4] We will discuss the representation of and computation with algebraic numbers in Section 7.3. The fields of real and complex algebraic numbers are denoted by $\mathbb{R}_{\mathbb{A}}$ and $\mathbb{C}_{\mathbb{A}}$, respectively. In general, partition functions are no counting functions (in the sense that their values are no integers) and hence they do not belong to the complexity class #P. For that reason, in the following results we only state #P-hardness and not completeness. As an easy upper bound, we note that partition functions over $\mathbb{R}_{\mathbb{A}}$ and $\mathbb{C}_{\mathbb{A}}$ belong to the complexity class $\mathrm{FP}^{\#\mathrm{P}}$ of all functional problems that can be solved by a polynomial time algorithm with an oracle to a problem in #P.

Let us turn to partition functions over the reals. We first observe that if $A \in \mathbb{R}_{\mathbb{A}}^{n\times n}$ is a symmetric matrix of row rank 1, then Z_A is easily computable in polynomial time. Indeed, write $A = \mathbf{a}^T\mathbf{a}$ for a (row) vector $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}_{\mathbb{A}}^n$. Then for every graph $G = (V, E)$,

$$Z_A(G) = \sum_{\sigma:V\to[n]} \prod_{vw\in E} a_{\sigma(v)}a_{\sigma(w)} = \sum_{\sigma:V\to[n]} \prod_{v\in V} a_{\sigma(v)}^{\deg(v)} = \prod_{v\in V} \sum_{i=1}^{n} a_i^{\deg(v)}. \tag{3.1}$$

Here $\deg(v)$ denotes the degree of a vertex v. The last term in (3.1), which only involves a polynomial number of arithmetic operations, can easily be evaluated in

[3] We count the empty graph and the graph K_1 with one vertex and no edges as bipartite graphs.

[4] In [**BG05**], the complexity classification of partition functions for non-negative real matrices (Theorem 3.2 of this article) was stated for arbitrary real matrices in one of the standard models of real number computation. However, the proof is faulty and can only be made to work for real algebraic numbers (cf. Section 7.3 of this article).

polynomial time. Thus partition functions of rank-1 matrices are easy to compute. It turns out that all easy partition functions of non-negative real matrices are based on a "rank-1" condition.

Let us call a matrix A *bipartite* if its underlying graph $G(A)$ is bipartite. After suitable permuting rows and columns, a bipartite matrix has the form

$$A = \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix},$$

where we call B and B^T the *blocks* of A. A similar argument as the one above shows that if B has row rank 1 then Z_A is computable in polynomial time. Note that we only need to compute $Z_A(G)$ for bipartite G, because $Z_A(G) = 0$ for non-bipartite G.

The *connected components* of a matrix A are the principal submatrices corresponding to the connected components of the underlying graph $G(A)$. Note that if A is a matrix with connected components $A_1, \dots, A_m$, then for every connected graph G it holds that $Z_A(G) = \sum_{j=1}^m Z_{A_j}(G)$. By the multiplicativity of partition functions, for a graph G with connected components $G_1, \dots, G_\ell$, we thus have $Z_A(G) = \prod_{i=1}^{\ell} \sum_{j=1}^m Z_{A_j}(G_i)$. This reduces the computation of Z_A to the computation of the Z_{A_j}.

Theorem 3.2 (Bulatov and Grohe [**BG05**]). *Let $A \in \mathbb{R}_{\mathbb{A}}^{n\times n}$ be a symmetric non-negative matrix. Then Z_A is computable in polynomial time if all components of A are either of row rank 1 or bipartite and with blocks of row rank 1. Otherwise, Z_A is #P-hard.*

Note that the theorem is consistent with Theorem 3.1, the special case for 0-1-matrices. We have already proved that Z_A is computable in polynomial time if all components of A are either of row rank 1 or bipartite and with blocks of row rank 1. The much harder proof that all other cases are #P-hard will be given in Sections 4–9 of this paper. As a by-product, we also obtain a proof of Theorem 3.1.

The rest of this section is a survey of further dichotomy results. We will not prove them. The following two examples show that if we admit negative entries in our matrices, then the rank-1 condition is no longer sufficient to explain tractability.

Example 3.3. *Consider the matrix $B = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ first introduced in Example 2.6. The row rank of this matrix is 2, yet we shall prove that Z_B is computable in polynomial time. Let $G = (V, E)$. Then*

$$\begin{aligned} Z_B(G) = \sum_{\sigma:V\to[2]} \prod_{vw\in E} B_{v,w} &= \sum_{\sigma:V\to[2]} \prod_{vw\in E} (-1)^{(\sigma(v)-1)\cdot(\sigma(w)-1)} \\ &= \sum_{\sigma:V\to[2]} (-1)^{\sum_{vw\in E}\sigma(v)\cdot\sigma(w)}. \end{aligned}$$

Hence $Z_B(G)$ is 2^N minus twice the number of mappings $\sigma : V \to \{0,1\}$ such that $\sum_{vw\in E} \sigma(v)\cdot\sigma(w)$ is odd. Here N is the number of vertices of G. Thus to compute Z_B, we need to determine the number of solutions of the quadratic equation

$$\sum_{vw\in E} x_v x_w = 1$$

over the 2*-element field* $\mathbb{F}_2$*. The number of solutions of a quadratic equation over* $\mathbb{F}_2$ *or any other finite field can be computed in polynomial time. This follows easily from standard normal forms for quadratic equations (see, for example,* [**LN97**]*, Section 6.2).*

EXAMPLE 3.4. *The* tensor product *of two matrices* $A \in \mathbb{S}^{m\times n}$ *and* $B \in \mathbb{S}^{k\times \ell}$ *is the* $m \cdot k \times n \cdot \ell$ *matrix*

$$A \otimes B := \begin{pmatrix} A_{1,1}\cdot B & \cdots & A_{1,n}\cdot B \\ \vdots & & \vdots \\ A_{m,1}\cdot B & \cdots & A_{m,n}\cdot B \end{pmatrix}.$$

It is easy to see that for all square matrices A, B *and all graphs* G *it holds that*

$$Z_{A\otimes B}(G) = Z_A(G)\cdot Z_B(G).$$

We can thus use the tensor product to construct new matrices with polynomial time computable partition functions. For example, the following three 4×4*-matrices have polynomial time computable partition functions:*

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \qquad \begin{pmatrix} -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 2 & 1 & 2 \\ 2 & 4 & 2 & 4 \\ 1 & 2 & -1 & -2 \\ 2 & 4 & -2 & -4 \end{pmatrix}.$$

Very roughly, all symmetric real matrices with a polynomial time computable partition function can be formed from matrices of rank 1 and matrices associated with quadratic equations over $\mathbb{F}_2$ (in a way similar to the matrix B of Example 3.3) by tensor products and similar constructions. The precise characterization of such matrices is very complicated, and it makes little sense to state it explicitly. In the following, we say that a class $\mathcal{F}$ of functions *exhibits an* FP – #P*-dichotomy* if all functions in $\mathcal{F}$ are either in FP or #P-hard. We say that $\mathcal{F}$ *exhibits an effective* FP – #P*-dichotomy* if in addition it is decidable if a given function in $\mathcal{F}$, represented for example by a matrix, is in FP or #P-hard.

THEOREM 3.5 (Goldberg, Grohe, Jerrum, Thurley [**GGJT10**]). *The class of partition functions of symmetric matrices over the reals exhibits an effective* FP – #P*-dichotomy.*

There are two natural ways to generalize this result to complex matrices. Symmetric complex matrices where studied by Cai, Chen, Lu [**CCL10a**], and Hermitian matrices by Thurley [**Thu09**].

THEOREM 3.6 (Cai, Chen, Lu [**CCL10a**], Thurley [**Thu09**]).

(1) *The class of partition functions of symmetric matrices over the complex numbers exhibits an effective* FP – #P*-dichotomy.*
(2) *The class of partition functions of Hermitian matrices exhibits an effective* FP – #P*-dichotomy.*

Beyond Hermitian matrices, partition functions of arbitrary, not necessarily symmetric matrices are much more difficult to handle. An FP – #P-dichotomy follows from Bulatov's Theorem 3.8 below, but it is difficult to understand how this dichotomy classifies directed graphs. Dyer, Goldberg and Paterson [**DGP07**] proved an effective FP – #P-dichotomy for the class of homomorphism counting functions

Z_H for directed acyclic graphs H; it is based on a complicated "rank-1" condition. A very recent result by Cai and Chen [**CC10**] establishes an effective dichotomy for partition functions Z_A on non-negative real-valued (i.e. not necessarily symmetric) matrices A. The complexity of hypergraph partition functions was studied by Dyer, Goldberg, and Jerrum [**DGJ08**], who proved the following theorem:

THEOREM 3.7 (Dyer, Goldberg, and Jerrum [**DGJ08**]). *For every r, the class of functions Z_A from the class $\mathcal{H}_r$ of r-uniform hypergraphs defined by non-negative symmetric functions $A : [n]^r \to \mathbb{R}_{\mathbb{A}}$ exhibits an effective* FP – #P-*dichotomy.*

Let us finally turn to homomorphism counting functions for arbitrary relational structures, or equivalently solution counting functions for constraint satisfaction problems. Creignou and Hermann [**CH96**] proved a dichotomy for the Boolean case, that is, for homomorphism counting functions of relational structures with just two elements. Dichotomies for the weighted Boolean case were proved in [**DGJ09**] for non-negative real weights, in [**BDG$^+$09**] for arbitrary real weights, and in [**CLX09**] for complex weights. Briquel and Koiran [**BK09**] study the problem in an algebraic computation model. The general (unweighted) case was settled by Bulatov:

THEOREM 3.8 (Bulatov [**Bul08**]). *The class of homomorphism counting functions for relational structures exhibits an* FP – #P-*dichotomy.*

Bulatov's proof uses deep results from universal algebra, and for some time it was unclear whether his dichotomy is effective. Dyer and Richerby [**DR10b**] gave an alternative proof which avoids much of the universal algebra machinery, and they could show that the dichotomy is decidable in NP [**DRar, DR10a**]. Extensions to the weighted case for nonnegative weights in $\mathbb{Q}$ have been given by Bulatov et al. [**BDG$^+$10**], and for nonnegative algebraic weights by Cai et al. [**CCL10b**].

In the context of holant functions complexity dichotomies have been obtained by Cai, Lu, and Xia [**CLX08, CLX09**]. Besides the result stated above, they considered holant functions on signature grids $\Omega = (G, \mathcal{F}, \pi)$, given that $\mathcal{F}$ is a set of symmetric functions satisfying certain additional conditions. A first a dichotomy [**CLX08**] is given for any of boolean symmetric functions $\mathcal{F}$ on planar bipartite $2,3$-regular graphs. In [**CLX09**] dichotomies are presented, assuming that the class $\mathcal{F}$ contains certain unary functions.

4. An Itinerary of the Proof of Theorem 3.2

From now on, we will work exclusively with partition functions defined on matrices with entries in one of the rings $\mathbb{R}_{\mathbb{A}}, \mathbb{Q}[X], \mathbb{Q}, \mathbb{Z}[X], \mathbb{Z}$, and we always let $\mathbb{S}$ denote one of these rings. For technical reasons, we will always assume that numbers in $\mathbb{R}_{\mathbb{A}}$ are given in *standard representation* in some algebraic extension field $\mathbb{Q}(\theta)$. That is, we consider numbers in $\mathbb{Q}(\theta)$ as vectors in a d-dimensional $\mathbb{Q}$-vectorspace, where d is the degree of $\mathbb{Q}(\theta)$ over $\mathbb{Q}$. It is well-known that for any finite set of numbers from $\mathbb{R}_{\mathbb{A}}$ we can compute a θ which constitutes the corresponding extension field (cf. [**Coh93**] p. 181). For further details see also the treatment of this issue in [**DGJ08, Thu09**].

By $\deg(f)$ we denote the *degree* of a polynomial f. For two problems P and Q we use $P \leq Q$ to denote that P is polynomial time Turing reducible to Q. Further, we write $P \equiv Q$ to denote that $P \leq Q$ and $Q \leq P$ holds.

An $m \times n$ matrix A is *decomposable*, if there are non-empty index sets $I \subseteq [m]$, $J \subseteq [n]$ with $(I, J) \neq [m] \times [n]$ such that $A_{ij} = 0$ for all $(i, j) \in \bar{I} \times J$ and all $(i, j) \in I \times \bar{J}$, where $\bar{I} := [m] \setminus I$ and $\bar{J} := [n] \setminus J$. A matrix is *indecomposable* if it is not decomposable. A *block* of A is a maximal indecomposable submatrix. Let A be an $m \times m$ matrix and $G := G(A)$ its underlying graph. Note that every connected component of G that is not bipartite corresponds to a block of A, and every connected component that is bipartite corresponds to two blocks B, B^T arranged as in $\begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}$.

The proof of Theorem 3.2 falls into two parts, corresponding to the following two lemmas.

Lemma 4.1 (Polynomial Time Solvable Cases). *Let $A \in \mathbb{S}^{m\times m}$ be symmetric. If each block of A has row rank at most 1 then* EVAL(A) *is polynomial time computable.*

We leave the proof of this lemma as an exercise for the reader. The essential ideas of its proof were explained in Section 3 before the statement of Theorem 3.2.

Lemma 4.2 (#P-hard Cases). *Let $A \in \mathbb{S}^{m\times m}$ be symmetric and non-negative. If A contains a block of row rank at least 2 then* EVAL(A) *is #P-hard.*

Theorem 3.2 directly follows from Lemmas 4.1 and 4.2. The remainder of this paper is devoted to a proof of Lemma 4.2.

For technical reasons, we need to introduce an extended version of partition functions with vertex weights. Several different restricted flavors of these will be used in the proof to come. Let $A \in \mathbb{R}_{\mathbb{A}}^{m\times m}$ be a symmetric matrix and $D \in \mathbb{R}_{\mathbb{A}}^{m\times m}$ a diagonal matrix. Let $G = (V, E)$ be some given graph. Recall that a *configuration* is a mapping $\sigma : V \to [m]$ which assigns a spin to *every* vertex of G. By contrast, a *pinning* of (vertices of) G with respect to A is a mapping $\phi : W \to [m]$ for some subset $W \subseteq V$. Whenever the context is clear, we simply speak of pinnings and configurations without mentioning the matrices A,D and the graph G explicitly. We define the *partition function* on G and ϕ by

$$Z_{A,D}(\phi, G) = \sum_{\phi\subseteq\sigma:V\to[m]} \prod_{uv\in E} A_{\sigma(u),\sigma(v)} \prod_{v\in V\setminus \mathrm{dom}(\phi)} D_{\sigma(v),\sigma(v)}$$

where dom(ϕ) denotes the domain of ϕ. Note that the sum is over all configurations $\sigma : V \to [m]$ which extend the fixed given pinning ϕ. In the presence of a pinning ϕ we denote the *weight* of the configuration σ by the following term

$$\prod_{uv\in E} A_{\sigma(u),\sigma(v)} \prod_{v\in V\setminus \mathrm{dom}(\phi)} D_{\sigma(v),\sigma(v)}.$$

Note that, for technical reasons, the terms $D_{\sigma(v),\sigma(v)}$ for $v \in \mathrm{dom}(\phi)$ are excluded from this weight. Whenever ϕ is empty in the sense that $\mathrm{dom}(\phi) = \emptyset$ then we say that it is *trivial*. In this case, its appearance in the above expression is vacuous. This is analogously true for D if it is the identity matrix. In either of these cases we omit the terms D (ϕ, respectively) in the expression. For example if $\mathrm{dom}(\phi) = \emptyset$ and $D = I_m$, then

$$Z_A(G) = \sum_{\sigma:V\to[m]} \prod_{uv\in E} A_{\sigma(u),\sigma(v)}.$$

The definitions of $Z_{A,D}(G)$ and $Z_A(\phi, G)$ are analogous. We define $\mathrm{EVAL}^{\mathsf{pin}}(A, D)$ as the computational problem of computing $Z_{A,D}(\phi, G)$ on input ϕ, G. Similarly $\mathrm{EVAL}(A, D)$ restricts the inputs to empty pinnings and $\mathrm{EVAL}^{\mathsf{pin}}(A)$ denotes the problem where D is the identity matrix.

We will now explain the overall structure of the proof of Lemma 4.2. In a first step we will see how we can augment our capabilities so as to fix some vertices of the input graphs, without changing the complexity of the problems under consideration (cf. Lemma 4.3). Then in the General Conditioning Lemma 4.4 we will show, that we can reduce the abundance of non-negative matrices to certain well-structured cases. From these we will show in two steps (Lemmas 4.5 and 4.6) how to obtain #P-hardness.

LEMMA 4.3 (Pinning Lemma). *Let $A \in \mathbb{R}_{\mathbb{A}}^{m\times m}$ be a symmetric non-negative matrix. Then*

$$\mathrm{EVAL}^{pin}(A) \equiv \mathrm{EVAL}(A).$$

A 1-*cell* in a matrix $A \in \mathbb{S}^{m\times m}$ is a submatrix A_{IJ} such that $A_{ij} = 1$ for all $(i, j) \in I \times J$ and $A_{ij} \neq 1$ for all $(i, j) \in (\bar{I} \times J) \cup (I \times \bar{J})$. For a number or an indeterminate X an X*-matrix* is a matrix whose entries are powers of X.

General Conditions. For a matrix $A \in \mathbb{S}^{m\times m}$ we define conditions

(A) A is symmetric positive and has $\operatorname{rank} A \geq 2$.

(B) A is an X-matrix for an indeterminate X.

(C) There is a $k \geq 2$, numbers $1 = m_0 < \ldots < m_k = m + 1$ and $I_i = [m_{i-1}, m_i - 1]$ for all $i \in [k]$ such that $A_{I_iI_i}$ is a 1-cell for every $i \in [k-1]$. The matrix $A_{I_kI_k}$ may or may not be a 1-cell. Furthermore, all 1-entries are contained in one of these 1-cells.

LEMMA 4.4 (General Conditioning Lemma). *Let $A \in \mathbb{S}^{m\times m}$ be a non-negative symmetric matrix which contains a block of rank at least* 2. *Then there is a $\mathbb{Z}[X]$-matrix A' satisfying conditions* **(A)**–**(C)** *such that*

$$\mathrm{EVAL}^{pin}(A') \leq \mathrm{EVAL}^{pin}(A).$$

If a matrix A satisfies the General Conditions two different cases can occur which we will treat separately in the following. The first case is the existence of at least two 1-cells.

LEMMA 4.5 (Two 1-Cell Lemma). *Let $A \in \mathbb{Z}[X]^{m\times m}$ be a positive symmetric matrix containing at least two* 1*-cells. Then* $\mathrm{EVAL}^{pin}(A)$ *is* #P*-hard.*

The second case is then the existence of only one 1-cell. The proof of this case is much more involved than the first one.

LEMMA 4.6 (Single 1-Cell Lemma). *Let $A \in \mathbb{Z}[X]^{m\times m}$ be a matrix which satisfies conditions* **(A)** – **(C)** *and has exactly one* 1*-cell. Then* $\mathrm{EVAL}^{pin}(A)$ *is* #P*-hard.*

Once these results have been derived, it will be easy to prove the main result.

PROOF OF LEMMA 4.2. Let $A \in \mathbb{R}_{\mathbb{A}}^{m\times m}$ be a non-negative symmetric matrix which contains a block of rank at least 2. By the General Conditioning Lemma 4.4 there is a matrix A' satisfying conditions **(A)**–**(C)** such that $\mathrm{EVAL}^{\mathsf{pin}}(A') \leq \mathrm{EVAL}^{\mathsf{pin}}(A)$. If A' contains a single 1-cell then $\mathrm{EVAL}^{\mathsf{pin}}(A')$ is #P-hard by Lemma 4.6. Otherwise, it is #P-hard by Lemma 4.5. In both cases this proves #P-hardness of $\mathrm{EVAL}(A)$ by means of Lemma 4.3. □

5. Evaluation and counting

Let A be an $m \times m$ matrix. Then, for every $q \in \mathbb{Q}$ we define the matrix $A^{(q)}$ by

$$A^{(q)}_{ij} = \begin{cases} (A_{ij})^q & \text{, if } A_{ij} \neq 0 \\ 0 & \text{, otherwise.} \end{cases}$$

The following lemma provides two basic reductions which form basic building blocks of many hardness proofs.

LEMMA 5.1. *Let $A \in \mathbb{S}^{m \times m}$ the following is true for every $p \in \mathbb{N}$*

(p-thickening): $\mathrm{EVAL}^{\mathit{pin}}(A^{(p)}) \leq \mathrm{EVAL}^{\mathit{pin}}(A)$.

(p-stretching): $\mathrm{EVAL}^{\mathit{pin}}(A^{p}) \leq \mathrm{EVAL}^{\mathit{pin}}(A)$.

PROOF. Let $G = (V, E)$ be a graph and ϕ a pinning. Let the *p-thickening* $G^{(p)}$ of G be the graph obtained from G by replacing each edge by p many parallel edges. The *p-stretching* G_p of G is obtained from G by replacing each edge by a path of length p. The reductions follow from the (easily verifiable) identities $Z_{A^{(p)}}(\phi, G) = Z_A(\phi, G^{(p)})$ and $Z_{A^p}(\phi, G) = Z_A(\phi, G_p)$. □

LEMMA 5.2. *Let $A \in \mathbb{S}^{m \times n}$ be non-negative. If A contains a block of rank ≥ 2 then so does AA^T.*

PROOF. Let A_{IJ} be a block of rank at least 2 in A and define $A' = AA^T$. We claim that A'_{II} contains a block of rank at least 2, which clearly implies the statement of the lemma.

For $i, i' \in I$ let $A_{i,*}$ and $A_{i',*}$ be two linearly independent rows. Since A_{IJ} is a block, there are indices $i = i_1, \dots, i_k = i'$ in I and $j_1, \dots j_{k-1} \in J$ such that $A_{i_\nu j_\nu} \neq 0$ and $A_{i_{\nu+1} j_\nu} \neq 0$ for all $\nu \in [k-1]$. Linear independence of $A_{i,*}$ and $A_{i',*}$ implies that there is a $\mu \in [k-1]$ such that also $A_{i_\mu,*}$ and $A_{i_{\mu+1},*}$ are linearly independent. We claim that the submatrix

$$\begin{pmatrix} A'_{i_\mu, i_\mu} & A'_{i_\mu, i_{\mu+1}} \\ A'_{i_{\mu+1}, i_\mu} & A'_{i_{\mu+1}, i_{\mu+1}} \end{pmatrix} = \begin{pmatrix} \langle A_{i_\mu,*}, A_{i_\mu,*} \rangle & \langle A_{i_\mu,*}, A_{i_{\mu+1},*} \rangle \\ \langle A_{i_{\mu+1},*}, A_{i_\mu,*} \rangle & \langle A_{i_{\mu+1},*}, A_{i_{\mu+1},*} \rangle \end{pmatrix}$$

is a witness for the existence of a block of rank at least two in A'. To see this note first that, by definition, all entries of this submatrix are positive, it thus remains to show that this submatrix has rank 2. Assume, for contradiction, that it has zero determinant, that is

$$\langle A_{i_\mu,*}, A_{i_\mu,*} \rangle \langle A_{i_{\mu+1},*}, A_{i_{\mu+1},*} \rangle = \langle A_{i_{\mu+1},*}, A_{i_\mu,*} \rangle^2.$$

The Cauchy Schwarz inequality therefore implies linear dependence of $A_{i_\mu,*}$ and $A_{i_{\mu+1},*}$. Contradiction. □

For A an $m \times m$ matrix and $\pi : [m] \to [m]$ a permutation, define $A_{\pi\pi}$ by

$$(A_{\pi\pi})_{ij} = A_{\pi(i)\pi(j)} \text{ for all } i, j \in [m].$$

The *Permutability Principle* states that for any evaluation problem on some matrix A, we may assume any simultaneous permutation of the rows and columns of this matrix. Its proof is straightforward.

LEMMA 5.3 (Permutability Principle). *Let $A, D \in \mathbb{S}^{m \times m}$ and $\pi : [m] \to [m]$ a permutation. Then* $\mathrm{EVAL}^{\mathit{pin}}(A, D) \equiv \mathrm{EVAL}^{\mathit{pin}}(A_{\pi\pi}, D_{\pi\pi})$.

We will make extensive use of interpolation; the following simple lemma is one instance.

LEMMA 5.4. *For some fixed $\theta \in \mathbb{R}_{\mathbb{A}}$ let $x_1, \ldots, x_n \in \mathbb{Q}(\theta)$ be pairwise different and non-negative reals. Let $b_1, \ldots, b_n \in \mathbb{Q}(\theta)$ be arbitrary such that*

$$b_j = \sum_{i=1}^{n} c_i x_i^j \text{ for all } j \in [n].$$

Then the coefficients $c_1, \ldots, c_n$ are uniquely determined and can be computed in polynomial time.

5.1. The Equivalence of $\mathbf{EVAL}^{\mathsf{pin}}(A)$ and $\mathbf{COUNT}^{\mathsf{pin}}(A)$. Let $A \in \mathbb{S}^{m \times m}$ be a matrix, $G = (V, E)$ a graph and ϕ a pinning. We define a set of *potential weights*

(5.1)

$$\mathcal{W}_A(G) := \left\{ \prod_{i,j \in [m]} A_{ij}^{m_{ij}} \mid \sum_{i,j \in [m]} m_{ij} = |E|, \text{ and } m_{ij} \geq 0, \text{ for all } i, j \in [m] \right\}.$$

For every $w \in \mathbb{S}$ define the value

$$N_A(G, \phi, w) := \left| \left\{ \sigma : V \to [m] \middle| \mid \phi \subseteq \sigma, \ w = \prod_{uv \in E} A_{\sigma(u)\sigma(v)} \right\} \right|.$$

By $\mathrm{COUNT}^{\mathsf{pin}}(A)$ we denote the problem of computing $N_A(G, \phi, w)$ for given G, ϕ and $w \in \mathbb{S}$. In analogy to the evaluation problems we write $\mathrm{COUNT}(A)$ for the subproblem restricted to instances with trivial pinnings. It turns out that these problems are computationally equivalent to the evaluation problems of partition functions.

LEMMA 5.5. *For every matrix $A \in \mathbb{S}^{m \times m}$ we have*

$$\mathrm{EVAL}^{\mathit{pin}}(A) \equiv \mathrm{COUNT}^{\mathit{pin}}(A) \quad \textit{and} \quad \mathrm{EVAL}(A) \equiv \mathrm{COUNT}(A).$$

PROOF. Let $G = (V, E)$ be a graph and ϕ a pinning. We have

$$Z_A(\phi, G) = \sum_{\phi \subseteq \sigma : V \to [m]} \prod_{uv \in E} A_{\sigma(u)\sigma(v)} = \sum_{w \in \mathcal{W}_A(G)} w \cdot N_A(G, \phi, w).$$

As the cardinality of $\mathcal{W}_A(G)$ is polynomial in the size of G this proves the reducibilities

$$\mathrm{EVAL}^{\mathsf{pin}}(A) \leq \mathrm{COUNT}^{\mathsf{pin}}(A) \quad \text{and} \quad \mathrm{EVAL}(A) \leq \mathrm{COUNT}(A).$$

For the backward direction let $G^{(t)}$ denote the graph obtained from G by replacing each edge by t parallel edges. We have

$$Z_A(\phi, G^{(t)}) = \sum_{\phi \subseteq \sigma : V \to [m]} \left(\prod_{uv \in E} A_{\sigma(u)\sigma(v)} \right)^t = \sum_{w \in \mathcal{W}_A(G)} w^t \cdot N_A(G, \phi, w).$$

Using an $\mathrm{EVAL}^{\mathsf{pin}}(A)$ oracle, we can evaluate this for $t = 1, \ldots, |\mathcal{W}_A(G)|$. Therefore, if $\mathbb{S}$ is one of $\mathbb{R}_{\mathbb{A}}, \mathbb{Q}, \mathbb{Z}$, the values $N_A(G, \phi, w)$ can be recovered in polynomial time by Lemma 5.4.

If $\mathbb{S}$ is one of $\mathbb{Z}[X]$ or $\mathbb{Q}[X]$, then let $f_t(X)$ denote $Z_A(\phi, G^{(t)})$. We obtain the following system of equations, for $t = 1, \ldots, |\mathcal{W}_A(G)|$,

$$f_t(X) = \sum_{w \in \mathcal{W}_A(G)} w^t(X) \cdot N_A(G, \phi, w). \tag{5.2}$$

Let $\det(X)$ denote the determinant of this system of equations. W.l.o.g., all the $w(X) \in \mathcal{W}_A(G)$ are non-zero polynomials, $\det(X)$, as it is a Vandermonde determinant in these $w(X)$, is itself a non-zero polynomial of the form

$$\det(X) = \prod_{w \in \mathcal{W}_A(G)} w(X) \cdot \prod_{w \neq w' \in \mathcal{W}_A(G)} (w(X) - w'(X)).$$

Let $\delta = 1 + \max\{\deg w(X) \mid w \in \mathcal{W}_A(G)\}$ and observe that each $w \in \mathcal{W}_A(G)$ has at most $\delta - 1$ roots. Further, each of the terms $w(X) - w'(X)$ of $\det(X)$ has degree at most $\delta - 1$ and thus the degree of $\det(X)$ is strictly smaller than $\delta' := \binom{|\mathcal{W}_A(G)|+1}{2} \cdot \delta$. Hence there is an integer $a \leq \delta'$ such that $\det(a)$ is non-zero. By equation (5.2) we obtain an invertible system of equations

$$f_t(a) = \sum_{w \in \mathcal{W}_A(G)} w^t(a) \cdot N_A(G, \phi, w). \tag{5.3}$$

The coefficients $N_A(G, \phi, w)$ can be now obtained in polynomial time by Lemma 5.4. This finishes the proof of $\mathrm{COUNT}^{\mathsf{pin}}(A) \equiv \mathrm{EVAL}^{\mathsf{pin}}(A)$. The proof for $\mathrm{COUNT}(A) \equiv \mathrm{EVAL}(A)$ also follows by the argument just presented, as the given pinnings remain unaffected. □

5.2. Dealing with Vertex Weights.

Lemma 5.6 (Theorem 3.2 in [**DG00**]). *Let $A \in \mathbb{R}_{\mathbb{A}}^{m \times m}$ be a symmetric matrix with non-negative entries such that every pair of rows in A is linearly independent. Let $D \in \mathbb{R}_{\mathbb{A}}^{m \times m}$ be a diagonal matrix of positive vertex weights. Then*

$$\mathrm{EVAL}^{pin}(A) \leq \mathrm{EVAL}^{pin}(A, D) \text{ and } \mathrm{EVAL}(A) \leq \mathrm{EVAL}(A, D).$$

After some preparation, we will prove this lemma in Section 5.2.2.

5.2.1. *Some Technical Tools.* The following is a lemma from [**DG00**] (Lemma 3.4).

Lemma 5.7. *Let $A \in \mathbb{R}_{\mathbb{A}}^{m \times m}$ be symmetric and non-singular, $G = (V, E)$ a graph, ϕ a pinning, and $F \subseteq E$. If we know the values*

$$f_r(G) = \sum_{\phi \subseteq \sigma : V \to [m]} c_A(\sigma) \prod_{uv \in F} A^r_{\sigma(u)\sigma(v)} \tag{5.4}$$

for all $r \in [(|F| + 1)^{m^2}]$, where c_A is a function depending on A but not on r. Then we can evaluate

$$\sum_{\phi \subseteq \sigma : V \to [m]} c_A(\sigma) \prod_{uv \in F} (I_m)_{\sigma(u)\sigma(v)}$$

in polynomial time.

Proof. As A is symmetric and non-singular, there is an orthogonal matrix P such that $P^T A P =: D$ is a diagonal matrix with non-zero diagonal. From

$A = PDP^T$ we have $A^r = PD^rP^T$ and thus every entry of A^r satisfies $A^r_{ij} = (PD^rP^T)_{ij} = \sum_{\mu=1}^m P_{\sigma(u)\mu}P_{\sigma(v)\mu}(D_{\mu\mu})^r$. Hence equation (5.4) can be rewritten as

$$f_r(G) \quad = \quad \sum_{\phi \subseteq \sigma : V \to [m]} c_A(\sigma) \prod_{uv \in F} \sum_{\mu=1}^m P_{\sigma(u)\mu}P_{\sigma(v)\mu}(D_{\mu\mu})^r$$

Define the set

$$\mathcal{W} = \left\{ \prod_{i=1}^m (D_{ii})^{\alpha_i} \mid 0 \leq \alpha_i \text{ for all } i \in [m], \ \sum_{i=1}^m \alpha_i = |F| \right\}$$

which can be constructed in polynomial time. We rewrite

$$f_r(G) = \sum_{w \in \mathcal{W}} c_w w^r.$$

for unknown coefficients c_w. By interpolation (cf. Lemma 5.4), we can recover these coefficients in polynomial time and can thus calculate $f_0(G) = \sum_{w \in \mathcal{W}} c_w$. We have

$$\begin{aligned} f_0(G) \quad &= \quad \sum_{\phi \subseteq \sigma : V \to [m]} c_A(\sigma) \prod_{uv \in F} \sum_{\mu=1}^m P_{\sigma(u)\mu}P_{\sigma(v)\mu}(D_{\mu\mu})^0 \\ &= \quad \sum_{\phi \subseteq \sigma : V \to [m]} c_A(\sigma) \prod_{uv \in F} I_{\sigma(u)\sigma(v)} \end{aligned}$$

which finishes the proof. □

The following two lemmas are restatements of those in [**DG00**] (see Lemma 3.6, 3.7 and Theorem 3.1).

LEMMA 5.8. *Let $A \in \mathbb{R}^{m \times m}$ be a symmetric matrix in which every pair of distinct rows is linearly independent. Let $D \in \mathbb{R}^{m \times m}$ be a diagonal matrix of positive vertex weights. Then every pair of rows in ADA is linearly independent. Furthermore there is an $0 < \epsilon < 1$ such that for all $i \neq j$*

$$|(ADA)_{ij}| \leq \epsilon \sqrt{(ADA)_{ii}(ADA)_{jj}}$$

PROOF. Define $Q = AD^{(1/2)}$. We have $ADA = AD^{(1/2)}D^{(1/2)}A^T = QQ^T$. That is $(ADA)_{ij} = \langle Q_{i,*}, Q_{j,*} \rangle$ every pair of rows in Q is linearly independent as it is linearly independent in A. Hence, by the Cauchy-Schwarz inequality

$$\langle Q_{i,*}, Q_{j,*} \rangle < \sqrt{\langle Q_{i,*}, Q_{i,*} \rangle \langle Q_{j,*}, Q_{j,*} \rangle}$$

which implies that the corresponding 2×2 submatrix of ADA defined by i and j has non-zero determinant. The existence of ϵ follows. □

LEMMA 5.9. *Let $A \in \mathbb{R}^{m \times m}$ be a symmetric non-negative matrix in which every pair of distinct rows is linearly independent. Let $D \in \mathbb{R}^{m \times m}$ be a diagonal matrix of positive vertex weights. Then there is a $p \in \mathbb{N}$ such that the matrix*

$$(ADA)^{(p)}$$

is non-singular.

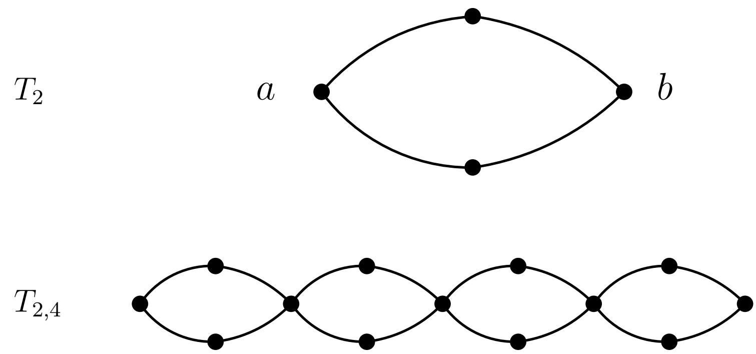


FIGURE 3. The graphs T_2 and $T_{2,4}$.

PROOF. Let $A' = ADA$ and consider the determinant

$$\det(A') = \sum_{\pi \in S_m} \operatorname{sgn}(\pi) \prod_{i=1}^{m} A'_{i\pi(i)}$$

where S_m is the set of permutations of $[m]$. For some $\pi \in S_m$ define $t(\pi) = |\{i \mid \pi(i) \neq i\}|$. Let ϵ be as in Lemma 5.8. Then

$$\prod_{i=1}^{m} |A'_{i\pi(i)}| \leq \epsilon^{t(\pi)} \prod_{i=1}^{m} \sqrt{A'_{ii}} \prod_{i=1}^{m} \sqrt{A'_{\pi(i)\pi(i)}} = \epsilon^{t(\pi)} \prod_{i=1}^{m} A'_{ii}. \tag{5.5}$$

Let **id** denote the trivial permutation. Then

$$\det((A')^{(p)}) \quad \geq \quad \left(\prod_{i=1}^{m} A'_{ii} \right)^p - \sum_{\pi \in S_m \setminus \{\mathbf{id}\}} \left(\prod_{i=1}^{m} A'_{i\pi(i)} \right)^p .$$

By equation (5.5), we have

$$m!\epsilon^p \left(\prod_{i=1}^{m} A'_{ii} \right)^p \quad \geq \quad \sum_{\pi \in S_m \setminus \{\mathbf{id}\}} \left(\prod_{i=1}^{m} A'_{i\pi(i)} \right)^p$$

and hence, as $0 < \epsilon < 1$, the matrix $(ADA)^{(p)}$ is non-singular for large enough p. □

5.2.2. *The proof of Lemma 5.6.* By Lemma 5.9 there is a $p \in \mathbb{N}$ such that $(ADA)^{(p)}$ is non-singular. We will fix such a p for the rest of the proof.

Let G, ϕ be an instance of $\text{EVAL}^{\mathsf{pin}}(A)$ with $G = (V, E)$ a graph. Construct from G a graph G' as follows. Define

$$V' \quad = \quad \{v_0, \dots, v_{d-1} \mid v \in V, d = d_G(v)\}$$

that is, for each vertex $v \in V$ we introduce $d(v)$ new vertices. Let ϕ' be the pinning which for every $v \in \text{dom}(\phi)$ satisfies $\phi'(v_0) = \phi(v)$. Let E''' be the set which contains, for each $v \in V$ the cycle on $\{v_0, \dots, v_{d_G(v)-1}\}$, i.e. $\{v_0v_1, \dots, v_{d_G(v)-1}v_0\} \subseteq E'''$. Let E'' be the set of edges, such that each edge from E incident with v is connected to exactly one of the v_i. Define $E' = E'' \cup E'''$ and denote by $G'_{p,r}$ the graph obtained from G' by replacing each edge in E''' with a distinct copy of the graph $T_{p,r}$ to be defined next.

To conveniently define the graph $T_{p,r}$ we will first describe another graph T_p. This graph consists of two distinguished vertices a and b connected by p many length 2 paths from a to b. Then $T_{p,r}$ is the series composition of r many copies of T_p. The construction is illustrated in Figure 3. We call a and b the "start" and "end" vertex of T_p and $T_{p,r}$ has start and end vertices induced by the start and end vertices of the series composition of T_p.

For a graph H with designated "start" and "end" vertex by $Z_{A,D}(i,j;H)$ we denote the partition function $Z_{A,D}(\phi, H)$ for ϕ such that it pins the start vertex of H to i and the end vertex to j. Note that, by definition, the vertex weights of the pinned vertices do not occur in this term.

Claim 1. Let $C = (ADA)^{(p)}$, then with $X = D^{(1/2)}$ and $Y = XCX$, we have for all $i,j \in [m]$ and $r \in \mathbb{N}$,

$$Z_{A,D}(i,j;T_{p,r}) = (X_{ii}X_{jj})^{-1}(Y^r)_{ij}. \tag{5.6}$$

Proof. Straightforwardly,

$$Z_{A,D}(i,j;T_p) = \left(\sum_{k=1}^{m} A_{ik}A_{kj}D_{kk}\right)^p = (ADA)^{(p)}_{ij} = C_{ij}$$

and therefore

$$\begin{aligned} Z_{A,D}(i,j;T_{p,r}) &= \sum_{\substack{\sigma:[r+1]\to[m]\\ \sigma(1)=i,\, \sigma(r+1)=j}} \prod_{k=1}^{r} Z_{A,D}(\sigma(k),\sigma(k+1);T_p) \prod_{k=2}^{r} D_{\sigma(k),\sigma(k)} \\ &= \sum_{\substack{\sigma:[r+1]\to[m]\\ \sigma(1)=i,\, \sigma(r+1)=j}} \prod_{k=1}^{r} C_{\sigma(k),\sigma(k+1)} \prod_{k=2}^{r} (X_{\sigma(k),\sigma(k)})^2 \\ &= (X_{ii}X_{jj})^{-1} \sum_{\substack{\sigma:[r+1]\to[m]\\ \sigma(1)=i,\\ \sigma(r+1)=j}} \prod_{k=1}^{r} X_{\sigma(k),\sigma(k)}C_{\sigma(k),\sigma(k+1)}X_{\sigma(k+1),\sigma(k+1)} \end{aligned}$$

By inspection, the last line equals the right hand side of equation (5.6) — as claimed. ⊣

Using the expression just obtained for $Z_{A,D}(i,j;T_{p,r})$, we will now rewrite $Z_{A,D}(\phi', G'_{p,r})$. To do so, we will, for all $v \in V$, count the indices of the vertices $v_0, \ldots, v_{d_G(v)-1}$ modulo $d_G(v)$ such that, in particular, $v_{d_G(v)} = v_0$. Define $\gamma = \prod_{v\in \mathrm{dom}(\phi')} D_{\phi'(v),\phi'(v)}$. We have for all $r \in \mathbb{N}$,

$$\begin{aligned} Z_{A,D}(\phi', G'_{p,r}) &= \gamma^{-1} \sum_{\phi' \subseteq \sigma: V' \to [m]} \prod_{uv\in E''} A_{\sigma(u)\sigma(v)} \prod_{v\in V'} D_{\sigma(v),\sigma(v)} \\ &\qquad \cdot \prod_{v\in V} \prod_{i=0}^{d_G(v)-1} Z_{A,D}(\sigma(v_i),\sigma(v_{i+1});T_{p,r}). \end{aligned}$$

As the vertices in V' are grouped according to the vertices in V we have

$$\prod_{v\in V'} D_{\sigma(v),\sigma(v)} = \prod_{v\in V} \prod_{i=0}^{d_G(v)-1} D_{\sigma(v_i),\sigma(v_i)}.$$

Using equation (5.6), we further see that for each $\sigma : V \to [m]$, the expression

$$\prod_{v \in V'} D_{\sigma(v),\sigma(v)} \prod_{v \in V} \prod_{i=0}^{d_G(v)-1} Z_{A,D}(\sigma(v_i), \sigma(v_{i+1}); T_{p,r})$$

turns into

$$\prod_{v \in V} \prod_{i=0}^{d_G(v)-1} D_{\sigma(v_i),\sigma(v_i)} \frac{(Y^r)_{\sigma(v_i),\sigma(v_{i+1})}}{X_{\sigma(v_i),\sigma(v_i)} X_{\sigma(v_{i+1}),\sigma(v_{i+1})}} = \prod_{v \in V} \prod_{i=0}^{d_G(v)-1} (Y^r)_{\sigma(v_i),\sigma(v_{i+1})}$$

Thus

$$\begin{aligned} Z_{A,D}(\phi', G'_{p,r}) &= \gamma^{-1} \sum_{\phi' \subseteq \sigma: V' \to [m]} \prod_{uv \in E''} A_{\sigma(u)\sigma(v)} \prod_{v \in V} \prod_{i=0}^{d_G(v)-1} (Y^r)_{\sigma(v_i)\sigma(v_{i+1})} \\ &= \gamma^{-1} \sum_{\phi' \subseteq \sigma: V' \to [m]} \prod_{uv \in E''} A_{\sigma(u)\sigma(v)} \prod_{uv \in E'''} (Y^r)_{\sigma(u)\sigma(v)} \end{aligned}$$

Given the EVAL(A, D) oracle, we can, in polynomial time, evaluate this expression for every r which is polynomial in the size of G. Therefore Lemma 5.7 implies that we can compute the value

$$Z = \gamma^{-1} \sum_{\phi' \subseteq \sigma: V' \to [m]} \prod_{uv \in E''} A_{\sigma(u)\sigma(v)} \prod_{uv \in E'''} I_{\sigma(u)\sigma(v)}$$

in polynomial time. The proof follows, if we can show that $\gamma \cdot Z = Z_A(\phi, G)$. To see this, note that, for every configuration $\sigma : V' \to [m]$ the corresponding weight $\prod_{uv \in E''} A_{\sigma(u)\sigma(v)} \prod_{uv \in E'''} I_{\sigma(u)\sigma(v)}$ in the above expression is zero unless the following holds: For all $v \in V$ we have $\sigma(v_0) = \ldots = \sigma(v_{d-1})$ for $d = d_G(v)$. for such a configuration, define a configuration $\sigma' : V \to [m]$ such that $\sigma(v) = \sigma(v_0)$ for every $v \in V$. It follows from the construction of G' that then

$$\prod_{uv \in E''} A_{\sigma(u),\sigma(v)} = \prod_{uv \in E} A_{\sigma'(u),\sigma'(v)}.$$

Since every configuration $\sigma' : V \to [m]$ arises this way, we have $\gamma \cdot Z = Z_A(\phi, G)$. This finishes the proof of $\text{EVAL}^{\mathsf{pin}}(A) \leq \text{EVAL}^{\mathsf{pin}}(A, D)$. Since the proof is correct also for empty input pinnings ϕ, this also proves EVAL$(A) \leq$ EVAL(A, D).

6. The Pinning Lemma

In this section we shall give the proof of the Pinning Lemma 4.3. Before we do this, however, we will introduce a technical reduction which will also be used later.

6.1. The Twin Reduction Lemma. For a symmetric $m \times m$ matrix A, we say that two rows $A_{i,*}$ and $A_{j,*}$ are *twins*, if $A_{i,*} = A_{j,*}$. A matrix A is *twin-free* if $A_{i,*} \neq A_{j,*}$ for all row indices $i \neq j$. The concept of twins induces an equivalence relation on the rows of A. Let $I_1, \ldots, I_k$ be the equivalence classes of this relation. The *twin resolvent* of A is the $k \times k$ matrix $[A]$, defined by

$$[A]_{i,j} = A_{\mu,\nu} \text{ for some } \mu \in I_i \text{ and } \nu \in I_j.$$

We say that $[A]$ is obtained from A by *twin reduction*. Further, we define the *twin resolution mapping* $\tau : [m] \to [k]$ of A in such a way that for all $\mu \in [m]$ we have $\mu \in I_{\tau(\mu)}$. That is, τ maps every $\mu \in [m]$ to the class I_j it is contained in. Hence

$$[A]_{\tau(i),\tau(j)} = A_{i,j} \text{ for all } i, j \in [m]. \tag{6.1}$$

To use twin reductions in the context of partition functions we need to clarify the effect which twin reduction on a matrix A induces on the corresponding diagonal matrix of vertex weights D. We will see that the diagonal $k \times k$ matrix $D^{[A]}$ defined in the following captures this effect

$$D^{[A]}_{i,i} = \sum_{\nu \in I_i} D_{\nu,\nu} \text{ for all } i \in [k]. \tag{6.2}$$

LEMMA 6.1 (Twin Reduction Lemma). *Let $A \in \mathbb{R}_{\mathbb{A}}^{m\times m}$ be non-negative and symmetric and $D \in \mathbb{R}_{\mathbb{A}}^{m\times m}$ a diagonal matrix of positive vertex weights. Let $[A]$ be the twin resolvent of A. Then*

$$\mathrm{EVAL}([A], D^{[A]}) \equiv \mathrm{EVAL}(A, D) \quad \textit{and} \quad \mathrm{EVAL}^{\mathsf{pin}}([A], D^{[A]}) \equiv \mathrm{EVAL}^{\mathsf{pin}}(A, D).$$

PROOF. Let τ be the twin-resolution mapping of A. Let $G = (V, E)$ be a graph and ϕ a pinning. Define $V' = V \setminus \mathrm{dom}(\phi)$. By the definition of τ we have

$$\begin{aligned} Z_{A,D}(\phi, G) &= \sum_{\phi \subseteq \sigma : V \to [m]} \prod_{uv \in E} A_{\sigma(u),\sigma(v)} \prod_{v \in V'} D_{\sigma(v),\sigma(v)} \\ &= \sum_{\phi \subseteq \sigma : V \to [m]} \prod_{uv \in E} [A]_{\tau\circ\sigma(u),\tau\circ\sigma(v)} \prod_{v \in V'} D_{\sigma(v),\sigma(v)}. \end{aligned}$$

For all configurations $\sigma : V \to [m]$ we have $\tau \circ \sigma : V \to [k]$. Hence, we can partition the configurations σ into classes according to their images under concatenation with τ and obtain

$$\begin{aligned} Z_{A,D}(\phi, G) &= \sum_{\sigma' : V \to [k]} \sum_{\substack{\phi \subseteq \sigma : V \to [m] \\ \tau\circ\sigma = \sigma'}} \prod_{uv \in E} [A]_{\sigma'(u),\sigma'(v)} \prod_{v \in V'} D_{\sigma(v),\sigma(v)} \\ &= \sum_{\tau\circ\phi \subseteq \sigma' : V \to [k]} \prod_{uv \in E} [A]_{\sigma'(u),\sigma'(v)} \cdot \Delta(\sigma'). \end{aligned}$$

Here, $\Delta(\sigma')$ is defined by

$$\Delta(\sigma') = \sum_{\substack{\phi \subseteq \sigma : V \to [m] \\ \tau\circ\sigma = \sigma'}} \prod_{v \in V'} D_{\sigma(v),\sigma(v)}.$$

Fix some $\sigma' : V \to [k]$, we will argue that

$$\Delta(\sigma') = \prod_{v \in V'} D^{[A]}_{\sigma'(v),\sigma'(v)} \tag{6.3}$$

For every configuration $\sigma : V \to [m]$ we have $\tau \circ \sigma = \sigma'$ if, and only if, ${\sigma'}^{-1}(\{i\}) = \sigma^{-1}(I_i)$ for all $i \in [k]$. Define, for each $i \in [k]$, the set $V_i := {\sigma'}^{-1}(\{i\})$ and the mapping $\phi_i := \phi \restriction_{\mathrm{dom}(\phi)\cap V_i}$. Then

$$\Delta(\sigma') = \sum_{\substack{\phi \subseteq \sigma : V \to [m] \\ \forall\, i \in [k]:\, \sigma(V_i) \subseteq I_i}} \prod_{v \in V'} D_{\sigma(v),\sigma(v)}$$

Define $V_i' := V_i \setminus \mathrm{dom}(\phi_i)$, then

$$\begin{aligned}\Delta(\sigma') &= \prod_{i=1}^{k} \sum_{\phi_i \subseteq \sigma_i : V_i \to I_i} \prod_{v \in V_i'} D_{\sigma_i(v),\sigma_i(v)} \\ &= \prod_{i=1}^{k} \prod_{v \in V_i'} \sum_{\nu \in I_i} D_{\nu,\nu} \\ &= \prod_{v \in V'} D^{[A]}_{\sigma'(v),\sigma'(v)}\end{aligned}$$

This proves equation (6.3). Therefore,

$$Z_{A,D}(\phi, G) = \sum_{\tau\circ\phi \subseteq \sigma' : V \to [k]} \prod_{uv \in E} [A]_{\sigma'(u),\sigma'(v)} \prod_{v \in V'} D^{[A]}_{\sigma'(v),\sigma'(v)} = Z_{[A],D^{[A]}}(\tau \circ \phi, G).$$

This witnesses the claimed reducibilities. □

6.2. Proof of the Pinning Lemma. We shall prove the Pinning Lemma 4.3 now. The proof of this Lemma relies on a result of Lovász [**Lov06**]. To state this result we need some further preparation. Let A, D be $m \times m$ matrices and A', D' be $n \times n$ matrices such that D and D' are diagonal. The pairs (A, D) and (A', D') are *isomorphic*, if there is a bijection $\alpha : [m] \to [n]$ such that $A_{ij} = A'_{\alpha(i)\alpha(j)}$ for all $i, j \in [m]$ and $D_{ii} = D'_{\alpha(i),\alpha(i)}$ for all $i \in [m]$. An *automorphism* is an isomorphism of (A, D) with itself.

We will moreover need to consider pinnings a bit differently than before. Rather than defining a pinning ϕ for some given graph G, it will be convenient in the following to fix pinnings and consider graphs which are compatible with these. To define this more formally, we fix $\phi : [k] \to [m]$ to denote our pinning. A *k-labeled graph* $G = (V, E)$ is then a graph whose vertex set satisfies $V \supseteq [k]$. In this way ϕ is compatible with every k-labeled graph.

LEMMA 6.2 (Lemma 2.4 in [**Lov06**]). *Let $A \in \mathbb{R}^{m\times m}$ be a non-negative symmetric and twin-free matrix and $D \in \mathbb{R}^{m\times m}$ a diagonal matrix of positive vertex weights. Let ϕ, ψ be pinnings. If $Z_{A,D}(\phi, G) = Z_{A,D}(\psi, G)$ for all k-labeled graphs G, then there is an automorphism α of (A, D) such that $\phi = \alpha(\psi)$.*

Utilizing this result, we can now prove the pinning result first for twin-free matrices.

LEMMA 6.3. *Let $A \in \mathbb{R}_{\mathbb{A}}^{m\times m}$ be non-negative, symmetric, and twin-free and $D \in \mathbb{R}_{\mathbb{A}}^{m\times m}$ a diagonal matrix of positive vertex weights. Then*

$$\mathrm{EVAL}^{\mathsf{pin}}(A, D) \equiv \mathrm{EVAL}(A, D).$$

PROOF. As $\mathrm{EVAL}(A, D) \leq \mathrm{EVAL}^{\mathsf{pin}}(A, D)$ holds trivially we only need to prove $\mathrm{EVAL}^{\mathsf{pin}}(A, D) \leq \mathrm{EVAL}(A, D)$.

Let $G = (V, E)$ and a pinning ϕ be an instance of $\mathrm{EVAL}^{\mathsf{pin}}(A, D)$. By appropriate permutation of the rows/columns of A and D (cf. Lemma 5.3) we may assume that $[k] = \mathrm{img}\ \phi \subseteq [m]$ for some $k \leq m$. Let $\hat{G} = (\hat{V}, \hat{E})$ be the graph obtained from G by collapsing the the sets $\phi^{-1}(i)$ for all $i \in [k]$. Formally, define a map

$$\gamma(v) = \begin{cases} i & , v \in \phi^{-1}(i) \text{ for some } i \in [k] \\ v & , \text{ otherwise} \end{cases}$$

Then $\hat{G}$ is a k-labeled multigraph (with possibly some self-loops) defined by

$$\begin{aligned}\hat{V} &= [k] \,\dot{\cup}\, (V \setminus \mathrm{dom}(\phi)) \\ \hat{E} &= \{\gamma(u)\gamma(v) \mid uv \in E\}.\end{aligned}$$

Recall that in partition functions of the form $Z_{A,D}(\phi, G)$ vertices pinned by ϕ do not contribute any vertex weights. Hence, $Z_{A,D}(\phi, G) = Z_{A,D}(\mathbf{id}_{[k]}, \hat{G})$ where $\mathbf{id}_{[k]}$ denotes the identity map on $[k]$. Call two mappings $\chi, \psi : [k] \to [m]$ *equivalent* if there is an automorphism α of (A, D) such that $\chi = \alpha \circ \psi$. Partition the mappings $\psi : [k] \to [m]$ into equivalence classes $I_1, \dots, I_c$ according to this definition and for all $i \in [c]$ fix some $\psi_i \in I_i$. Assume furthermore, that $\psi_1 = \mathbf{id}_{[k]}$. Clearly for any two χ, ψ from the same equivalence class, we have $Z_{A,D}(\chi, F) = Z_{A,D}(\psi, F)$ for every graph F. Therefore, for every graph G',

$$Z_{A,D}(G') = \sum_{i=1}^{c} Z_{A,D}(\psi_i, G') \cdot \left(\sum_{\psi \in I_i} \prod_{v \in \mathrm{dom}(\psi)} D_{\psi(v)\psi(v)} \right) \tag{6.4}$$

Define, for each $i \in [c]$ the value $c_i = \left(\sum_{\psi \in I_i} \prod_{v \in \mathrm{dom}(\psi)} D_{\psi(v)\psi(v)} \right)$. We claim the following

Claim 2. Let $I \subseteq [c]$ be a set of cardinality at least 2 such that $1 \in I$. Assume that we can, for every k-labeled graph G', compute the value

$$\sum_{i \in I} c_i \cdot Z_{A,D}(\psi_i, G'). \tag{6.5}$$

Then there is a proper subset $I' \subset I$ which contains 1 such that we can compute, for every k-labeled graph G'', the value

$$\sum_{i \in I'} c_i \cdot Z_{A,D}(\psi_i, G''). \tag{6.6}$$

This claim will allow us to finish the proof. To see this, note first that by equation (6.4) we can compute the value (6.5) for $I = [c]$ and $G' = \hat{G}$. Thus after at most c iterations of Claim 2 we arrive at $c_1 \cdot Z_{A,D}(\psi_1, \hat{G})$. Further, c_1 is effectively computable in time depending only on D and therefore we can compute $Z_{A,D}(\mathbf{id}_{[k]}, \hat{G}) = Z_{A,D}(\phi, G)$. This proves the reducibility $\mathrm{EVAL}^{\mathsf{pin}}(A, D) \leq \mathrm{EVAL}(A, D)$.

Proof Of Claim 2. Assume that we can compute the value given in (6.5). Lemma 6.2 implies that for every pair $i \neq j \in I$ there is a k-labeled graph Γ such that

$$Z_{A,D}(\psi_i, \Gamma) \neq Z_{A,D}(\psi_j, \Gamma). \tag{6.7}$$

Fix such a pair $i \neq j \in I$ and a graph Γ satisfying this equation. Note that this graph can be computed effectively in time depending only on A, D and ψ_i, ψ_j. Let G^s denote the graph obtained from G by iterating s times the k-labeled product of G with itself. We can thus compute

$$\sum_{i \in I} c_i \cdot Z_{A,D}(\psi_i, G'\Gamma^s) = \sum_{i \in I}^{c} c_i Z_{A,D}(\psi_i, G') \cdot Z_{A,D}(\psi_i, \Gamma)^s. \tag{6.8}$$

Partition I into classes $J_0, \ldots, J_z$ such that for every $\nu \in [0, z]$ we have $i', j' \in J_\nu$ if, and only if, $Z_{A,D}(\psi_{i'}, \Gamma) = Z_{A,D}(\psi_{j'}, \Gamma)$. Since one of these sets J_ν contains 1 and all of these are proper subsets of I, it remains to show that we can compute, for each $\nu \in [0, z]$, the value

$$\sum_{i' \in J_\nu} c_i Z_{A,D}(\psi_{i'}, G').$$

To prove this, define $x_\nu := Z_{A,D}(\psi_{i'}, \Gamma)$ for each $\nu \in [z]$ and an $i' \in J_\nu$. Equation (6.8) implies that we can compute

$$\sum_{\nu=0}^{z} x_\nu^s \left(\sum_{i' \in J_\nu} c_{i'} Z_{A,D}(\psi_{i'}, G') \right).$$

One of the values x_ν might be zero. Assume therefore w.l.o.g. that $x_0 = 0$, then evaluating the above for $s = 1, \ldots, z$ yields a system of linear equations, which by Lemma 5.4 can be solved in polynomial time such that we can recover the values $\sum_{i' \in J_\nu} c_{i'} Z_{A,D}(\psi_{i'}, G')$ for each $\nu \geq 1$. Using equation (6.5) we can thus also compute the value

$$\left(\sum_{i \in I} c_i \cdot Z_{A,D}(\psi_i, G') \right) - \sum_{\nu=1}^{z} \left(\sum_{i' \in J_\nu} c_{i'} Z_{A,D}(\psi_{i'}, G') \right) = \sum_{i' \in J_0} c_{i'} Z_{A,D}(\psi_{i'}, G').$$

□

The proof of the Pinning Lemma now follows easily from Lemmas 6.1 and 6.3.

Proof of the Pinning Lemma 4.3. Fix the twin resolvent $A' = [A]$ of A and let $D' = D^{[A]}$. It suffices to show

$$\text{EVAL}^{\mathsf{pin}}(A', D') \equiv \text{EVAL}(A', D'). \tag{6.9}$$

To see this note that by the Twin Reduction Lemma 6.1 we then have the chain of reductions

$$\text{EVAL}^{\mathsf{pin}}(A, D) \equiv \text{EVAL}^{\mathsf{pin}}(A', D') \equiv \text{EVAL}(A', D') \equiv \text{EVAL}(A, D).$$

The proof of equation (6.9) follows from Lemma 6.3. □

7. The General Conditioning Lemma

7.1. Dealing with $\{0, 1\}$ Matrices. As a technical prerequisite, we need a part of the result of [**DG00**] (Theorem 1.1 in there). We call a block *non-trivial* if it contains a non-zero entry.

Lemma 7.1 (#H-Coloring Lemma). *Let A be a symmetric connected and bipartite $\{0, 1\}$-matrix with underlying non-trivial block B. If B contains a zero entry then the problem* $\text{EVAL}^{\mathsf{pin}}(A)$ *is* #P*-hard.*

Proof. We will start with the following claim which captures the main reduction. Call a matrix A *powerful* if it is a symmetric connected and bipartite $\{0, 1\}$-matrix whose underlying block B contains a zero entry.

CLAIM 3. Let A be a powerful matrix with underlying $m \times n$ block B. If either $n > 2$ or $m > 2$ then there is a powerful matrix A' with underlying $m' \times n'$ block B' such that $2 \le m' \le m$ and $2 \le n' \le n$, at least one of these inequalities is strict and

$$\mathrm{EVAL}^{\mathsf{pin}}(A') \le \mathrm{EVAL}^{\mathsf{pin}}(A).$$

Before we prove the claim, let us see how it helps in proving the lemma. Let A be as in the statement of the lemma. Iterating Claim 3 for a finite number of steps we arrive at $\mathrm{EVAL}^{\mathsf{pin}}(A') \le \mathrm{EVAL}^{\mathsf{pin}}(A)$ such that the block B' underlying A' is a 2×2 matrix of the form (up to permutation of rows/columns)

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

For every empty pinning ϕ and bipartite graph G with c connected components $Z_{A'}(\phi, G)$ equals 2^c times the number of independent sets of G. This straightforwardly gives rise to a reduction from the problem of counting independent sets in bipartite graphs which is well-known to be #P-hard (see [**PB83**]).

Proof of Claim 3. As B is a block with a zero entry, there are indices i, j, k, l such that $B_{ik} = B_{il} = B_{jk} = 1$ and $B_{jl} = 0$. Fix these indices and let $I = \{\nu \mid B_{i\nu} = 1\}$ and $J = \{\mu \mid B_{\mu k} = 1\}$ be the sets of indices of 1-entries in row i and column k. Let A^* be the connected bipartite matrix with underlying block B_{IJ}. We will show that

$$\mathrm{EVAL}^{\mathsf{pin}}(A^*) \le \mathrm{EVAL}^{\mathsf{pin}}(A). \tag{7.1}$$

To see this, let G, ϕ be an instance of $\mathrm{EVAL}^{\mathsf{pin}}(A^*)$. We shall consider only the case here that $G = (U \cup W, E)$ is connected and bipartite with bipartition U, W and that ϕ pins a vertex $a \in U$ to a row of B_{IJ}. All other cases follow similarly.

Define G' as the graph obtained from G by adding two new vertices u' and w' and connect every vertex of U by an edge to w' and every vertex of W by a edge to u'. Let ϕ' be the pinning obtained from ϕ by adding $u' \mapsto i$ and $w' \mapsto k$. We have $Z_{A^*}(\phi, G) = Z_A(\phi', G')$ which yields a reduction witnessing equation (7.1).

Before we proceed, we need another

CLAIM 4. Let A^+ be a powerful matrix with underlying $m^+ \times n^+$ block B^+. There is a twin-free powerful matrix A'' with underlying $m'' \times n''$ block B'' such that $2 \le m'' \le m^+$ and $2 \le n'' \le n^+$ and

$$\mathrm{EVAL}^{\mathsf{pin}}(A'') \le \mathrm{EVAL}^{\mathsf{pin}}(A^+).$$

PROOF. This is a straightforward combination of the Twin Reduction Lemma 6.1 and Lemma 5.6. ⊣

Combining equation (7.1) and Claim 4 we arrive at $\mathrm{EVAL}^{\mathsf{pin}}(A'') \le \mathrm{EVAL}^{\mathsf{pin}}(A)$ for a powerful twin-free matrix A''. The block B'' underlying A'' has some dimension $m'' \times n''$ such that $2 \le m'' \le m$ and $2 \le n'' \le n$. Further, up to permutation of rows/columns, this block has the following form

$$B'' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 0 & \dots & * \\ \vdots & & \ddots & \\ 1 & * & \dots & \end{pmatrix}$$

To prove Claim 3 it suffices to devise a reduction witnessing

$$\text{EVAL}^{\mathsf{pin}}(A') \leq \text{EVAL}^{\mathsf{pin}}(A'') \tag{7.2}$$

for a powerful matrix A' such that the block B' underlying A' has either fewer rows or columns than B''.

To devise such a reduction, assume that B'' has at least three rows (the case that B'' has at least three columns is symmetric). As B'' is twin-free, there must be entries $B''_{2a} \neq B''_{3a}$. Assume w.l.o.g. that $B''_{2a} = 0$ and $B''_{3a} = 1$. As B'' is twin-free we further have that $B''_{3b} = 0$ for some b. Let $K = \{\kappa \mid B''_{3\kappa} = 1\}$ be the indices of non-zero entries of $B''_{3,*}$. Define $B' = B''_{*,K}$ and let A' be the connected bipartite matrix with underlying block B'. It remains to devise the desired reduction.

Let G, ϕ be an instance of $\text{EVAL}^{\mathsf{pin}}(A')$. As before, we shall consider only the case that $G = (U \cup W, E)$ is connected and bipartite with bipartition U, W and that ϕ pins a vertex $a \in U$ to a row of B'. All other cases follow similarly.

Define G' as the graph obtained from G by adding one new vertex u' and connect every vertex of W by a edge to u'. Let ϕ' be the pinning obtained from ϕ by adding $u' \mapsto 3$. This yields a reduction witnessing equation (7.2) and hence finishes the proof of Claim 3. □

7.2. From General Matrices to Positive Matrices. In the first step of the proof of the General Conditioning Lemma 4.4 we will see how to restrict attention to matrices with positive entries.

LEMMA 7.2 (The Lemma of the Positive Witness). *Let $A \in \mathbb{S}^{m \times m}$ be a symmetric non-negative matrix containing a block of row rank at least* 2. *Then there is an $\mathbb{S}$-matrix A' satisfying condition* **(A)**, *that is A' is positive symmetric of row rank at least* 2, *such that*

$$\text{EVAL}^{\mathit{pin}}(A') \leq \text{EVAL}^{\mathit{pin}}(A).$$

The proof of this lemma relies on the elimination of zero entries. We do this in two steps, first (in the next Lemma) we pin to components. Afterwards, we will see how to eliminate zero entries within a component.

LEMMA 7.3 (Component Pinning). *Let $A \in \mathbb{S}^{m \times m}$ be a symmetric matrix. Then for each component C of A we have*

$$\text{EVAL}^{\mathit{pin}}(C) \leq \text{EVAL}^{\mathit{pin}}(A).$$

PROOF. Let G, ϕ be the input to $\text{EVAL}^{\mathsf{pin}}(C)$ for some component C of some order $m' \times m'$. Let $G_1, \dots, G_\ell$ be the components of G and $\phi_1, \dots, \phi_\ell$ the corresponding restrictions of ϕ. Since

$$Z_C(\phi, G) = \prod_{i=1}^{\ell} Z_C(\phi_i, G_i)$$

we may assume w.l.o.g that G is connected which we shall do in the following. Let v be a vertex in G which is not pinned by ϕ. Define, for each $i \in [m]$, ϕ_i as the pinning obtained from ϕ by adding $v \mapsto i$. We have

$$Z_C(\phi, G) = \sum_{i=1}^{m} Z_C(\phi_i, G).$$

Thus it suffices to compute $Z_C(\phi_i, G)$ which can be obtained straightforwardly as these values equal $Z_A(\phi_i, G)$. □

LEMMA 7.4 (Zero-Free Block Lemma). *Let $A \in \mathbb{S}^{m\times m}$ be a symmetric matrix. Then either* $\mathrm{EVAL}^{\mathsf{pin}}(A)$ *is* #P*-hard or no block of A contains zero entries.*

PROOF. Let A'' be the matrix obtained from A by replacing each non-zero entry by 1. Let C be a component of A'' whose underlying block B contains a zero entry. For every given graph G and pinning ϕ each configuration σ which contributes a non-zero weight to $Z_A(\phi, G)$ contributes a weight 1 to $Z_{A''}(\phi, G)$. Thus by $\mathrm{EVAL}^{\mathsf{pin}}(A) \equiv \mathrm{COUNT}^{\mathsf{pin}}(A)$ (cf. Lemma 5.5) and component pinning (cf. Lemma 7.3) it is easy to see that

$$\mathrm{EVAL}^{\mathsf{pin}}(C) \leq \mathrm{EVAL}^{\mathsf{pin}}(A'') \leq \mathrm{EVAL}^{\mathsf{pin}}(A).$$

If C is bipartite, the result follows from the #H-Coloring Lemma 7.1.

If C is not bipartite, we need a bit of additional work. Note that in this case $C = B$. Let A' be the connected bipartite $2r \times 2r$ matrix with underlying block B. If we can show that

$$\mathrm{EVAL}^{\mathsf{pin}}(A') \leq \mathrm{EVAL}^{\mathsf{pin}}(B)$$

then the result follows from the #H-Coloring Lemma 7.1.

Let G, ϕ be an instance of $\mathrm{EVAL}^{\mathsf{pin}}(A')$. For simplicity, assume that $G = (U \cup W, E)$ is connected and bipartite. If ϕ is empty, then we have $Z_{A'}(G) = 2 \cdot Z_B(G)$. Otherwise, $Z_{A'}(\phi, G) = 0$ unless ϕ maps all elements of $\mathrm{dom}(\phi) \cap U$ to $[r]$ and all entries of $\mathrm{dom}(\phi) \cap W$ to $[r+1, 2r]$ or vice versa. Since both cases are symmetric, we consider only the first one. Let $\phi' : \mathrm{dom}(\phi) \to [2r]$ be the pinning which agrees with ϕ on $\mathrm{dom}(\phi) \cap U$ and for all $w \in \mathrm{dom}(\phi) \cap W$ satisfies $\phi'(w) = \phi(w) - r$. By inspection we have

$$Z_{A'}(\phi, G) = Z_B(\phi', G').$$

□

PROOF OF LEMMA 7.2. Let $A \in \mathbb{S}^{m\times m}$ contain a block B of rank at least 2. By 2-stretching (cf. Lemma 5.1) we have $\mathrm{EVAL}^{\mathsf{pin}}(A^2) \leq \mathrm{EVAL}^{\mathsf{pin}}(A)$ and Lemma 5.2 guarantees that A^2 contains a component BB^T which has rank at least 2. By component pinning (Lemma 7.3) we have $\mathrm{EVAL}^{\mathsf{pin}}(BB^T) \leq \mathrm{EVAL}^{\mathsf{pin}}(A^2)$. If BB^T contains no zero entries, we let $A' = BB^T$.

Otherwise let $A' \in \mathbb{N}^{2\times 2}$ be a matrix satisfying the conditions of the lemma. As BB^T contains a zero entry, Lemma 7.4 implies that $\mathrm{EVAL}^{\mathsf{pin}}(BB^T)$ is #P-hard and thus $\mathrm{EVAL}^{\mathsf{pin}}(A') \leq \mathrm{EVAL}^{\mathsf{pin}}(BB^T)$ finishing the proof. □

7.3. From Positive Matrices to X-matrices. Let A be a matrix which satisfies condition **(A)**. We will now see, how to obtain a matrix A' which additionally satisfies **(B)**. That is, we will prove the following lemma.

LEMMA 7.5 (X-Lemma). *Let $A \in \mathbb{S}^{m\times m}$ be a matrix satisfying condition* **(A)**. *Then there is an $\mathbb{S}$-matrix A' satisfying conditions* **(A)** *and* **(B)** *such that*

$$\mathrm{EVAL}^{pin}(A') \leq \mathrm{EVAL}^{pin}(A).$$

To prepare the proof, we present some smaller lemmas which will also be useful in later sections.

LEMMA 7.6 (Prime Elimination Lemma). *Let $A \in \mathbb{Z}^{m\times m}$ ($A \in \mathbb{Z}[X]^{m\times m}$) and p be a prime number (an irreducible polynomial). Let A' be the matrix obtained from A by replacing all entries divisible by p with 0. Then*

$$\mathrm{EVAL}^{pin}(A') \leq \mathrm{EVAL}^{pin}(A).$$

PROOF. By Lemma 5.5 it suffices to give a reduction witnessing $\mathrm{EVAL}^{\mathsf{pin}}(A') \leq \mathrm{COUNT}^{\mathsf{pin}}(A)$, which we will construct in the following. Let $G = (V, E)$ be a given graph ϕ a pinning. As $\mathcal{W}_{A'}(G) = \mathcal{W}_A(G) \setminus \{w \mid w \text{ divisible by } p\}$, we have $\mathcal{W}_{A'}(G) \subseteq \mathcal{W}_A(G)$. Moreover $N_{A'}(G, \phi, w) = N_A(G, \phi, w)$ for all $w \in \mathcal{W}_{A'}(G)$ which implies

$$Z_{A'}(\phi, G) = \sum_{w \in \mathcal{W}_{A'}(G)} w \cdot N_A(G, \phi, w).$$

The values $N_A(G, \phi, w)$ can be obtained directly using the $\mathrm{COUNT}^{\mathsf{pin}}(A)$ oracle. □

Let p be a prime number (an irreducible polynomial, respectively) and $a \in \mathbb{Z}$ ($a \in \mathbb{Z}[X]$, resp.). Define

$$a|_p = \begin{cases} p^{\max\{k \geq 0 | p^k \text{ divides } a\}} & , \text{ if } a \neq 0 \\ 0 & , \text{ otherwise.} \end{cases}$$

For a matrix A the matrix $A|_p$ is then defined by replacing each entry A_{ij} with $A_{ij}|_p$.

LEMMA 7.7 (Prime Filter Lemma). *Let $A \in \mathbb{Z}^{m\times m}$ ($A \in \mathbb{Z}[X]^{m\times m}$) and p be a prime number (an irreducible polynomial). Then*

$$\mathrm{EVAL}^{pin}(A|_p) \leq \mathrm{EVAL}^{pin}(A).$$

PROOF. By Lemma 5.5 it suffices to give a reduction witnessing $\mathrm{EVAL}^{\mathsf{pin}}(A|_p) \leq \mathrm{COUNT}^{\mathsf{pin}}(A)$, which we will construct in the following. For a given graph $G = (V, E)$ and a pinning ϕ we have

$$Z_{A|_p}(\phi, G) = \sum_{w \in \mathcal{W}_A(G)} w|_p \cdot N_A(G, \phi, w).$$

The values $N_A(G, \phi, w)$ can be obtained directly using a $\mathrm{COUNT}^{\mathsf{pin}}(A)$ oracle. □

LEMMA 7.8 (Renaming Lemma). *Let $p \in \mathbb{Z}[X] \setminus \{-1, 0, 1\}$ and $A \in \mathbb{Z}[X]^{m\times m}$ a p-matrix. Let $q \in \mathbb{Z}[X]$ and define $A' \in \mathbb{Z}[X]^{m\times m}$ by*

$$A'_{ij} = \begin{cases} q^l & , \textit{there is an } l \geq 0 \textit{ s.t. } A_{ij} = p^l \\ 0 & , \textit{otherwise} \end{cases}$$

That is, A' is the matrix obtained from A by substituting powers of p with the corresponding powers of q. Then

$$\mathrm{EVAL}^{pin}(A') \leq \mathrm{EVAL}^{pin}(A).$$

PROOF. Consider A' as a function $A'(Y)$ in some indeterminate Y. We have $A = A'(p)$. Let $\ell_{\max}$ be the maximum power of Y occurring in $A'(Y)$. For every graph G and pinning ϕ, the value $f(Y) := Z_{A'(Y)}(\phi, G)$ is a polynomial in Y of maximum degree $|E| \cdot \ell_{\max}$. By t-thickening (cf. Lemma 5.1), using an $\mathrm{EVAL}^{\mathsf{pin}}(A)$

oracle, we can compute the values $f(p^t)$ for $t = 1, \ldots, |E| \cdot \ell_{\max}$. Thus, by interpolation (cf. Lemma 5.4) we can compute the value $f(q)$ for a q as given in the statement of the lemma. By $f(q) = Z_{A'(q)}(\phi, G)$ this proves the claimed reducibility. □

LEMMA 7.9 (Prime Rank Lemma). *Let $A \in \mathbb{Z}^{m\times m}$ ($A \in \mathbb{Z}[X]^{m\times m}$) contain a block of row rank at least 2. There is a prime number (an irreducible polynomial) p such that $A|_p$ contains a block of row rank at least 2.*

PROOF. Let $A_{i,*}$ and $A_{i',*}$ be linearly independent rows from a block of A. Assume, for contradiction, that for all primes (irreducible polynomials, resp.) p every block in $A|_p$ has rank at most 1. We have, for all $j \in [m]$,

$$A_{ij} = \prod_p A_{ij}|_p \quad \text{and} \quad A_{i'j} = \prod_p A_{i'j}|_p$$

where the products are over all primes (irreducible polynomials, resp.) dividing an entry of A. By assumption, there are $\alpha_p, \beta_p \in \mathbb{Z}$ (in $\mathbb{Z}[X]$, resp.) such that $\alpha_p \cdot A_{i,*}|_p = \beta_p \cdot A_{i',*}|_p$ for all primes (irreducible polynomials). Therefore,

$$A_{ij} \prod_p \alpha_p = \prod_p \alpha_p A_{ij}|_p = \prod_p \beta_p A_{i'j}|_p = A_{i'j} \cdot \prod_p \beta_p \text{ for all } j \in [m].$$

And hence, $A_{i',*}$ and $A_{i,*}$ are linearly dependent — a contradiction. □

Dealing with algebraic numbers. The following lemma tells us that the structure of the numbers involved in computations of partition functions on matrices with algebraic entries is already captured by matrices with natural numbers as entries.

LEMMA 7.10 (The Arithmetical Structure Lemma). *Let $A \in \mathbb{R}_{\mathbb{A}}^{m\times m}$ be symmetric and non-negative. There is a matrix A' whose entries are natural numbers such that*

$$\text{EVAL}^{\mathsf{pin}}(A') \equiv \text{EVAL}^{\mathsf{pin}}(A).$$

If further A contains a block of rank at least 2 then this is also true for A'.

We have to introduce some terminology. Let $B = \{b_1, \ldots, b_n\} \subseteq \mathbb{R}_{\mathbb{A}}$ be a set of positive numbers. The set B is called *multiplicatively independent*, if for all $\lambda_1, \ldots, \lambda_n \in \mathbb{Z}$ the following holds: if $b_1^{\lambda_1} \cdots b_n^{\lambda_n}$ is a root of unity then $\lambda_1 = \ldots = \lambda_n = 0$. In all other cases we say that B is *multiplicatively dependent.* We say that a set S of positive numbers is *effectively representable* in terms of B, if for given $x \in S$ we can compute $\lambda_1, \ldots, \lambda_n \in \mathbb{Z}$ such that $x \cdot b_1^{\lambda_1} \cdots b_n^{\lambda_n} = 1$. A set B is an *effective representation system* for a set S, if S is effectively representable in terms of B and B is multiplicatively independent.

We need a result from [**Ric01**] which we rephrase a bit for our purposes.

LEMMA 7.11 (Theorem 2 in [**Ric01**]). *Let $a_1, \ldots, a_n \in \mathbb{Q}(\theta)$ be positive real numbers given in standard representation, each of description length at most s. There is a matrix $A \in \mathbb{Z}^{n\times n}$ such that, for vectors $\lambda \in \mathbb{Z}^n$ we have*

$$\prod_{i=1}^{n} a_i^{\lambda_i} = 1 \textit{ if, and only if, } A \cdot \lambda = 0. \tag{7.3}$$

The description length of A is bounded by a computable function in n and s.

This result straightforwardly extends to an algorithm solving the multiplicative independence problem for algebraic numbers.

Corollary 7.12. *Let $a_1, \dots, a_n \in \mathbb{Q}(\theta)$ be positive reals given in standard representation. There is an algorithm which decides if there is a non-zero vector $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{Z}^n$ such that*

$$\prod_{i=1}^{n} a_i^{\lambda_i} = 1. \tag{7.4}$$

Furthermore, if it exists, the algorithm computes such a vector λ.

Lemma 7.13. *Let $S \subseteq \mathbb{R}_{\mathbb{A}}$ be a set of positive numbers. There is an effective representation system $B \subseteq \mathbb{R}_{\mathbb{A}}$ of positive numbers for S which can be computed effectively from S.*

Proof. We shall start with the following

Claim 5. If S is multiplicatively dependent then there is a set $B' \subseteq \mathbb{R}_{\mathbb{A}}$ of non-negative numbers such that $|B'| < |S|$ and S is effectively representable by B'.

Proof. Let $S = \{b_1, \dots, b_n\}$ then Corollary 7.12 implies that we can compute a non-zero vector $\lambda \in \mathbb{Z}^n$ such that $b_1^{\lambda_1} \cdots b_n^{\lambda_n} = 1$. We can easily make sure that at least one of the λ_i is larger than zero. Assume therefore w.l.o.g. that $\lambda_1 > 0$. Fix a set $B' = \{b'_2, \dots, b'_n\}$ where each b'_i is the positive real λ_1-th root of b_i, that is $(b'_i)^{\lambda_1} = b_i$. Then

$$b_1^{\lambda_1} \cdot \left(\prod_{i=2}^{n} (b'_i)^{\lambda_i} \right)^{\lambda_1} = 1 \quad \text{and hence} \quad b_1 \cdot \prod_{i=2}^{n} (b'_i)^{\lambda_i} = 1.$$

All operations are computable and effective representation of S by B' follows. ⊣

Apply Claim 5 recursively on S. Since the empty set is multiplicatively independent, after at most finitely many steps, we find an effective representation system B for S. □

Proof of Lemma 7.10. Let S be the set of non-zero entries of A. By means of Lemma 7.13 we can compute an effective representation system B for S. However, with respect to our model of computation we need to be a bit careful, here: assume that $S \subseteq \mathbb{Q}(\theta)$ for some primitive element θ. The application of Lemma 7.13 does not allow us to stipulate that $B \subseteq \mathbb{Q}(\theta)$. But in another step of pre-computation, we can compute another primitive element θ' for the elements of B such that $B \subseteq \mathbb{Q}(\theta')$ (c.f [**Coh93**]). Then we may consider all computations as taking place in $\mathbb{Q}(\theta')$.

Assume that $B = \{b_1, \dots, b_n\}$, then every non-zero entry of A has a unique computable representation

$$A_{ij} = \prod_{\nu=1}^{n} b_\nu^{\lambda_{ij\nu}}.$$

Let $p_1, \dots, p_\beta$ be $\beta = |B|$ distinct prime numbers and define A' as the matrix obtained from A by replacing in each non-zero entry A_{ij} the powers of $b \in B$ by the corresponding powers of primes, that is,

$$A'_{ij} = \prod_{\nu=1}^{n} p_\nu^{\lambda_{ij\nu}}.$$

Recall the definition of $\mathcal{W}_A(G)$ in equation (5.1). For each $w \in \mathcal{W}_A(G)$ we can, in polynomial time compute a representation $w = \prod_{i,j} A_{ij}^{m_{ij}}$ as powers of elements in

S. The effective representation of S in terms of B extends to $\mathcal{W}_A(G)$ being effectively representable by B. Moreover, as S depends only on A, the representation of each $w \in \mathcal{W}_A(G)$ is even polynomial time computable. We have

$$Z_A(\phi, G) = \sum_{w \in \mathcal{W}_A(G)} w \cdot N_A(G, \phi, w)$$

In particular, for each $w \in \mathcal{W}_A(G)$, we can compute unique $\lambda_{w,1}, \dots, \lambda_{w,n} \in \mathbb{Z}$ such that $w \cdot b_1^{\lambda_{w,1}} \cdots b_n^{\lambda_{w,n}} = 1$. Define functions f and g such that for every $w \in \mathcal{W}_A(G)$ we have

$$f(w) = \prod_{\nu=1}^{n} p_\nu^{\lambda_{w,\nu}} \quad \text{and} \quad g(w) = \prod_{\nu=1}^{n} b_\nu^{\lambda_{w,\nu}}.$$

Thus we obtain

$$Z_{A'}(\phi, G) = \sum_{w \in \mathcal{W}_A(G)} w \cdot \frac{f(w)}{g(w)} \cdot N_A(G, \phi, w).$$

This yields a reduction for $\mathrm{EVAL}^{\mathsf{pin}}(A') \leq \mathrm{EVAL}^{\mathsf{pin}}(A)$. The other direction follows by

$$Z_A(\phi, G) = \sum_{w' \in \mathcal{W}_{A'}(G)} w' \cdot \frac{g(w)}{f(w)} \cdot N_{A'}(G, \phi, w').$$

This also proves the reducibilities $\mathrm{EVAL}(A') \equiv \mathrm{EVAL}(A)$ since the input pinnings remain unaffected. To finish the proof it remains to consider the case that A contains a block of rank at least 2. We have to show that A' has this property as well. Let us now argue that A' contains a block of rank at least 2. Let $A_{i,*}$ and $A_{i',*}$ be linearly independent rows from a block of A. Assume, for contradiction, that $A'_{i',*} = \alpha \cdot A'_{i,*}$ for some α. Let J be the set of indices j such that $A'_{ij} \neq 0$. For each $j \in J$ we have

$$\alpha = A'_{i'j} \cdot (A'_{ij})^{-1} = \prod_{\nu=1}^{n} p_\nu^{\lambda_{i'j\nu} - \lambda_{ij\nu}}.$$

Hence, for $\beta = b_1^{\lambda_{i'j1} - \lambda_{ij1}} \cdots b_n^{\lambda_{i'jn} - \lambda_{ijn}}$ we obtain $A_{i',*} = \beta \cdot A_{i,*}$ — a contradiction. □

7.3.1. *Proof of the X-Lemma 7.5.* Let $A \in \mathbb{S}^{m \times m}$ be a matrix satisfying condition **(A)**. Recall that $\mathbb{S}$ is one of $\mathbb{R}_{\mathbb{A}}, \mathbb{Q}, \mathbb{Z}, \mathbb{Z}[X]$ or $\mathbb{Q}[X]$.

If $\mathbb{S} = \mathbb{R}_{\mathbb{A}}$ then the entries of A are all positive real values. Thus Lemma 7.10 implies that there is a positive matrix $A' \in \mathbb{Z}^{m \times m}$ of rank at least 2 such that $\mathrm{EVAL}^{\mathsf{pin}}(A') \equiv \mathrm{EVAL}^{\mathsf{pin}}(A)$.

If A is a matrix of entries in $\mathbb{Q}$ ($\mathbb{Q}[X]$, respectively) then let λ be the lowest common denominator of (coefficients of) entries in A. For a given graph $G = (V, E)$ and pinning ϕ we have

$$Z_{\lambda A}(\phi, G) = \lambda^{|E|} Z_A(\phi, G).$$

The matrix $\lambda \cdot A$ is a matrix with entries in $\mathbb{Z}$ ($\mathbb{Z}[X]$, respectively).

It remains to prove the Lemma for the case that $\mathbb{S}$ is either $\mathbb{Z}$ or $\mathbb{Z}[X]$. By the Prime Rank Lemma 7.9 there is a prime (irreducible polynomial) p such that $A|_p$

contains a block of rank at least 2. Fix such a p and define $A' = A|_p$. By the Prime Filter Lemma 7.7 we have

$$\mathrm{EVAL}^{\mathsf{pin}}(A') \leq \mathrm{EVAL}^{\mathsf{pin}}(A).$$

Furthermore, by the Renaming Lemma 7.8 we may assume that A' is an X-matrix. This completes the proof.

7.4. From X-matrices to the General Conditioning Lemma.

LEMMA 7.14. *Let $A \in \mathbb{Z}[X]^{m\times m}$ be symmetric and positive such that not all 1-entries of A are contained in 1-cells. Then $\mathrm{EVAL}^{pin}(A)$ is #P-hard.*

PROOF. If not all 1-entries are contained in 1-cells, then there are i, j, k, l such that $A_{ik} = A_{il} = A_{jk} = 1$ and $A_{jl} \neq 1$. Let A' be obtained from A by replacing each entry not equal to 1 by 0. We have $\mathrm{EVAL}^{\mathsf{pin}}(A') \leq \mathrm{EVAL}^{\mathsf{pin}}(A)$ by the Prime Elimination Lemma 7.6.

By construction $A_{ik} = A_{il} = A_{jk} = 1$ and $A_{jl} = 0$. Thus A' contains a block with zero entries and $\mathrm{EVAL}^{\mathsf{pin}}(A')$ is #P-hard by Lemma 7.4. □

PROOF OF THE GENERAL CONDITIONING LEMMA 4.4. By Lemma 7.2, there is a matrix C' which satisfies **(A)** such that $\mathrm{EVAL}^{\mathsf{pin}}(C') \leq \mathrm{EVAL}^{\mathsf{pin}}(A)$. Then by the X-Lemma 7.5, there is a matrix C which satisfies **(A)** and **(B)** such that $\mathrm{EVAL}^{\mathsf{pin}}(C) \leq \mathrm{EVAL}^{\mathsf{pin}}(C')$. Lemma 7.14 implies that $\mathrm{EVAL}^{\mathsf{pin}}(C)$ is #P-hard if not all 1-entries are contained in 1-cells. In this case we let A' be some 2×2 matrix satisfying conditions **(A)**–**(C)**.

Assume therefore that all 1-entries of C are contained in 1-cells. If C contains exactly one 1-cell then the proof follows by the symmetry of C'. We only have to make sure that we can permute the entries of C such that condition **(C)** is satisfied. This is guaranteed by the Permutability Principle 5.3.

Assume therefore that C contains more than one 1-cell. Define an X-matrix $C^* = C^2|_X$.

CLAIM 6. For every 1-cell C_{KL} of C the principal submatrix C^*_{KK} is a 1-cell of C^*.

PROOF. Note that $C^*_{ij} = 1$ only if there is an ℓ such that $C_{i\ell} = C_{j\ell} = 1$. This proves the claim. ⊣

Now C^* has all 1-entries in principal 1-cells and is of rank at least 2 by the fact that there are at least two 1-cells in C. A reduction witnessing $\mathrm{EVAL}^{\mathsf{pin}}(C^*) \leq \mathrm{EVAL}^{\mathsf{pin}}(C)$ is given by applying 2-stretching (cf. Lemma 5.1) and the Prime Filter Lemma 7.7 in this order. □

8. The Two 1-Cell Lemma

The **(T1C)** – Conditions for Matrices with two 1-cells. We define two additional conditions.

(T1C – A) A has at least two 1-cells.

(T1C – B) All diagonal entries of A are 1.

A 1-*row* (1-*column*) in a matrix A is a row (column) which contains a at least one 1 entry. We call all other rows (columns) *non*-1-*rows* (*non*-1-*columns*).

LEMMA 8.1 (1-Row-Column Lemma). *Let $A \in \mathbb{Z}[X]^{m\times m}$ be a positive and symmetric X-matrix. Let A' be obtained from A by removing all non-1-rows and non-1-columns. Then*

$$\mathrm{EVAL}^{pin}(A') \le \mathrm{EVAL}^{pin}(A).$$

PROOF. For every $i \in [m]$ let c_i denote the number of 1-entries in row $A_{i,*}$. And let A'' be the matrix defined by $A''_{ij} = c_i c_j A'_{ij}$ for all $i,j \in [m]$.

We start with a reduction witnessing $\mathrm{EVAL}^{\mathsf{pin}}(A'') \le \mathrm{EVAL}^{\mathsf{pin}}(A)$. Let $G = (V,E)$, ϕ be an instance of $\mathrm{EVAL}^{\mathsf{pin}}(A'')$. Let Δ be the maximum degree of X in A, let $k = \Delta\cdot|E|+1$ and define a graph $G' = (V',E')$ by

$$\begin{aligned} V' &= \{v, v^1 \mid v \in V\} \\ E' &= E \cup \{e_v^1,\dots,e_v^k \mid v \in V,\ e_v^i = vv^1,\ \forall i \in [k]\}. \end{aligned}$$

We have

$$\begin{aligned} Z_A(\phi,G') &= \sum_{\phi\subseteq\sigma:V'\to[m]} \prod_{uv\in E'} A_{\sigma(u),\sigma(v)} \\ &= \sum_{\phi\subseteq\sigma:V'\to[m]} \prod_{uv\in E} A_{\sigma(u),\sigma(v)} \prod_{v\in V} \left(A_{\sigma(v),\sigma(v^1)}\right)^k \end{aligned}$$

That is, for every σ the degree (as a polynomial in X) of the weight expression

$$\prod_{uv\in E} A_{\sigma(u),\sigma(v)} \prod_{v\in V} \left(A_{\sigma(v),\sigma(v^1)}\right)^k$$

is smaller than k if, and only if, $A_{\sigma(v)\sigma(v^1)} = 1$ holds for every $v \in V$. That is, if a configuration σ maps a vertex of V to a non-1-row, then the degree of the weight expression of σ will always be at least k. On the other hand, for each configuration σ mapping G to A there are exactly $c_1^{|\sigma^{-1}(1)|}\cdot\ldots c_m^{|\sigma^{-1}(m)|}$ many configurations $\sigma' : V' \to [m]$ of G' extending σ in such a way that their weight is of degree smaller than k.

By the polynomial time equivalence of $\mathrm{COUNT}^{\mathsf{pin}}(A)$ and $\mathrm{EVAL}^{\mathsf{pin}}(A)$ it suffices to reduce $\mathrm{EVAL}^{\mathsf{pin}}(A'') \le \mathrm{COUNT}^{\mathsf{pin}}(A)$. We can do this by computing the values $w\cdot N_A(G',\phi,w)$ for all weights w of degree smaller than k. Then, with $A' = A''|_X$, the remaining step $\mathrm{EVAL}^{\mathsf{pin}}(A') \le \mathrm{EVAL}^{\mathsf{pin}}(A'')$ is given by the Prime Filter Lemma 7.7. □

LEMMA 8.2. *Let $A \in \mathbb{Z}[X]^{m\times m}$ be a matrix which satisfies* **(A)** *and contains at least two* 1*-cells. Then there is a matrix A' satisfying conditions* **(A)**–**(C)** *and both* **(T1C)** *conditions such that*

$$\mathrm{EVAL}^{pin}(A') \le \mathrm{EVAL}^{pin}(A).$$

PROOF. Let $i,j,i',j' \in [m]$ be witnesses for the existence of two 1-cells in A in the sense that $A_{ij} = A_{i'j'} = 1$ but $A_{i'j} \ne 1$ and let p be an irreducible polynomial which divides $A_{i'j}$. Let B be the matrix obtained from $A|_p$ by replacing all powers of p with the corresponding powers of X. Then by Prime Filter Lemma 7.7 and Renaming Lemma 7.8, we have $\mathrm{EVAL}^{\mathsf{pin}}(B) \le \mathrm{EVAL}^{\mathsf{pin}}(A)$. Note that B satisfies condition **(B)** and it satisfies **(A)** as A does. We may assume that all 1-entries of B are contained in 1-cells, because otherwise $\mathrm{EVAL}^{\mathsf{pin}}(B)$ would be #P-hard by Lemma 7.14.

As it could possibly be the case that not all 1-cells of B are on the diagonal, we form $B' = B^2|_X$.

CLAIM 7. For every 1-cell B_{KL} of B the principal submatrix B'_{KK} is a 1-cell of B'.

PROOF. Note that $B'_{ij} = 1$ only if there is an ℓ such that $B_{i\ell} = B_{j\ell} = 1$. This proves the claim. ⊣

By this claim, since B contains at least two 1-cells, the matrix B' does so as well and therefore it satisfies **(A)**. Condition **(B)** is satisfied by definition. We have $\mathrm{EVAL}^{\mathsf{pin}}(B') \leq \mathrm{EVAL}^{\mathsf{pin}}(B)$ by application of 2-stretching (cf. Lemma 5.1) and the Prime Filter Lemma 7.7 in this order. Applying the 1-Row-Column Lemma 8.1 on B' then yields $\mathrm{EVAL}^{\mathsf{pin}}(A') \leq \mathrm{EVAL}^{\mathsf{pin}}(B')$ for a matrix A' which (up to permuting rows and columns) has the desired properties. □

The *cells* of a matrix A satisfying conditions **(A)**–**(C)** are the submatrices $A_{I_iI_j}$ for I_i, I_j as defined in condition **(C)**. We call such a matrix A a *cell matrix*, if in each of its cells $A_{I_iI_j}$ all entries are equal.

LEMMA 8.3. *Let $A \in \mathbb{Z}[X]^{m\times m}$ satisfy conditions* **(A)** – **(C)** *and both* **(T1C)** *conditions. Then there is a cell matrix $C \in \mathbb{Z}[X]^{m\times m}$ which also satisfies* **(A)** – **(C)** *and both* **(T1C)** *conditions such that*

$$\mathrm{EVAL}^{\mathsf{pin}}(C) \leq \mathrm{EVAL}^{\mathsf{pin}}(A).$$

PROOF. We define a sequence of matrices with $A_0 = A$ and for all $\nu \in \mathbb{N}$ we let $A_{\nu+1} = A_\nu^2|_X$. As A satisfies conditions **(A)** – **(C)** and both **(T1C)** conditions, this is true for all matrices in the sequence. Further we have $\mathrm{EVAL}^{\mathsf{pin}}(A_{\nu+1}) \leq \mathrm{EVAL}^{\mathsf{pin}}(A_\nu)$ for all ν, by applying 2-stretching (cf. Lemma 5.1) and the Prime Filter Lemma 7.7 in this order.

By definition, $\deg((A_{\nu+1})_{ij}) = \min\{\deg((A_\nu)_{ik}) + \deg((A_\nu)_{jk}) \mid k \in [m]\}$ for all $i, j \in [m]$. As condition **(T1C − B)** implies $\deg(A_{jj}) = 0$ for all $j \in [m]$, we obtain

$$\deg((A_{\nu+1})_{ij}) \leq \deg((A_\nu)_{ij}) \text{ for all } \nu.$$

That is, the degrees of the entries of A_ν are non-increasing with ν and thus there is a μ such that $A_{\mu+1} = A_\mu$. To finish the proof we will show that $C = A_\mu$ is a cell matrix. Let C_{IJ} be a cell of C which is not a 1-cell. Let $i \in I$ and $j \in J$ be such that $\deg(C_{ij})$ is minimal. By definition we have $C = C^2|_X$ and therefore, for all $j' \in J$,

$$\deg(C_{ij'}) = \deg((C^2|_X)_{ij'}) = \min\{\deg(C_{ik}) + \deg(C_{jk}) \mid k \in [m]\} \leq \deg(C_{ij}).$$

The inequality follows from the fact that $C_{jj} = 1$. Then by the minimality of $\deg(C_{ij})$ we have $\deg(C_{ij'}) = \deg(C_{ij})$. Analogous reasoning on $i' \in I$ yields $\deg(C_{ij}) = \deg(C_{i'j'})$ for all $i' \in I, j' \in J$. Thus C is a cell matrix. □

8.1. #P-hardness.

LEMMA 8.4. *Let $\delta \in \mathbb{N}$ and A' be a matrix of the form*

$$A' = \begin{pmatrix} 1 & \cdots & 1 & 2^\delta & \cdots & 2^\delta \\ \vdots & & \vdots & \vdots & & \vdots \\ 1 & \cdots & 1 & 2^\delta & \cdots & 2^\delta \\ 2^\delta & \cdots & 2^\delta & 1 & \cdots & 1 \\ \vdots & & \vdots & \vdots & & \vdots \\ 2^\delta & \cdots & 2^\delta & 1 & \cdots & 1 \end{pmatrix}$$

Then $\mathrm{EVAL}^{\mathit{pin}}(A)$ *is* #P*-hard.*

PROOF. Let $[A']$ be the twin resolvent of A'. By the Twin Reduction Lemma 6.1 we obtain $\mathrm{EVAL}([A'], D) \equiv \mathrm{EVAL}(A')$ where, for $a, b, \delta \geq 1$, the matrices $[A']$ and D satisfy

$$[A'] = \begin{pmatrix} 1 & 2^\delta \\ 2^\delta & 1 \end{pmatrix} \text{ and } D = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$$

We have $\mathrm{EVAL}([A']) \leq \mathrm{EVAL}([A'], D)$ by Lemma 5.6. Therefore it remains to show that $\mathrm{EVAL}([A'])$ is #P-hard. To see this, let $G = (V, E)$ be a graph and for two sets of vertices $U, W \subseteq V$ let $e(U, W)$ denote the number of edges in G between U and W. We have

$$Z_A(G) = \sum_{\sigma: V \to [m]} \prod_{uv \in E} A_{\sigma(u),\sigma(v)} \qquad = \sum_{\sigma: V \to [m]} 2^{\delta \cdot e(\sigma^{-1}(1), \sigma^{-1}(2))}$$

Let c_i be the number of $\sigma : V \to [m]$ with weight $2^{\delta i}$ then we have.

$$Z_A(G) \quad = \quad \sum_{i=0}^{|E|} c_i 2^{\delta i}$$

By $\mathrm{EVAL}(A) \equiv \mathrm{COUNT}(A)$ we can determine the coefficients c_i. Let ν be maximum such that $c_\nu \neq 0$. Then c_ν is the number of maximum cardinality cuts in G. Therefore, this yields a reduction from the problem #MAXCUT of counting maximum cardinality cuts — a problem well known to be #P-hard (this follows, for example, from the work of Simon [**Sim77**]). □

LEMMA 8.5. *Let $A \in \mathbb{Z}[X]^{m \times m}$ be a cell matrix satisfying conditions* **(A)**–**(C)** *and both* **(T1C)** *conditions. Then* $\mathrm{EVAL}^{\mathit{pin}}(A)$ *is* #P*-hard.*

PROOF. Let $\delta = \min\{\deg(A_{ij}) \mid A_{ij} \neq 1,\ i, j \in [m]\}$ and $\Delta = \max\{\deg(A_{ij}) \mid A_{ij} \neq 1,\ i, j \in [m]\}$. Let A_{IJ} be a cell of A with entries X^δ and define $A' = A_{(I \cup J)(I \cup J)}$ which, by symmetry, the definition of the cells and **(C)** has the form

$$A' = A_{(I \cup J)(I \cup J)} = \begin{pmatrix} 1 & \cdots & 1 & X^\delta & \cdots & X^\delta \\ \vdots & & \vdots & \vdots & & \vdots \\ 1 & \cdots & 1 & X^\delta & \cdots & X^\delta \\ X^\delta & \cdots & X^\delta & 1 & \cdots & 1 \\ \vdots & & \vdots & \vdots & & \vdots \\ X^\delta & \cdots & X^\delta & 1 & \cdots & 1 \end{pmatrix}$$

We will show first that $\mathrm{EVAL}^{\mathsf{pin}}(A') \leq \mathrm{EVAL}^{\mathsf{pin}}(A)$. For a graph $G = (V, E)$ and a pinning ϕ of $\mathrm{EVAL}^{\mathsf{pin}}(A')$, define a graph $G' = (V', E')$ as follows. Let $k = |E|\cdot\Delta+1$, we obtain G' from G by adding two apices which are connected to each vertex of V by edges with multiplicity k, that is,

$$\begin{aligned} V' &= V\dot{\cup}\{x, y\} \\ E' &= E \cup \{(xv)^k, (yv)^k \mid v \in V\}. \end{aligned}$$

Let furthermore ϕ' be the extension of ϕ to G' by adding $x \mapsto i_0$, $y \mapsto j_0$ for some $i_0 \in I$ and $j_0 \in J$. Consider a configuration $\sigma : V' \to [m]$ with $\phi' \subseteq \sigma$. For some appropriate $K(\sigma)$ we have

$$\prod_{uv\in E'} A_{\sigma(u),\sigma(v)} = X^{K(\sigma)} \prod_{uv\in E} A_{\sigma(u),\sigma(v)}.$$

Further, if $\sigma(V) \subseteq I \cup J$ then $K(\sigma) = k|V|\delta$ and otherwise, by the definition of δ and the cell structure of A, we have $K(\sigma) \geq k|V|\delta + k\delta$. In particular, as $\deg\left(\prod_{uv\in E} A_{\sigma(u),\sigma(v)}\right) \leq \Delta \cdot |E| < k$ we have

The degree of σ is strictly less than $k|V|\delta + k$ iff $\sigma(V) \subseteq I \cup J$.

We can thus compute $Z_{A'}(\phi, G)$ using a $\mathrm{COUNT}^{\mathsf{pin}}(A)$ oracle, by determining the values $w \cdot N_A(G', \phi', w)$ for all w of degree at most $k|V|\delta + k - 1$. This yields a reduction $\mathrm{EVAL}^{\mathsf{pin}}(A') \leq \mathrm{EVAL}^{\mathsf{pin}}(A)$ by the polynomial time equivalence of $\mathrm{COUNT}^{\mathsf{pin}}(A)$ and $\mathrm{EVAL}^{\mathsf{pin}}(A)$.

Let A'' be the matrix obtained from A' by substituting X with 2. Trivially $\mathrm{EVAL}(A'') \leq \mathrm{EVAL}^{\mathsf{pin}}(A')$ and $\mathrm{EVAL}(A'')$ is #P-hard by Lemma 8.4. □

Proof of the Two-1-Cell Lemma. Let $A \in \mathbb{Z}[X]^{m\times m}$ be a positive symmetric matrix containing at least two 1-cells. That is, A satisfies condition **(A)**. Application of Lemmas 8.2, 8.3 and 8.5 in this order yields the result. □

9. The Single 1-Cell Lemma

Lemma 9.1 (Symmetrized 1-Row Filter Lemma). *Let $A \in \mathbb{Z}[X]^{m\times m}$ be a matrix satisfying conditions* **(A)**–**(C)** *and containing exactly one 1-cell. Let C be obtained from A by removing all non-1-rows. Then*

$$\mathrm{EVAL}^{pin}(CC^T) \leq \mathrm{EVAL}^{pin}(A).$$

Proof. As A satisfies conditions **(A)** – **(C)** and contains a single 1-cell we see that there is an $r < m$ such that with $I = [r]$ the principal submatrix A_{II} forms this single 1-cell. Recall that $\mathrm{EVAL}^{\mathsf{pin}}(A^2) \leq \mathrm{EVAL}^{\mathsf{pin}}(A)$ by 2-stretching (cf. Lemma 5.1), and we have

$$(A^2)_{ij} = \sum_{k=1}^{m} A_{ik}A_{jk} = \sum_{k=1}^{r} A_{ik}A_{jk} + \sum_{k=r+1}^{m} A_{ik}A_{jk}.$$

Therefore, for $i, j \in [r]$ we have $(A^2)_{ij} = r + \sum_{k=r+1}^{m} A_{ik}A_{jk}$, i.e. a polynomial with a constant term. On the other hand, if either $i \notin I$ or $j \notin I$ then A^2_{ij} is divisible by X. Further we have $C = A_{[r][m]}$ and thus $CC^T = (A^2)_{[r][r]}$. Hence CC^T is the submatrix of A^2 which consists of exactly those entries not divisible by X. Recall that for any graph G and pinning ϕ, we have

$$Z_{A^2}(\phi, G) = \sum_{w \in \mathcal{W}_{A^2}(G)} w \cdot N_{A^2}(G, \phi, w)$$

and by the above

$$Z_{CC^T}(\phi, G) = \sum_{\substack{w \in \mathcal{W}_{A^2}(G) \\ X \text{ does not divide } w}} w \cdot N_{A^2}(G, \phi, w)$$

Filtering the weights w appropriately by means of the $\mathrm{EVAL}^{\mathsf{pin}}(A) \equiv \mathrm{COUNT}^{\mathsf{pin}}(A)$ correspondence, we can devise a reduction witnessing $\mathrm{EVAL}^{\mathsf{pin}}(CC^T) \leq \mathrm{EVAL}^{\mathsf{pin}}(A)$. □

Single 1-Cell Conditions **(S1C)**.. We define some further conditions for matrices $A \in \mathbb{Z}[X]^{m \times m}$ satisfying conditions **(A)**–**(C)**. Define $\delta_{ij} = \deg(A_{ij})$ for all $i, j \in [m]$ and let $r := \min\{i \in [m] \mid A_{i1} > 1\}$. That is, r is the smallest index such that $\delta_{1r} = \delta_{r1}$ is greater than zero and it exists because A has rank at least 2.

(S1C–A) A has exactly one 1-cell.

(S1C–B) $A_{1,*} = \ldots = A_{r-1,*}$, i.e. the first $r-1$ rows of A are identical.

(S1C–C) For $i \in [r, m]$ we have $\delta_{1k} \leq \delta_{ik}$ for all $k \in [m]$.

Lemma 9.2. *Let $A \in \mathbb{Z}[X]^{m \times m}$ be a matrix satisfying conditions* **(A)**–**(C)** *and* **(S1C–A)**. *Then there is a matrix A' satisfying* **(A)**–**(C)** *and all* **(S1C)** *conditions such that*

$$\mathrm{EVAL}^{\mathit{pin}}(A') \leq \mathrm{EVAL}^{\mathit{pin}}(A).$$

Proof. The matrix A already satisfies **(A)**–**(C)** and **(S1C–A)**. Observe first that if A satisfies **(S1C–B)** as well, a matrix A' whose existence we want to prove, can be defined as follows. For $t \in \mathbb{N}$, let $A' = A'(t)$ be given by

$$A'_{ij} = A_{ij}(A_{ii}A_{jj})^t \text{ for all } i, j \in [m].$$

Clearly, for all $t \in \mathbb{N}$ the matrix $A'(t)$ satisfies conditions **(A)**–**(C)**, **(S1C–A)** and **(S1C–B)**.

We claim that there is a $t \in \mathbb{N}$ such that $A'(t)$ also satisfies condition **(S1C–C)**. To see this, note first that the degrees of A satisfy $\min\{\delta_{rr}, \ldots, \delta_{mm}\} > 0$ as all 1-entries of A are contained in the single 1-cell $A_{[r-1][r-1]}$. Therefore, there is a t such that $\delta_{1j} \leq t \cdot \min\{\delta_{rr}, \ldots, \delta_{mm}\}$ for all $j \in [m]$. Fix such a t and note that

$$\deg(A'_{ij}) = \delta_{ij} + t \cdot \delta_{ii} + t \cdot \delta_{jj}.$$

Thus, with $\delta_{11} = 0$ we see that for all $j \in [m]$ and $i \in [r, m]$

$$\deg(A'_{1j}) = \delta_{1j} + t \cdot \delta_{jj} \leq t \cdot \delta_{ii} + t \cdot \delta_{jj} \leq \deg(A'_{ij}).$$

This proves that A' satisfies condition **(S1C–C)**. Reducibility is given as follows

Claim 8. For all $t \in \mathbb{N}$ we have $\mathrm{EVAL}^{\mathsf{pin}}(A'(t)) \leq \mathrm{EVAL}^{\mathsf{pin}}(A)$.

Proof. Let $G = (V, E)$, ϕ be an instance of $\mathrm{EVAL}^{\mathsf{pin}}(A')$, Let $G' = (V, E')$ be the graph obtained from G by adding t many self-loops to each vertex. Then $Z_{A'}(\phi, G) = Z_A(\phi, G')$, witnessing the claimed reducibility. ⊣

It remains to show how to obtain condition **(S1C–B)**. We give the proof by induction on m. For $m = 2$ this is trivial. If $m \geq r > 2$ assume that A does not satisfy **(S1C–B)**. Define $C = A_{[r-1][m]}$, i.e. the matrix consisting of the first $r-1$ rows of A. By the Symmetrized 1-Row Filter Lemma 9.1 we have $\mathrm{EVAL}^{\mathsf{pin}}(CC^T) \leq \mathrm{EVAL}^{\mathsf{pin}}(A)$. Further, C has rank at least 2 as $C_{i1} = 1$ for all $i \in [r-1]$ but the rows of C are not identical and therefore CC^T has rank at least 2, as well (see Lemma 5.2).

Application of the General Conditioning Lemma 4.4 yields $\mathrm{EVAL}^{\mathsf{pin}}(C') \leq \mathrm{EVAL}^{\mathsf{pin}}(CC^T)$ for a $k \times k$ matrix satisfying **(A)**–**(C)** such that $k \leq r-1$. If C' has at least two 1-cells we apply the Two 1-Cell Lemma 4.5 to obtain #P-hardness of $\mathrm{EVAL}^{\mathsf{pin}}(C')$. Therefore we can chose some A' satisfying the conditions of the Lemma such that $\mathrm{EVAL}^{\mathsf{pin}}(A') \leq \mathrm{EVAL}^{\mathsf{pin}}(C')$. If C' has only one 1-cell then the proof follows by the induction hypothesis, as C' has order $k \leq r-1$. □

Definition of $A^{[k]}$ and $C^{[k]}$. For the remainder of this section we fix $A \in \mathbb{Z}[X]^{m\times m}$ satisfying conditions **(A)**–**(C)** and all **(S1C)** conditions. Furthermore, δ_{ij} for all $i, j \in [m]$ and r are defined for A as in the Single 1-Cell Conditions. For $k \in \mathbb{N}$ define the matrix $A^{[k]}$ by

$$A^{[k]}_{ij} = A_{ij}(A_{1j})^{k-1} \text{ for all } i, j \in [m]. \tag{9.1}$$

Let further, $C^{[k]}$ be defined by

$$C^{[k]} = A^{[k]}(A^{[k]})^T. \tag{9.2}$$

Lemma 9.3. *Let A and $C^{[k]}$ be defined as above. Then*

$$\mathrm{EVAL}^{pin}(C^{[k]}) \leq \mathrm{EVAL}^{pin}(A).$$

Proof. Let $G = (V, E)$ and a pinning ϕ be an instance of $\mathrm{EVAL}^{\mathsf{pin}}(C^{[k]})$. Define a graph $G' = (V', E')$ as follows:

$$\begin{aligned} V' &:= V \dot\cup \{z\} \dot\cup \{v_e \mid e \in E\} \\ E' &:= \{uv_e, v_e v \mid e = uv \in E\} \dot\cup \{(v_e z)^{2k-2} \mid e \in E\}. \end{aligned}$$

The edges $v_e z$ have multiplicity $2k-2$. Let ϕ' be the extension of ϕ defined by $\phi \cup \{z \mapsto 1\}$. Let $\sigma \supseteq \phi'$ be a configuration on G' and A. Then its weight equals

$$\begin{aligned} \prod_{uv\in E'} A_{\sigma(u),\sigma(v)} &= \prod_{uv\in E}\sum_{\nu=1}^{m} A_{\sigma(u),\nu}A_{\sigma(v),\nu}(A_{1,\nu})^{2k-2} \\ &= \prod_{uv\in E} C^{[k]}_{\sigma(u),\sigma(v)}. \end{aligned}$$

The last equality follows directly from the definition of $C^{[k]}$. Hence $\mathrm{EVAL}^{\mathsf{pin}}(C^{[k]}) \leq \mathrm{EVAL}^{\mathsf{pin}}(A)$ as required. □

We need some further notation to deal with polynomials. For $f \in \mathbb{Q}[X]$ and $\lambda \in \mathbb{C}$ let $\mathsf{mult}(\lambda, f)$ denote the *multiplicity* of λ in f if λ is a root of f, and $\mathsf{mult}(\lambda, f) = 0$, otherwise. The *k-th root* of $\lambda \in \mathbb{C}$ is the k-element set $\lambda^{1/k} = \{\mu \in$

$\mathbb{C} \mid \mu^k = \lambda\}$. Slightly abusing notation, $\lambda^{1/k}$ will denote any element from this set. For every root λ of $C_{11}^{[1]}$, every $r \leq j \leq m$ and $k \in \mathbb{N}$ we define

$$m(\lambda, j) = \min\{ \ \mathsf{mult}(\lambda^{1/k}, C_{1j}^{[k]}) \mid k \geq 1\}. \tag{9.3}$$

The following Lemma is the technical core of this section.

LEMMA 9.4 (Single-1-Cell Technical Core). *Let $C^{[k]}$ be defined as above. The following is true for all $j \in [r, m]$.*

(1) *For any root λ of $C_{11}^{[1]}$ and all $k \in \mathbb{N}$ we have*
  (1a) $\mathsf{mult}(\lambda^{1/k}, C_{11}^{[k]}) = \mathsf{mult}(\lambda, C_{11}^{[1]}) \geq m(\lambda, j)$.
  (1b) $\mathsf{mult}(\lambda^{1/k}, C_{jj}^{[k]}) \geq m(\lambda, j)$.

(2) *If row $A_{1,*}$ and $A_{j,*}$ are linearly dependent then for any root λ of $C_{11}^{[1]}$ and all $k \in \mathbb{N}$ we have*

$$\mathsf{mult}(\lambda^{1/k}, C_{11}^{[k]}) = \mathsf{mult}(\lambda^{1/k}, C_{1j}^{[k]}) = \mathsf{mult}(\lambda^{1/k}, C_{jj}^{[k]}).$$

(3) *If row $A_{1,*}$ and $A_{j,*}$ are linearly independent there is a root λ of $C_{11}^{[1]}$ such that*

$$\mathsf{mult}(\lambda, C_{11}^{[1]}) > m(\lambda, j).$$

We omit the proof of this lemma, due to space restrictions. For the proof we refer the reader to the full version of this paper [**GT11**] (or alternatively [**BG05**], and [**Thu09**]). We shall only show, how to prove the Single-1-Cell Lemma, provided that the above holds.

The following basic facts[5] about polynomials will be used frequently in the following.

LEMMA 9.5. *Let $f \in \mathbb{Q}[X]$ and $\lambda \in \mathbb{C}$.*

(1) *There exists a unique (up to a scalar factor) irreducible polynomial $p_\lambda \in \mathbb{Q}[X]$ such that λ is a root of p_λ. If $f(\lambda) = 0$ then p_λ divides f.*
(2) *If $\mathsf{mult}(\lambda, f) = s$ then $f = p_\lambda^s \bar{f}$ for some $\bar{f} \in \mathbb{Q}[X]$ which satisfies $\bar{f}(\lambda) \neq 0$.*
(3) *If $f(\lambda) = 0$ then $\lambda^{1/k}$ is a root of $f(X^k)$ and $\mathsf{mult}(\lambda^{1/k}, f(X^k)) = \mathsf{mult}(\lambda, f(X))$.*

PROOF OF THE SINGLE 1-CELL LEMMA 4.6. Let $A \in \mathbb{Z}[X]^{m \times m}$ satisfy conditions **(A)**–**(C)** and **(S1C–A)**. By Lemma 9.2 we may assume w.l.o.g. that A indeed satisfies all **(S1C)** conditions. By the Two–1–Cell Lemma 4.5, it suffices to prove the existence of a positive symmetric matrix C which contains at least two 1-cells such that

$$\text{EVAL}^{\mathsf{pin}}(C) \leq \text{EVAL}^{\mathsf{pin}}(A).$$

Recall the definition of $C^{[k]}$ in equation (9.2) and the definition of r in the **(S1C)** conditions. We prove a technical tool.

CLAIM 9. There is a root λ of $C_{11}^{[1]}$, an index $j \in [r, m]$ and $k \in \mathbb{N}$ such that

(1) $\mathsf{mult}(\lambda^{1/k}, C_{1j}^{[k]}) < \mathsf{mult}(\lambda^{1/k}, C_{11}^{[k]})$.
(2) For all $i \in [r, m]$ we have $\mathsf{mult}(\lambda^{1/k}, C_{1j}^{[k]}) \leq \mathsf{mult}(\lambda^{1/k}, C_{ii}^{[k]})$.

---

[5] cf., for example, Chapter IV in [**Lan02**]

Proof. Choose λ and $j \in [r, m]$ such that $m(\lambda, j)$ is minimal for all choices of λ and j which satisfy

$$A_{1,*} \text{ and } A_{j,*} \text{ are linearly independent and } \mathsf{mult}(\lambda, C_{11}^{[1]}) > m(\lambda, j). \tag{9.4}$$

Note that these λ and j exist by Lemma 9.4(3). Choose $k \in \mathbb{N}$ such that, by equation (9.3), $m(\lambda, j) = \mathsf{mult}(\lambda^{1/k}, C_{1j}^{[k]})$. This proves (1) by Lemma 9.4(1a).

To prove (2) fix $i \in [r, m]$. If $A_{1,*}$ and $A_{i,*}$ are linearly independent our choice of λ and j implies that $m(\lambda, i) \geq m(\lambda, j)$ and therefore

$$\mathsf{mult}(\lambda^{1/k}, C_{ii}^{[k]}) \geq m(\lambda, i) \geq m(\lambda, j) = \mathsf{mult}(\lambda^{1/k}, C_{1j}^{[k]})$$

where the first inequality holds by Lemma 9.4(1b).

If otherwise $A_{1,*}$ and $A_{i,*}$ are linearly dependent then

$$\mathsf{mult}(\lambda^{1/k}, C_{ii}^{[k]}) = \mathsf{mult}(\lambda^{1/k}, C_{11}^{[k]}) \geq m(\lambda, j) = \mathsf{mult}(\lambda^{1/k}, C_{1j}^{[k]}).$$

The first equality holds by Lemma 9.4(2) and the inequality is true by Lemma 9.4(1a). $\dashv$

Choose j, λ, k as in Claim 9. Define $t = \mathsf{mult}(\lambda^{1/k}, C_{1j}^{[k]})$. Let p_λ be an irreducible polynomial such that $p_\lambda(\lambda) = 0$. Let $s := \min\{ \mathsf{mult}(\lambda^{1/k}, C_{ab}^{[k]}) \mid a, b \in [m]\}$ and define the positive symmetric matrix

$$C := \frac{1}{p_\lambda^s} C^{[k]}|_{p_\lambda}.$$

Note that $s \leq t$ by definition. We consider two cases. If $s = t$ then $C_{1j} = C_{j1} = 1$ but by Claim 9(1) we have $C_{11} \neq 1$. Therefore, C contains at least two 1-cells.

If otherwise $s < t$, fix witnesses $a, b \in [m]$ with $\mathsf{mult}(\lambda^{1/k}, C_{ab}^{[k]}) = s$.

Claim 10. Either $a \geq r$ or $b \geq r$.

Proof. Recall that by condition **(S1C–B)** the first $r - 1$ rows of A — and hence those of $A^{[k]}$ — are identical. By the definition $C^{[k]} = A^{[k]}(A^{[k]})^T$ (cf. equation (9.2)) we have

$$C_{ij}^{[k]} = \sum_{\nu=1}^{m} A_{i\nu}^{[k]} A_{j\nu}^{[k]} = \sum_{\nu=1}^{m} A_{i\nu} A_{j\nu} (A_{1\nu})^{2k-2}$$

In particular, for all $a', b' \in [r-1]$ we have $C_{a'b'}^{[k]} = C_{11}^{[k]}$ and hence $\mathsf{mult}(\lambda^{1/k}, C_{a'b'}^{[k]}) = \mathsf{mult}(\lambda^{1/k}, C_{11}^{[k]}) > \mathsf{mult}(\lambda^{1/k}, C_{ab}^{[k]})$ which proves the claim. $\dashv$

Combining this claim with Claim 9(2) and the fact that $0 < t = \mathsf{mult}(\lambda^{1/k}, C_{1j}^{[k]})$, we have either $C_{aa} \neq 1$ or $C_{bb} \neq 1$. As $C_{ab} = C_{ba} = 1$ by the definition of a, b we see that C contains at least two 1-cells. We have $\text{EVAL}^{\mathsf{pin}}(C^{[k]}) \leq \text{EVAL}^{\mathsf{pin}}(A)$ by Lemma 9.3. Further $\text{EVAL}^{\mathsf{pin}}(C^{[k]}|_{p_\lambda}) \leq \text{EVAL}^{\mathsf{pin}}(C^{[k]})$ by the Prime Filter Lemma 7.7. Then $\text{EVAL}^{\mathsf{pin}}(C) \leq \text{EVAL}^{\mathsf{pin}}(A)$ follows from the fact that

$$Z_{C^{[k]}|_{p_\lambda}}(\phi, G) = p_\lambda^{s \cdot |E|} Z_C(\phi, G) \text{ for all graphs } G = (V, E) \text{ and pinnings } \phi.$$

$\square$

References

[Aig07] M. Aigner, *A course in enumeration*, Springer-Verlag, 2007.

[Aus07] P. Austrin, *Towards sharp inapproximability for any 2-CSP*, Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science, 2007, pp. 307–317.

[BDG+09] A. Bulatov, M. E. Dyer, L. A. Goldberg, M. Jalsenius, and D. Richerby, *The complexity of weighted boolean #csp with mixed signs*, Theoretical Computer Science **410** (2009), 3949–3961.

[BDG+10] A. Bulatov, M. E. Dyer, L. A. Goldberg, M. Jalsenius, M. Jerrum, and D. Richerby, *The complexity of weighted and unweighted #csp*, CoRR **abs/1005.2678** (2010).

[BDGM09] A. Bulatov, V. Dalmau, M. Grohe, and D. Marx, *Enumerating homomorphisms*, Proceedings of the 26th Annual Symposium on Theoretical Aspects of Computer Science (S. Albers and J.-Y. Marion, eds.), 2009, pp. 231–242.

[BG05] A. Bulatov and M. Grohe, *The complexity of partition functions*, Theoretical Computer Science **348** (2005), 148–186.

[BK09] I. Briquel and P. Koiran, *A dichotomy theorem for polynomial evaluation*, Proceedings of the 34th International Symposium on Mathematical Foundations of Computer Science (R. Královic and D. Niwinski, eds.), Lecture Notes in Computer Science, vol. 5734, Springer Verlag, 2009, pp. 187–198.

[BKJ05] A. Bulatov, A. Krokhin, and P. Jeavons, *Classifying the complexity of constraints using finite algebras*, SIAM Journal on Computing **34** (2005), 720–742.

[BKN09] L. Barto, M. Kozik, and T. Niven, *The CSP dichotomy holds for digraphs with no sources and no sinks (a positive answer to a conjecture of bang-jensen and hell)*, SIAM Journal on Computing **38** (2009), 1782–1802.

[Bul06] A. Bulatov, *A dichotomy theorem for constraint satisfaction problems on a 3-element set.*, Journal of the ACM **53** (2006), 66–120.

[Bul08] ———, *The complexity of the counting constraint satisfaction problem*, Proceedings of the 35th International Colloquium on Automata, Languages and Programming, Part I (Luca Aceto, Ivan Damgård, Leslie Ann Goldberg, Magnús M. Halldórsson, Anna Ingólfsdóttir, and Igor Walukiewicz, eds.), Lecture Notes in Computer Science, vol. 5125, Springer-Verlag, 2008, pp. 646–661.

[CC10] J. Cai and X. Chen, *A decidable dichotomy theorem on directed graph homomorphisms with non-negative weights*, To appear in Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS 2010), 2010.

[CCL10a] J. Cai, X. Chen, and P. Lu, *Graph homomorphisms with complex values: A dichotomy theorem*, ICALP (1), 2010, pp. 275–286.

[CCL10b] ———, *Non-negative weighted #csps: An effective complexity dichotomy*, CoRR **abs/1012.5659** (2010).

[CH96] N. Creignou and M. Hermann, *Complexity of generalized satisfiability counting problems*, Information and Computation **125** (1996), no. 1, 1–12.

[CLX08] J. Cai, P. Lu, and M. Xia, *Holographic algorithms by fibonacci gates and holographic reductions for hardness*, FOCS, IEEE Computer Society, 2008, pp. 644–653.

[CLX09] ———, *Holant problems and counting CSP*, Proceedings of the 41st ACM Symposium on Theory of Computing, 2009, pp. 715–724.

[Coh93] H. Cohen, *A course in computational algebraic number theory*, Graduate Texts in Mathematics, vol. 138, Springer-Verlag, Berlin, 1993. MR MR1228206 (94i:11105)

[DG00] M. Dyer and C. Greenhill, *The complexity of counting graph homomorphisms*, Random Structures and Algorithms **17** (2000), no. 3–4, 260–289.

[DGJ08] M. E. Dyer, L. A. Goldberg, and M. Jerrum, *A complexity dichotomy for hypergraph partition functions*, CoRR **abs/0811.0037** (2008).

[DGJ09] ———, *The complexity of weighted boolean csp*, SIAM J. Comput. **38** (2009), no. 5, 1970–1986.

[DGP07] M. E. Dyer, L. A. Goldberg, and M. Paterson, *On counting homomorphisms to directed acyclic graphs*, Journal of the ACM **54** (2007), no. 6.

[DJKK08] V. Deineko, P. Jonsson, M. Klasson, and A. Krokhin, *The approximability of max csp with fixed-value constraints*, Journal of the ACM **55** (2008), no. 4, 1–37.

[DR10a] M. E. Dyer and D. Richerby, *The complexity of #csp*, CoRR **abs/1003.3879** (2010).

[DR10b] ———, *On the complexity of #csp*, STOC (L. J. Schulman, ed.), ACM, 2010, pp. 725–734.

[DRar] ———, *The #csp dichotomy is decidable*, STACS, 2011, to appear.

[FLS07] M. Freedman, L. Lovász, and A. Schrijver, *Reflection positivity, rank connectivity, and homomorphism of graphs*, Journal of the American Mathematical Society **20** (2007), 37–51.

[FV98] T. Feder and 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** (1998), 57–104.

[GGJT10] L. A. Goldberg, M. Grohe, M. Jerrum, and M. Thurley, *A complexity dichotomy for partition functions with mixed signs*, SIAM J. Comput. **39** (2010), no. 7, 3336–3402.

[Gro07] M. Grohe, *The complexity of homomorphism and constraint satisfaction problems seen from the other side*, Journal of the ACM **54** (2007), 1–24.

[GT11] M. Grohe and M. Thurley, *Counting homomorphisms and partition functions*, CoRR **abs/1104.0185** (2011).

[HN90] P. Hell and J. Nešetřil, *On the complexity of H-coloring*, Journal of Combinatorial Theory, Series B **48** (1990), 92–110.

[Isi25] E. Ising, *Beitrag zur Theorie des Ferromagnetismus*, Zeitschrift fur Physik **31** (1925), 253–258.

[KMZ11] T. Kotek, J.A. Makowsky, and B. Zilber, *On Counting Generalized Colorings*. This Volume.

[Lad75] R. E. Ladner, *On the structure of polynomial time reducibility*, J. ACM **22** (1975), no. 1, 155–171.

[Lan02] S. Lang, *Algebra*, third ed., Graduate Texts in Mathematics, vol. 211, Springer-Verlag, New York, 2002. MR MR1878556 (2003e:00003)

[LN97] R. Lidl and H. Niederreiter, *Finite fields*, 2nd ed., Encyclopedia of Mathematics and its Applications, vol. 20, Cambridge University Press, Cambridge, 1997.

[Lov06] L. Lovász, *The rank of connection matrices and the dimension of graph algebras*, European Journal of Combinatorics **27** (2006), 962–970.

[PB83] J. S. Provan and M. O. Ball, *The complexity of counting cuts and of computing the probability that a graph is connected*, SIAM J. Comput. **12** (1983), no. 4, 777–788.

[Rag08] P. Raghavendra, *Optimal algorithms and inapproximability results for every CSP?*, Proceedings of the 40th ACM Symposium on Theory of Computing, 2008, pp. 245–254.

[Ric01] D. Richardson, *Multiplicative independence of algebraic numbers and expressions*, J. Pure Appl. Algebra **164** (2001), no. 1-2, 231–245, Effective methods in algebraic geometry (Bath, 2000). MR MR1854340 (2002j:11079)

[Sch09] A. Schrijver, *Graph invariants in the spin model*, Journal of Combinatorial Theory, Series B **99** (2009), 502–511.

[Sim77] J. Simon, *On the difference between one and many (preliminary version)*, ICALP (Arto Salomaa and Magnus Steinby, eds.), Lecture Notes in Computer Science, vol. 52, Springer, 1977, pp. 480–491.

[SS07] H. Schnoor and I. Schnoor, *Enumerating all solutions for constraint satisfaction problems*, Proceedings of the 24th Annual Symposium on Theoretical Aspects of Computer Science (W. Thomas and P. Weil, eds.), Lecture Notes in Computer Science, vol. 4393, Springer-Verlag, 2007, pp. 694–705.

[Sze07] B. Szegedy, *Edge coloring models and reflection positivity*, Journal of the American mathematical Society **20** (2007), 969–988.

[Thu09] M. Thurley, *The complexity of partition functions*, Ph.D. thesis, Humboldt Universität zu Berlin, 2009.

[Val79] L. G. Valiant, *The complexity of computing the permanent*, Theor. Comput. Sci. **8** (1979), 189–201.

[Val08] ———, *Holographic algorithms*, SIAM J. Comput. **37** (2008), no. 5, 1565–1594.

Humboldt Universität zu Berlin, Berlin, Germany
E-mail address: grohe@informatik.hu-berlin.de

Centre de Recerca Matemàtica, Bellaterra, Spain
E-mail address: marc.thurley@googlemail.com

This page intentionally left blank

Contemporary Mathematics
Volume **558**, 2011

Some examples of universal and generic partial orders

Jan Hubička and Jaroslav Nešetřil

Abstract. We survey structures endowed with natural partial orderings and prove their universality. These partial orders include partial orders on sets of words, partial orders formed by geometric objects, grammars, polynomials and homomorphism order for various combinatorial objects.

1. Introduction

For given class $\mathcal{K}$ of countable partial orders we say that class $\mathcal{K}$ contains an *embedding-universal* (or simply *universal*) structure $(U, \leq_U)$ if every partial order $(P, \leq_P) \in \mathcal{K}$ can be found as induced suborder of $(U, \leq_U)$ (or in other words, there exists embedding from $(P, \leq_P)$ to $(U, \leq_U)$).

A partial order $(P, \leq_P)$ is *ultrahomogeneous* (or simply *homogeneous*), if every isomorphism of finite suborders of $(P, \leq_P)$ can be extended to an automorphism of $(P, \leq_P)$.

A partial order $(P, \leq_P)$ is *generic* if it is both *ultrahomogeneous* and *universal*.

The generic objects can be obtained from the Fraïssé limit [**4**]. But it is important that often these generic objects (despite their apparent complexity and universality) admit a concise presentation. Thus for example the Rado graph (i.e. countable universal and homogeneous undirected graph) can be represented in various ways by elementary properties of sets or finite sequences, number theory or even probability. Similar concise representations were found for some other generic objects such as all undirected ultrahomogeneous graphs [**9**] or the Urysohn space [**7**]. The study of generic partial order also motivated this paper and we consider representation of the generic partial order in Section 3.

The notion of finite presentation we interpret here broadly as a succinct representation of an infinite set, succint in the sense that elements are finite models with relations induced by "compatible mappings" (such as homomorphisms) between the corresponding models. This intuitive definition suffices as we are interested in the (positive) examples of such representations.

1991 *Mathematics Subject Classification.* Primary 06A06, 06A11; Secondary 05C38.

Key words and phrases. universal partial order, generic partial order, path homomorphisms, homomorphism order, order of convex sets, words order, Conway numbers.

The Institute for Theoretical Computer Science (ITI) is supported as project 1M0545 by the Ministry of Education of the Czech Republic.

A finite presentation of the generic partial order is given in [**9**] —however this construction is quite complicated. This paper gives a more streamlined construction and relates it to Conway surreal numbers (see Section 3).

In Section 2 we present several simple constructions which yield (countably) universal partial orders. Such objects are interesting on their own and were intensively studied in the context of universal algebra and categories. For example, it is a classical result of Pultr and Trnková [**20**] that finite graphs with the homomorphism order are countably universal quasiorder. Extending and completing [**9**] we give here several constructions which yields to universal partial orders. These constructions include:

(1) The order $(\mathcal{W}, \leq_{\mathcal{W}})$ on sets of words in the alphabet $\{0, 1\}$.
(2) The dominance order on the binary tree $(\mathcal{B}, \leq_{\mathcal{B}})$.
(3) The inclusion order of finite sets of finite intervals $(\mathcal{I}, \leq_{\mathcal{I}})$.
(4) The inclusion order of convex hulls of finite sets of points in the plane $(\mathcal{C}, \leq_{\mathcal{C}})$.
(5) The order of piecewise linear functions on rationals $(\mathcal{F}, \leq_{\mathcal{F}})$.
(6) The inclusion order of periodic sets $(\mathcal{S}, \subseteq)$.
(7) The order of sets of truncated vectors (generalization of orders of vectors of finite dimension) $(\mathcal{T}V, \leq_{\mathcal{T}V})$.
(8) The orders implied by grammars on words $(\mathcal{G}, \leq_{\mathcal{G}})$.
(9) The homomorphism order of oriented paths $(\mathcal{P}, \leq_{\mathcal{P}})$.

Note that with universal partial orders we have more freedom (than with the generic pratial order) and as a consequence we give a perhaps surprising variety of finite presentations.

We start with a simple representation by means of finite sets of binary words. This representation seems to capture properties of such a universal partial order very well and it will serve as our "master" example. In most other cases we prove the universality of some particular partial order by finding a mapping from the words representation into the structure in question. This technique will be shown in several applications in the next sections. While some of these structures are known be universal, see e.g. [**5, 16, 8**], in several cases we can prove the universality in a new, we believe, much easier way. The embeddings of structures are presented in Figure 1 (ones denoted by dotted lines are not presented in this paper, but references are given).

At this point we would like to mention that the (countable) universality is an essentially finite problem as it can be formulated as follows: By an *on-line representation* of a class $\mathcal{K}$ of partial orders in a partial order $(P, \leq_P)$, we mean that one can construct an embedding $\varphi : R \to P$ of any partial order $(R, \leq_R)$ in class $\mathcal{K}$ under the restriction that the elements of R are revealed one by one. The on-line representation of a class of partial orders can be considered as a game between two players A and B (usually Alice and Bob). Player B chooses a partial order $(P, \leq_P)$ in the class $\mathcal{K}$, and reveals the elements of P one by one to player A (B is a bad guy). Whenever an element of x of P is revealed to A, the relations among x and previously revealed elements are also revealed. Player A is required to assign a vertex $\varphi(x)$—before the next element is revealed—such that φ is an embedding of the suborder induced by $(P, \leq_P)$ on the already revealed elements of $(R, \leq_R)$. Player A wins the game if he succeeds in constructing an embedding φ.

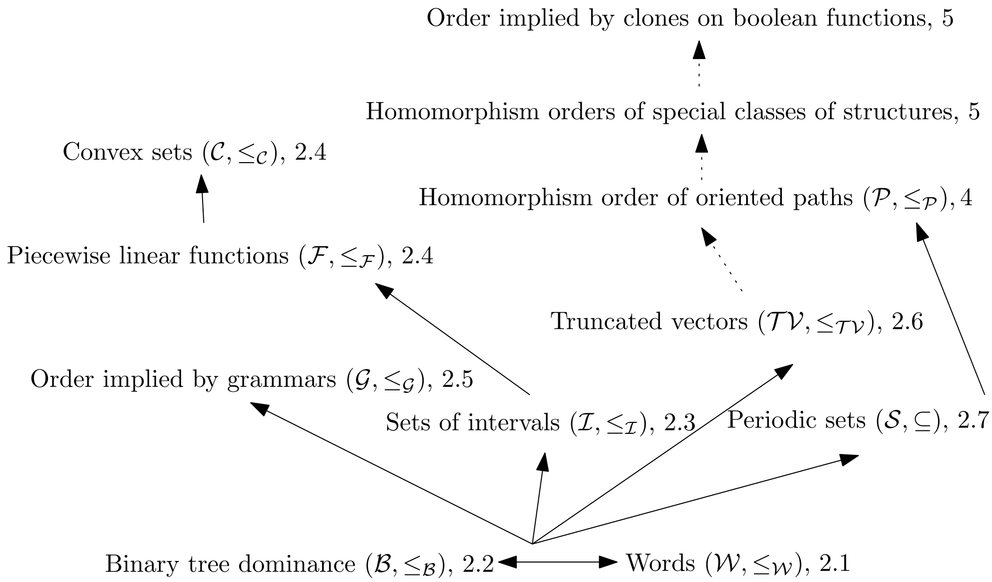


FIGURE 1. Embeddings presented in this paper

The class $\mathcal{K}$ of partial orders is on-line representable in the partial order $(P, \leq_P)$ if player A has a winning strategy.

On-line representation (describing winning strategy of A) is a convenient way of showing the universality of given partial order. In particular it transforms problem of embedding countable structures into a finite problem of extending the existing partial embedding by next element.

We say that a partial order $(P, \leq_P)$ has the *extension property* if the following holds: for any finite mutually disjoint subsets $L, G, U \subseteq P$ there exist a vertex $v \in P$ such that $v' <_P v$ for each $v' \in L$, $v <_P v'$ for each $v' \in G$ and neither $v \leq_P v'$ nor $v' \leq_P v$ for each $v' \in U$. The extension property is a stronger form of on-line representability of any partial order. Using a zig-zag argument it is easy to show that a partial order having the extension property is homogeneous (and thus generic).

In Section 3 we describe a finite representation of the generic partial order related to Conway's surreal numbers. Somewhat surprisingly, this is the only known finite presentation of the generic partial order [**9**]. The constructions of universal partial orders are easier, but they are often not generic. We discuss reasons why other structures fail to be ultrahomogeneous. In particular we will look for gaps in the partial order. Recall that the *gap* in a partial order $(P, \leq_P)$ is a pair of elements $v, v' \in P$ such that $v <_{\mathcal{B}} v'$. A partial order having no gaps is called *dense*. We will show examples of universal partial orders both with gaps and without gaps but still failing to be generic.

2. Examples of Universal Partial Orders

To prove the universality of a given partially ordered set is often a difficult task [**5, 20, 10, 16**]. However, the individual proofs, even if developed independently, use similar tools. We demonstrate this by isolating a "master" construction (in

Section 2.1). This construction is then embedded into partial orders defined by other structures (as listed above). We shall see that the representation of this particular order is flexible enough to simplify further embeddings.

2.1. Word representation. The set of all words over the alphabet $\Sigma = \{0,1\}$ is denoted by $\{0,1\}^*$. For words W, W' we write $W \leq_w W'$ if and only if W' is an initial segment (left factor) of W. Thus we have, for example, $\{011000\} \leq_w \{011\}$ and $\{010111\} \not\leq_w \{011\}$.

Definition 2.1. Denote by $\mathcal{W}$ the class of all finite subsets A of $\{0,1\}^*$ such that no distinct words W, W' in A satisfy $W \leq_w W'$. For $A, B \in \mathcal{W}$ we put $A \leq_{\mathcal{W}} B$ when for each $W \in A$ there exists $W' \in B$ such that $W \leq_w W'$.

Obviously $(\mathcal{W}, \leq_{\mathcal{W}})$ is a partial order (antisymmetry follows from the fact that A is an antichain in the order $\leq_w$).

Definition 2.2. For a set A of finite words denote by $\min A$ the set of all minimal words in A (i.e. all $W \in A$ such that there is no $W' \in A$ satisfying $W' <_w W$).

Now we show that there is an on-line embedding of any finite partial order to $(\mathcal{W}, \leq_{\mathcal{W}})$. Let $[n]$ be the set $\{1, 2, \ldots, n\}$. The partial orders will be restricted to those whose vertex sets are sets $[n]$ (for some $n > 1$) and the vertices will always be embedded in the natural order. Given a partial order $([n], \leq_P)$ let $([i], \leq_{P_i})$ denote the partial order induced by $([n], \leq_P)$ on the set of vertices $[i]$.

Our main construction is the function Ψ mapping partial orders $([n], \leq_P)$ to elements of $(\mathcal{W}, \leq_{\mathcal{W}})$ defined as follows:

Definition 2.3.
Let $L([n], \leq_P)$ be the union of all $\Psi([m], \leq_{P_m})$, $m < n$, $m \leq_P n$.
Let $U([n], \leq_P)$ be the set of all words W such that W has length n, the last letter is 0 and for each $m < n, n \leq_P m$ there is a $W' \in \Psi([m], \leq_{P_m})$ such that W is an initial segment of W'.
Finally, let $\Psi([n], \leq_P)$ be $\min(L([n], \leq_P) \cup U([n], \leq_P))$.
In particular, $L([1], \leq_P) = \emptyset, U([1], \leq_P) = \{0\}, \Psi([1], \leq_P) = \{0\}$.

The main result of this section is the following:

Theorem 2.4. *Given a partial order $([n], \leq_P)$ we have:*

(1) *For every $i, j \in [n]$,*

$$i \leq_P j \text{ if and only if } \Psi([i], \leq_{P_i}) \leq_{\mathcal{W}} \Psi([j], \leq_{P_j})$$

and

$$\Psi([i], \leq_{P_i}) = \Psi([j], \leq_{P_j}) \text{ if and only if } i = j.$$

(This says that the mapping $\Phi(i) = \Psi([i], \leq_{P_i})$ is an embedding of $([n], \leq_P)$ into $(\mathcal{W}, \leq_{\mathcal{W}})$),

(2) *For every $S \subseteq [n]$ there is a word W of length n such that for each $k \leq n$, $\{W\} \leq_{\mathcal{W}} \Psi([k], \leq_{P_k})$ if and only if either $k \in S$ or there is a $k' \in S$ such that $k' \leq_P k$.*

The on-line embedding Φ is illustrated by the following example:

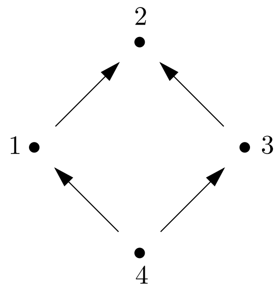


FIGURE 2. The partial order $([4], \leq_P)$.

EXAMPLE 2.5. The partial order $([4], \leq_P)$ depicted in Figure 2 has the following values of $\Psi([k], \leq_{P_k}), k = 1, 2, 3, 4$:

$$\begin{array}{lll} L([1], \leq_{P_1}) = \emptyset, & U([1], \leq_{P_1}) = \{0\}, & \Psi([1], \leq_{P_1}) = \{0\}, \\ L([2], \leq_{P_2}) = \{0\}, & U([2], \leq_{P_2}) = \{00, 10\}, & \Psi([2], \leq_{P_2}) = \{0, 10\}, \\ L([3], \leq_{P_3}) = \emptyset, & U([3], \leq_{P_3}) = \{000, 100\}, & \Psi([3], \leq_{P_3}) = \{000, 100\}, \\ L([4], \leq_{P_4}) = \emptyset, & U([4], \leq_{P_4}) = \{0000\}, & \Psi([4], \leq_{P_4}) = \{0000\}. \end{array}$$

PROOF (OF THEOREM 2.4). We proceed by induction on n.

The theorem obviously holds for $n = 1$.

Now assume that the theorem holds for every partial order $([i], \leq_{P_i})$, $i = 1, \dots, n-1$.

We first show that (2) holds for $([n], \leq_P)$. Fix $S \subseteq \{1, 2, \dots, n\}$. Without loss of generality assume that for each $m \leq n$ such that there is an $m' \in S$ with $m' \leq_P m$, we also have $m \in S$ (i.e. S is closed upwards). By the induction hypothesis, there is a word W of length $n-1$ such that for each $n' < n$, $\{W\} \leq_{\mathcal{W}} \Psi([n'], \leq_{P_{n'}})$ if and only if $n' \in S$. Given the word W we can construct a word W' of length n such that $\{W'\} \leq_{\mathcal{W}} \Psi([n'], \leq_{P_{n'}})$ if and only if $n' \in S$. To see this, consider the following cases:

(1) $n \in S$
 - (a) $\{W\} \leq_{\mathcal{W}} \Psi([n], \leq_P)$. Put $W' = W0$. Since $\{W'\} \leq_{\mathcal{W}} \{W\}$, W' obviously has the property.
 - (b) $\{W\} \not\leq_{\mathcal{W}} \Psi([n], \leq_P)$. In this case we have $m \in S$ for each $m < n, n \leq_P m$, and thus $\{W\} \leq_{\mathcal{W}} \Psi([m], \leq_{P_m})$. By the definition of $\leq_{\mathcal{W}}$, for each such m we have $W'' \in \Psi([m], \leq_{P_m})$ such that W'' is an initial segment of W. This implies that $W0$ is in $U([n], \leq_P)$ and thus $\{W\} \leq_{\mathcal{W}} \Psi([n], \leq_P)$, a contradiction.

(2) $n \notin S$
 - (a) $\{W\} \not\leq_{\mathcal{W}} \Psi([n], \leq_P)$. In this case we can put either $W' = W0$ or $W' = W1$.
 - (b) $\{W\} \leq_{\mathcal{W}} \Psi([n], \leq_P)$. We have $\{W\} \not\leq_{\mathcal{W}} L([n], \leq_P)$—otherwise we would have $\{W\} \leq_{\mathcal{W}} \Psi([m], \leq_{P_m}) \leq_{\mathcal{W}} \Psi([n], \leq_P)$ for some $m < n$ and thus $n \in S$. Since $U([n], \leq_P)$ contains words of length n whose last digit is 0 putting $W' = W1$ gives $\{W'\} \not\leq_{\mathcal{W}} U([n], \leq_P)$ and thus also $\{W'\} \not\leq_{\mathcal{W}} \Psi([m], \leq_{P_m})$.

This finishes the proof of property (2).

Now we prove (1). We only need to verify that for $m = 1, 2, \dots, n-1$ we have $\Psi([n], \leq_P) \leq_{\mathcal{W}} \Psi([m], \leq_{P_m})$ if and only if $n \leq_P m$ and $\Psi([m], \leq_{P_m}) \leq_{\mathcal{W}} \Psi([n], \leq_P)$ if and only if $m \leq_P n$. The rest follows by induction. Fix m and consider the following cases:

(1) *$m \leq_P n$ implies $\Psi([m], \leq_{P_m}) \leq_{\mathcal{W}} \Psi([n], \leq_P)$*: This follows easily from the fact that every word in $\Psi([m], \leq_{P_m})$ is in $L([n], \leq_P)$ and the initial segment of each word in $L([n], \leq_P)$ is in $\Psi([n], \leq_P)$.

(2) *$n \leq_P m$ implies $\Psi([n], \leq_P) \leq_{\mathcal{W}} \Psi([m], \leq_{P_m})$*: $U([n], \leq_P)$ is a maximal set of words of length n with last digit 0 such that $U([n], \leq_P) \leq_{\mathcal{W}} \Psi([m'], \leq_{P_{m'}})$ for each $m' < n, n \leq_P m'$, in particular for $m' = m$. It suffices to show that $L([n], \leq_P) \leq_{\mathcal{W}} \Psi([m], \leq_{P_m})$. For $W \in L([n], \leq_P)$, we have an m'', $m'' \leq_P n \leq_P m$, such that $W \in \Psi([m''], \leq_{P_{m''}})$. From the induction hypothesis $\Psi([m''], \leq_{P_{m''}}) \leq_{\mathcal{W}} \Psi([m], \leq_{P_m})$—in particular the initial segment of W is in $\Psi([m], \leq_{P_m})$.

(3) *$\Psi([m], \leq_{P_m}) \leq_{\mathcal{W}} \Psi([n], \leq_P)$ implies $m \leq_P n$*: Since $U([n], \leq_P)$ contains words longer than any word of m, we have $\Psi([m], \leq_{P_m}) \leq_{\mathcal{W}} L([n], \leq_P)$. By (2) for $S = \{m\}$ there is a word W such that $\{W\} \leq_{\mathcal{W}} \Psi([m'], \leq_{P_{m'}})$ if and only if $m \leq_P m'$. Since $\{W\} \leq_{\mathcal{W}} L([n], \leq_P)$, we have an m' such that $m \leq_P m' \leq_P n$.

(4) *$\Psi([n], \leq_P) \leq_{\mathcal{W}} \Psi([m], \leq_{P_m})$ implies $n \leq_P m$*: We have $\Psi([n], \leq_P) \leq_{\mathcal{W}} \Psi([m], \leq_{P_m})$. By (2) for $S = \{n\}$ there is a word W such that $\{W\} \leq_{\mathcal{W}} \Psi([m'], \leq_{P_{m'}})$ if and only if $n \leq_P m'$. Since $\{W\} \leq_{\mathcal{W}} \Psi([m], \leq_{P_m})$ we also have $n \leq_P m$.

□

Corollary 2.6. The partial order $(\mathcal{W}, \leq_{\mathcal{W}})$ is universal.

Note that $\mathcal{W}$ fails to be a ultrahomogeneous partial order. For example the empty set is the minimal element. $\mathcal{W}$ is also not dense as shown by the following example:

$$A = \{0\}, B = \{00, 01\}.$$

This is not unique gap—we shall characterize all gaps in $(\mathcal{W}, \leq_{\mathcal{W}})$ after reformulating it in a more combinatorial setting in Section 2.2.

2.2. Dominance in the countable binary tree. As is well known, the Hasse diagram of the partial order $(\{0,1\}^*, \leq_w)$ forms a complete binary tree T_u of infinite depth. Let r be its root vertex (corresponding to the empty word). Using T_u we can reformulate our universal partial order as:

Definition 2.7. The vertices of $(\mathcal{B}, \leq_{\mathcal{B}})$ are finite sets S of vertices of T_u such that there is no vertex $v \in S$ on any path from r to $v' \in S$ except for v'. (Thus S is a finite antichain in the order of the tree T.)

We say that $S' \leq_{\mathcal{B}} S$ if and only if for each path from r to $v \in S$ there is a vertex $v' \in S'$.

Corollary 2.8. The partially ordered set $(\mathcal{B}, \leq_{\mathcal{B}})$ is universal.

Proof. $(\mathcal{B}, \leq_{\mathcal{B}})$ is just a reformulation of $(\mathcal{W}, \leq_{\mathcal{W}})$ and thus both partial orders are isomorphic. □

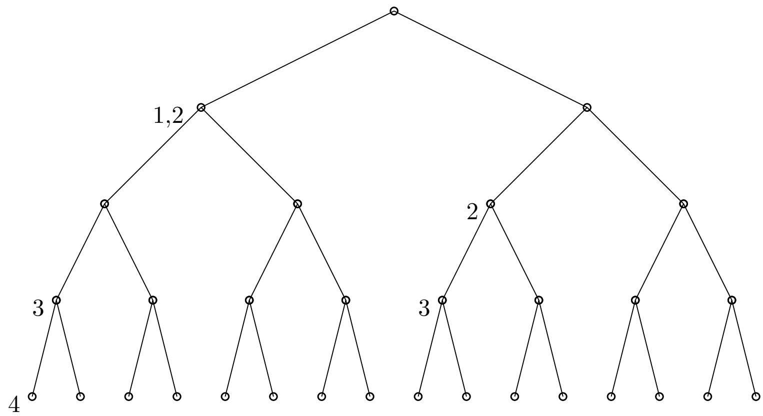


FIGURE 3. Tree representation of $([4], \leq_P)$ (Figure 2).

Figure 3 shows a portion of the tree T representing the same partial order as in Figure 2.

The partial order $(\mathcal{B}, \leq_{\mathcal{B}})$ offers perhaps a better intuitive understanding as to how the universal partial order is built from the very simple partial order $(\{0,1\}^*, \leq_w)$ by using sets of elements instead of single elements. Understanding this makes it easy to find an embedding of $(\mathcal{W}, \leq_{\mathcal{W}})$ (or equivalently $(\mathcal{B}, \leq_{\mathcal{B}})$) into a new structure by first looking for a way to represent the partial order $(\{0,1\}^*, \leq_w)$ within the new structure and then a way to represent subsets of $\{0,1\}^*$. This idea will be applied several times in the following sections.

Now we characterize gaps.

PROPOSITION 2.9. $S < S'$ is a gap in $(\mathcal{B}, \leq_{\mathcal{B}})$ if and only if there exists an $s' \in S'$ such that

(1) there is a vertex $s \in S$ such that both sons s_0, s_1 of s in the tree T are in S',
(2) $S \setminus \{s_0, s_1\} = S' \setminus \{s\}$.

This means that all gaps in $\mathcal{B}$ result from replacing a member by its two sons.

PROOF. Clearly any pair $S < S'$ satisfying (1), (2) is a gap (as any $S \leq_{\mathcal{B}} S'' \leq_{\mathcal{B}} S'$ has to contain $S' \setminus \{s\}$, and either s or the two vertices s_0, s_1).

Let $S \leq_{\mathcal{B}} S'$ be a gap. If there are distinct vertices s'_1 and s'_2 in S' and $s_1, s_2 \in S$ are such that $s_i \leq s'_i$, i=1,2, then S'' defined as $min(S \setminus \{s_1\}) \cup \{S'_1\}$ satisfies $S <_{\mathcal{B}} S'' <_{\mathcal{B}} S'$.

Thus there is only one $S' \in S' \setminus S$ such that $s' > s$ for an $s \in S$. However then there is only one such s' (so if s_1, s_2 are distinct then $S < S \setminus \{s_2\} < S'$). Moreover it is either $s = s'0$ or $s = s'1$. Otherwise $S < S'$ would not be a gap. $\square$

The abundance of gaps indicates that $(\mathcal{B}, \leq_{\mathcal{B}})$ (or $(\mathcal{W}, \leq_{\mathcal{W}})$) are redundant universal partial orders. This makes them, in a way, far from being generic, since the generic partial order has no gaps. The next section has a variant of this partial order avoiding this problem. On the other hand gaps in partial orders are interesting and are related to dualities, see [**21, 19**].

2.3. Intervals. We show that the vertices of $(\mathcal{W}, \leq_{\mathcal{W}})$ can be coded by geometric objects ordered by inclusion. Since we consider only countable structures we restrict ourselves to objects formed from rational numbers.

While the interval on rationals ordered by inclusion can represent infinite increasing chains, decreasing chains or antichains, obviously this interval order has dimension 2 and thus fails to be universal. However considering multiple intervals overcomes this limitation:

DEFINITION 2.10. The vertices of $(\mathcal{I}, \leq_{\mathcal{I}})$ are finite sets S of closed disjoint intervals $[a, b]$ where a, b are rational numbers and $0 \leq a < b \leq 1$.

We put $A \leq_{\mathcal{I}} B$ when every interval in A is covered by some interval of B.

In the other words elements of $(\mathcal{I}, \leq_{\mathcal{I}})$ are finite sets of pairs of rational numbers. $A \leq_{\mathcal{I}} B$ holds if for every $[a, b] \in A$, there is an $[a', b'] \in B$ such that $a' \leq a$ and $b \leq b'$.

DEFINITION 2.11. A word $W = w_1 w_2 \dots w_t$ on the alphabet $\{0, 1\}$ can be considered as a number $0 \leq n_W \leq 1$ with ternary expansion:

$$n_W = \sum_{i=1}^{t} w_i \frac{1}{3^i}.$$

For $A \in \mathcal{W}$, the representation of A in $\mathcal{I}$ is then the following set of intervals:

$$\Phi_{\mathcal{I}}^{\mathcal{W}}(A) = \{[n_W, n_W + \frac{2}{3^{|W|+1}}] : W \in A\}.$$

The use of the ternary base might seem unnatural—indeed the binary base would suffice. The main obstacle to using the later is that the embedding of $\{00, 01\}$ would be two intervals adjacent to each other overlapping in single point. One would need to take special care when taking the union of such intervals—we avoid this by using ternary numbers.

LEMMA 2.12. *$\Phi_{\mathcal{I}}^{\mathcal{W}}$ is an embedding of $(\mathcal{W}, \leq_{\mathcal{W}})$ into $(\mathcal{I}, \leq_{\mathcal{I}})$.*

PROOF. It is sufficient to prove that for W, W' there is an interval $[n_W, n_W + \frac{1}{3^{|W|}}]$ covered by an interval $[n_{W'}, n_{W'} + \frac{1}{3^{|W'|}}]$ if and only if W' is initial segment of W. This follows easily from the fact that intervals represent precisely all numbers whose ternary expansion starts with W with the exception of the upper bound itself. □

EXAMPLE 2.13. The representation of $([4], \leq_P)$ as defined by Figure 2 in $(\mathcal{I}, \leq_{\mathcal{I}})$ is:

$$\begin{array}{lclcl}
\Phi_{\mathcal{I}}^{\mathcal{W}}(\Psi([1], \leq_{P_1})) & = & \Phi_{\mathcal{I}}^{\mathcal{W}}(\{0\}) & = & \{(0, \frac{2}{3^2})\}, \\
\Phi_{\mathcal{I}}^{\mathcal{W}}(\Psi([2], \leq_{P_2})) & = & \Phi_{\mathcal{I}}^{\mathcal{W}}(\{0, 10\}) & = & \{(0, \frac{2}{3^2}), (\frac{1}{3}, \frac{1}{3} + \frac{2}{3^3})\}, \\
\Phi_{\mathcal{I}}^{\mathcal{W}}(\Psi([3], \leq_{P_3})) & = & \Phi_{\mathcal{I}}^{\mathcal{W}}(\{000, 100\}) & = & \{(0, \frac{2}{3^4}), (\frac{1}{3}, \frac{1}{3} + \frac{2}{3^4})\}, \\
\Phi_{\mathcal{I}}^{\mathcal{W}}(\Psi([4], \leq_{P_4})) & = & \Phi_{\mathcal{I}}^{\mathcal{W}}(\{0000\}) & = & \{(0, \frac{2}{3^5})\}.
\end{array}$$

COROLLARY 2.14. The partial order $(\mathcal{I}, \leq_{\mathcal{I}})$ is universal.

The partial order $(\mathcal{I}, \leq_{\mathcal{I}})$ differs significantly from $(\mathcal{W}, \leq_{\mathcal{W}})$ by the following:

PROPOSITION 2.15. The partial order $(\mathcal{I}, \leq_{\mathcal{I}})$ has no gaps (is dense).

Proof. Take $A, B \in \mathcal{I}$, $A <_{\mathcal{I}} B$. Because all the intervals in both A and B are closed and disjoint, there must be at least one interval I in B that is not fully covered by intervals of A (otherwise we would have $B \leq_{\mathcal{I}} A$). We may construct an element C from B by shortening the interval I or splitting it into two disjoint intervals in a way such that $A <_{\mathcal{I}} C <_{\mathcal{I}} B$ holds. $\square$

Consequently the presence (and abundance) of gaps in most of the universal partial orders studied is not the main obstacle when looking for representations of partial orders. It is easy to see that $(\mathcal{I}, \leq_{\mathcal{I}})$ is not generic.

By considering a variant of $(\mathcal{I}, \leq_{\mathcal{I}})$ with open (instead of closed) intervals we obtain a universal partial order $(\mathcal{I}', \leq_{\mathcal{I}'})$ with gaps. The gaps are similar to the ones in $(\mathcal{B}, \leq_{\mathcal{B}})$ created by replacing interval (a, b) by two intervals (a, c) and (c, d). Half open intervals give a quasi-order containing a universal partial order.

2.4. Geometric representations. The representation as a set of intervals might be considered an artificially constructed structure. Partial orders represented by geometric objects are studied in [**1**]. It is shown that objects with n "degrees of freedom" cannot represent all partial orders of dimension $n+1$. It follows that convex hulls used in the representation of the generic partial order cannot be defined by a constant number of vertices. We will show that even the simplest geometric objects with unlimited "degrees of freedom" can represent a universal partial order.

Definition 2.16. Denote by $(\mathcal{C}, \leq_{\mathcal{C}})$ the partial order whose vertices are all convex hulls of finite sets of points in $\mathbb{Q}^2$, ordered by inclusion.

This time we will embed $(\mathcal{I}, \leq_{\mathcal{I}})$ into $(\mathcal{C}, \leq_{\mathcal{C}})$.

Definition 2.17. For every $A \in \mathcal{I}$ denote by $\Phi^{\mathcal{I}}_{\mathcal{C}}(A)$ the convex hull generated by the points:

$$(a, a^2), (\frac{a+b}{2}, ab), (b, b^2), \text{ for every } (a, b) \in A.$$

See Figure 4 for the representation of the partial order in Figure 2.

Theorem 2.18. *$\Phi^{\mathcal{I}}_{\mathcal{C}}$ is an embedding of $(\mathcal{I}, \leq_{\mathcal{I}})$ to $(\mathcal{C}, \leq_{\mathcal{C}})$.*

Proof. All points of the form (x, x^2) lie on a convex parabola $y = x^2$. The points $(\frac{a+b}{2}, ab)$ are the intersection of two tangents of this parabola at the points (a, a^2) and (b, b^2). Consequently all points in the construction of $\Phi^{\mathcal{I}}_{\mathcal{C}}(A)$ lie in a convex configuration.

We have (x, x^2) in the convex hull $\Phi^{\mathcal{I}}_{\mathcal{C}}(A)$ if and only if there is $[a, b] \in A$ such that $a \leq x \leq b$. Thus for $A, B \in \mathcal{I}$ we have $\Phi^{\mathcal{I}}_{\mathcal{C}}(A) \leq_{\mathcal{C}} \Phi^{\mathcal{I}}_{\mathcal{C}}(B)$ implies $A \leq_{\mathcal{I}} B$.

To see the other implication, observe that the convex hull of (a, a^2), $(\frac{a+b}{2}, ab)$, (b, b^2) is a subset of the convex hull of (a', a'^2), $(\frac{a'+b'}{2}, a'b')$, (b', b'^2) for every $[a, b]$ that is a subinterval of $[a', b']$. $\square$

We have:

Corollary 2.19. The partial order $(\mathcal{C}, \leq_{\mathcal{C}})$ is universal.

Remark 2.20. Our construction is related to Venn diagrams. Consider the partial order $([n], \leq_P)$. For the empty relation $\leq_P$ the representation constructed by $\Phi^{\mathcal{I}}_{\mathcal{C}}(\Phi^{\mathcal{W}}_{\mathcal{I}}(\Psi([n], \emptyset)))$ is a Venn diagram, by Theorem 2.4 (2). Statement 2 of Theorem 2.4 can be seen as a Venn diagram condition under the constraints imposed by $\leq_P$.

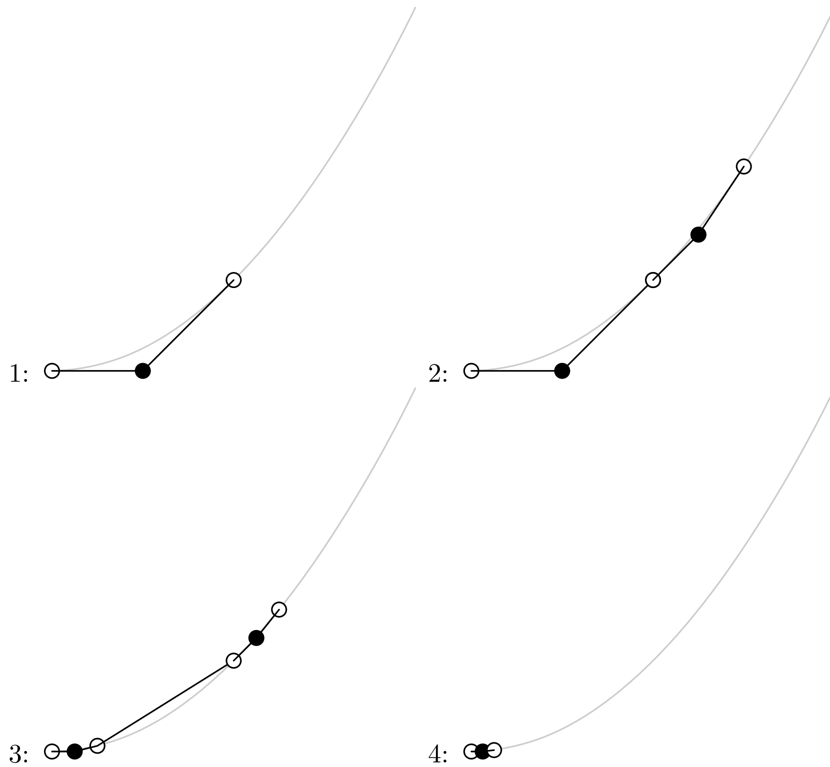


FIGURE 4. Representation of the partial order $([4], \leq_P)$ in $(\mathcal{C}, \leq_\mathcal{C})$.

The same construction can be applied to functions, and stated in a perhaps more precise manner.

COROLLARY 2.21. Consider the class $\mathcal{F}$ of all convex piecewise linear functions on the interval $(0, 1)$ consisting of a finite set of segments, each with rational boundaries. Put $f \leq_\mathcal{F} g$ if and only if $f(x) \leq g(x)$ for every $0 \leq x \leq 1$. Then the partial order $(\mathcal{F}, \leq_\mathcal{F})$ is universal.

Similarly the following holds:

THEOREM 2.22. *Denote by $\mathcal{O}$ the class of all finite polynomials with rational coefficients. For $p, q \in \mathcal{O}$, put $p \leq_\mathcal{O} q$ if and only if $p(x) \leq q(x)$ for $x \in (0, 1)$. The partial order $(\mathcal{O}, \leq_\mathcal{O})$ is universal.*

The proof of this theorem needs tools of mathematical analysis The proof of this theorem needs more involved tools of mathematical analysis and it will appear elsewhere (jointly with Robert Šámal).

2.5. Grammars. The rewriting rules used in a context-free grammar can be also used to define a universal partially ordered set.

DEFINITION 2.23. The vertices of $(\mathcal{G}, \leq_{\mathcal{G}})$ are all words over the alphabet $\{\downarrow, \uparrow, 0, 1\}$ created from the word 1 by the following rules:

$$\begin{array}{rcl} 1 & \to & \downarrow 11 \uparrow, \\ 1 & \to & 0. \end{array}$$

$W \leq_{\mathcal{G}} W'$ if and only if W can be constructed from W' by:

$$\begin{array}{rcl} 1 & \to & \downarrow 11 \uparrow, \\ 1 & \to & 0, \\ \downarrow 00 \uparrow & \to & 0. \end{array}$$

$(\mathcal{G}, \leq_{\mathcal{G}})$ is a quasi-order: the transitivity of $\leq_{\mathcal{G}}$ follows from the composition of lexical transformations.

DEFINITION 2.24. Given $A \in \mathcal{W}$ construct $\Phi_{\mathcal{G}}^{\mathcal{W}}$ as follows:

(1) $\Phi_{\mathcal{G}}^{\mathcal{W}}(\emptyset) = 0$.
(2) $\Phi_{\mathcal{G}}^{\mathcal{W}}(\{\text{empty word}\}) = 1$.
(3) $\Phi_{\mathcal{G}}^{\mathcal{W}}(A)$ is defined as the concatenation $\downarrow \Phi_{\mathcal{G}}^{\mathcal{W}}(A_0)\Phi_{\mathcal{G}}^{\mathcal{W}}(A_1) \uparrow$, where A_0 is created from all words of A starting with 0 with the first digit removed and A_1 is created from all words of A starting with 1 with the first digit removed.

EXAMPLE 2.25. The representation of $([4], \leq_P)$ as defined by Figure 2 in $(\mathcal{G}, \leq_{\mathcal{G}})$ is as follows (see also the correspondence with the $\mathcal{B}$ representation in Figure 3):

$$\begin{array}{lclcl} \Phi_{\mathcal{G}}^{\mathcal{W}}(\Psi([1], \leq_{P_1})) & = & \Phi_{\mathcal{G}}^{\mathcal{W}}(\{0\}) & = & \downarrow 10 \uparrow, \\ \Phi_{\mathcal{G}}^{\mathcal{W}}(\Psi([2], \leq_{P_2})) & = & \Phi_{\mathcal{G}}^{\mathcal{W}}(\{0, 10\}) & = & \downarrow 1 \downarrow 10 \uparrow \uparrow, \\ \Phi_{\mathcal{G}}^{\mathcal{W}}(\Psi([3], \leq_{P_3})) & = & \Phi_{\mathcal{G}}^{\mathcal{W}}(\{000, 100\}) & = & \downarrow \downarrow \downarrow 10 \uparrow 0 \uparrow \downarrow \downarrow 10 \uparrow 0 \uparrow \uparrow, \\ \Phi_{\mathcal{G}}^{\mathcal{W}}(\Psi([4], \leq_{P_4})) & = & \Phi_{\mathcal{G}}^{\mathcal{W}}(\{0000\}) & = & \downarrow \downarrow \downarrow \downarrow 10 \uparrow 0 \uparrow 0 \uparrow 0 \uparrow. \end{array}$$

We state the following without proof as it follows straightforwardly from the definitions.

PROPOSITION 2.26. For $A, B \in \mathcal{W}$ the inequality $A \leq_{\mathcal{W}} B$ holds if and only if $\Phi_{\mathcal{G}}^{\mathcal{W}}(A) \leq_{\mathcal{G}} \Phi_{\mathcal{G}}^{\mathcal{W}}(B)$.

$(\mathcal{G}, \leq_{\mathcal{G}})$ is a quasi-order. We have:

COROLLARY 2.27. The quasi-order $(\mathcal{G}, \leq_{\mathcal{G}})$ contains a universal partial order.

2.6. Multicuts and Truncated Vectors. A universal partially ordered structure similar to $(\mathcal{W}, \leq_{\mathcal{W}})$, but less suitable for further embeddings, was studied in [**5, 16, 10**]. While the structures defined in these papers are easily shown to be equivalent, their definition and motivations were different. [**5**] contains the first finite presentation of universal partial order. [**16**] first used the notion of on-line embeddings to (1) prove the universality of the structure and (2) as intermediate structure to prove the universality of the homomorphism order of multigraphs. The motivation for this structure came from the analogy with Dedekind cuts and thus its members were called *multicuts*. In [**10**] an essentially equivalent structure with the inequality reversed was used as an intermediate structure for the stronger result showing the universality of oriented paths. This time the structure arises in the context of orders of vectors (as the simple extension of the orders of finite dimension represented by finite vectors of rationals) resulting in name *truncated vectors*.

We follow the presentation in [**10**].

Definition 2.28. Let $\vec{v} = (v_1, \dots, v_t)$, $\vec{v}' = (v'_1, \dots, v'_{t'})$ be 0–1 vectors. We put:

$$\vec{v} \leq_{\vec{v}} \vec{v}' \text{ if and only if } t \geq t' \text{ and } v_i \geq v'_i \text{ for } i = 1, \dots, t'.$$

Thus we have e.g. $(1,0,1,1,1) <_{\vec{v}} (1,0,0,1)$ and $(1,0,0,1) >_{\vec{v}} (1,1,1,1)$. An example of an infinite descending chain is e.g.

$$(1) >_{\vec{v}} (1,1) >_{\vec{v}} (1,1,1) >_{\vec{v}} \dots.$$

Any finite partially ordered set is representable by vectors with this ordering: for vectors of a fixed length we have just the reverse ordering of that used in the (Dushnik-Miller) dimension of partially ordered sets, see e.g. [**21**].

Definition 2.29. We denote by $\mathcal{T}V$ the class of all finite vector-sets. Let $\vec{V}$ and $\vec{V}'$ be two finite sets of 0–1 vectors. We put $\vec{V} \leq_{\mathcal{T}V} \vec{V}'$ if and only if for every $\vec{v} \in \vec{V}$ there exists a $\vec{v}' \in \vec{V}'$ such that $\vec{v} \leq_{\vec{v}} \vec{v}'$.

For a word W on the alphabet $\{0,1\}$ we construct a vector $\vec{v}(W)$ of length $2|W|$ such that $2n$-th element of vector $\vec{v}(W)$ is 0 if and only if the n-th character of W is 0, and the $(2n+1)$-th element of the vector $\vec{v}(W)$ is 1 if and only if the n-th character of W is 0.

It is easy to see that $W \leq_{\mathcal{W}} W'$ if and only if $\vec{v}(W) \leq_{\vec{v}} \vec{v}(W')$. The embedding $\Phi^{\mathcal{W}}_{\mathcal{T}V} : (\mathcal{W}, \leq_{\mathcal{W}}) \to (\mathcal{T}V, \leq_{\mathcal{T}V})$ is constructed as follows:

$$\Phi^{\mathcal{W}}_{\mathcal{T}V} = \{\vec{v}(W), W \in A\}.$$

For our example $([4], \leq_P)$ in Figure 2 we have embedding:

$$\begin{array}{lclcl}
\Phi^{\mathcal{W}}_{\mathcal{T}V}(\Psi([1], \leq_{P_1})) & = & \Phi^{\mathcal{W}}_{\mathcal{T}V}(\{0\}) & = & \{(0,1)\}, \\
\Phi^{\mathcal{W}}_{\mathcal{T}V}(\Psi([2], \leq_{P_2})) & = & \Phi^{\mathcal{W}}_{\mathcal{T}V}(\{0,10\}) & = & \{(0,1),(1,0,0,1)\}, \\
\Phi^{\mathcal{W}}_{\mathcal{T}V}(\Psi([3], \leq_{P_3})) & = & \Phi^{\mathcal{W}}_{\mathcal{T}V}(\{000,100\}) & = & \{(0,1,0,1,0,1),(1,0,0,1,0,1)\}, \\
\Phi^{\mathcal{W}}_{\mathcal{T}V}(\Psi([4], \leq_{P_4})) & = & \Phi^{\mathcal{W}}_{\mathcal{T}V}(\{0000\}) & = & \{(0,1,0,1,0,1,0,1)\}.
\end{array}$$

Corollary 2.30. The quasi-order $(\mathcal{T}V, \leq_{\mathcal{T}V})$ contains a universal partial order.

The structure $(\mathcal{T}V, \leq_{\mathcal{T}V})$ as compared to $(\mathcal{W}, \leq_{\mathcal{W}})$ is more complicated to use for further embeddings: the partial order of vectors is already a complex finite-universal partial order. The reason why the structure $(\mathcal{T}V, \leq_{\mathcal{T}V})$ was discovered first is is that it allows a remarkably simple on-line embedding that we outline now.

Again we restrict ourselves to the partial orders whose vertex sets are the sets $[n]$ (for some $n > 1$) and we will always embed the vertices in the natural order. The function Ψ' mapping partial orders $([n], \leq_P)$ to elements of $(\mathcal{T}V, \leq_{\mathcal{T}V})$ is defined as follows:

Definition 2.31. Let $\vec{v}([n], \leq_P) = (v_1,\, v_2,\, \dots,\, v_n)$ where $v_m = 1$ if and only if $n \leq_{\mathcal{P}} m$, $m \leq n$, otherwise $v_m = 0$.

Let

$$\Psi'([n], \leq_P) = \{\vec{v}([m], \leq_{P_m}) : m \in P, m \leq n, m \leq_{\mathcal{P}} n\}.$$

For our example in Figure 2 we get a different (and more compact) embedding:

$$\begin{array}{lclclcl}
\vec{v}(1) & = & (1), & \Psi'([1], \leq_{P_1}) & = & \{(1)\}, \\
\vec{v}(2) & = & (0,1), & \Psi'([2], \leq_{P_2}) & = & \{(1),(0,1)\}, \\
\vec{v}(3) & = & (1,0,1), & \Psi'([3], \leq_{P_3}) & = & \{(1,0,1)\}, \\
\vec{v}(4) & = & (1,1,1,1), & \Psi'([4], \leq_{P}) & = & \{(1,1,1,1)\}.
\end{array}$$

THEOREM 2.32. *Fix the partial order* $([n], \leq_P)$. *For every* $i, j \in [n]$,

$$i \leq_P j \text{ if and only if } \Psi'([i], \leq_{P_i}) \leq_{\mathcal{T}V} \Psi'([j], \leq_{P_j})$$

and

$$\Psi'([i], \leq_{P_i}) = \Psi'([j], \leq_{P_j}) \text{ if and only if } i = j.$$

(Or in the other words, the mapping $\Phi'(i) = \Psi'([i], \leq_{P_i})$ *is the embedding of* $([n], \leq_P)$ *into* $(\mathcal{T}V, \leq_{\mathcal{T}V})$*).*

The proof can be done via induction analogously as in the second part of the proof of Theorem 2.4. See our paper [**10**]. The main advantage of this embedding is that the size of the answer is $O(n^2)$ instead of $O(2^n)$.

2.7. Periodic sets. As the last finite presentation we mention the following what we believe to be very elegant description. Consider the partial order defined by inclusion on sets of integers. This partial order is uncountable and contains every countable partial order. We can however show the perhaps surprising fact that the subset of all periodic subsets (which has a very simple and finite description) is countably universal.

DEFINITION 2.33. $S \subseteq \mathbb{Z}$ is *p-periodic* if for every $x \in S$ we have also $x + p \in S$ and $x - p \in S$.

For a periodic set S with period p denote by the *signature* $s(p, S)$ a word over the alphabet $\{0, 1\}$ of length p such that n-th letter is 1 if and only if $n \in S$.

By $\mathcal{S}$ we denote the class of all sets $S \subseteq \mathbb{Z}$ such that S is 2^n-periodic for some n.

Clearly every periodic set is determined by its signature and thus $(\mathcal{S}, \subseteq)$ is a finite presentation. We consider the ordering of periodic sets by inclusion and prove:

THEOREM 2.34. *The partial order* $(\mathcal{S}, \subseteq)$ *is universal.*

PROOF. We embed $(\mathcal{W}, \leq_{\mathcal{W}})$ into $(\mathcal{S}, \subseteq)$ as follows: For $A \in \mathcal{W}$ denote by $\Phi_{\mathcal{S}}^{\mathcal{W}}(A)$ the set of integers such that $n \in \Phi_{\mathcal{S}}^{\mathcal{W}}(A)$ if and only if there is $W \in A$ and the $|A|$ least significant digits of the binary expansion of n forms a reversed word W (when the binary expansion has fewer than $|W|$ digits, add 0 as needed).

It is easy to see that $\Phi_{\mathcal{S}}^{\mathcal{W}}(A)$ is 2^n-periodic, where n is the length of longest word in W, and $\Phi_{\mathcal{S}}^{\mathcal{W}}(A) \subseteq \Phi_{\mathcal{S}}^{\mathcal{W}}(A')$ if and only if $A \leq_{\mathcal{W}} A'$. □

$(\mathcal{S}, \subseteq)$ is dense, but it fails to have the 3-extension property: there is no set strictly smaller than the set with signature 01 and greater than both sets with signatures 0100 and 0010.

3. Generic Poset and Conway numbers

One of the striking (and concise) incarnations of the generic Rado graph is provided by the set theory: vertices of $\mathcal{R}$ are all sets in a fixed countable model $\mathfrak{M}$ of the theory of finite sets, and the edges correspond to pairs $\{A, B\}$ for which either $A \in B$ or $B \in A$. In [**9**] we aimed for a similarly concise representation of a generic partial order. That appeared to be a difficult task and we had to settle for the weaker notion of "finite presentation". At present [**9**] is the only finite presentation of the generic partial order. This is related to Conway surreal numbers [**12, 2**].

In this section, for completeness we give the finite presentation of the generic partial order as shown in [**9**]. This construction is of independent interest as one can give a finite presentation of the rational Urysohn space along the same lines [**7**]. We work in a fixed countable model $\mathfrak{M}$ of the theory of finite sets extended by a single atomic set $\acute{O}$. To represent ordered pairs (M_L, M_R), we use following notation:

$$M_L = \{A; A \in M, \acute{O} \notin A\};$$
$$M_R = \{A; (A \cup \{\acute{O}\}) \in M, \acute{O} \notin A\}.$$

Definition 3.1. Define the partially ordered set $(\mathcal{P}_\in, \leq_\in)$ as follows:
The elements of $\mathcal{P}_\in$ are all sets M with the following properties:

(1) (correctness)
  (a) $\acute{O} \notin M$,
  (b) $M_L \cup M_R \subset \mathcal{P}_\in$,
  (c) $M_L \cap M_R = \emptyset$.
(2) (ordering property) $(\{A\} \cup A_R) \cap (\{B\} \cup B_L) \neq \emptyset$ for each $A \in M_L, B \in M_R$,
(3) (left completeness) $A_L \subseteq M_L$ for each $A \in M_L$,
(4) (right completeness) $B_R \subseteq M_R$ for each $B \in M_R$.

The relation of $\mathcal{P}_\in$ is denoted by $\leq_\in$ and it is defined as follows: We put $M <_\in N$ if

$$(\{M\} \cup M_R) \cap (\{N\} \cup N_L) \neq \emptyset.$$

We write $M \leq_\in N$ if either $M <_\in N$ or $M = N$.

The class $\mathcal{P}_\in$ is non-empty (as $M = \emptyset = (\,\emptyset \mid \emptyset\,) \in \mathcal{P}_\in$). (Obviously the correctness property holds. Since $M_L = \emptyset$, $M_R = \emptyset$, the ordering property and completeness properties follow trivially.)

Here are a few examples of non-empty elements of the structure $\mathcal{P}_\in$:

$$(\,\emptyset \mid \emptyset\,),$$
$$(\,\emptyset \mid \{(\,\emptyset \mid \emptyset\,)\}\,),$$
$$(\,\{(\,\emptyset \mid \emptyset\,), (\,\emptyset \mid \{(\,\emptyset \mid \emptyset\,)\}\,)\} \mid \emptyset\,).$$

It is a non-trivial fact that $(\mathcal{P}_\in, \leq_\in)$ is a partially ordered set. This will be proved after introducing some auxiliary notions:

Definition 3.2. Any element $W \in (A \cup A_R) \cap (B \cup B_L)$ is called a *witness* of the inequality $A <_\in B$.

Definition 3.3. The *level of* $A \in \mathcal{P}_\in$ is defined as follows:

$$\begin{array}{rcl} l(\emptyset) & = & 0, \\ l(A) & = & max(l(B) : B \in A_L \cup A_R) + 1 \text{ for } A \neq \emptyset. \end{array}$$

We observe the following facts (which follow directly from the definition of $\mathcal{P}_\in$):

Fact 3.4. $X <_\in A <_\in Y$ for every $A \in \mathcal{P}_\in$, $X \in A_L$ and $Y \in A_R$.

Fact 3.5. $A \leq_\in W^{AB} \leq_\in B$ for any $A <_\in B$ and witness W^{AB} of $A <_\in B$.

Fact 3.6. Let $A <_\in B$ and let W^{AB} be a witness of $A <_\in B$. Then $l(W^{AB}) \leq \min(l(A), l(B))$, and either $l(W^{AB}) < l(A)$ or $l(W^{AB}) < l(B)$.

First we prove transitivity.

LEMMA 3.7. *The relation $\leq_\in$ is transitive on the class $\mathcal{P}_\in$.*

PROOF. Assume that three elements A, B, C of $\mathcal{P}_\in$ satisfy $A <_\in B <_\in C$. We prove that $A <_\in C$ holds. Let W^{AB} and W^{BC} be witnesses of the inequalities $A <_\in B$ and $B <_\in C$ respectively. First we prove that $W^{AB} \leq_\in W^{BC}$. We distinguish four cases (depending on the definition of the witness):

(1) $W^{AB} \in B_L$ and $W^{BC} \in B_R$.
In this case it follows from Fact 3.4 that $W^{AB} <_\in W^{BC}$.

(2) $W^{AB} = B$ and $W^{BC} \in B_R$.
Then W^{BC} is a witness of the inequality $B <_\in W^{BC}$ and thus $W^{AB} <_\in W^{BC}$.

(3) $W^{AB} \in B_L$ and $W^{BC} = B$.
The inequality $W^{AB} \leq_\in W^{BC}$ follows analogously to the previous case.

(4) $W^{AB} = W^{BC} = B$ (and thus $W^{AB} \leq_\in W^{BC}$).

In the last case B is a witness of the inequality $A <_\in C$. Thus we may assume that $W^{AB} \neq_\in W^{BC}$. Let W^{AC} be a witness of the inequality $W^{AB} <_\in W^{BC}$. Finally we prove that W^{AC} is a witness of the inequality $A <_\in C$. We distinguish three possibilities:

(1) $W^{AC} = W^{AB} = A$.
(2) $W^{AC} = W^{AB}$ and $W^{AC} \in A_R$.
(3) $W^{AC} \in W^{AB}_R$, then also $W^{AC} \in A_R$ from the completeness property.

It follows that either $W^{AC} = A$ or $W^{AC} \in A_R$. Analogously either $W^{AC} = C$ or $W^{AC} \in C_L$ and thus W^{AC} is the witness of inequality $A <_\in C$. □

LEMMA 3.8. *The relation $<_\in$ is strongly antisymmetric on the class $\mathcal{P}_\in$.*

PROOF. Assume that $A <_\in B <_\in A$ is a counterexample with minimal $l(A) + l(B)$. Let W^{AB} be a witness of the inequality $A <_\in B$ and W^{BA} a witness of the reverse inequality. From Fact 3.5 it follows that $A \leq_\in W^{AB} \leq_\in B \leq_\in W^{BA} \leq_\in A \leq_\in W^{AB}$. From the transitivity we know that $W^{AB} \leq_\in W^{BA}$ and $W^{BA} \leq_\in W^{AB}$.

Again we consider 4 possible cases:

(1) $W^{AB} = W^{BA}$.
From the disjointness of the sets A_L and A_R it follows that $W^{AB} = W^{BA} = A$. Analogously we obtain $W^{AB} = W^{BA} = B$, which is a contradiction.

(2) Either $W^{AB} = A$ and $W^{BA} = B$ or $W^{AB} = B$ and $W^{BA} = A$.
Then a contradiction follows in both cases from the fact that $l(A) < l(B)$ and $l(B) < l(A)$ (by Fact 3.6).

(3) $W^{AB} \neq A$, $W^{AB} \neq B$, $W^{AB} \neq W^{BA}$.
Then $l(W^{AB}) < l(A)$ and $l(W^{AB}) < l(B)$. Additionally we have $l(W^{BA}) \leq l(A)$ and $l(W^{BA}) \leq l(B)$ and thus A and B is not a minimal counter example.

(4) $W^{BA} \neq A$, $W^{BA} \neq B$, $W^{AB} \neq W^{BA}$.
The contradiction follows symmetrically to the previous case from the minimality of $l(A) + l(B)$.

□

THEOREM 3.9. *$(\mathcal{P}_\in, \leq_\in)$ is a partially ordered set.*

PROOF. Reflexivity of the relation follow directly from the definition, transitivity and antisymmetry follow from Lemmas 3.7 and 3.8. □

Now we are ready to prove the main result of this section:

THEOREM 3.10. $(\mathcal{P}_\in, \leq_\in)$ *is the generic partially ordered set for the class of all countable partial orders.*

First we show the following lemma:

LEMMA 3.11. $(\mathcal{P}_\in, \leq_\in)$ *has the extension property.*

PROOF. Let M be a finite subset of the elements of $\mathcal{P}_\in$. We want to extend the partially ordered set induced by M by the new element X. This extension can be described by three subsets of M: M_- containing elements smaller than X, M_+ containing elements greater than X, and M_0 containing elements incomparable with X. Since the extended relation is a partial order we have the following properties of these sets:

I. Any element of M_- is strictly smaller than any element of M_+,
II. $B \leq_\in A$ for no $A \in M_-$, $B \in M_0$,
III. $A \leq_\in B$ for no $A \in M_+$, $B \in M_0$,
IV. M_-, M_+ and M_0 form a partition of M.

Put

$$\overline{M_-} = \bigcup_{B \in M_-} B_L \cup M_-,$$

$$\overline{M_+} = \bigcup_{B \in M_+} B_R \cup M_+.$$

We verify that the properties I., II., III., IV. still hold for sets $\overline{M_-}$, $\overline{M_+}$, M_0.

ad I. We prove that any element of $\overline{M_-}$ is strictly smaller than any element of $\overline{M_+}$:

Let $A \in \overline{M_-}, A' \in \overline{M_+}$. We prove $A <_\in A'$. By the definition of $\overline{M_-}$ there exists $B \in M_-$ such that either $A = B$ or $A \in B_L$. By the definition of $\overline{M_+}$ there exists $B' \in M_+$ such that either $A' = B'$ or $A' \in B'_R$. By the definition of $<_\in$ we have $A \leq_\in B$, $B <_\in B'$ (by I.) and $B' \leq_\in A'$ again by the definition of $<_\in$. It follows $A <_\in A'$.

ad II. We prove that $B \leq_\in A$ for no $A \in \overline{M_-}$, $B \in M_0$:

Let $A \in \overline{M_-}, B \in M_0$ and let $A' \in M_-$ satisfy either $A = A'$ or $A \in A'_L$. We know that $B \not\leq_\in A'$ and as $A \leq_\in A'$ we have also $B \not\leq_\in A$.

ad III. To prove that $A \leq_\in B$ for no $A \in \overline{M_+}$, $B \in M_0$ we can proceed similarly to ad II.

ad IV. We prove that $\overline{M_-}$, $\overline{M_+}$ and M_0 are pairwise disjoint:

$\overline{M_-} \cap \overline{M_+} = \emptyset$ follows from I. $\overline{M_-} \cap M_0 = \emptyset$ follows from II. $\overline{M_+} \cap M_0 = \emptyset$ follows from III.

It follows that $A = (\,\overline{M_-} \mid \overline{M_+}\,)$ is an element of $\mathcal{P}_\in$ with the desired inequalities for the elements in the sets M_- and M_+.

Obviously each element of M_- is smaller than A and each element of M_+ is greater than A.

It remains to be shown that each $N \in M_0$ is incomparable with A. However we run into a problem here: it is possible that $A = N$. We can avoid this problem by first considering the set:

$$M' = \bigcup_{B \in M} B_R \cup M.$$

It is then easy to show that $B = (\, \emptyset \mid M' \,)$ is an element of $\mathcal{P}_\in$ strictly smaller than all elements of M.

Finally we construct the set $A' = (\, A_L \cup \{B\} \mid A_R \,)$. The set A' has the same properties with respect to the elements of the sets M_- and M_+ and differs from any set in M_0. It remains to be shown that A' is incomparable with N.

For contrary, assume for example, that $N <_\in A'$ and $W^{NA'}$ is the witness of the inequality. Then $W^{NA'} \in \overline{M_-}$ and $N \leq_\in W^{NA'}$. Recall that $N \in M_0$. From IV. above and the definition of A' it follows that $N <_\in W^{NA'}$. From *ad* III. above it follows that there is no choice of elements with $N <_\in W^{NA'}$, a contradiction.

The case $N >_\in A'$ is analogous. □

Proof. Proof of Theorem 3.10 follows by combining Lemma 3.11 and fact that extension property imply both universality and ultrahomogeneity of the partial order. □

Example 3.12. Consider partial order $(P, \leq_P)$ depicted in Figure 2. The function c embedding $(P, \leq_P)$ to $(\mathcal{P}_\in, \leq_{\mathcal{P}_\in})$ can be defined as:

$$\begin{aligned}
c(1) &= (\, \emptyset \mid \emptyset \,),\\
c(2) &= (\, \emptyset \mid \{(\, \emptyset \mid \emptyset \,)\} \,),\\
c(3) &= (\, \{(\, \emptyset \mid \emptyset \,), (\, \emptyset \mid \{(\, \emptyset \mid \emptyset \,)\} \,)\} \mid \emptyset \,),\\
c(4) &= (\, \{(\, \emptyset \mid \{(\, \emptyset \mid \emptyset \,)\} \,)\} \mid \{(\, \{(\, \emptyset \mid \emptyset \,), (\, \emptyset \mid \{(\, \emptyset \mid \emptyset \,)\} \,)\} \mid \emptyset \,)\} \,).
\end{aligned}$$

3.1. Remark on Conway's surreal numbers. Recall the definition of surreal numbers, see [**12**]. (For a recent generalization see [**3**]). Surreal numbers are defined recursively together with their linear order. We briefly indicate how the partial order $(\mathcal{P}_\in, \leq_{\mathcal{P}_\in})$ fits into this scheme.

Definition 3.13. A surreal number is a pair $x = \{x^L | x^R\}$, where every member of the sets x^L and x^R is a surreal number and every member of x^L is strictly smaller than every member of x^R.

We say that a surreal number x is less than or equal to the surreal number y if and only if y is not less than or equal to any member of x^L and any member of y^R is not less than or equal to x.

We will denote the class of surreal numbers by $\mathbb{S}$.

$\mathcal{P}_\in$ may be thought of as a subset of $\mathbb{S}$ (we recursively add $\circlearrowleft$ to express pairs x^L, x^R). The recursive definition of $A \in \mathcal{P}_\in$ leads to the following order which we define explicitly:

Definition 3.14. For elements $A, B \in \mathcal{P}_\in$ we write $A \leq_\mathbb{S} B$, when there is no $l \in A_L$ such that $B \leq_\mathbb{S} l$ and no $r \in B_R$ such that $r \leq_\mathbb{S} A$.

$\leq_\mathbb{S}$ is a linear order of $\mathcal{P}_\in$ and it is the restriction of Conway's order. It is in fact a linear extension of the partial order $(\mathcal{P}_\in, \leq_\in)$:

Theorem 3.15. *For any $A, B \in \mathcal{P}_\in$, $A <_\in B$ implies $A <_\mathbb{S} B$.*

PROOF. We proceed by induction on $l(A) + l(B)$.

For empty A and B the theorem holds as they are not comparable by $<_{\in}$.

Let $A <_{\in} B$ with W^{AB} as a witness. If $W^{AB} \neq A, B$, then $A <_{\mathbb{S}} W^{AB} <_{\mathbb{S}} B$ by induction. In the case $A \in B_L$, then $A <_{\mathbb{S}} B$ from the definition of $<_{\mathbb{S}}$. $\square$

4. Universality of Graph Homomorphisms

Perhaps the most natural order between finite models is induced by homomorphisms. The universality of the homomorphism order for the class of all finite graphs was first shown by [**20**].

Numerous other classes followed (see e. g. [**20**]) but planar graphs (and other topologically restricted classes) presented a problem.

The homomorphism order on the class of finite paths was studied in [**19**]. It has been proved it is a dense partial order (with the exception of a few gaps which were characterized; these gaps are formed by all core-path of height ≤ 4). [**19**] also rises (seemingly too ambitious) question whether it is a universal partial order. This has been resolved in [**8, 10**] by showing that finite oriented paths with homomorphism order are universal. In this section we give a new proof of this result. The proof is simpler and yields a stronger result (see Theorem 4.11).

Recall that an *oriented path* P of length n is any oriented graph (V, E) where $V = \{v_0, v_1, \ldots, v_n\}$ and for every $i = 1, 2, \ldots, n$ either $(v_{i-1}, v_i) \in E$ or $(v_i, v_{i-1}) \in E$ (but not both), and there are no other edges. Thus an oriented path is any orientation of an undirected path.

Denote by $(\mathcal{P}, \leq_{\mathcal{P}})$ the class of all finite paths ordered by homomorphism order. Given paths $P = (V, E)$, $P' = (V', E')$, a *homomorphism* is a mapping $\varphi : V \to V'$ which preserves edges:

$$(x, y) \in E \implies (\varphi(x), \varphi(y)) \in E'.$$

For paths P and P' we write $P \leq_{\mathcal{P}} P'$ if and only if there is homomorphism $\varphi : P \to P'$

To show the universality of oriented paths, we will construct an embedding of $(\mathcal{S}, \subseteq)$ to $(\mathcal{P}, \leq_{\mathcal{P}})$. Recall that the class $\mathcal{S}$ denotes the class of all periodic subsets of $\mathbb{Z}$ (see Section 2.7). This is a new feature, which gives a new, more streamlined and shorter proof of the [**8**]. The main difference of the proof in [**8, 10**] and the one presented here is the use of $(\mathcal{S}, \subseteq)$ as the base of the representation instead of $(\mathcal{T}V, \leq_{\mathcal{T}V})$. The linear nature of graph homomorphisms among oriented paths make it very difficult to adapt many-to-one mapping involved in $\leq_{\mathcal{T}V}$. The cyclic mappings of $(\mathcal{S}, \subseteq)$ are easier to use.

Let us introduce terms and notations that are useful when speaking of homomorphisms between paths. (We follow standard notations as e.g. in [**6, 19**].)

While oriented paths do not make a difference between initial and terminal vertices, we will always consider paths in a specific order of vertices from the initial to the terminal vertex. We denote the initial vertex v_0 and the terminal vertex v_n of P by $in(P)$ and $term(P)$ respectively. For a path P we will denote by $\overleftarrow{P}$ the flipped path P with order of vertices $v_n, v_{n-1}, \ldots, v_0$. For paths P and P' we denote by PP' the path created by the concatenation of P and P' (i.e. the disjoint union of P and P' with $term(P)$ identified with $in(P')$).

The *length* of a path P is the number of edges in P. The *algebraic length* of a path P is the number of forwarding minus the number of backwarding edges in

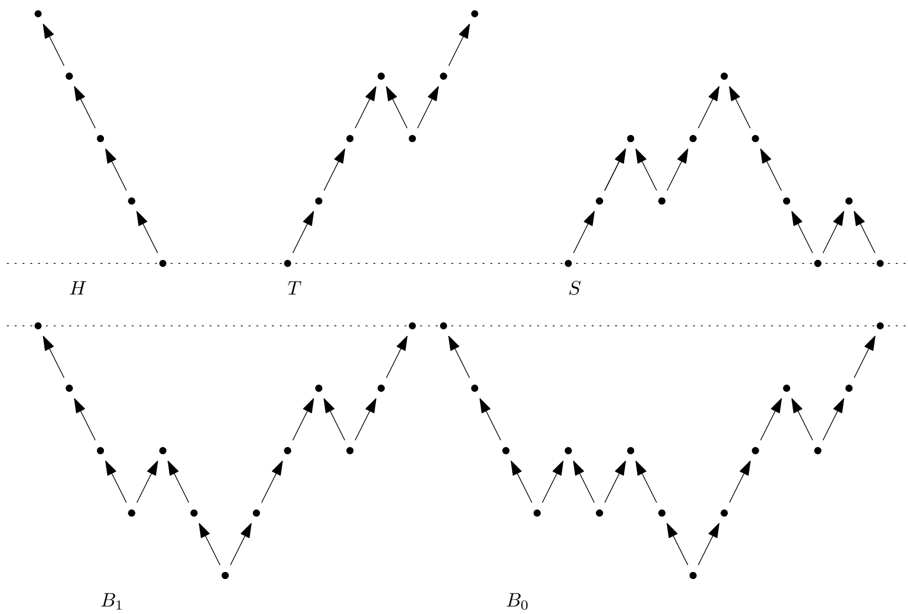


FIGURE 5. Building blocks of $p(W)$.

P. Thus the algebraic length of a path could be negative. The *level* $l_P(v_i)$ of v_i is the algebraic length of the subpath $(p_0, p_1, \ldots, p_i)$ of P. The *distance* between vertices p_i and p_j, $d_P(p_i, p_j)$, is given by $|j - i|$. The *algebraic distance,* $a_P(p_i, p_j)$, is $l_P(v_j) - l_P(v_i)$.

Denote by $\varphi : P \to P'$ a homomorphism from path P to P'. Observe that we always have $d_P(p_i, p_j) \leq d_{P'}(\varphi(p_i), \varphi(p_j))$ and $a_P(p_i, p_j) = a_{P'}(\varphi(p_i), \varphi(p_j))$. We will construct paths in such a way that every homomorphism φ between path P and P' must map the initial vertex of P to the initial vertex of P' and thus preserve levels of vertices (see Lemma 4.4 bellow).

The basic building blocks if our construction are the paths shown in Figure 5 (H stands for *head*, T for *tail*, B for *body* and S for *šipka*—arrow in Czech language). Their initial vertices appear on the left, terminal vertices on the right. Except for H and T the paths are balanced (i.e. their algebraic length is 0). We will construct paths by concatenating copies of these blocks. H will always be the first path, T always the last. (The dotted line in Figure 5 and Figure 6 determines vertices with level -4.)

DEFINITION 4.1. Given a word W on the alphabet $\{0, 1\}$ of length 2^n, we assign path $p(W)$ recursively as follows:

(1) $p(0) = B_0$.
(2) $p(1) = B_1$.
(3) $p(W) = p(W_1)S\overleftarrow{p(W_2)}$ where W_1 and W_2 are words of length 2^{n-1} such that $W = W_1W_2$.

Put $\overline{p}(W) = Hp(W)T$.

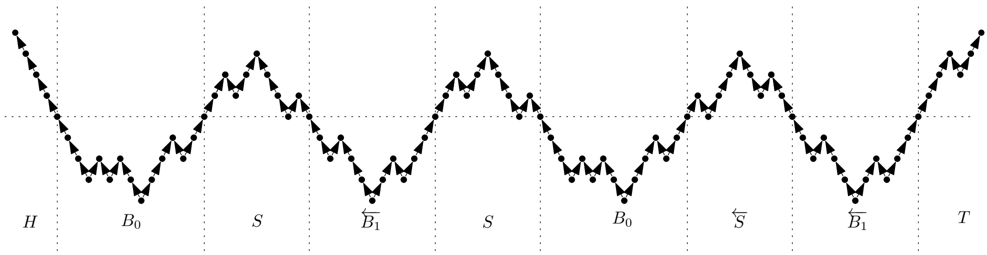


FIGURE 6. $\overline{p}(0110)$.

EXAMPLE 4.2. For a periodic set S, $s(4, S) = 0110$, we construct $\overline{p}(s(4, S))$ in the following way:

$$\begin{aligned} p(0) &= B_0, \\ p(1) &= B_1, \\ p(01) &= B_0 S \overleftarrow{B_1}, \\ p(10) &= B_1 S \overleftarrow{B_0}, \\ p(0110) &= B_0 S \overleftarrow{B_1} S B_0 \overleftarrow{S} \overleftarrow{B_1}, \\ \overline{p}(0110) &= H B_0 S \overleftarrow{B_1} S B_0 \overleftarrow{S} \overleftarrow{B_1} T. \end{aligned}$$

See Figure 6.

The key result of our construction is given by the following:

PROPOSITION 4.3. Fix a periodic set S of period 2^k and a periodic set S' of period $2^{k'}$. There is a homomorphism

$$\varphi : \overline{p}(s(2^k, S)) \to \overline{p}(s(2^{k'}, S'))$$

if and only if $S \subseteq S'$ and $k' \leq k$.

If a homomorphism φ exists, then φ maps the initial vertex of $\overline{p}(s(2^k, S))$ to the initial vertex of $\overline{p}(s(2^{k'}, S'))$. If $k' = k$ then φ maps the terminal vertex of $\overline{p}(s(2^k, S))$ to the terminal vertex of $\overline{p}(s(2^{k'}, S'))$. If $k' < k$ then φ maps the terminal vertex of $\overline{p}(s(2^k, S))$ to the initial vertex of $\overline{p}(s(2^{k'}, S'))$.

Prior to the proof of Proposition 4.3 we start with observations about homomorphisms between our special paths.

LEMMA 4.4. *Any homomorphism $\varphi : \overline{p}(W) \to \overline{p}(W')$ must map the initial vertex of $\overline{p}(W)$ to the initial vertex of $\overline{p}(W')$.*

PROOF. $\overline{p}(W)$ starts with the monotone path of 7 edges. The homomorphism φ must map this path to a monotone path in $\overline{p}(W')$. The only such subpath of $\overline{p}(W')$ is formed by first 8 vertices of $\overline{p}(W')$.

It is easy to see that φ cannot flip the path: If φ maps the initial vertex of $\overline{p}(W)$ to the 8th vertex of $\overline{p}(W')$ then $\overline{p}(W)$ has vertices at level -8 and because homomorphisms must preserve algebraic distances, they must map to the vertex of level 1 in $\overline{p}(W')$ and there is no such vertex in $\overline{p}(W')$. $\square$

LEMMA 4.5. *Fix words W, W' of the same length 2^k. Let φ be a homomorphism $\varphi : p(W) \to p(W')$. Then φ maps the initial vertex of $p(W)$ to the initial vertex of $p(W')$ if and only if φ maps the terminal vertex of $p(W)$ to the terminal vertex of $p(W')$.*

PROOF. We proceed by induction on length of W:

For $W = i$ and $W' = j$, $i, j \in \{0, 1\}$ we have $p(W) = B_i$ and $p(W') = B_j$. There is no homomorphism $B_1 \to B_0$. The unique homomorphism $B_0 \to B_1$ has the desired properties. The only homomorphism $B_0 \to B_0$ is the isomorphism $B_0 \to B_0$.

In the induction step put $W = W_0W_1$ and $W' = W_0'W_1'$ where W_0, W_1, W_0', W_1' are words of length 2^{k-1}. We have $p(W) = p(W_0)S\overleftarrow{p(W_1)}$ and $p(W') = p(W_0')S\overleftarrow{p(W_1')}$.

First assume that φ maps $in(p(W))$ to $in(p(W'))$. Then φ clearly maps $p(W_0)$ to $p(W_0')$ and thus by the induction hypothesis φ maps $term(p(W_0))$ to $term(p(W_0'))$. Because the vertices of S are at different levels than the vertices of the final blocks B_0 or B_1 of $p(W_0')$, a copy of S that follows in $p(W)$ after $p(W_0)$ must map to a copy of S that follows in $p(W')$ after $p(W_0')$. Further φ cannot flip S and thus φ maps $term(S)$ to $term(S)$. By same argument φ maps $p(W_1)$ to $p(W_1')$. The initial vertex of $p(W_1)$ is the terminal vertex of $p(W)$ and it must map to the initial vertex of $p(W_1')$ and thus also the terminal vertex of $p(W')$.

The second possibility is that φ maps $term(p(W))$ to $term(p(W'))$. This can be handled similarly (starting from the terminal vertex of paths in the reverse order). $\square$

LEMMA 4.6. *Fix periodic sets S, S' of the same period 2^k. There is a homomorphism*

$$\varphi : p(s(2^k, S)) \to p(s(2^k, S'))$$

mapping $in(p(s(2^k, S)))$ to $in(p(s(2^k, S')))$ if and only if $S \subseteq S'$.

PROOF. If $S \subseteq S'$ then the Lemma follows from the construction of $p(s(2^k, S))$. Every digit 1 of $s(2^k, S)$ has a corresponding copy of B_1 in $p(s(2^k, S))$ and every digit 0 has a corresponding copy of B_0 in $p(s(2^k, S))$. It is easy to build a homomorphism φ by concatenating a homomorphism $B_0 \to B_1$ and identical maps of S, B_0 and B_1.

In the opposite direction, assume existence of homomorphism φ from $p(s(2^k, S))$ to $p(s(2^k, S'))$. By the assumption and Lemma 4.5, φ must be map $term(p(s(2^k, S)))$ to $term(p(s(2^k, S')))$. Because S use vertices at different levels than B_0 and B_1, all copies of S must be mapped to copies of S. Similarly copies of B_0 and B_1 must be mapped to copies of B_0 or B_1. If $S \not\subseteq S'$ then there is position i such that i-th letter of $s(2^k, S)$ is 1 and i-th letter of $s(2^k, S')$ is 0. It follows that the copy of B_1 corresponding to this letter would have to map to a copy of B_0. This contradicts with the fact that there is no homomorphism $B_1 \to B_0$. $\square$

LEMMA 4.7 (folding). *For a word W of length 2^k, there is a homomorphism*

$$\varphi : \overline{p}(WW) \to \overline{p}(W)$$

mapping $in(\overline{p}(WW))$ to $in(\overline{p}(W))$ and $term(\overline{p}(WW))$ to $in(\overline{p}(W))$.

Proof. By definition

$$\overline{p}(WW) = Hp(W)S\overleftarrow{p(W)}T$$

and

$$\overline{p}(W) = Hp(W)T.$$

The homomorphism φ maps the first copy of $p(W)$ in $\overline{p}(WW)$ to a copy of $p(W)$ in $\overline{p}(W)$, a copy of S is mapped to T such that the terminal vertex of S maps to the initial vertex of T and thus it is possible to map a copy of $\overleftarrow{p(W)}$ in $\overline{p}(WW)$ to the same copy of $p(W)$ in $\overline{p}(W)$. $\square$

We will use the folding Lemma iteratively. By composition of homomorphisms there is also homomorphism $p(WWWW) \to p(WW) \to p(W)$. (From the path constructed from 2^k copies of W to $p(W)$.)

Proof (of Proposition 4.3). Assume the existence of a homomorphism φ as in Proposition 4.3. First observe that $k' \leq k$ (if $k < k'$ then there is a copy of T in $\overline{p}(s(2^k, S))$ would have to map into the middle of $\overline{p}(s(2^{k'}, S'))$, but there are no vertices at the level 0 in $\overline{p}(s(2^{k'}, S'))$ except for the initial and terminal vertex).

For $k = k'$ the statement follows directly from Lemma 4.6.

For $k' < k$ denote by W'' the word that consist of $2^{k-k'}$ concatenations of W'. Consider a homomorphism φ' from $p(W)$ to $p(W'')$ mapping $in(p(W))$ to $in(p(W''))$. W and W'' have the same length and such a homomorphism exists by Lemma 4.6 if and only if $S \subseteq S'$. Applying Lemma 4.7 there is a homomorphism $\varphi'' : p(W'') \to p(W')$. A homomorphism φ can be obtained by composing φ' and φ''. It is easy to see that any homomorphism $\overline{p}(W) \to \overline{p}(W')$ must follow the same scheme of "folding" the longer path $\overline{p}(W)$ into $\overline{p}(W')$ and thus there is a homomorphism φ if and only if $S \subseteq S'$. We omit the details. $\square$

For a periodic set S denote by $S^{(i)}$ the inclusion maximal periodic subset of S with period i. (For example for $s(4, S) = 0111$ we have $s(2, S^{(2)}) = 01$.)

Definition 4.8. For $S \in \mathcal{S}$ let i be the minimal integer such that S has period 2^i. Let $\Phi^{\mathcal{S}}_{\mathcal{P}}(S)$ be the concatenation of the paths

$$H,$$

$$\overline{p}(s(1, S^{(1)}))\overleftarrow{\overline{p}(s(1, S^{(1)}))},$$

$$\overline{p}(s(2, S^{(2)}))\overleftarrow{\overline{p}(s(2, S^{(2)}))},$$

$$\overline{p}(s(4, S^{(4)}))\overleftarrow{\overline{p}(s(4, S^{(4)}))},$$

$$\dots,$$

$$\overline{p}(s(2^{i-1}, S^{(2^{i-1})}))\overleftarrow{\overline{p}(s(2^{i-1}, S^{(2^{i-1})}))},$$

$$\overline{p}(s(2^i, S))\overleftarrow{\overline{p}(s(2^i, S))}.$$

Theorem 4.9. *$\Phi^{\mathcal{S}}_{\mathcal{P}}(v)$ is an embedding of $(\mathcal{S}, \subseteq)$ to $(\mathcal{P}, \leq_{\mathcal{P}})$.*

PROOF. Fix S and S' in $\mathcal{S}$ of periods 2^i and $2^{i'}$ respectively.

Assume that $S \subseteq S', i > i'$. Then the homomorphism $\varphi : \Phi^{\mathcal{S}}_{\mathcal{P}}(S) \to \Phi^{\mathcal{S}}_{\mathcal{P}}(S')$ can be constructed via the concatenation of homomorphisms:

$$
\begin{gathered}
H \to H,\\
\overline{p}(s(1,S^{(1)}))\overleftarrow{\overline{p}(s(1,S^{(1)}))} \to \overline{p}(s(1,S'^{(1)}))\overleftarrow{\overline{p}(s(1,S'^{(1)}))},\\
\overline{p}(s(1,S^{(2)}))\overleftarrow{\overline{p}(s(1,S^{(2)}))} \to \overline{p}(s(2,S'^{(2)}))\overleftarrow{\overline{p}(s(2,S'^{(2)}))},\\
\dots,\\
\overline{p}(s(2^{i'-1},S^{(2^{i'-1})}))\overleftarrow{\overline{p}(s(2^{i'-1},S^{(2^{i'-1})}))} \to \overline{p}(s(2^{i'-1},S'^{(2^{i'-1})}))\overleftarrow{\overline{p}(s(2^{i'-1},S'^{(2^{i'-1})}))},\\
\overline{p}(s(2^{i'},S^{(2^{i'})}))\overleftarrow{\overline{p}(s(2^{i'},S^{(2^{i'})}))} \to \overline{p}(s(2^{i'},S')),\\
\overline{p}(s(2^{i'+1},S^{(2^{i'+1})}))\overleftarrow{\overline{p}(s(2^{i'+1},S^{(2^{i'+1})}))} \to \overline{p}(s(2^{i'},S')),\\
\dots,\\
\overline{p}(s(2^{i},S))\overleftarrow{\overline{p}(s(2^{i'},S))} \to \overline{p}(s(2^{i'},S')).
\end{gathered}
$$

Individual homomorphisms exists by Proposition 4.3. For $i \leq i'$ the construction is even easier.

In the opposite direction assume that there is a homomorphism $\varphi : \Phi^{\mathcal{S}}_{\mathcal{P}}(S) \to \Phi^{\mathcal{S}}_{\mathcal{P}}(S')$. $\Phi^{\mathcal{S}}_{\mathcal{P}}(S)$ starts by two concatenations of H and thus a long monotone path and using a same argument as in Lemma 4.4, φ must map the initial vertex of $\Phi^{\mathcal{S}}_{\mathcal{P}}(S)$ to the initial vertex of $\Phi^{\mathcal{S}}_{\mathcal{P}}(S')$. It follows that φ preserves levels of vertices. It follows that for every $k = 1, 2, 4, \dots, 2^i$, φ must map $\overline{p}(s(k, S^{(k)}))$ to $\overline{p}(s(k', S'^{(k')}))$ for some $k' \leq k, k' = 1, 2, 4, \dots, 2^{i'}$. By application of Proposition 4.3 it follows that $S^{(k)} \subseteq S'^{(k')}$. In particular $S \subseteq S'^{(k')}$. This holds only if $S \subseteq S'$. $\square$

THEOREM 4.10 ([**8**]). *The quasi order $(\mathcal{P}, \leq_{\mathcal{P}})$ contains universal partial order.*

In fact our new proof of Corollary 4.10 gives the following strengthening for rooted homomorphisms of paths. A *plank* (P, r) is an oriented path rooted at the initial vertex $r = in(P)$. Given planks (P, r) and (P', r'), a homomorphism $\varphi : (P, r) \to (P', r')$ is a homomorphism $P \to P'$ such that $\varphi(r) = r'$.

THEOREM 4.11. *The quasi order formed by all planks ordered by the existence of homomorphisms contains a universal partial order.*

5. Related results

The universality of oriented paths implies the universality of the homomorphism order of many naturally defined classes of structures (such as undirected planar or series-parallel graphs) ordered by homomorphism via the indicator construction (see [**10**], [**18**]). By similar techniques the universality of homomorphism the order on labelled partial orders is shown in [**13**].

Lehtonen and Nešetřil [**14**] consider also the partial order defined on boolean functions in the following way. Each clone $\mathcal{C}$ on a fixed base set A determines a quasiorder on the set of all operations on A by the following rule: f is a $\mathcal{C}$-minor of g if f can be obtained by substituting operations from $\mathcal{C}$ for the variables of g. Using embedding homomorphism order on hypergraphs, it can be shown that a clone $\mathcal{C}$ on $\{0, 1\}$ has the property that the corresponding $\mathcal{C}$ minor partial order is universal if and only if $\mathcal{C}$ is one of the countably many clones of clique functions or the clone of self-dual monotone functions (using the classification of Post classes).

It seems that in most cases the homomorphism order of classes of relational structures is either universal or fails to be universal for very simple reasons (such as the absence of infinite chains or anti-chains). [**18**] look for minimal minor closed classes of graphs that are dense and universal. They show that $(\mathcal{P}, \leq_{\mathcal{P}})$ is a unique minimal class of oriented graphs which is both universal and dense. Moreover, they show a dichotomy result for any minor closed class $\mathcal{K}$ of directed trees. $\mathcal{K}$ is either universal or it is well-quasi-ordered. Situation seems more difficult for the case of undirected graphs, where such minimal classes are not known and only partial result on series-parallel graphs was obtained.

6. Acknowledgement

We thank Andrew Goodall and Robert Šámal for help and several remarks which improved the quality of this paper. We thank to the anonymous referee for an excellent job.

References

1. N. Alon, E. Scheinerman: Degrees of freedom versus dimension for containment orders, Order 5 (1988) 11-16. 2.
2. J. H. Conway: On numbers and games, London Math. Soc. Monographs, Academic press, 1976.
3. P. Ehrlich: Number Systems with Simplicity Hiearchies. A generalization of Conway's Theory of Surreal Numbers, J. of Symbolic Logic 66, 3 (2001), 1231–1258.
4. R. Fraïssé: Théorie des relations, North Holland, 1986.
5. Z. Hedrlín: On universal partly oriented sets and classes, J. Algebra 11 (1969), 503–509.
6. P. Hell, N. Nešetřil: Graphs and Homomorphisms, Oxford University Press, Oxford, 2004.
7. J. Hubička J. Nešetřil: A Finite Presentation of the rational Urysohn Space, Topology and Applications 155 (14) (2008), 1483–1492.
8. J. Hubička, J. Nešetřil: Finite Paths are Universal, Order 21 (2004), 181–200.
9. J. Hubička, J. Nešetřil: Finite presentation of homogeneous graphs, posets and Ramsey classes, Israel J. Math 149 (2005), 21–44.
10. J. Hubička, J. Nešetřil: On universal posets represented by means of trees and other simple graphs, European J. Comb. 26 (2005), 765–778.
11. A. S. Kechris, V. G. Pestov, S. Todorcevic: Fraïssé Limits, Ramsey Theory, and Topological Dynamics of Automorphism Groups, to appear in GAFA. Geom. Funct. Anal., 15 (2005), 106-189.
12. D. E. Knuth: Surreal Numbers, Addison Wesley, 1974.
13. E. Lehtonen: Labeled posets are universal, European J. Comb., 29 (2) (2008), 493–506.
14. E. Lehtonen, J. Nešetřil, Minors of Boolean functions with respect to clique functions and hypergraph homomorphisms, European J. Comb. 31(2010) 1981–1995.
15. J. Nešetřil: For graphs there are only four types of hereditary Ramsey Classes, J. Comb. Th. B, 46(2), (1989), 127–132.
16. J. Nešetřil: On Universality of Set Systems. KAM–DIMATIA Series 491, Charles University, 2000.
17. J. Nešetřil: Ramsey Theory. In: Handbook of Combinatorics (ed. R. L. Graham, M. Grötschel, L. Lovász), Elsevier, 1995, 1331–1403.
18. J. Nešetřil, Y. Nigussie, Minimal universal and dense minor closed classes, European Journal of Combin. 27 (2006), 1159–1171.
19. J. Nešetřil, X. Zhu, Path Homomorphisms. Math. Proc. Comb. Phil. Soc. (1996), 207–220.
20. A. Pultr, V. Trnková, Combinatorial, Algebraic and Topological Representations of Groups, Semigroups and Categories, North Holland, 1980.
21. W. T. Trotter: Dimension Theory John Hopkins Univ. Press, 1992.

Department of Applied Mathematics and Institute of Theoretical Computer sciences (ITI), Charles University, Malostranské nám. 25, 118 00 Praha 1, Czech Republic
E-mail address: hubicka@kam.mff.cuni.cz

Department of Applied Mathematics and Institute of Theoretical Computer sciences (ITI), Charles University, Malostranské nám. 25, 118 00 Praha 1, Czech Republic
E-mail address: nesetril@kam.mff.cuni.cz

This page intentionally left blank

Contemporary Mathematics
Volume **558**, 2011

Two problems on homogeneous structures, revisited

Gregory Cherlin

Abstract. We take up Peter Cameron's problem of the classification of countably infinite graphs which are homogeneous as metric spaces in the graph metric [**Cam98**]. We give an explicit catalog of the known examples, together with results supporting the conjecture that the catalog may be complete, or nearly so.

We begin in Part I with a presentation of Fraïssé's theory of amalgamation classes and the classification of homogeneous structures, with emphasis on the case of homogeneous metric spaces, from the discovery of the Urysohn space to the connection with topological dynamics developed in [**KPT05**]. We then turn to a discussion of the known metrically homogeneous graphs in Part II. This includes a 5-parameter family of homogeneous metric spaces whose connections with topological dynamics remain to be worked out. In the case of diameter 4, we find a variety of examples buried in the tables at the end of [**Che98**], which we decode and correlate with our catalog.

In the final Part we revisit an old chestnut from the theory of homogeneous structures, namely the problem of approximating the generic triangle free graph by finite graphs. Little is known about this, but we rephrase the problem more explicitly in terms of finite geometries. In that form it leads to questions that seem appropriate for design theorists, as well as some questions that involve structures small enough to be explored computationally. We also show, following a suggestion of Peter Cameron (1996), that while strongly regular graphs provide some interesting examples, one must look beyond this class in general for the desired approximations.

1. Introduction

The core of the present article is a presentation of the known *metrically homogeneous graphs*: these are the graphs which, when viewed as metric spaces in the graph metric, are homogeneous metric spaces.

1.1. Homogeneous Metric Spaces, Fraïssé Theory, and Classification. A metric space M is said to be *homogeneous* if every isometry between finite subsets of M is induced by an isometry taking M onto itself. An interesting and early

1991 *Mathematics Subject Classification.* Primary 03C10; Secondary 03C13, 03C15, 05B25, 20B22, 20B27.

Key words and phrases. homogeneous structure, permutation group, finite model property, model theory, amalgamation, extreme amenability, Ramsey theory.

Supported by NSF Grant DMS-0600940.

example is the Urysohn space U [**Ury25, Ury27**] found in the summer of 1924, the last product of Urysohn's short but intensely productive life. While the problem of Fréchet that prompted this construction concerned universality rather than homogeneity, Urysohn took particular notice of this homogeneity property in his initial letter to Hausdorff [**Huš08**], a point repeated in much the same terms in the posthumous announcement [**Ury25**]. We will discuss this in §2.

¿From the point of view of Fraïssé's later theory of *amalgamation classes* [**Fra54**], the essential point in Urysohn's construction is that finite metric spaces can be amalgamated: if M_1, M_2 are finite metric spaces whose metrics agree on their common part $M_0 = M_1 \cap M_2$, then there is a metric on $M_1 \cup M_2$ extending the given metrics; and more particularly, the same applies if we limit ourselves to metric spaces with a rational valued metric.

Fraïssé's theory facilitates the construction of infinite homogeneous structures of all sorts, which are then universal in various categories, and the theory is often used to that effect. This gives a construction of Rado's universal graph [**Rad64**], generalized by Henson to produce universal K_n-free graphs [**Hen71**], and uncountably many quite similar homogeneous directed graphs [**Hen72**]. A variant of the same construction also yields uncountably many homogeneous nilpotent groups and commutative rings [**CSW93**]. More subtly, Fraïssé's of amalgamation classes can be used to *classify* homogeneous structures of various types: homogeneous graphs [**LW80**], homogeneous directed graphs [**Sch79, Lac84, Che88, Che98**], the finite and the imprimitive infinite graphs with two edge colors [**Che99**], the homogeneous partial orders with a vertex coloring by countably many colors [**TrT08**], and even homogeneous permutations [**Cam03**].

There is a remarkable 3-way connection involving the theory of amalgamation classes, structural Ramsey theory, and topological dynamics, developed in [**KPT05**]. In this setting the Urysohn space appears as one of the natural examples, but more familiar combinatorial structures come in on an equal footing. The Fraïssé theory, and its connection with topological dynamics, is the subject of §3.

In §4 we conclude Part I with a discussion of the use of Fraïssé's theory to obtain classifications of all the homogeneous structures in some limited classes. Noteworthy here is the classification by Lachlan and Woodrow of the homogeneous graphs [**LW80**], which plays an important role in Part II.

1.2. Metrically Homogeneous Graphs. Part II is devoted to a classification problem for a particular class of homogeneous metric spaces singled out by Peter Cameron, an ambitious generalization of the case of homogeneous graphs treated by Lachlan and Woodrow. Every connected graph is a metric space in the graph metric, and Peter Cameron raised the question of the classification of the graphs for which the associated metric spaces are homogeneous [**Cam98**]. Such graphs are referred to as *metrically homogeneous* or *distance homogeneous*. This condition is much stronger than the condition of distance transitivity which is familiar in finite graph theory; the complete classification of the finite distance transitive graphs is much advanced and actively pursued. Cameron raised this issue in the context of his "census" of the very rich variety of countably infinite distance transitive graphs, and we develop Cameron's "census" into what may reasonably be considered a "catalog." The most striking feature of our catalog is a 4-parameter family of metrically homogeneous graphs determined by constraints on triangles.

For all of the known metrically homogeneous graphs, the next order of business would be to explore the Ramsey theoretic properties of their associated metric spaces, as well as the behavior of their automorphism groups in the setting of topological dynamics, in the spirit of §3.

We believe that our catalog may be complete, though we are very far from having a proof of that, or a clear strategy for one. But we will show that as far as certain natural classes of extreme examples are concerned, this catalog is complete. The catalog, and a statement of the main results about it, will be found in §5.

There are, a priori, three kinds of metrically homogeneous graphs which are sufficiently exceptional to merit separate investigation.

First, there are those for which the graph induced on the neighbors of a given vertex is exceptional. Here the distinction between the generic and exceptional cases is furnished, very conveniently, by the Lachlan/Woodrow classification. When a graph is metrically homogeneous, the induced graph on the neighborhood of any vertex is homogeneous as a graph, and its isomorphism type is independent of the vertex selected as base point. So this induced graph provides a convenient invariant which falls under the Lachlan/Woodrow classification. Those induced graphs which can occur within graphs in our known 4-parameter family are treated as non-exceptional, while the others are considered as exceptional: according to the Lachlan/Woodrow classification, these are the ones which do not contain an infinite independent set, and the imprimitive ones.

Next, there are the imprimitive ones, which carry a nontrivial equivalence relation invariant under the automorphism group.

A third class of metrically homogeneous graphs meriting separate consideration may be described in terms of the *minimal constraints* on the graph, which are the minimal finite integer-valued metric spaces which cannot be embedded isometrically into the graph with the graph metric. There is a profusion of metrically homogeneous graphs in which the minimal constraints all have order at most 3 (i.e., exactly 3, together with a possible bound on the diameter), and one would expect their explicit classification to be an important step in the construction of an appropriate catalog. Our catalog is complete in this third sense—it contains all the metrically homogeneous graphs whose minimal constraints have order at most 3. This last point proved elusive.

In the construction of our catalog, we began by determining those in the first class, that is the metrically homogeneous graphs whose neighborhood graphs are exceptional in the sense indicated above. These turn out to be of familiar types, namely the homogeneous graphs in the ordinary sense [**LW80**], the finite ones given in [**Cam80**], and the natural completion of the class of tree-like graphs considered in [**Mph82**].

Turning to the imprimitive case, we find that this cannot really be separated from the "generic" case. With few exceptions, a primitive metrically homogeneous graph is either bipartite or is of "antipodal" type, which in our context comes down to the following, after some analysis: the graph has finite diameter δ, and an "antipodality" relation $d(x, y) = \delta$ defines an involutory automorphism $y = \alpha(x)$ of the graph. In the bipartite case there is a reduction to an associated graph on each half of the bipartition. The antipodal case does not have a neat reduction: we will discuss the usual "folding" operation applied in such cases, and show that it does not preserve metric homogeneity. Our catalog predicts that the bipartite and antipodal graphs occur largely within the generic family by specializing some of the numerical parameters to extreme values, with the proviso that the treatment of side conditions of Henson's type varies slightly in the antipodal case. We investigated some special classes of bipartite graphs using the standard reduction and found some examples that looked curious at the time but find their natural place in the present catalog as metrically homogeneous graphs of generic type.

The class of graphs determined by constraints of order 3 appears to be the key to the classification of the remaining metrically homogeneous graphs. Examples of such graphs are found in [**Cam98**], and we took as our own starting point the class of graphs $\Gamma^\delta_{K,C,\mathcal{S}}$ defined as follows. We introduce the class $\mathcal{A}^\delta_{K,C,\mathcal{S}}$ of all finite integer-valued metric spaces of diameter at most δ, in which there are no metric triangles with odd perimeter less than $2K+1$, and there are no metric triangles with perimeter at least C. Here we take $1 \le K \le \delta$ or $K = \infty$, $2\delta+1 \le C \le 3\delta+1$. The extreme values of K correspond to no constraint or the bipartite case, while the extreme values of C correspond to the antipodal case or no constraint. Finally, $\mathcal{S}$ is a set of $(1,\delta)$-spaces, that is, spaces in which only the distances 1 and δ occurs (here $\delta \ge 3$), which we take as additional forbidden substructures. This is the natural extension of Henson's use of forbidden cliques to our setting. It turns out that this misses a significant source of examples determined by constraints of order 3, and when we came belatedly to test the catalog on this point we were led finally to a more complicated but similar family which will be denoted $\Gamma^\delta_{K_1,K_2;C_0,C_1;\mathcal{S}}$ in which the parameters δ, K_1, C_0, C_1 and the set $\mathcal{S}$ are much as in the previous case, but the parameter K_2 is a little more exotic: it forbids the occurrence of certain triangles of odd perimeter P, but only those satisfying

$$P > 2K_2 + d(a,b)$$

for some pair of vertices a, b. Of course the parameters $\delta, K_1, K_2, C_0, C_1$ and the set $\mathcal{S}$ must satisfy some auxiliary conditions to provide an amalgamation class. We will lay out the precise conditions on the parameters in detail. We will observe also that as far as the parameters $\delta, K_1, K_2, C_0, C_1$ are concerned, our conditions are expressible in Presburger arithmetic, and that this is to be expected, given that the set of constraints involved is itself definable in Presburger arithmetic from the 5 given numerical parameters. We also deal with an antipodal variation in which the set $\mathcal{S}$ of side constraints is modified.

We will present our catalog in §5. The main point is to state the conditions on the auxiliary parameters in full. We will not give complete proofs of existence: this comes down to the amalgamation property for the classes we associate with admissible choices of the parameters. We take this up in §6, where we confine ourselves largely to a presentation of the amalgamation method. The essential point here is that one may determine amalgamations by determining one distance at a time, and in that case the range of values available for that distance (subject to the triangle inequality) is a non-empty interval. Then the various constraints associated with our parameters and the set $\mathcal{S}$ may restrict the size and parity of the desired distance further, and one must show that some suitable value remains in all cases. We generally deal with the set $\mathcal{S}$ by avoiding the values 1 and δ entirely. As the antipodal case is somewhat different from the rest, we give more detail in that case.

In §7 we discuss the imprimitive case, showing that with minor exceptions these are bipartite or antipodal (and in the latter case, of a very restricted type). We also classify the antipodal bipartite graphs of odd diameter.

For any metrically homogeneous bipartite graph Γ, there is an associated graph $B\Gamma$ on either half of the bipartition of Γ which is again metrically homogeneous. If Γ is an antipodal bipartite graph of odd diameter, our classification implies that when $B\Gamma$ is in our catalog, Γ is as well. When the diameter is even, the associated graph $B\Gamma$ is again antipodal, but not bipartite. Again one should aim to show that when $B\Gamma$ is in the catalog, then Γ is as well, but we did not look into that.

In §8 we turn to the proof that the exceptional metrically homogeneous graphs (the ones for which the induced graph on the neighbors of a fixed vertex either contains no infinite independent set, or is imprimitive) all lie in our catalog.

In §9 we look at another class of metrically homogeneous bipartite graphs, in terms of the structure of the associated graph $B\Gamma$. It is natural to take up the case in which $B\Gamma$ is itself exceptional. We show in this case that Γ is in the catalog. We note that there are some cases in which Γ falls under the "generic" case in our catalog and $B\Gamma$ is exceptional. These occur up through diameter 5, with $B\Gamma$ a homogeneous graph.

Lastly, we look back in §10 to some examples that can be found in tables in the Appendix to [**Che98**] in a very different form. There we listed all amalgamation classes corresponding to primitive infinite homogeneous structures with four orbits on pairs of distinct elements, all self-paired (in other words, four nontrivial 2-types, all of them symmetric). There are 27, and within that list one finds 17 which can be interpreted as metric spaces, some in more than one way, corresponding to 20 metrically homogeneous graphs. So we decode that list and recast it in terms which allow a direct comparison with our catalog, which does indeed contain all of these examples. We remark that with most of these 27 examples understood as metrically homogeneous graphs, one might take another look at finding a framework that accounts for the remaining ones.

The net result of our explorations has been to turn up nothing new on the sporadic side of the classification, but to broaden considerably our conception of the generic case. Given the structure of the resulting catalog, the natural way to think about a proof of its completeness is in the following terms, using the theory of amalgamation classes (which we will review in Part I). Let us use the term "generic type" for metrically homogeneous graphs which are not already in the catalog as exceptions of one kind or another.

(1) Show that any amalgamation class of finite metric spaces associated with a metrically homogeneous graph of generic type involves exactly the same triangles (subspaces of order 3) as one of our generic classes determined by triangle constraints.
(2) Show that any amalgamation class of finite metric spaces associated with a metrically homogeneous graph of generic type whose triangle constraints are the same as some catalogued graph of generic type, is in fact in the catalog.

In practice a proof may involve an elaborate induction in which the two sides of the issue become mixed together, but any step in the proof is likely to target just one of the two issues. An equivalent statement of the first point would be that for any amalgamation class $\mathcal{A}$ of finite metric spaces associated with a metrically homogeneous graph of generic type, the associated class $\mathcal{A}'$ of finite metric spaces A such that every triangle in A belongs to $\mathcal{A}$ is itself an amalgamation class. This is inherently plausible, but not something which one would aim to prove by a direct argument.

Our sense of these problems is that they are both difficult. We are also convinced of the correctness of the classification as far as diameter 3; the work of Amato and Macpherson [**AMp10**] covers the antipodal case and the case of triangle free metrically homogeneous graphs (in our notation, $K_1 > 1$). Identifying the compatible combinations of constraints on triangles is easy in this case, and the number of cases is reasonable. We have convinced ourselves that the full classification can be completed in diameter 3 by direct methods, and that the outcome is consistent with the catalog.

1.3. Is the generic triangle free graph pseudofinite? In Part III we turn our attention to another problem suggested by the study of homogeneous structures. In its general form, the problem is to find a testable criterion for a homogeneous structure to be *pseudofinite* (that is, a model of the theory of all finite structures). For example, the universal homogeneous triangle free graph will be pseudofinite if and only if for each n we can find a finite triangle free graph with the following two properties:

(i) any maximal independent set of vertices has order at least n; and
(ii) for any set A of n independent vertices, and any subset B of A, there is a vertex adjacent to all vertices of B and to no vertices of $A \setminus B$.

It is known that the universal homogeneous graph is pseudofinite, because the associated problem on finite graphs is easily solved using random graphs. But this problem remains open for the universal homogeneous triangle free graph—and the more general problem is likely to remain in the shade till this particular instance is settled, one way or another.

We will discuss what is known about this problem in the triangle free case for extremely small values of n, namely $n = 3$ or 4. There is not a great deal known, but it is worth noticing that the problem can be rephrased in terms of finite geometries, and that some concrete problems emerge that design theorists may be able to make something of. So here I aim less at "model theoretic methods in combinatorics," and more at the hope that combinatorial methods can shed more light on this problem arising in model theory.

Some nice examples are known for the case $n = 3$, notably the Higman-Sims graph (as observed by Simon Thomas), as well as an infinite family constructed by Michael Albert, again with $n = 3$ [**SWS93**, p. 447]. We still have no example with $n = 4$, but we will suggest that the case $n = 3$ is worth much closer scrutiny, and raises problems that seem relatively accessible and which have a design-theoretic flavor. We also point out a hierarchy of conditions between the cases $n = 3$ and $n = 4$ which seems to us to represent a steep climb at each level. In the case $n = 3$ we have many examples which are degenerate in a precise sense, and which can be varied quite freely, while the other examples known are subgraphs of the Higman-Sims graph.

Thus we have the following key problem.

PROBLEM 1. *Is there a finite triangle free graph with the following properties, which is not a subgraph of the Higman-Sims graph?*

(i) *any maximal independent set of vertices has order at least* 3*; and*

($ii_{3,2}$) *for any set A of* 3 *independent vertices, and any subset B of A, there are at least two vertices adjacent to all vertices of B and to no vertices of $A \setminus B$.*

This may be phrased equivalently in terms of finite combinatorial geometries and the Higman-Sims geometry on 22 points, as we shall see. In §11 we give an interpretation of the general problem of approximating the generic triangle free graph by finite ones in terms of combinatorial geometries. In §12.1 we explore the connection with strongly regular graphs. Following a suggestion made privately in 1996 by Peter Cameron, we show that there is no strongly regular graph which meets our conditions for $n = 4$. My notes on the matter are long gone, but my impression now is that the case $n = 4$ is something of a squeaker. In coming back to this I found the explicit formulas in [**Big09**] helpful. To take the next step and bound the size of a strongly regular graph satisfying our conditions for $n = 3$ seems to involve the central problems of the field. But perhaps the experts can do something clever in that direction.

In §12.4, we show by elementary and direct analysis directly from the definitions that any graph satisfying our conditions for $n = 4$ will have minimal degree at least 66. One might expect this lower bound to translate into an impressive lower bound on the total number of vertices, but I don't see that.

The case $n = 3$ without any assumption of strong regularity is taken up in earnest in §13, in terms of geometries rather than graphs. Just as the Higman-Sims graph on 100 vertices is more readily seen in terms of the Higman-Sims geometry on 22 points (itself a 1-point extension of a projective plane over a field of order 4), one can describe nontrivial examples in terms of geometries on relatively few points. In particular, the smallest geometry which does not fall into the class we call "Albert Geometries" lives on a set of just 8 points. There is not much general theory to be seen in the present state of knowledge, but there is one very basic open question. In a combinatorial geometry we have a set of points and a set of blocks, the blocks being sets of points. In each of the known geometries associated with the case $n = 3$, with very few exceptions, there is a block of order 2. The main question is whether there is any geometry associated with the case $n = 3$ in which the minimum block size is greater than 2, other than geometries embedding into the Higman-Sims geometry.

In §13.2 we give a family of geometries satisfying our conditions for $n = 3$, having a unique block of order 2, and with the next smallest block size arbitrarily large. Getting rid of that last block of order 2 seems to put a wholly different complexion on the matter. One would expect design theorists to be able to say something substantial about this situation, one way or another.

Bonato has explored similar problems in the context of graphs and tournaments [**Bon09, Bon10**]. Here one knows by probabilistic arguments that finite graphs or tournaments with analogous properties exist, but in looking for the minimal size one lands in somewhat similar territory, and again certain aspects of design theory come into consideration, including strongly regular graphs, skew Hadamard matrices, and Paley graphs or tournaments.

The theory of homogeneous structures has many other aspects that we will not touch upon, many connected with the study of the automorphism groups of homogeneous structures, e.g. the small index property and reconstruction of structures from their automorphism groups [**DNT86, HHLS93, KT01, Rub94, Tru92**], group theoretic issues [**Tru85, Tru03, Tru09**], and the classification of reducts of homogeneous structures [**Tho91, Tho96**], which is tied up with structural Ramsey theory. We mention also the extensive and readable survey [**Mph11**] by Macpherson which covers a number of directions not touched on here.

There is also an elaborate theory due to Lachlan treating stable homogeneous relational structures systematically as limits of finite structures, and by the same token giving a very general analysis of the finite case [**Lac86b**].

The subject of homogeneity falls under the much broader heading of "oligomorphic permutation groups", that is the study of infinite permutation groups having only finitely many orbits on n-tuples for each n. As the underlying set is infinite, this property has the flavor of a very strong transitivity condition, and leads to a very rich theory [**Cam90, Cam97**].

1.4. Acknowledgment. This article has benefited enormously from a careful reading by the referee, particularly in Part II. The first draft of Part II described work in progress at the time; I learned a good deal from writing it, but not enough to rewrite it immediately. The form of the catalog given here (specifically, the classes $\mathcal{A}^{\delta}_{K_1,K_2;C_0,C_1;\mathcal{S}}$ mentioned above) was still a couple of months off at that point. Having a robust catalog of examples in Part II has certainly helped matters, but the referee's response to a variation of that first draft was also very helpful, as I was finding my way to what should be a much clearer account than the original.

Contents

Part I. Homogeneous Structures and Amalgamation Classes

2. Urysohn's space

2.1. A little history. A special number of *Topology and its Applications* (vol. 155) contains the proceedings of a conference on the Urysohn space (Beer-Sheva, 2006). A detailed account of the circumstances surrounding the discovery of that space, shortly before a swimming accident took Urysohn's life off the coast of Brittany, can be found in the first section of [**Huš08**], which we largely follow here. There is also an account of Urysohn's last days in [**GK09**], which provides additional context.

Urysohn completed his habilitation in 1921, and his well known contributions to topology were carried out in the brief interval between that habilitation and his fatal accident on August 17th, 1924. Fréchet raised the question of the existence of a universal complete separable metric space (one into which any other should embed isometrically) in an article published in the American Journal of Mathematics in 1925 with a date of submission given as August 21, 1924. Fréchet had communicated his question to Aleksandrov and Urysohn some time before that, and an announcement of a solution is contained in a letter from Aleksandrov and Urysohn to Hausdorff dated August 3, 1924, a letter to which Hausdorff replied in detail on August 11. The letter from Aleksandrov and Urysohn is quoted in the original German in [**Huš08**]. In that letter, the announcement of the construction of a universal complete separable metric space is followed immediately by the remark: "...and in addition [it] satisfies a quite powerful condition of homogeneity: the latter being, that it is possible to map the whole space onto itself (isometrically) so as to carry an arbitrary finite set M into an equally arbitrary set M_1, congruent to the set M." The letter goes on to note that this pair of conditions, universality together with homogeneity, actually characterizes the space constructed up to isometry. This comment on the property of homogeneity is highlighted in very similar terms in the published announcement [**Ury25**].

Urysohn's construction proceeds in two steps. He first constructs a space U_0, now called the *rational Urysohn space*, which is universal in the category of countable metric spaces with rational-valued metric. This space is constructed as a limit of *finite* rational-valued metric spaces, and Urysohn takes its completion U as the solution to Fréchet's problem.

It is the *rational* Urysohn space which fits neatly into the general framework later devised by Fraïssé [**Fra54**]. A countable structure is called *homogeneous* if any isomorphism between finitely generated substructures is induced by an automorphism of M. If we construe metric spaces as structures in which the metric defines a weighted complete graph, with the metric giving the weights, then finitely generated substructures are just finite subsets with the inherited metric, and isomorphism is isometry. Other examples of homogeneity arise naturally in algebra, such as vector spaces (which may carry forms—symplectic, orthogonal, or unitary), or algebraically closed fields. We will mainly be interested in relational systems, that is combinatorial structures in which "f.g. substructure" simply means "finite subset, with the induced structure." But Fraïssé's general theory, to which we turn in the next section, does allow for the presence of functions.

2.2. Topological dynamics and the Urysohn space. The Urysohn space, or rather its group of isometries, turns up in topological dynamics as an example of the phenomenon of *extreme amenability.* A topological group is said to be extremely amenable if any continuous action on a compact space has a fixed point. The isometry group of Urysohn space is shown to be extremely amenable in [**Pes02**], and subsequently the general theory of [**KPT05**] showed that the isometry group of the *ordered rational Urysohn space* (defined in the next section) is also extremely amenable. The general theory of [**KPT05**] requires Fraïssé's setup, but we quote one of the main results in advance:

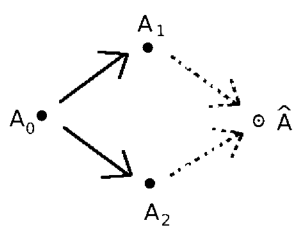
THEOREM 1. [**KPT05**, Theorem 2] *The extremely amenable closed subgroups of the infinite symmetric group* Sym_∞ *are exactly the groups of the form* $\mathrm{Aut}(F)$, *where* F *is the* Fraïssé limit *of a Fraïssé order class with the Ramsey property.*

We turn now to the Fraïssé theory.

3. Fraïssé Classes and the Ramsey Property

3.1. Amalgamation Classes. It is not hard to show that any two countable homogeneous structures of a given type will be isomorphic if and only if they have the same isomorphism types of f.g. substructures. This uses a "back-and-forth" construction, as in the usual proof that any two countable dense linear orders are isomorphic, which is indeed a particular instance. In view of this uniqueness, it is natural to look for a characterization of countable homogeneous structures directly in terms of the associated class $\mathrm{Sub}(M)$ of f.g. structures embedding into M. Fraïssé identified the relevant properties:

I. $\mathrm{Sub}(M)$ is hereditary (closed downward, and under isomorphism): in other words, if A is in the class, then any f.g. structure isomorphic with a substructure of A is in the class;
II. There are only countably many isomorphism types represented in $\mathrm{Sub}(M)$;
III. $\mathrm{Sub}(M)$ has the *joint embedding* and *amalgamation* properties: if A_1, A_2 are f.g. substructures in the class, and A_0 embeds into A_1, A_2 isomorphically via embeddings f_1, f_2, then there is a structure $\hat{A}$ in the class with embeddings $A_1, A_2 \to \hat{A}$, completing the diagram



The joint embedding property is the case in which A_0 is empty, which should be treated as a distinct condition if one does not allow empty structures.

A key point is that amalgamation follows from homogeneity: taking A_0 to be a subset of A_1, and embedded into A_2, apply an automorphism of the ambient structure to move the image of A_0 in A_2 back to A_0, and A_2 to some isomorphic structure A_2' containing A_0; then the structure generated by $A_1 \cup A_2'$ will serve as an amalgam.

Conversely, if $\mathcal{A}$ is a class of structures with properties $(I - III)$, then there is a countable homogeneous structure M, unique up to isomorphism, for which

$\mathrm{Sub}(M) = \mathcal{A}$. This homogeneous structure M is called the *Fraïssé limit* of the class $\mathcal{A}$. Thus with $\mathcal{A}$ taken as the class of finite linear orders, the Fraïssé limit is isomorphic to the rational order; with $\mathcal{A}$ the class of all finite graphs, the Fraïssé limit is Rado's universal graph [**Rad64**]; and with $\mathcal{A}$ the class of finite rational-valued metric spaces, after checking amalgamation, we take the Fraïssé limit to get the rational Urysohn space. The computation that checks amalgamation can be found in Urysohn's own construction, though not phrased as such.

So we see that Fraïssé's theory is at least a ready source of "new" homogeneous structures, and we now give a few more examples in the same vein. Starting with the class of all partial orders, we obtain the "generic" countable partial order (to call it merely "dense," as in the linear case, would be to understate its properties). Or starting with the class of finite triangle free graphs, we get the "generic" triangle free graph, and similarly for any n the generic K_n-free graph will be obtained [**Hen71**]. The amalgamation procedure here is simply graph theoretical union, and the special role of the complete graphs here is due to their *indecomposability* with respect to this amalgamation procedure: a complete graph which embeds into the graph theoretical union of two graphs (with no additional edges permitted) must embed into one of the two. The generalization to the case of directed graphs is immediate: amalgamating via the graph theoretical union, the indecomposable directed graphs are the tournaments. So we associate with any set of tournaments $\mathcal{T}$ the Fraïssé class of finite directed graphs *omitting* $\mathcal{T}$—i.e., with no directed subgraph isomorphic to one in $\mathcal{T}$. The corresponding Fraïssé limits are the generic $\mathcal{T}$-free graphs considered by Henson [**Hen72**]. The richness of the construction is confirmed by showing that $2^{\aleph_0}$ directed graphs arise in this way, because of the existence of an *infinite antichain* in the class of finite tournaments, that is an infinite set $\mathcal{X}$ of finite tournaments, which are pairwise incomparable under embedding, so that each subset of $\mathcal{X}$ gives rise to a different Fraïssé limit. A suitable construction of such an antichain is given by Henson [**Hen72**]. A parallel construction in the category of commutative rings provides, correspondingly, uncountably many homogeneous commutative rings [**CSW93**].

The structure of the infinite antichains of finite tournaments has been investigated further, but has not been fully elucidated. Any such antichain lies over one which is minimal in an appropriate sense, and after some close analysis by Latka [**Lat94, Lat03, Lat02**] a general finiteness theorem emerged [**ChL00**] to the effect that for any fixed k, there is a finite set of minimal antichains which will serve as universal witnesses for any collection of finite tournaments determined by k constraints (forbidden tournaments) which allows an infinite antichain. This means that whenever such an antichain is present, one of the given antichains is also present, up to a finite difference. But there is still no known *a priori* bound for the number of antichains required, as a function of k. Even the question as to whether the number of antichains needed is bounded by a computable function of k remains open.

In the terminology of [**KPT05**], the notion of Fraïssé class is taken to incorporate a further condition of *local finiteness*, meaning that all f.g. structures are finite. This may be viewed as a strengthening of condition (II). This convention is in force, in particular, in the statement of Theorem 1, which we now elucidate further.

3.2. Order Classes. The Fraïssé classes that occur in the theorem of Kechris-Pestov-Todorcevic above are *order classes:* this means that the structures considered are equipped with a distinguished relation $<$ representing a linear order. Thus in this theorem nothing is said about graphs, directed graphs, or metric spaces, but rather their ordered counterparts: ordered graphs, ordered directed graphs, ordered metric spaces. In particular the *ordered rational Urysohn space* is, by definition, the homogeneous ordered rational valued metric space delivered by Fraïssé's theory. As there is no connection between the order and the metric the necessary amalgamation may be carried out separately in both categories.

In the most straightforward, and most common, applications of the Fraïssé theory there is often some notion of "free amalgamation" in use. In the case of order classes amalgamation cannot be entirely canonical, as some "symmetry breaking" is inevitable. But there are also finite homogeneous structures—such as the pentagon graph, or 5-cycle—for which the theory of amalgamation classes is not illuminating, and the amalgamation procedure consists largely of the forced identification of points.

When one passes from the *construction* of examples to their systematic *classification,* there is typically some separation between the determination of more or less sporadic examples, and the remaining cases described naturally by the Fraïssé theory. Such a classification has only been carried out in a few cases, and perhaps a more nuanced picture will appear eventually. But this simple picture continues to guide our expectations for the classification of metrically homogeneous graphs, considered in the next Part, and so far everything we have seen is consistent with that picture in this case.

Below we will say something more about how the theory of [**KPT05**] applies in the absence of order, but first we complete the interpretation of Theorem 1 by discussing the second key property required.

3.3. The Ramsey Property. The ordinary Ramsey theorem is expressed in Hungarian notation by the symbolism:

$$\forall k, m, n\, \exists N : N \to (n)^m_k$$

meaning that for given k, m, n, there is N so that: for any coloring of increasing m-tuples from $A = \{1, \dots, N\}$ by k colors, there is a subset B of cardinality n which is monochromatic with respect to the coloring.

Structural Ramsey theory deals with a locally finite hereditary class $\mathcal{A}$ of finite structures of fixed type, which on specialization to the case of the class $\mathcal{L}$ of finite linear orders will degenerate to the usual Ramsey theory. In general, given two structures A, B in $\mathcal{A}$, write $\binom{B}{A}$ for the class of induced substructures of B isomorphic to A. This gives $\binom{N}{n}$ an appropriate meaning if $\mathcal{A} = \mathcal{L}$, namely increasing sequences of length n from an ordered set of size N.

We may then use the Hungarian notation

$$M \to (B)^A_k$$

to mean that whenever we have a coloring of of $\binom{M}{A}$ by k colors, there is a copy of B inside M which is monochromatic with respect to the induced coloring of $\binom{B}{A}$. And the *Ramsey property* will be:

$$\forall A, B \in \mathcal{A}\, \forall k\, \exists M \in \mathcal{A} : M \to (B)^A_k$$

So the Ramsey property for $\mathcal{L}$ is the usual finite Ramsey theorem.

Ramsey theory for Fraïssé classes is a subtle matter, but a highly developed one. In [**HN05**] it is shown that the Ramsey property implies the amalgamation property, by a direct argument. What one would really like to classify are the Fraïssé classes with the Ramsey property, but according to [**HN05**] the most promising route toward that is via classification of amalgamation classes first, and then the identification of the Ramsey classes.

For unordered graphs, the only instances of the Ramsey property that hold are those for which the subgraphs being colored are complete graphs K_n, or their complements [**NR75b**]. But the collection of finite ordered graphs does have the Ramsey property [**NR77a, NR77b, AH78**].

To illustrate the need for an ordering, consider colorings of the graph $A = K_1 + K_2$, a graph on 3 vertices with one edge, and let $B = 2 \cdot K_2$ be the disjoint sum of two complete graphs of order 2. If we order B in any way, we may color the copies of A in B by three colors according to the relative position of the isolated vertex of A, with respect to the other two vertices, namely before, after, or between them. Then B, with this coloring, cannot be monochromatic. Thus we can never have a graph G satisfying $G \to (B)^A_3$, since given such a graph G we would first order G, then define a coloring of copies of A in G as above, using this order, and there could be no monochromatic copy of B.

The topological significance of the Ramsey property for the ordered rational Urysohn space only emerged in [**KPT05**], and the appropriate structural Ramsey theorem was proved "on demand" by Nešetřil [**Neš07**].

At this point, we have collected all the notions needed for Theorem 1. Before we leave this subject, we note that the theory of [**KPT05**] also exhibits a direct connection between topological dynamics and the more classical examples of the Fraïssé theory (lacking a built-in order). The following is a fragment of Theorem 5 of [**KPT05**].

THEOREM 2. *Let G be the automorphism group of one of the following countable structures M:*

(1) *The random graph;*
(2) *The generic K_n-free graph, $n \geq 2$;*
(3) *The rational Urysohn space.*

Let L be the space of all linear orderings of M, with its compact topology as a closed subset of $2^{M\times M}$. Then under the natural action of G on L, L is the universal minimal compact flow *for G.*

The *minimality* here means that there is no proper closed invariant subspace; and the *universality* means that this is the largest such minimal flow (projecting on to any other). Again, Theorem 2 has an abstract formulation in terms of Fraïssé theory [**KPT05**]. The following is a special case.

THEOREM 3 ([**KPT05**, Theorem 4]). *Let $\mathcal{A}$ be a (locally finite) Fraïssé class and let $\mathcal{A}^+$ be the class of ordered structures $(K,<)$ with $K \in \mathcal{A}$. Suppose that $\mathcal{A}^+$ is a Fraïssé class with the order property and the Ramsey property. Let M be the Fraïssé limit of $\mathcal{A}$, $G = \mathrm{Aut}(M)$, and $\mathcal{L}$ the space of linear orderings of M, equipped with the compact topology inherited by inclusion into $2^{M\times M}$. Then under the natural action of G on $\mathcal{L}$, the space $\mathcal{L}$ is the universal minimal compact flow for G.*

Here one has in mind the case in which amalgamation in $\mathcal{A}$ does not require any identification of vertices (strong amalgamation); then $\mathcal{A}^+$ is certainly an amalgamation class. The *order property* is the following additional condition: given $A \in \mathcal{A}$, there is $B \in \mathcal{A}$ such that under any ordering on A, and any ordering on B, there is some order preserving isomorphic embedding of A into B. This is again a property which must be verified when needed, and is known in the cases cited. To see an example where the order property does not hold, consider the class $\mathcal{A}$ of finite equivalence relations. Any equivalence relation B may be ordered so that its classes are intervals; thus the order property fails.

We turn next to the problem of classifying homogeneous structures of particular types. Here again the Fraïssé theory provides the key.

4. Classification

The homogeneous structures of certain types have been completely classified, notably homogeneous graphs [**LW80**], homogeneous tournaments [**Lac84**], homogeneous tournaments with a coloring by finitely many colors and homogeneous directed graphs [**Che88**], homogeneous partial orders with a coloring by countably many colors [**TrT08**], and homogeneous permutations [**Cam03**]: this last is a less familiar notion, that we will enlarge upon. There is also work on the classification of homogeneous 3-hypergraphs [**LT95, AL95**], and on graphs with two colors of edges [**Lac86a, Che99**], the latter covering only the finite and imprimitive cases: this uncovers some sporadic examples, but the main problem remains untouched in this class.

4.1. Homogeneous Permutations. Cameron observed that permutations have a natural interpretation as structures, and that when one adopts that point of view the model theoretic notion of embedding is the appropriate one. A finite permutation may naturally be viewed as a finite structure consisting of two linear orderings. This is equivalent to a pair of bijections between the structure and the set $\{1, \ldots, n\}$, n being the cardinality, and thus to a permutation. In this setting, an embedding of one permutation into another is an occurrence in the second of a *permutation pattern* corresponding to the first, so that this formalism meshes nicely with the very active subject of permutations omitting specified patterns ("pattern classes").

By a direct analysis, Cameron showed that there are just 6 homogeneous permutations, in this sense, up to isomorphism: the trivial permutation of order 1, the identity permutation of $\mathbb{Q}$ or its reversal, the class corresponding to the lexicographic order on $\mathbb{Q} \times \mathbb{Q}$, where the second order agrees with the first in one coordinate and reverses the first in the other coordinate, and the generic permutation (corresponding to the class of all finite permutations). The existence of the generic permutation is immediate by the Fraïssé theory.

As amalgamations of linear orders are tightly constrained, the classification of the amalgamation classes of permutations is quite direct. Cameron also observes that it would be natural to generalize from structures with two linear orders to an arbitrary finite number, but I do not know of any further progress on this interesting question.

4.2. Homogeneous Graphs. This is the case that really launched the classification project. The classification of homogeneous graphs by Lachlan and Woodrow involves an ingenious inductive setup couched directly in terms of amalgamation classes of finite graphs. We will need the results of that classification later, when discussing metrically homogeneous graphs. Indeed, homogeneous graphs are just the diameter two (or less) case of metrically homogeneous graphs. Furthermore, in any metrically homogeneous graph, the graph induced on the neighbors of a point is a homogeneous graph, and we will find it useful to consider the possibilities individually, or more exactly to distinguish the "exceptional" and "generic" cases, and to treat the exceptional ones on an ad hoc basis.

In our catalog of the homogeneous graphs we will use the following notation. The graph K_n is a complete graph of order n, allowing $n = \infty$, which stands for $\aleph_0$ in this context. We write I_n for the complement of K_n, that is an independent set of vertices of order n, and $m \cdot K_n$ for the disjoint sum of m copies of K_n, again allowing m and n to become infinite. Bearing in mind that the complement of a homogeneous graph is again a homogeneous graph, we arrange the list as follows.

I. *Degenerate cases,* K_n or I_n; these are actually homogeneous structures for a simpler language (containing just the equality symbol).

II. *Imprimitive homogeneous graphs,* $m \cdot K_n$ and their complements, where $m, n \geq 2$. The complement of $m \cdot K_n$ is complete n-partite with parts of constant size.

III. *Primitive, nondegenerate, homogeneous finite graphs* (highly exceptional): the pentagon or 5-cycle C_5, and a graph on 9 points which may be described as the line graph of the complete bipartite graph $K_{3,3}$, or the graph theoretic square of K_3. These graphs are isomorphic with their complements.

IV. *Primitive, nondegenerate, infinite homogeneous graphs,* with which the classification is primarily concerned: Henson's generic graphs omitting K_n, and their complements, generic omitting I_n, as well as the generic or random graph Γ_∞ (Rado's graph) corresponding to the class of all finite graphs. Rado's graph is isomorphic with its complement.

In this setting there is no difficulty identifying the degenerate and imprimitive examples, and little difficulty in identifying the remaining finite ones by an inductive analysis. Since the class of homogeneous graphs is closed under complementation, the whole classification comes down to the following result.

Theorem 4 ([**LW80**, Theorem 2′, paraphrased]). *Let* Γ *be a homogeneous non-degenerate primitive graph containing an infinite independent set, as well as the complete graph* K_n. *Then* Γ *contains every finite graph not containing a copy of* K_{n+1}.

Let us see that this completes the classification in the infinite, primitive, nondegenerate case. As the graph Γ under consideration is infinite, by Ramsey's theorem it contains either K_∞ or I_∞, and passing to the complement if necessary, we may suppose the latter. So if K_n embeds in Γ and K_{n+1} does not, then Theorem 4 says that Γ is the corresponding Henson graph, while if K_n embeds in Γ for all n, Theorem 4 then says that it is the Rado graph.

The method of proof is by induction on the order N of the finite graph G which we wish to embed in Γ. The difficulty is that on cursory inspection, Theorem 4 does not at all lend itself to such an inductive proof. Lachlan and Woodrow show that as sometimes happens in such cases, a stronger statement may be proved by induction. Their strengthening is on the extravagant side, and involves some additional technicalities, but it arises naturally from the failure of the first try at an inductive argument. So let us first see what difficulties appear in a direct attack.

Let G be a graph of order N, not containing K_{n+1}, and let Γ be the homogeneous graph under consideration. We aim to show that G embeds in Γ, proceeding by induction on N. Pick a vertex v of G. If v is isolated, or if v is adjacent to the remaining vertices of G, we will need some special argument (even more so later, once we strengthen our inductive claim). For example, if v is adjacent to the remaining vertices of G, then we have an easy case: we identify v with any vertex of Γ, we consider the graph Γ_1 on the vertices adjacent to v in Γ, and after verifying that Γ_1 inherits all hypotheses on Γ (with n replaced by $n-1$) we can conclude directly by induction on n. If the vertex v is isolated, the argument will be less immediate, but still quite manageable.

Turning now to the main case, when the vertex v has both a neighbor and a non-neighbor in G, matters are considerably less simple. Let G_0 be the graph induced on the other vertices of G, and let a, b be vertices of G_0 chosen with a adjacent to v, and b not. At this point, we must build an amalgamation diagram which forces a copy of G into Γ, and we hope to get the factors of the diagram by induction on N, which of course does not quite work. It goes like this.

Let A be the graph obtained from G by deleting v and a, and let B be the graph obtained from G by deleting v and b. Let H_0 be the disjoint sum $A+B$ of A and B, and form two graphs $H_1 = H_0 \cup \{u\}$ and $H_2 = H_0 \cup \{c\}$, with edge relations as follows. The vertex u plays the role of v, and is therefore related to A and B as v is in G. The vertex c plays a more ambiguous role, as a or b, and is related to A as a is, and to B as b is.

Suppose for the moment that copies of H_1, H_2 occur in Γ. Then so does an amalgam $H_1 \cup H_2$ over H_0, and in that amalgam either u is joined to c, which may then play the role of a, with the help of A, or else u is not joined to c, and then c may play the role of b, with the help of B. In either case, a copy of G is forced into Γ.

What may be said about the structure of H_1 and H_2? These are certainly too large to be embedded into Γ *by induction,* but they have a simple structure: H_2 is the free amalgam of $A \cup \{a\}$ with $B \cup \{b\}$ under the identification of a with b, and H_1 is constructed similarly, over u. Here each factor (e.g., $A \cup \{a\}$, $A \cup \{b\}$ in the case of H_2) embeds into Γ by induction, but we need also the sum of the two factors over a common vertex. This leads to the following definitions.

DEFINITION 4.1.

(1) A *pointed graph* (G, v) is a graph G with a distinguished vertex v.
(2) The *pointed sum* of two pointed graphs (G, v) and (H, w) is the graph obtained from the disjoint sum $G + H$ by identifying the base points.
(3) Let $\mathcal{A}(n)$ be the set of finite graphs belonging to every amalgamation class which contains K_n, I_∞, the path of order 3, and its complement (the last two eliminate imprimitive cases).
(4) Let $\mathcal{A}^*(n)$ be the set of finite graphs G such that for any vertex v of G, and any pointed graph (H, w) with $H \in \mathcal{A}(n)$, the pointed sum $(G, v) + (H, w)$ belongs to $\mathcal{A}(n)$.

Notice that $\mathcal{A}^*(n)$ is contained in $\mathcal{A}(n)$ for trivial reasons, simply taking for (H, w) the pointed graph of order 1. Now we can state the desired strengthening of Theorem $2'$.

THEOREM 5 ([**LW80**, Lemma 6]). *For any n, if G is a finite graph omitting K_{n+1}, then G belongs to $\mathcal{A}^*(n)$.*

With this definition, the desired inductive proof actually goes through. Admittedly the special cases encountered in our first run above become more substantial the second time around. As a result, this version of the main theorem will be preceded by 5 other preparatory lemmas required to support the final induction. However the process of chasing one's tail comes to an end at this point.

4.3. Homogeneous Tournaments. In Lachlan's classification of the homogeneous tournaments [**Lac84**] two new ideas occur, which later turned out to be sufficient to carry out the full classification of the homogeneous directed graphs [**Che98**], with suitable orchestration. A byproduct of that later work was a more efficient organization of the case of tournaments, given in [**Che88**]. The main idea introduced at this stage was a certain use of Ramsey's theorem that we will describe in full. The second idea arises naturally at a later stage as one works through the implications of the first; it involves an enlargement of the setting beyond tournaments, where much as in the case of the Lachlan/Woodrow argument, the point is to find an inductive framework large enough to carry through an argument that leads somewhat beyond the initial context of homogeneous tournaments.

It turns out that there are only 5 homogeneous tournaments, four of them of a special type which are easily classified, and the last one fully generic. The whole difficulty comes in the characterization of this last tournament as the only homogeneous tournament of general type, in fact the only one containing a specific tournament of order 4 called $[T_1, C_3]$. In this notation, T_1 is the tournament of order 1, C_3 is a 3-cycle, and $[T_1, C_3]$ is the tournament consisting of a vertex (T_1) dominating a copy of C_3. So the analog of Theorem $2'$ of [**LW80**] is the following.

THEOREM 6 ([**Lac84**]). *Let T be a countable homogeneous tournament containing a tournament isomorphic with $[T_1, C_3]$. Then every finite tournament embeds into T.*

We now give the classification of the homogeneous tournaments explicitly, and indicate the reduction of that classification to Theorem 6.

A *local order* is a tournament with the property that for any vertex v, the tournaments induced on the sets $v^+ = \{u : v \to u\}$ and $v^- = \{u : u \to v\}$ are both transitive (i.e., given by linear orders). Equivalently, these are the tournaments not embedding $[T_1, C_3]$ or its dual $[C_3, T_1]$. There is a simple structure theory for the local orders, which we will not go into here. But the result is that there are exactly four homogeneous local orders, two of them finite: the trivial one of order 1, and the 3-cycle C_3. The infinite homogeneous local orders are the rational order $(\mathbb{Q}, <)$ and a very similar generic local order, which can be realized equally concretely.

Now a tournament T which does not contain a copy of $[T_1, C_3]$ can easily be shown to be of the form $[S, L]$ where S is a local order whose vertices all dominate a linear order L; here S or L may be empty. Indeed, if T is homogeneous, one of the two must be empty, and in particular T is a local order. Thus if the homogeneous tournament T contains a copy of $[T_1, C_3]$ then it contains a copy of $[C_3, T_1]$ as well, and it remains only to prove Theorem 6.

At this point, the following interesting technical notion comes into the picture. If $\mathcal{A}$ is an amalgamation class, let $\mathcal{A}^*$ be the set of finite tournaments T such that every tournament T^* of the following form belongs to $\mathcal{A}$: $T^* = T \cup L$, L is linear, and every pattern of edges between T and L is permitted. Theorem 6 is equivalent to the following rococo variation.

THEOREM 7. *If $\mathcal{A}$ is an amalgamation class containing $[T_1, C_3]$, then $\mathcal{A}^*$ is an amalgamation class containing $[T_1, C_3]$.*

That $\mathcal{A}^*$ is an amalgamation class is simply an exercise in the definitions, but worth working through to see why the definition of $\mathcal{A}^*$ takes the particular form that it does (because linear orders have strong amalgamation). The deduction of Theorem 6 from Theorem 7 is also immediate, as we will now verify.

Assuming Theorem 7, we argue by induction on $N = |A|$ that any finite tournament A belongs to any amalgamation class $\mathcal{A}$ containing $[T_1, C_3]$. Take any vertex v of A and let A_0 be the tournament induced on the remaining vertices. By induction, A_0 belongs to every amalgamation class containing $[T_1, C_3]$; in particular $A_0 \in \mathcal{A}^*$. Since A is the extension of A_0 by a single vertex, and since a single vertex constitutes a linear tournament, then from $A_0 \in \mathcal{A}^*$ we derive $A \in \mathcal{A}$, and we are done.

Note the progress which has been made. In Theorem 6 we consider arbitrary tournaments; in Theorem 7 we consider only linear extensions of $[T_1, C_3]$. Now a further reduction comes in, and eventually the statement to be proved reduces to a finite number of specific instances of Theorem 6 which can be proved individually. But we have not yet encountered the leading idea of the argument, which comes in at the next step.

4.4. The Ramsey Argument. We introduce another class closely connected with $\mathcal{A}^*$.

DEFINITION 4.2.

1. For tournaments A, B we define the *composition* $A[B]$ as usual as the tournament derived from A by replacing each vertex of A by a copy of B, with edges determined within each copy of B as in B, and between each copy of B, as in A. The composition of two tournaments is a tournament.

2. If T is any tournament, a *stack* of copies of T is a composition $L[T]$ with L linear.

3. If $\mathcal{A}$ is an amalgamation class of finite tournaments, let $\mathcal{A}^{**}$ be the set of tournaments T such that every tournament T^* of the following form belongs to $\mathcal{A}$: $T^* = L[T] \cup \{v\}$ is an extension of some stack of copies of T by one more vertex.

The crucial point here is the following.

FACT 4.3. *Let $\mathcal{A}$ be an amalgamation class of finite tournaments, and T a finite tournament in $\mathcal{A}^{**}$. Then T belongs to $\mathcal{A}^*$.*

We will not give the argument here. It is a direct application of the Ramsey theorem, given explicitly in [**Lac84**] and again in [**Che88, Che98**]. The idea is that one may amalgamate a large number of one point extensions of a long stack of copies of T so that in any amalgam, the additional points contain a long linear tournament, and one of the copies of T occurring in the stack will hook up with that linear tournament in any previously prescribed fashion desired.

This leads to our third, and nearly final, version of the main theorem.

THEOREM 8. *Let $\mathcal{A}$ be an amalgamation class of finite tournaments containing $[T_1, C_3]$. Then C_3 belongs to $\mathcal{A}^{**}$.*

Notice that a stack of copies of $[T_1, C_3]$ embeds in a longer stack of copies of C_3, so that Theorem 8 immediately implies the same result for $[T_1, C_3]$. Since we already saw that $\mathcal{A}^{**} \subseteq \mathcal{A}^*$, Theorem 8 implies Theorem 7. In view of the very simple structure of a stack of copies of T, we are almost ready to prove Theorem 8 by induction on the length of the stack. Unfortunately the additional vertex v occurring in $T^* = L[T] \cup \{v\}$ complicates matters, and leads to a further reformulation of the statement.

At this point, it is convenient to return from the language of amalgamation classes to the language of structures. So let the given amalgamation class correspond to the homogeneous tournament $\mathbb{T}$, and let $T = L[C_3] \cup \{v\}$ be a 1-point extension of a finite stack of 3-cycles. Theorem 8 says that T embeds into $\mathbb{T}$. It will be simpler to strengthen the statement slightly, as follows.

Let a be an arbitrary vertex in $\mathbb{T}$, and consider

$$\mathbb{T}_1 = a^+ = \{v : a \to v\} \text{ and } \mathbb{T}_2 = a^- = \{v : v \to a\}$$

separately. We claim that T embeds into $\mathbb{T}$ with $L[C_3]$ embedding into $\mathbb{T}_1$, and with the vertex v going into $\mathbb{T}_2$. This now sets us up for an inductive argument in which we consider a single 3-cycle C in $\mathbb{T}_1$, and the parts $\mathbb{T}_1'$ and $\mathbb{T}_2'$ defined relative to C as follows: $\mathbb{T}_1'$ consists of the vertices of $\mathbb{T}_1$ dominated by the three vertices of C, and $\mathbb{T}_2'$ consists of the vertices v' of $\mathbb{T}_2$ which relate to C as the specified vertex v does. What remains at this point is to clarify what we know, initially, about $\mathbb{T}_1$ and $\mathbb{T}_2$, and to show that these properties are inherited by $\mathbb{T}_1'$ and $\mathbb{T}_2'$ (in particular, $\mathbb{T}_2'$ should be nonempty!). This then allows an inductive argument to run smoothly.

At this point we have traded in the tournament $\mathbb{T}$ for a richer structure $(\mathbb{T}_1, \mathbb{T}_2)$ consisting of a tournament with a distinguished partition into two sets. The homogeneity of $\mathbb{T}$ will give us the homogeneity of $(\mathbb{T}_1, \mathbb{T}_2)$ in its expanded language. Such structures will be called 2-tournaments, and the particular class of 2-tournaments arising here will be called *ample tournaments*. The main inductive step in the proof of Theorem 8 will be the claim that an ample tournament $(\mathbb{T}_1, \mathbb{T}_2)$ gives rise to an ample tournament $(\mathbb{T}_1', \mathbb{T}_2')$ if we fix a 3-cycle in $\mathbb{T}_1$ and pass to the subsets considered above.

We will not dwell on this last part. The main steps in the proof are the reduction to Theorem 8, and then the realization that we should step beyond the class of homogeneous tournaments to the class of homogeneous 2-tournaments, to find a setting which is appropriately closed under the construction corresponding to the inductive step of the argument. This then leaves us concerned only about the base of the induction, which reduces to a small number of specific claims about tournaments of order not exceeding 6. Once the problem is finitized, it can be settled by explicit amalgamation arguments.

Lachlan's Ramsey theoretic argument functions much the same way in the context of directed graphs as it does for tournaments, and comes more into its own there, as it is not a foregone conclusion that Ramsey's theorem will necessarily produce a linear order; but it will produce something, and modifying the definition of $\mathcal{A}^*$ to allow for this additional element of vagueness, things proceed much as they did before.

In [**Che98**] there is also a treatment of the case of homogeneous graphs using the ideas of [**Lac84**] in place of the methods of [**LW80**]. This cannot be said to be a simplification, having roughly the complexity of the original proof, but it is a viable alternative, and the proof of the classification of homogeneous directed graphs is more or less a combination of the ideas of the tournament classification with the ideas which appear in a treatment of homogeneous graphs by this second method.

This ends our general survey of the general theory of amalgamation classes and its application to classification results. It is not clear how much further those ideas can be taken. The proofs are long and ultimately computational even when the final classifications have a reasonably simple form, and at this level of generality one has good methods but no very general theory. In the finite case, one has an excellent theory due to Lachlan making good use of model theoretic ideas. These were Lachlan's words in [**Lac86b**]:

> *The situation can be summarized as follows: Finite homogeneous structures are well understood. Stable homogeneous structures turn out to be just the unions of chains of finite ones. Thus, understanding stable homogeneous structures goes hand in hand with understanding finite ones. Beyond this, some special cases have been investigated successfully, but almost no general results have been obtained.*

That assessment stands today as far as the theory of homogeneous structures is concerned. A generalization of the theory of finite homogeneous structures beyond the homogeneous framework was also envisioned by Lachlan, and came to fruition, based on a combination of permutation group theory and model theory [**KLM89, ChH03**].

It seems to be impossible to say at this point how much farther one can go with the methods of classification for homogeneous structures currently available. While we have few general results, we also have no known limitations on the method. The existence of $2^{\aleph_0}$ homogeneous directed graph was once taken as such a limitation, a point of view I shared till seeing the classification of homogeneous tournaments. At that point, there was a tension between the existence of $2^{\aleph_0}$ known examples of homogeneous digraphs and the fact that there was no clear obstruction to the use of Lachlan's methods in this case. This conflict was resolved in favor of Lachlan's methods [**Che98**]. So we know less now than we thought we did originally. It still seems doubtful to me that the methods of Lachlan and Woodrow can be pushed very far beyond their current range, but on the other hand we have not actually found any concrete evidence of their limitations, or indeed any homogeneous relational structures that are not readily accounted for as either occurring in nature or coming naturally from the Fraïssé theory.

We believe that the classification of the metrically homogeneous graphs will provide another case in which some sporadic examples can be accounted for as growing naturally out of one or another special phenomenon, and the remainder fall neatly into the theory of amalgamation classes of "generic" type. The evidence for this is admittedly thin—the catalog of known types reached its present form after the first draft of the present paper was complete, and a coherent plan for a proof of its completeness still does not exist. But the catalog strikes me as satisfactorily robust now, and whenever one has a catalog with a clear division of exceptional and generic cases, one has some reason to expect the existing theory to be adequate to a proof of its completeness, with the proviso that there is often a striking disparity between the complexity of the catalog and the complexity of the resulting proof.

With all this in mind—or out of mind, as the case may be—we will turn in the next part to a catalog of the known metrically homogeneous graphs, and a full discussion of those which are "non-generic" in one of a number of senses. As we mentioned earlier, we will not give full existence proofs for the metrically homogeneous graphs in our catalog, though we will give some of the leading ideas and full proofs in special cases. The main class we present depends on four numerical parameters satisfying some simple linear inequalities and parity constraints, in more than one possible combination. We can account for those inequalities and parity constraints on abstract grounds: they are connected heuristically with quantifier elimination in Presburger arithmetic (§5.4).

In another direction, we think the generalization of Cameron's classification of homogeneous permutations to the case of structures equipped with k linear orders (also called k-dimensional permutations [**Wat07**, §5.9]) is another attractive classification problem. Here one can easily make a catalog of the "natural" examples but it is unclear whether one should expect that catalog to be complete. This problem seems to us to have something to do with the existing theory of weakly o-minimal structures. But we will not explore the matter here.

Part II. Metrically Homogeneous Graphs

5. Metrically homogeneous graphs: A catalog

5.1. The Classification Problem. Any connected graph may be considered as a metric space under the graph metric, and if the associated metric space is homogeneous then the graph is said to be *metrically homogeneous*[1] [**Cam98**]). Cameron asked whether this class of graphs can be *completely classified,* and gave some examples of constructions via the Fraïssé theory of amalgamation classes.

We believe that such a classification can be given. As a first step, we will give a catalog of all the known metrically homogeneous graphs, with the expectation that this catalog is complete or nearly so. That catalog is the focus of the present part. It consists of a few graphs of exceptional types, and two "generic" families which are best understood in terms of the Fraïssé theory of amalgamation classes.

Since the main examples in the catalog are presented in terms of amalgamation classes, it is necessary to check the amalgamation property for the classes we define. This is not trivial, and as there are a number of distinct cases to consider, it will not be covered in detail. The main point of §6 will be to lay out explicitly the amalgamation procedure followed. We will leave for another occasion a detailed proof that this procedure succeeds for all the classes considered. We will go into more detail in the discussion of a variant appropriate to the so-called antipodal case, as this is distinct perturbation of the general case.

The second order of business is to show that this catalog is reasonably complete. This is largely a byproduct of the way the catalog was constructed. There are two natural notions of "exceptional" metrically homogeneous graph. In addition, there

[1] A considerably weaker notion occurs in the geometry literature under the name *metrically homogeneous set*

is a natural Fraïssé style construction whose main ingredient is an amalgamation class determined entirely by constraints on triangles. The most difficult point to work through was the determination of this last class of examples. This caught us by surprise; the analogous step in previous problems of this type has been straightforward. In the present case, the conditions on an amalgamation class determined by constraints of order 3 depend on five numerical parameters (one of which is the diameter δ, which we always have lurking in the background). We denote the resulting classes by $\mathcal{A}^\delta_{K_1,K_2;C_0,C_1}$ where the numerical parameters K_1, K_2 are used to specify which metric triangles of small odd perimeter are forbidden, and the parameters C_0, C_1 are used to specify which triangles of large perimeter, even or odd respectively, are forbidden. When we come to the details it will be seen that the parameter K_2 is used in a more subtle way than the other three parameters. We more or less guessed the role of K_1, C_0, C_1 at the outset, except that we expected $C_1 = C_0 \pm 1$, and hence we worked with just two parameters, $K = K_1$ and $C = \min(C_0, C_1)$. The role of K_2 came as a surprise and we took the first examples found to be sporadic.

In the next subsection we will lay out our notions of "exceptional" metrically homogeneous graph and "generic" metrically homogeneous graph explicitly, and indicate the way the catalog was devised, before actually giving the catalog. One of our two notions of "exceptional" metrically homogeneous graph (the imprimitive case) turns out to lie mainly on the generic side in the catalog as it now stands.

Cameron made a number of fundamental observations on the classification problem in [**Cam98**]. He noted that the Lachlan/Woodrow classification is the diameter 2 case. He pointed out that Fraïssé constructions give graphs of any fixed diameter whose associated metric spaces are analogous to the Urysohn space, but with bounded integral distance, and that there is a bipartite variant of this construction. He also noted related work by Komjáth, Mekler and Pach which turns out to be very closely connected with the construction of graphs of generic type [**KMP88**], specifically with the classes $\mathcal{A}^\delta_{K_1,K_2;C_0,C_1}$. Cameron also observed that one may forbid cliques in the manner of Henson. When this is generalized a little more one gets the classes $\mathcal{A}^\delta_{K_1,K_2;C_0,C_1;\mathcal{S}}$ where $\mathcal{S}$ is a set of side constraints slightly more general than Henson's cliques.

Also noteworthy in this regard is the classification by Macpherson of the infinite, locally finite distance transitive graphs [**Mph82**], which occupies a privileged position among the exceptional entries in the catalog (slightly generalized). The finite case was dealt with in [**Cam80**],

Ongoing work by Amato and Macpherson [**AMp10**] sheds considerable light on the case of diameter 3, and similar methods appear to suffice to confirm the completeness of the catalog in diameter 3. In particular, with this small diameter the potentially problematic antipodal case can be handled directly and is covered in [**AMp10**]. The role of the generalized Henson constraints is very clear in that work, and is the main focus of that article. In the form I have seen, that article treats the antipodal and triangle free cases.

Some examples of metrically homogeneous graphs can be extracted from tables given at the end of [**Che98**]. These tables present all the primitive metrically homogeneous graphs of diameter 3 or 4 which can be defined by forbidding a set of triangles, excluding those of diameter 3 in which none of the forbidden triangles involve the distance 2 (where there is a notion of free amalgamation which seemed not very interesting). At the time I produced those tables, it never crossed my mind that a significant number of them could be construed as metric spaces. But the 27 examples originally listed with 4 nontrivial self-paired orbits on pairs (i.e., 4 nontrivial symmetric 2-types) give rise to 20 distinct metric spaces, each derived from a metrically homogeneous graph by taking distance 1 as the edge relation. Some of the examples listed have no interpretation as metric spaces, while others can be interpreted as metric spaces in two distinct ways. Some of these examples lie outside the catalog based on the families $\mathcal{A}^{\delta}_{K,C,S}$ but are consistent with the catalog as it now stands, as indeed they must be, since we have proved the completeness of our catalog in its present form for examples determined entirely by constraints of order 3. We will tabulate the relevant examples and give a translation from the notation of [**Che98**] to the notation of our catalog.

There are good methods, but there is no clear strategy, for an eventual proof of completeness of the catalog. In Part I we have indicated some of the methods which have been applied to similar problems, notably in the classification of the homogeneous connected graphs by Lachlan and Woodrow. Since the homogeneous connected graphs are the metrically homogeneous connected graphs of diameter at most 2, the Lachlan/Woodrow classification appears as the point of departure for Cameron's problem. In the case of diameter 2 the fact that the triangle inequality is vacuous is helpful when applying Lachlan's method. With higher diameters the emphasis shifts toward the use of the triangle inequality, but Lachlan's Ramsey argument (discussed in Part I) retains considerable power.

After presenting our notions of exceptional and generic type metrically homogeneous graphs, we will state the main facts known to us regarding the completeness of the catalog, and then present the catalog itself. The remainder of this Part will then deal with proofs of our results in the exceptional case, as well as some discussion of the generic case, in considerably less detail.

Most of the metrically homogeneous graphs considered are taken to be connected, a point occasionally mentioned. In an inductive analysis, disconnected metrically homogeneous subgraphs may come into play. In such cases we focus mainly but not exclusively on their connected components.

5.2. Exceptional Cases. Our basic strategy in designing the catalog presented below was to try to ensure that the following three types of connected metrically homogeneous graph Γ were all adequately covered.

- **Exceptional:** The induced graph on the neighbors of a fixed vertex of Γ is exceptional (in a sense specified below).
- **Imprimitive:** The graph Γ is imprimitive; that is, it carries a nontrivial equivalence relation invariant under $\mathrm{Aut}(\Gamma)$.
- **3-constrained:** The class of finite metric spaces which embed isometrically into Γ can be specified in terms of forbidden substructures of order 3.

Let us take up these three possibilities separately. We will mention various results about them along the way, but leave the precise statements to follow the presentation of the catalog.

First, the class of connected metrically homogeneous graphs which we officially declare to be "exceptional" is defined as follows. Let Γ be a metrically homogeneous graph, and $v \in \Gamma$ a fixed vertex. Let $\Gamma_1 = \Gamma_1(v)$ be the graph induced on the neighbors of v. Then Γ_1 is a homogeneous graph, and its isomorphism type is independent of the choice of base vertex v. As Γ_1 is homogeneous, it can be found in the list of Lachlan/Woodrow presented in §4.2, which we should keep in mind throughout.

The cases in which the induced graph Γ_1 is an independent set, a Henson graph, or the Rado graph are all associated with natural constructions of Fraïssé type and do not belong on the exceptional side. We will put all other possibilities for Γ_1 on the exceptional side: these are the imprimitive or finite cases, and the complements of the Henson graphs. More abstractly, they are the cases in which Γ_1 is either imprimitive or contains no infinite independent set. We give the classification of the metrically homogeneous graphs Γ with Γ_1 exceptional as Theorem 10. These turn out to be of familiar types.

The second class of graphs we looked into were the imprimitive graphs. Here the situation is at first quite close to what is known in the finite case for distance transitive graphs, where it goes under the name of Smith's Theorem [**AH06, Smi71**], though eventually the analysis diverges, losing some ground in the infinite case, but with the much stronger hypothesis of metric homogeneity providing considerable compensation. It is easy to show, as in the finite case, that the imprimitive metrically homogeneous graphs are either bipartite, or *antipodal*, or possibly both. Here a graph is antipodal if its diameter δ is finite, and the relation $d(x, y) = \delta$ (or 0) defines an equivalence relation. For metrically homogeneous graphs it turns out that with minor exceptions the equivalence classes in the antipodal case have order 2, and that the pairing defined by the relation $d(x, y) = \delta$ defines an involutory automorphism.

This analysis is useful but does not lead to a complete classification for reasons that will become quite clear in a moment. There is a general reduction from the bipartite case to the nonbipartite (but possibly antipodal) case. One considers the graph $B\Gamma$ induced on one half of the bipartition by taking as the edge relation $d(x, y) = 2$. This is then a metrically homogeneous graph in its own right, and we may consider the case in which $B\Gamma$ is exceptional. In fact, we give a complete classification of the bipartite metrically homogeneous graphs for which $B\Gamma_1$ is not the Rado graph as Theorem 13. We also characterize the antipodal bipartite metrically homogeneous graphs of odd diameter in Theorem 12. It would also be reasonable at this point to try to complete the reduction of the bipartite case to the nonbipartite (possibly antipodal) case by proving the existence and uniqueness of Γ with $B\Gamma$ falling under the remaining cases; or, if necessary, at least to prove that result for $B\Gamma$ falling within the catalog of known examples.

The third class of metrically homogeneous graphs requiring special attention leads us to a conjectured description of the graphs of generic type. Informally, the graphs of generic type are those which come from amalgamation classes using natural methods of amalgamation approximating some notion of free amalgamation. In our case we are dealing with classes of metric spaces with an integer valued metric, typically with a bound δ on the diameter, and for which any geodesic of total length at most δ is allowed to occur. In that case, if the class $\mathcal{A}$ in question has the amalgamation property, the associated homogeneous metric space carries the graph structure given by the edge relation $d(x, y) = 1$, and the metric coincides with the graph metric (a point made in [**Cam98**]). As the class $\mathcal{A}$ must be hereditary (downward closed), it may be specified by giving a set of forbidden subspaces; that is, we specify a collection $\mathcal{C}$ of finite metric spaces, and then let $\mathcal{A}$ be the class of finite metric spaces X such that no space in $\mathcal{C}$ embeds isometrically into X. We single out for attention the case in which all constraints in $\mathcal{C}$ have order 3. We note that the triangle inequality is already a set of constraints of this type. Furthermore, imprimitivity is itself an example of a constraint given by a set of forbidden metric triangles.

Our first idea was to consider the classes $\mathcal{A}^{\delta}_{K;C}$ given by constraints of the following type: there are no triangles of odd perimeter less than $2K+1$, and no triangles of perimeter C or more. Here the perimeter is the sum of the three distances between pairs of points (and the term "triangle" refers to metric triangles, that is, to metric spaces with three points). This idea is suggested by known examples in the theory of universal graphs described by Komjáth, Mekler, and Pach [**KMP88**], and a proof of the amalgamation property in such cases is straightforward. We may combine this construction with the idea of omitting cliques as follows. With δ the diameter, let a $(1,\delta)$-*space* be any finite metric space in which only the distances 1 and δ occur. We will be concerned only with the case in which $\delta \geq 3$, in which case such a space consists of equivalence classes which are cliques with respect to the edge relation $d(x,y)=1$, with distinct classes separated by the maximal distance δ. It turns out that one may usually avoid both the minimal and the maximal values of the distance in completing amalgamation diagrams, and thus we can generalize the Henson construction to get amalgamation classes of the form $\mathcal{A}_{K;C;\mathcal{S}}$ with $\mathcal{S}$ a set of $(1,\delta)$-spaces. Here one must pay a little attention to the interaction of K, C, and $\mathcal{S}$ to ensure that the amalgamation property holds.

As a test of our original catalog, we set out to prove that the pattern of triangles occurring in any metrically homogeneous graph would be that of one of the classes $\mathcal{A}^{\delta}_{K;C}$. Once we saw that this was false it was not immediately clear whether the exceptions were sporadic. In the end the class with five parameters $\mathcal{A}^{\delta}_{K_1,K_2;C_0,C_1}$ emerged along with its Hensonian variations $\mathcal{A}^{\delta}_{K_1,K_2;C_0,C_1;\mathcal{S}}$, and a further antipodal variant, by solving a much simpler problem: identify all amalgamation classes determined by constraints of order 3. It still needs to be shown that the collection of triangles not embedding in an arbitrary metrically homogeneous graph agrees with the collection associated with some graph in our extended catalog. We would view this as a significant step toward a proof of completeness of the catalog.

The case of imprimitive graphs has been analyzed to the point at which the remaining cases should fall under the generic case, corresponding to extreme values of the parameters: in the bipartite case we forbid all triangles of odd perimeter, and in the antipodal case we forbid all triangles of perimeter greater than 2δ (as we shall see). But the antipodal case involves another variation on the Henson construction, so it appears separately in the catalog.

5.3. The Catalog.

NOTATION 5.1. If Γ is metrically homogeneous then for $v \in \Gamma$ we denote by $\Gamma_i(v)$ the set of vertices at distance i from v, with the induced metric; this is a homogeneous metric space, but it does not necessarily come from a graph metric, and in fact the distance 1 may not even be represented in $\Gamma_i(v)$. Since the isomorphism type of $\Gamma_i(v)$ is independent of the choice of v, we often write Γ_i rather than $\Gamma_i(v)$.

If the distance 1 is represented in Γ_i and Γ_i is connected, then the metric on Γ_i is the graph metric (see [**Cam98**]).

Our catalog uses specialized notations and constructions which will be explained in §5.4.

CATALOG

I. $\delta \le 2$ (Cf. §4.2)
   (a) Finite primitive: C_5, $L[K_{3,3}]$
   (b) Defective or imprimitive: $m \cdot K_n$, $K_m[I_m]$.
   (c) Infinite primitive, not defective: G_n, G_n^c, G_∞.

II. $\delta \ge 3$, Γ_1 finite or imprimitive.
   (a) An n-gon with $n \ge 6$.
   (b) Antipodal double cover of one of the graphs C_5, $L[K_{3,3}]$, or a finite independent set.
   (c) A tree-like graph $T_{r,s}$ as described by Macpherson in [**Mph82**], where $2 \le r, s \le \infty$, and if $s = \infty$ then $r \ge 3$.

III. Γ_1 infinite and primitive
   (a) $T_{2,\infty}$, the infinitely branching regular tree.
   (b) The generic antipodal graph omitting K_n: $\Gamma^\delta_{a,n}$, where either $\delta \ge 4$, or $\delta = 3$ and $n = 3$ or ∞.
   (c) The generic graph $\Gamma^\delta_{K_1,K_2;C_0,C_1;\mathcal{S}}$ associated with the class $\mathcal{A}^\delta_{K_1,K_2;C_0,C_1;\mathcal{S}}$, for an admissible choice of parameters $K_1, K_2; C_0, C_1; \mathcal{S}$.

The next order of business is to explain the following notions referred to in the catalog, after which we will state the main results relating to completeness of the catalog.

(1) The treelike graphs $T_{r,s}$ of [**Mph82**] (generalized to allow r or s to be infinite);
(2) The antipodal double cover;
(3) The notation

$$\Gamma^\delta_{K_1,K_2;C_0,C_1;\mathcal{S}}$$

for graphs constructed via the Fraïssé theory from a suitable amalgamation class, and the antipodal variations denoted by $\Gamma^\delta_{a,n}$, as well as the precise conditions for admissibility of the parameters.

5.4. Three Constructions.

DEFINITION 5.2 (The treelike graphs $T_{r,s}$.). For $2 \le r, s \le \infty$, we may construct an *r-tree of s-cliques* $T_{r,s}$ as follows. Take a tree $T(r,s)$ partitioned into two sets of vertices A, B, so that each vertex of A has r neighbors, all in B, and each vertex of B has s neighbors, all in A. Consider the graph induced on A with edge relation given by "$d(u,v) = 2$". This is $T_{r,s}$ (and the corresponding graph on B is $T_{s,r}$).

LEMMA 5.3. *For any r, s the tree $T(r,s)$ is homogeneous as a metric space with a fixed partition into two sets, and the graph $T_{r,s}$ is metrically homogeneous.*

PROOF. For any finite subset A of a tree T, one can see that the metric structure on A induced by T determines the structure of the convex closure of A, the smallest subtree of T containing A. Given that, a map between two finite subsets of $T(r,s)$ that respects the partition will extend first to the convex closures and then to the whole of $T(r,s)$.

This applies in particular to the two halves of $T(r,s)$. □

Now we turn to some "doubling" constructions.

DEFINITION 5.4.

(1) The *double cover* $\Gamma = 2 * G$ of G is the graph on $V(G) \times \mathbb{Z}_2$ with edges given by $(u, i) \sim (v, j)$ iff

$$\begin{cases} u \sim v & \text{if } i = j \\ u \not\sim v \text{ and } u \neq v & \text{if } i \neq j \end{cases}$$

(2) The *antipodal double cover* $\Gamma = \hat{G}$ of G is the double cover of the graph G^* obtained from G by adding one additional vertex $*$ adjacent to all vertices of G.

(3) Let Γ be a graph of diameter δ. The *bipartite double cover* of Γ is the graph with vertex set $V(\Gamma) \times \mathbb{Z}_2$ and edge relation $\sim$ given by $(u, i) \sim (v, j)$ iff:

$$d(u, v) = \delta \text{ and } i \neq j$$

Finally, we give the explicit definition of the classes $\mathcal{A}^\delta_{K_1,K_2;C_0,C_1}$.

We will write $\mathcal{M}^\delta$ for the class of all finite, integer valued metric spaces in which all lengths are bounded by δ. A triangle is a metric space containing three points. The type of a triangle is the triple (i, j, k) of distances realized in the triangle, taken in any order. The perimeter of a triangle of type i, j, k is the sum $i + j + k$.

DEFINITION 5.5. For $1 \leq K_1 \leq K_2 \leq \delta$ (or $K_1 = \infty$, $K_2 = 0$) and for $2\delta + 1 \leq C_0, C_1 \leq 3\delta + 2$, we define

(1) $\mathcal{A}^\delta_{K_1,K_2}$ is the subclass of $\mathcal{M}^\delta$ with forbidden triangles of types (i, j, k) with $P = i + j + k$ odd and either
   (a) $P < 2K_1 + 1$; or
   (b) $P > 2K_2 + \min(i, j, k)$

(2) $\mathcal{B}^\delta_{C_0,C_1}$ is the subclass of $\mathcal{M}^\delta$ with forbidden triangles of types (i, j, k) where $P = i + j + k$ satisfies

$$P \geq C_\ell, \quad P \equiv \ell \mod 2$$

(3) $\mathcal{A}^\delta_{K_1,K_2;C_0,C_1} = \mathcal{A}^\delta_{K_1,K_2} \cap \mathcal{B}^\delta_{C_0,C_1}$.

DEFINITION 5.6. A choice of parameters $\delta, K_1, K_2, C_0, C_1, \mathcal{S}$ is *admissible* if the following conditions are satisfied, where we write $C = \min(C_0, C_1)$ and $C' = \max(C_0, C_1)$.

- $\delta \geq 3$
- Either $1 \leq K_1 \leq K_2 \leq \delta$ or $K_1 = \infty$ and $K_2 = 0$
- $2\delta + 1 \leq C < C' \leq 3\delta + 2$, with one of C, C' even and the other odd.
- $\mathcal{S}$ is a set of finite $(1, \delta)$-spaces of order at least 3, and one of the following combinations of conditions holds:

(1) $K_1 = \infty$:

$$K_2 = 0,\ C_1 = 2\delta + 1, \text{ and}$$

$$\mathcal{S} \text{ is } \begin{cases} \text{empty} & \text{if } \delta \text{ is odd, or } C_0 \leq 3\delta \\ \text{a set of } \delta\text{-cliques} & \text{if } \delta \text{ is even, } C_0 = 3\delta + 2 \end{cases}$$

(2) $K_1 < \infty$ and $C \leq 2\delta + K_1$:

$$C = 2K_1 + 2K_2 + 1,\ K_1 + K_2 \geq \delta, \text{ and } K_1 + 2K_2 \leq 2\delta - 1$$

If $C' > C+1$ then $K_1 = K_2$ and $3K_2 = 2\delta - 1$.
If $K_1 = 1$ then $\mathcal{S}$ is empty.

(3) $K_1 < \infty$, and $C > 2\delta + K_1$:

$K_1 + 2K_2 \geq 2\delta - 1$ and $3K_2 \geq 2\delta$.
If $K_1 + 2K_2 = 2\delta - 1$ then $C \geq 2\delta + K_1 + 2$.
If $C' > C+1$ then $C \geq 2\delta + K_2$.
If $K_2 = \delta$ then $\mathcal{S}$ cannot contain a triangle of type $(1, \delta, \delta)$.
If $K_1 = \delta$ then $\mathcal{S}$ is empty.
If $C = 2\delta + 2$, then $\mathcal{S}$ is empty.

Note that if $\delta = \infty$ then C_0, C_1 should be omitted and the $(1, \delta)$-spaces are just cliques.

We claim of course that these admissibility conditions are precisely the conditions required on our parameters to ensure that the corresponding class is an amalgamation class. We will not prove this here, though we will present the amalgamation procedure which works when the parameters meet these conditions.

We also have some antipodal variations on the classes $\mathcal{A}^\delta_{K_1,K_2;C_0,C_1;\mathcal{S}}$ to present. The antipodal case falls under our formalism for appropriate values of the numerical parameters, but the additional constraint set $\mathcal{S}$ is somewhat different. On one hand, the $(1, \delta)$-spaces of order at least 3 with the distance δ present are already excluded by the antipodality condition, so we are only concerned with cliques. On the other hand, if K_n does not embed into an antipodal graph, then none of the graphs obtained by replacing k vertices of K_n by antipodal vertices can embed. So we make the following definition, which takes into account some further constraints when $\delta = 3$.

Definition 5.7. Let $\delta \geq 4$ be finite and $2 \leq n \leq \infty$, or $\delta = 3$ and $n = \infty$. Then

(1) $\mathcal{A}^\delta_a = \mathcal{A}^\delta_{1,\delta-1;\,2\delta+2,2\delta+1;\,\emptyset}$ is the set of finite integral metric spaces in which no triangle has perimeter greater than 2δ.
(2) $\mathcal{A}^\delta_{a,n}$ is the subset of $\mathcal{A}^\delta_a$ containing no subspace of the form $I_2^{\delta-1}[K_k, K_\ell]$ with $k + \ell = n$; here $I_2^{\delta-1}$ denotes a pair of vertices at distance $\delta - 1$ and $I_2^{\delta-1}[K_k, K_\ell]$ stands for the corresponding composition, namely a graph of the form $K_k \cup K_\ell$ with K_k, K_ℓ cliques (at distance 1), and $d(x, y) = \delta - 1$ for $x \in K_k$, $y \in K_\ell$. In particular, with $k = n$, $\ell = 0$, this means K_n does not occur.

One may make a general observation about the form of the admissibility conditions in Definition 5.6. Take $\mathcal{S} = \emptyset$, so that we consider $\mathcal{A}^\delta_{K_1,K_2;C_0,C_1;\emptyset}$ as a 5-parameter family of classes of finite metric structures. Looking over the conditions on the five parameters $\delta, K_1, K_2, C_0, C_1$ given above, we observe the following.

The condition $(*)$ "*$\mathcal{A}^\delta_{K_1,K_2;C_0,C_1}$ is an amalgamation class*"
is expressible in Presburger arithmetic.

If we knew this fact on *a priori* grounds we would then interpret the admissibility conditions as the result of expressing the property $(*)$ in quantifier free terms in a language suitable for quantifier elimination in Presburger arithmetic, namely a language permitting the formation of linear combinations of variables with integer coefficients, and with predicates for congruence modulo a fixed integer (above, parity will suffice).

To cast some light on this heuristically, consider the following three conditions, whose precise meaning will require some elucidation.

($C1$) The family of constraints defining $\mathcal{A}^{\delta}_{K_1,K_2;C_0,C_1;\emptyset}$ is uniformly definable in Presburger arithmetic, in the parameters $\delta, K_1, K_2, C_0, C_1$.

($C2$) For any fixed k, the condition *"The family $\mathcal{A}^{\delta}_{K_1,K_2;C_0,C_1;\emptyset}$ has the k-amalgamation property"* is definable in Presburger arithmetic as a property of $\delta, K_1, K_2, C_0, C_1$.

($C3$) The condition *"$\mathcal{A}^{\delta}_{K_1,K_2;C_0,C_1}$ is an amalgamation class"* is expressible in Presburger arithmetic as a property of $\delta, K_1, K_2, C_0, C_1$.

Let us clarify the meaning of these statements before considering their relationship. Since we vary δ, the language here is a binary language with predicates $R_i(x,y)$ for $i \in \mathbb{Z}$, $i \geq 0$. We consider only structures in which every ordered pair of points satisfies exactly one of the relations R_i (with R_0 the equality relation); we may restrict our attention to symmetric relations, but this is not essential.

The family of constraints defining $\mathcal{A}^{\delta}_{K_1,K_2;C_0,C_1;\emptyset}$ is the the set of minimal finite structures not in the class, which in the case at hand are structures of order at most 3, and include all pairs x, y violating the bound δ, all failures of symmetry, and all triples violating the triangle inequality, as well as the more specific constraints associated with K_1, K_2, C_0, C_1. We may identify a structure A whose universe is $\{1, \ldots, s\}$ with the s^2-tuple

$$(d(i,j) : 1 \leq i, j \leq s)$$

where we use the metric notation $d(i,j)$ to denote the unique subscript d such that $R_d(i,j)$ holds in A. With $n = s^2$ we arrange this n-tuple in a definite order. Then the constraints of order s associated with K_1, K_2, C_0, C_1 become a subset of $\mathbb{N}^n$ and with s bounded we are dealing with a finite number of such sets. So we may say that the set of constraints defining $\mathcal{A}^{\delta}_{K_1,K_2;C_0,C_1;\emptyset}$ is uniformly definable in Presburger arithmetic from the parameters if the encoded constraint sets in $\mathbb{N}^n$ are so definable (for $s = 2, 3$). This gives our first statement a definite sense.

In condition ($C2$) we consider the *k-amalgamation* property, which is the amalgamation condition for pairs of structures A_1, A_2 of orders $k' < k$, over a base of order $k' - 1$; in other words, we require the ability to complete amalgamation diagrams in which the relation of one pair of points a_1, a_2 remains to be determined, and the total number of points involved is at most k. In our second condition, we require that the set of parameter 5-tuples $(\delta, K_1, K_2, C_0, C_1)$ for which the associated class has the k-amalgamation property should be definable in Presburger arithmetic, for each fixed value of k.

The third condition is similar, but refers to the full amalgamation property.

These three conditions ($C1$)–($C3$) are all satisfied by our family $\mathcal{A}^{\delta}_{K_1,K_2;C_0,C_1;\emptyset}$, which has the additional property that it contains every amalgamation class of finite metric spaces which is determined by a set of constraints of order at most 3 (and 3 seems to be something of a magic number in terms of capturing a good deal of the sporadic side in classifications of homogeneous structures for finite binary languages). The following questions arise: is this kind of definability to be expected, not only in the present instance, but more generally; and could our analysis be materially simplified using automated techniques to perform an appropriate quantifier elimination in Presburger arithmetic? We do not have answers to these questions, but we point out the following: the first property is interesting, but evident in the case at hand; the first property implies the second; and the second makes the third seem highly probable. We add a word on each of these three points.

The uniform definability in Presburger arithmetic of the constraint sets associated with the classes $\mathcal{A}^{\delta}_{K_1,K_2;C_0,C_1;\emptyset}$ is evident by inspection of the definition, which is already in that form. In particular, the triangle inequality is expressed by a formula of Presburger arithmetic. We had expected to define an appropriate family of amalgamation classes by imposing conditions of triangles involving only the size and parity of their perimeters, but this was inadequate. Our first try, a 3-parameter family $\mathcal{A}^{\delta}_{K,C}$, had that simple form, but did not cover all relevant amalgamation classes. The modification required in going to the 5-parameter family brought an additional term of Presburger arithmetic into the definition. We do *not* have any heuristic which explains why we are able to define a single family of classes uniformly in Presburger arithmetic, in terms of a bounded number of numerical parameters, which captures all amalgamation classes defined by constraints of order at most 3.

Now let us consider the implication $(C1) \implies (C2)$. Consider the amalgamation property for amalgams of order *precisely* k in which the desired structure is completely determined apart from the relations holding between one pair of elements a, b. This amalgamation property is directly expressed by a formula of the following form, with $n = k^2$:

$$\forall x_1 \forall x_2 \cdots \forall x_{n-2}\, \exists y_1 \exists y_2 \phi(x_1, \ldots, x_{n-2}, y_1, y_2, \delta, K_1, K_2, C_0, C_1)$$

where ϕ simply states that the corresponding structure satisfies the required constraints, which will be expressible in Presburger arithmetic if condition $(C1)$ holds. Of course in symmetric structures we will require $y_1 = y_2$ and we may dispense with one existential quantifier (and several universal ones).

The displayed formula is very far from a quantifier free condition, and is not really useful for our purposes until it is expressed directly in terms of the parameters $\delta, K_1, K_2, C_0, C_1$. The admissibility conditions on the parameters given in Definition 5.6 express the 5-amalgamation property for the case of our particular family $\mathcal{A}^{\delta}_{K_1,K_2;C_0,C_1;\emptyset}$. Indeed, in deriving these conditions we considered only certain key amalgamation diagrams involving at most 5 elements; and eventually we found that we had enough conditions to derive the full amalgamation property. Thus the analysis that produced the admissibility conditions showed that they are equivalent to k-amalgamation for all $k \geq 5$. So at that point the obvious condition $(C1)$ explained the general form of our conditions, corresponding to $k = 5$ in condition $(C2)$, and we understood that our work to that point could be construed as carrying through an elimination procedure for a specific formula of Presburger arithmetic.

Of course, to pass from $(C2)$ to $(C3)$ in general requires finding a value of k_0 such that the quantifier-free form of condition $(C2)$ is independent of k for all $k \geq k_0$. It seems intrinsically reasonable that when we have a bound on the sizes of the constraints involved, there is also a bound k such that k-amalgamation implies full amalgamation. So there is some expectation that $(C2)$ will lead to $(C3)$, and thus that $(C1)$ will lead to $(C3)$.

Bearing in mind that Presburger arithmetic is decidable while diophantine problems over $\mathbb{Z}$ are in general undecidable, we may wonder whether the general classification problem for homogeneous structures in a finite relational language involves conditions which are largely expressible in Presburger arithmetic, or perhaps conditions of a more general diophantine type. One approach to this question would be to look more carefully at the examples tabulated in §14, giving 27 amalgamation classes determined by constraints on triangles whose Fraïssé limit is primitive, of which an impressive proportion (17) can be interpreted as metrically homogeneous graphs. We have just described a natural embedding of those 17 metrically homogeneous graphs into a 5-parameter family of graphs characterized by conditions expressible in Presburger arithmetic. One might ask for a natural extension of this 5-parameter family to another one given by similar conditions, in which all 27 examples for the case of 4 symmetric 2-types are found. Observe however that the collection of homogeneous structures for a given language is invariant under permutations of the language, while a useful definable relation in Presburger arithmetic should not be invariant under arbitrary permutations of the base set! So there is a prior issue of "symmetry breaking." Even the 17 examples which can be interpreted as metrically homogeneous graphs give rise to 20 examples after breaking the symmetry (§10.1).

5.5. Statement of Results. Our first two theorems can be taken as a summary of the way the catalog was constructed.

THEOREM 9. *Let $\delta \geq 3$, $1 \leq K_1 \leq K_2 \leq \delta$ or $K_1 = \infty$, $2\delta+1 \leq C_0, C_1 \leq 3\delta+2$ with C_0 even and C_1 odd, and $\mathcal{S}$ a set of $(1, \delta)$-spaces occurring in $\mathcal{A}^{\delta}_{K_1,K_2;C_0,C_1}$. Then the class*

$$\mathcal{A}^{\delta}_{K_1,K_2;C_0,C_1;\mathcal{S}}$$

is an amalgamation class if and only if the parameters $\delta, K_1, K_2, C_0, C_1, \mathcal{S}$ are admissible, or $K_1 = 1$, $K_2 = \delta - 1$, $C_0 = 2\delta + 2$, and $C_1 = 2\delta + 1$, and $\mathcal{S} = \emptyset$.

There is a point of notation that needs to be explained here, as it is possible for two sets of parameters to define the same class, with one set admissible and the other not. We make canonical choices of parameters, taking for example K_1 to be the least k for which there is an odd triangle of perimeter $2k + 1$, and K_2 the greatest, with values $K_1 = \infty$ and $K_2 = 0$ if there is none. In particular if $C_1 = 2\delta + 1$ this forces $K_2 < \delta$. One defines C_0, C_1 similarly as the least values strictly above 2δ such that perimeters of appropriate parity above the given bound are forbidden. Thus if there are no odd triangles, then $C_1 = 2\delta + 1$. And of course, we take only minimal constraints in $\mathcal{S}$ which are not consequences of the other constraints.

Note that when $K_1 = 1$, $K_2 = \delta - 1$, $C_0 = 2\delta + 2$, and $C_1 = 2\delta + 1$, we have the condition $\mathcal{S} = \emptyset$ here, but we also have the amalgamation classes $\mathcal{A}^{\delta}_{a,n}$ with the same parameters K_1, K_2, C_0, C_1 and a modified constraint set $\mathcal{S}$.

THEOREM 10. *Let Γ be a connected metrically homogeneous graph and suppose that Γ_1 is either imprimitive or contains no infinite independent set. Then Γ_1 is in our catalog under case I or II.*

We will prove Theorem 10 in §8, but our discussion of Theorem 9 in §6 is confined primarily to a detailed description of the amalgamation procedure. It takes some additional calculation to see that this procedure produces a metric meeting the required constraints in all cases.

Now let us turn to the case in which Γ is imprimitive. The first point is a very general fact which is quite familiar in the finite case but does not depend on finiteness, and which is also discussed in [**AMp10**].

FACT (cf. [**AH06**, Theorem 2.2]). *Let Γ be an imprimitive connected distance transitive graph of degree at least 3. Then Γ is either bipartite or antipodal (possibly both).*

The next result goes some distance toward bringing the antipodal case under control. It is proved in §7.2.

THEOREM 11. *Let Γ be a connected metrically homogeneous and antipodal graph, of diameter $\delta \geq 3$. Then for each vertex $u \in \Gamma$, there is a unique vertex $u' \in \Gamma$ at distance δ from u, and we have the "antipodal law"*

$$d(u,v) = \delta - d(u',v) \text{ for } u,v \in \Gamma$$

In particular, the map $u \mapsto u'$ is an automorphism of Γ.

In the study of finite distance transitive graphs, there are good reductions in both of the imprimitive cases—bipartite as well as antipodal—back (eventually) to the primitive case. The reduction in the bipartite case is straightforward in our context as well, but we will see that there is no simple reduction in the antipodal case, in the category of metrically homogeneous graphs. Still Theorem 11 suggests that the situation can be viewed as a modest elaboration beyond the primitive case.

We made some further explorations of the bipartite case, emphasizing extreme behavior, so as not to overlook anything obvious that would belong in the catalog. This led to the following results.

THEOREM 12. *Let Γ be a connected metrically homogeneous graph of odd diameter $\delta = 2\delta' + 1$ which is both antipodal and bipartite. Then $B\Gamma$ is connected, and Γ is the bipartite double cover of $B\Gamma$. The graph $B\Gamma$ is a metrically homogeneous graph with the following properties:*

1. *$B\Gamma$ has diameter δ';*
2. *No triangle in $B\Gamma$ has perimeter greater than $2\delta' + 1$;*
3. *$B\Gamma$ is not antipodal.*

Conversely, for any metrically homogeneous graph G with the three stated properties, there is a unique antipodal bipartite graph Γ of diameter $2\delta' + 1$ such that $B\Gamma \cong G$.

By inspection we find the following.

COROLLARY 12.1. *Let Γ be a connected metrically homogeneous graph of odd diameter $\delta = 2\delta' + 1$ which is both antipodal and bipartite. Suppose that $B\Gamma$ is in the catalog. Then Γ is in the catalog, and the relevant pairs $(B\Gamma, \Gamma)$ are as follows.*

1. *$B\Gamma$ complete of order n, $3 \leq n \leq \infty$, and Γ the bipartite double of an independent set of order $n-1$ (the bipartite complement of a perfect matching between two sets of order n).*
2. *(C_n, C_{2n}) with C_n an n-cycle, and n odd.*
3. *$(G_3^c, \Gamma^5_{\infty,0;12,13;\emptyset})$.*
4. *$(\Gamma^{\delta'}_{1,\delta';2\delta'+2,2\delta'+3;\emptyset}, \Gamma^{\delta}_{\infty,0;2\delta+2,2\delta+1;\emptyset})$ with $\delta' \geq 3$.*

In particular, for Γ antipodal and bipartite of odd diameter $\delta \leq 5$, since the Lachlan/Woodrow classification is complete through diameter 2, the corollary applies (Corollary 12.2, §7.2).

Another natural case in the bipartite context is the following. The structure of Γ_1 tells us nothing in the case of a bipartite graph when Γ_1 is infinite, but we may single out for special attention the metrically homogeneous bipartite graphs for which Γ_1 is infinite while $B\Gamma$ is itself exceptional. Here too everything conforms to the catalog.

THEOREM 13. *Let Γ be a connected, bipartite, and metrically homogeneous graph, of diameter at least 3, and degree at least 3, and with Γ_1 infinite. Then either $B\Gamma_1$ is isomorphic to the Rado graph, or $B\Gamma$ and Γ are in the catalog under one of the following headings.*

(1) *$B\Gamma \cong T_{\infty,\infty}$, and Γ is an infinitely branching tree $T_{2,\infty}$.*
(2) *$B\Gamma \cong K_\infty$, and Γ has diameter 3, with Γ either the complement of a perfect matching, or the generic bipartite graph $\Gamma^3_{\infty,0;10,7;\emptyset}$.*
(3) *$B\Gamma \cong K_\infty[I_2]$, Γ has diameter 4, and $\Gamma \cong \Gamma^4_{\infty,0;10,9;\emptyset}$ is the generic antipodal bipartite graph of diameter 4.*
(4) *$B\Gamma \cong G^c_n$, the complement of a Henson graph, for some $n \geq 3$, and $\Gamma \cong \Gamma^4_{\infty,0;14,9;\{I^{(4)}_n\}}$ the generic bipartite graph in which there is no set of n vertices which are pairwise at distance 4.*
(5) *$B\Gamma \cong G^c_3$, $\Gamma \cong \Gamma^5_{\infty,0;12,11;\emptyset}$ antipodal bipartite of diameter 5 as in Corollary 12.1.*

One could explore the imprimitive case more, and perhaps find something else of a similar sort that needs to be added to the catalog. But after that the question becomes how to approach a proof of its completeness. This is governed by the structure of the catalog. To show that the catalog is complete, we would need to show on the one hand that any example not falling under one of the exceptional headings has the same constraints on triangles as one of our standard amalgamation classes, and then that for each metrically homogeneous graphs with one of the specified patterns of forbidden triangles, any further minimal forbidden configurations must be $(1, \delta)$-spaces (or the analogous constraints appropriate to the antipodal case). Some instances of the latter type of analysis were given in [**AMp10**].

It is no doubt useful to divide the classification problem into the two parts just mentioned, but both parts appear challenging, and they are likely to become intertwined in an inductive treatment of the problem.

In proving the completeness of the catalog one first takes δ finite. There is something more to be checked when δ is infinite. Once one has a degree of control on the triangles involved it becomes easy to reduce from the case of infinite diameter to finite diameter. And that degree of control may well follow quickly from the classification in finite diameter, but we have not looked closely at that.

The rest of this Part is devoted to the proofs of Theorems 10 and 11–13. In the case of Theorem 9 we will confine ourselves to an indication of the amalgamation procedure used when the parameters are admissible.

6. Constructions of metrically homogeneous graphs

We are concerned here with the use of Fraïssé constructions to produce metrically homogeneous graphs of generic type, including some bipartite and antipodal cases.

6.1. The Main Construction. Recall that in the definition of the class $\mathcal{A}^\delta_{K_1,K_2;C_0,C_1;\mathcal{S}}$ of finite integral metric spaces of diameter δ with constraints K_1, K_2, C_0, C_1, and $\mathcal{S}$, the parameter K_1 controls the presence of triangles with small odd perimeter, the parameters C_0, C_1 control those of large perimeter, $\mathcal{S}$ is a set of $(1,\delta)$-spaces, and the parameter K_2 is used to further control the presence of odd triangles, excluding all those whose perimeter P satisfies

$$P > 2K_2 + d(a,b)$$

for some choice of a, b in the triangle. If $K_2 = \delta$ then this condition is vacuous. The admissible combinations of these parameters are described in Definition 5.6. In that presentation the parity of C_0 and C_1 plays less of a role than their relative sizes, so we introduce the notation $C = \min(C_0, C_1)$ and $C' = \max(C_0, C_1)$. The simplest cases are those in which $C' = C + 1$: we exclude all triangles of perimeter C or greater. We use the values $3\delta+1$ or $3\delta+2$ (with appropriate parity) to indicate that the constraints corresponding to C_0 or C_1 are absent.

The main result concerning amalgamation for classes of type $\mathcal{A}^\delta_{K_1,K_2;C_0,C_1;\mathcal{S}}$ can then be phrased as follows.

Theorem (9). *Let $\delta \geq 3$, $1 \leq K_1 \leq K_2 \leq \delta$ or $K_1 = \infty$, $2\delta+1 \leq C_0, C_1 \leq 3\delta+2$ with C_0 even and C_1 odd, and $\mathcal{S}$ a set of $(1,\delta)$-spaces occurring in $\mathcal{A}^\delta_{K_1,K_2;\,C_0,C_1}$. Then the class*

$$\mathcal{A}^\delta_{K_1,K_2;\,C_0,C_1;\mathcal{S}}$$

is an amalgamation class if and only if the parameters $\delta, K_1, K_2, C_0, C_1, \mathcal{S}$ are admissible.

There is also a variation associated with the antipodal case and a modified constraint set $\mathcal{S}$ (Theorem 14, below).

The proof that the admissibility constraint is *necessary* (for classes specified in this way) requires many constructions of specific amalgamation diagrams, mostly of order 5, which have no completion when the parameters lie outside the admissible range. The proof that in the admissible cases amalgamation can be carried out involves the specification of a particular amalgamation procedure varying in detail according to the different cases occurring within the definition of admissibility, and depending further on some details of the amalgamation problem itself, followed by additional computation to verify that the proposed procedure actually works. We will describe the procedure itself in detail and leave the rest for another occasion.

For the proof of amalgamation, it suffices to consider amalgamation diagrams of the special form $A_1 = A_0 \cup \{a_1\}$, $A_2 = A_0 \cup \{a_2\}$, that is with only one distance $d(a_1, a_2)$ needing to be determined; we call these *2-point* amalgamations. Furthermore, any distance lying between

$$d^-(a_1,a_2) = \max_{x \in A_0}(|d(a_1,x) - d(a_2,x)|)$$

and

$$d^+(a_1,a_2) = \min_{x \in A_0}(d(a_1,x) + d(a_2,x))$$

will give at least a pseudometric (and if $d^-(a_1, a_2) = 0$, we will identify a_1 and a_2).

When $C' = C + 1$ we have the requirement that all triangles have perimeter at most $C - 1$ and therefore we consider a third value

$$\tilde{d}(a_1, a_2) = \min_{x \in A_0} (C - 1 - [d(a_1, x) + d(a_2, x)])$$

In this case the distance $i = d(a_1, a_2)$ must satisfy:

$$d^-(a_1, a_2) \leq i \leq \min(d^+(a_1, a_2), \tilde{d}(a_1, a_2))$$

Similarly when $K_1 = \infty$ then as there are no odd triangles we modify the definition of $\tilde{d}$ as follows:

$$\tilde{d}(a_1, a_2) = \min_{x \in A_0} (C_0 - 2 - [d(a_1, x) + d(a_2, x)])$$

The amalgamation procedure is then given by the following rules for completing a 2-point amalgamation diagram $A_i = A_0 \cup \{a_i\}$ $(i = 1, 2)$. We will assume throughout that $d^-(a_1, a_2) > 0$, as otherwise we may simply identify a_1 with a_2. We will also write:

$$i^- = d^-(a_1, a_2),\ i^+ = d^+(a_1, a_2),\ \tilde{\imath} = \tilde{d}(a_1, a_2)$$

and we seek a suitable value for $i = d(a_1, a_2)$.

(1) If $K_1 = \infty$:

Then the parity of $d(a_1, x) + d(a_2, x)$ is independent of the choice of $x \in A_0$.

If $\mathcal{S}$ is empty then any value i with $i^- \leq i \leq i^+$ and of the correct parity will do.

If $\mathcal{S}$ is nonempty (and irredundant) then $\mathcal{S}$ consists of a δ-clique, and δ is even. In particular $\delta \geq 4$. In this case take $d(a_1, a_2) = i$ with $1 < i < \delta$ and with i of the correct parity. We may take $i = i^+$ if this is less than δ, $i = i^-$ if this is greater than 1, and otherwise, when $i^- \leq 1$ and $i^+ \geq \delta$, take $i = 2$ or 3 of the correct parity.

(2) If $K_1 < \infty$ and $C \leq 2\delta + K_1$:
- (a) If $C' = C + 1$ then:
  - (i) If $\min(i^+, \tilde{\imath}) \leq K_2$ let $d(a_1, a_2) = \min(i^+, \tilde{\imath})$. Otherwise:
  - (ii) If $i^- \geq K_1$ let $d(a_1, a_2) = i^-$. Otherwise:
  - (iii) Let $d(a_1, a_2) = K_2$.
- (b) If $C' > C + 1$ then:
  - (i) If $i^+ < K_2$ let $d(a_1, a_2) = i^+$. Otherwise:
  - (ii) If $d^- > K_2$ let $d(a_1, a_2) = i^-$. Otherwise:
  - (iii) Take $d(a_1, a_2) = K_2$ unless there is $x \in A_0$ with $d(a_1, x) = d(a_2, x) = \delta$, in which case take $d(a_1, a_2) = K_2 - 1$.

(3) If $K_1 < \infty$ and $C > 2\delta + K_1$:
- (a) If $i^- > K_1$, let $d(a_1, a_2) = i^-$.

(b) Otherwise:
  (i) If $C' = C + 1$:
    (A) If $i^+ \le K_1$ let $d(a_1, a_2) = \min(i^+, \tilde{\imath})$. Otherwise:
    (B) Let $d(a_1, a_2) = K_1$ unless we have one of the following:
      There is $x \in A_0$ with $d(a_1, x) = d(a_2, x)$, and $K_1 + 2K_2 = 2\delta - 1$; or $K_1 = 1$.
      In these cases, take $d(a_1, a_2) = K_1 + 1$.
  (ii) If $C' > C + 1$:
    If $i^+ < K_2$ let $d(a_1, a_2) = i^+$.
    Otherwise, let $d(a_1, a_2) = \min(K_2, C - 2\delta - 1)$.

We note some extreme cases. With $K_1 = \infty$ we are dealing with bipartite graphs; with $C = 2\delta + 1$ we are dealing with antipodal graphs. With $K_1 > 1$ we have the case $\Gamma_1 \cong I_\infty$; with $K_1 = 1$ and $\mathcal{S} = \{K_n\}$, we have the case of Γ_1 the generic K_n-free graph.

LEMMA 6.1. *Let Γ be a metrically homogeneous graph of diameter δ. Then Γ is antipodal if and only if no triangle has perimeter greater than 2δ.*

PROOF. The bound on perimeter immediately implies antipodality.

For the converse, let (a, b, c) be a triangle with $d(a, b) = i$, $d(a, c) = j$, $d(b, c) = k$, and let a' be the antipodal point to a. Then the triangle (a', b, c) has distances $\delta - i$, $\delta - j$, and k, and the triangle inequality yields $i + j + k \le 2\delta$, as claimed. □

6.2. An Antipodal Variation. We consider modifications of our definitions which allow us to include constraints on cliques when the associated graph is antipodal and $K_1 = 1$. In this case the associated amalgamation class may require constraints which are neither triangles nor $(1, \delta)$-spaces.

DEFINITION 6.2. Let $\delta \ge 4$ be finite and $2 \le n \le \infty$. Then

(1) $\mathcal{A}^\delta_a = \mathcal{A}^\delta_{1,\delta-1;\,2\delta+2,2\delta+1;\,\emptyset}$ is the set of finite metric spaces in which no triangle has perimeter greater than 2δ.
(2) $\mathcal{A}^\delta_{a,n}$ is the subset of $\mathcal{A}^\delta_a$ containing no subspace of the form $K_k \cup K_\ell$ with K_k, K_ℓ cliques, $k + \ell = n$, and $d(x, y) = \delta - 1$ for $x \in K_k$, $y \in K_\ell$. In particular, K_n does not occur.

THEOREM 14. *If $\delta \ge 4$ is finite and $2 \le n \le \infty$, then $\mathcal{A}^\delta_{a,n}$ is an amalgamation class. If $n \ge 3$ then the associated Fraïssé limit is a connected antipodal metrically homogeneous graph which is said to be generic for the specified constraints.*

Here the parameter n stands in place of the set $\mathcal{S}$; since there are no triangles of perimeter greater than 2δ, the only relevant $(1, \delta)$-spaces are 1-cliques. In these graphs Γ_1 is the generic graph omitting K_n.

We will give the proof of amalgamation for these particular classes in detail. The following lemma is helpful.

LEMMA 6.3. *Let δ be fixed, and let A be a finite metric space with no triangle of perimeter greater than δ. Then there is a unique "antipodal" extension $\hat{A}$ of A, up to isometry, to a metric space satisfying the same condition, in which every vertex is paired with an antipodal vertex at distance δ, and every vertex not in A is antipodal to one in A.*

If A is in $\mathcal{A}^\delta_{a,n}$, then $\hat{A}$ is in the same class.

PROOF. The uniqueness is clear: let

$$B = \{a \in A : \text{There is no } a' \in A \text{ with } d(a, a') = \delta\}$$

and introduce a set of new vertices $B' = \{b' : b \in B\}$. Let $\hat{A} = A \cup B'$ as a set. Then there is a unique symmetric function on $\hat{A}$ extending the metric on A, with $d(x, b') = \delta - d(x, b)$ for $x \in \hat{A}$, $b \in B$.

So the issue is one of existence, and for that we may consider the problem of extending A one vertex at a time, that is to $A \cup \{b'\}$ with $b \in B$, as the rest follows by induction.

We need to show that the canonical extension of the metric on A to a function d on $A \cup \{b'\}$ is in fact a metric, satisfies the antipodal law for δ, and also satisfies the constraints corresponding to n.

The triangle inequality for triples (b', a, c) or (a, b', c) corresponds to the ordinary triangle inequality for (b, a, c) or the bound on perimeter for (a, b, c) respectively, and the bound on perimeter for triangles (a, b', c) follows from the triangle inequality for (a, b, c).

Now suppose $n < \infty$ and b' belongs to a configuration $K_k \cup K_\ell$ with $k + \ell = n$ and $d(x, y) = \delta - 1$ for $x \in K_k$, $y \in K_\ell$. We may suppose that $b' \in K_k$; then $K_k \setminus \{b'\} \cup (K_\ell \cup \{b\})$ provides a copy of $K_{k-1} \cup K_{\ell+1}$ of forbidden type. □

LEMMA 6.4. *If $\delta \geq 4$ is finite, $2 \leq n \leq \infty$, then $\mathcal{A}^\delta_{a,n}$ is an amalgamation class.*

PROOF. We consider a two-point amalgam with $A_i = A_0 \cup \{a_i\}$, $i = 1, 2$. If $d(a_2, x) = \delta$ for some $x \in A_0$ then there is a canonical amalgam $A_1 \cup A_2$ embedded in $\hat{A}_1$. So we will suppose $d(a_i, x) < \delta$ for $i = 1, 2$ and $x \in A_0$. Then applying Lemma 6.3 to A_1 and A_2, we may suppose that for every vertex $v \in A_0$ there is an antipodal vertex $v' \in A_0$.

We claim that any metric d on $A_1 \cup A_2$ extending the given metrics d^i on A_i will satisfy the antipodal law for δ. So with d such a metric, consider a triangle of the form (a_0, a_1, a_2) with $a_0 \in A_0$. By the triangle inequality for (a_1, a_0', a_2) we have

$$d(a_1, a_2) \leq 2\delta - [d(a_1, a_0) + d(a_2, a_0)]$$

and this is the desired bound on perimeter.

We know by our general analysis that any value r for $d(a_1, a_2)$ with

$$d^-(a_1, a_2) \leq r \leq d^+(a_1, a_2)$$

will give us a metric, and in the bipartite case we will want r to have the same parity as $d^-(a_1, a_2)$ (or equivalently, $d^+(a_1, a_2)$).

To deal with the the constraints involving the parameter n, it is sufficient to avoid the values $r = 1$ and $r = \delta - 1$. But $d^+(a_1, a_2) > 1$, and $d^-(a_1, a_2) < \delta - 1$, so we may take r equal to one of these two values unless we have

$$d^-(a_1, a_2) = 1, d^+(a_1, a_2) \geq \delta - 1$$

In this case, we take r to be some intermediate value, and as $\delta > 3$, there is one. □

7. Imprimitive Graphs

7.1. Smith's Theorem. We now turn to Smith's Theorem, which provides a general analysis of the imprimitive case, following [**AH06**] (cf. [**BCN89, Smi71**]). This result applies to imprimitive distance transitive graphs (that is, the homogeneity condition is assumed to hold for pairs of vertices), and even more generally in the finite case. There are three points to this theory in the finite case: (1) the imprimitive graphs are of two extreme types, bipartite or antipodal; (2) associated with each type there is a reduction (folding or halving) to a potentially simpler graph; (3) with few exceptions, the reduced graph is primitive. Among the exceptions that need to be examined are the graphs which are both antipodal and bipartite. As our hypothesis of metric homogeneity is *not* preserved by the folding operation in general, we lose a good deal of the force of (2) and a corresponding part of (3). On the other hand, we will see that metric homogeneity implies that with trivial exceptions, in antipodal graphs the antipodal equivalence classes have order two. In that sense, antipodal graphs are not far from primitive, but there is no simple reduction to the primitive case, and the full classification is best thought of as a variation on the primitive case, to be handled by similar methods.

We will first take up the explicit form of Smith's Theorem given in [**AH06**], restricting ourselves to the distance transitive case. If Γ is a distance transitive graph, then any binary relation R invariant under $\mathrm{Aut}(\Gamma)$ is a union of relations R_i defined by $d(x,y)=i$, $R=\bigcup_{i\in I} R_i$. We denote by $\langle t\rangle$ the union $\bigcup_{t|i} R_i$ taken over the multiples of t. The first point is the following.

FACT 7.1 (cf. [**AH06**, Theorem 2.2]). *Let Γ be a connected distance transitive graph of diameter δ, and let E be a congruence of Γ.*

(1) *$E=\langle t\rangle$ for some t.*
(2) *If $2<t<\delta$, then Γ has degree 2.*
(3) *If $t=2$ then either Γ is bipartite, or Γ is a complete regular multipartite graph, of diameter 2.*

In particular, if the degree of Γ is at least 3, then Γ is either bipartite or antipodal (and possibly both).

Of course, the exceptional case of diameter 2 has already been noticed within the Lachlan/Woodrow classification, where it occurs as the complement of $m\cdot K_n$, with $m,n\le\infty$.

If Γ is a connected distance transitive bipartite graph, we write $B\Gamma$ for the graph induced on either of the two equivalence classes for the congruence $\langle 2\rangle$; these are isomorphic, with respect to the edge relation R_2: $d(x,y)=2$. This is called a *halved graph* for Γ, and Γ is a *doubling* of $B\Gamma$. If Γ is a connected distance transitive antipodal graph of diameter δ (necessarily finite), then $A\Gamma$ denotes the graph induced on the quotient Γ/R_δ by the edge relation: C_1 is adjacent to C_2 iff there are $u_i\in C_i$ with (u_1,u_2) an edge of Γ. This is called a *folding* of Γ, and Γ is called an *antipodal cover* of $A\Gamma$. In our context, the halving construction is more useful than the folding construction.

FACT 7.2 (cf. [**AH06**, Theorem 2.3]). *Let Γ be a connected metrically homogeneous bipartite graph. Then $B\Gamma$ is metrically homogeneous.*

PROOF. Since $\mathrm{Aut}(\Gamma)$ preserves the equivalence relation whose classes are the two halves of Γ, the homogeneity condition is inherited by each half. □

Some insight into the folding construction is afforded by the following.

LEMMA 7.3. *Let Γ be a connected distance transitive antipodal graph of diameter δ, and let C_1,C_2 be two equivalence classes for the antipodality relation R_δ. Then the set of distances $d(u,v)$ for $u\in C_1$, $v\in C_2$ is a pair of the form $\{i,\delta-i\}$ (which is actually a singleton if $i=\delta/2$ with δ even).*

Proof. Since there are geodesics (u, v, u') with $d(u, v) = i$, $d(v, u') = \delta - i$, and $d(u, u') = \delta$, whenever we have $d(u, v) = i$ we also have $d(u', v) = \delta - i$ for some u' antipodal to u, by distance transitivity.

We claim that for $u \in C_1$, $v, w \in C_2$, with $d(u, v) = i$, $d(u, w) = j$ and $i, j \leq \delta/2$, we have $i = j$. If $i < j$, then $d(v, w) \leq i + j < \delta$ and hence $v = w$, a contradiction.

Thus for $u \in C_1$ the set of distances $d(u, v)$ with v in C_2 has the form $\{i, \delta - i\}$, and for v in C_2 the same applies with respect to C_1, with the same pair of values. This implies our claim. □

Corollary 7.3.1 (cf. [**AH06**, Proposition 2.4]). *Let Γ be a connected distance transitive antipodal graph of diameter δ, and consider the graph $A\Gamma$ as a metric space. If $u, v \in \Gamma$ with $d(u, v) = i$, then in $A\Gamma$ the corresponding points $\bar{u}$, $\bar{v}$ lie at distance* $\min(i, \delta - i)$.

Proof. Replacing v by v' with $d(u, v) = \delta - i$, we may suppose $i = \min(i, \delta - i)$. We have $d(\bar{u}, \bar{v}) \leq d(u, v)$.

Let $j = d(\bar{u}, \bar{v})$ and lift a path of length j from $\bar{u}$ to $\bar{v}$ to a walk $(u, \ldots, v^*)$ in Γ. If $v^* = v$ then $d(u, v) \leq j$ and we are done. Otherwise, $\delta = d(v, v^*) \leq d(v^*, u) + d(u, v) \leq j + i$, and as $i, j \leq \delta/2$, we find $i = j = \delta/2$. □

This implies in particular that the folding of an antipodal metrically homogeneous graph of diameter at most 3 is complete, and that the folding of any connected distance transitive antipodal graph is distance transitive. But now consider the "generic" antipodal graph $\Gamma^{\delta}_{1,\delta-1;2\delta+2,2\delta+1;\emptyset}$ of diameter $\delta \geq 4$. Let $P = (u_0, \ldots, u_\delta)$ be a geodesic in Γ, and let $C = (v_0, \ldots, v_\delta, v_0)$ be an isometrically embedded cycle. On P we have $d(u_i, u_j) = |i - j|$ and on C we have $d(v_i, v_j) = \min(|i - j|, \delta - |i - j|)$, so the images of these graphs in $A\Gamma$ are isometric. We claim however that there is no automorphism of $A\Gamma$ taking one to the other.

Let $\delta_1 = \lfloor \delta/2 \rfloor$ and $\delta' = \lceil \delta_1/2 \rceil$. Then there is a vertex w_C in Γ at distance precisely δ' from every vertex of C. Hence the same applies in the folded graph $A\Gamma$ to the image of C. We claim that this does not hold for the image of P. Supposing the contrary, there would be a vertex w_P in Γ whose distance from each vertex of P is either δ' or $\delta - \delta'$. On the other hand, the distances $d(w_P, u_i)$, $d(w_P, u_{i+1})$ can differ by at most 1, and $\delta - \delta' > \delta' + 1$ since $\delta \geq 4$, so the distance $d(w_P, u_i)$ must be independent of i. However $d(u_0, w_P) = \delta - d(u_\delta, w_P)$, which would mean $\delta = 2\delta'$, while in fact $\delta > 2\delta'$.

Still, we can get a decent grasp of the antipodal case in another way.

7.2. The Antipodal Case. All graphs considered under this heading are connected and of finite diameter. Our main goal is the following.

Theorem (11). *Let Γ be a connected metrically homogeneous and antipodal graph, of diameter $\delta \geq 3$. Then for each vertex $u \in \Gamma$, there is a unique vertex $u' \in \Gamma$ at distance δ from u, and we have the law*

$$d(u, v) = \delta - d(u', v)$$

for $u, v \in \Gamma$. In particular, the map $u \mapsto u'$ is an automorphism of Γ.

For $v \in \Gamma$, $\Gamma_i(v)$ denotes the graph induced on the vertices at distance i from v, and since the isomorphism type is independent of v, this will sometimes be denoted simply by Γ_i, when the choice of v is immaterial. In particular, $\Gamma_\delta \cong I_n^{(\delta)}$, a set of n vertices mutually at distance δ. Our claim is that $n = 1$.

We begin with a variation of Lemma 7.3.

LEMMA 7.4. *Let Γ be metrically homogeneous and antipodal, of diameter δ. Suppose $u, u' \in \Gamma$, and $d(u, u') = \delta$. Then for $i < \delta/2$, the relation R_δ defines a bijection between $\Gamma_i(u)$ and $\Gamma_i(u')$, while $\Gamma_{\delta/2}(u) = \Gamma_{\delta/2}(u')$.*

PROOF. First suppose $i < \delta/2$, and $v \in \Gamma_i(u)$. We work with the equivalence classes C_1, C_2 of u and v respectively, with respect to the relation R_δ. As $d(u, v) = i$, $d(u, u') = \delta$, $i \le \delta/2$, and $d(v, u') \in \{i, \delta - i\}$, we have $d(v, u') = \delta - i$.

Now (v, u') extends to a geodesic (v, u', v') with $d(v, v') = \delta$, $d(u', v') = i$, and we claim that v' is unique. If (v, u', v'') is a second such geodesic then we have $d(v, v') = d(v, v'') = \delta$ and $d(v', v'') \le 2(\delta - i) < \delta$, so $v' = v''$.

Thus we have a well-defined function from $\Gamma_i(u)$ to $\Gamma_i(u')$, and interchanging u, u' we see that this is a bijection.

Now if $i = \delta/2$, apply Lemma 7.3: with $i = \delta/2$, the set $\{i, \delta - i\}$ is a singleton. $\square$

Taking $n > 1$, we will first eliminate some small values of δ.

LEMMA 7.5. *Let Γ be a metrically homogeneous antipodal graph, of diameter $\delta \ge 3$, and let $\Gamma_\delta \cong K_n^{(\delta)}$ with $1 < n \le \infty$. Then $\delta \ge 5$.*

PROOF. First suppose that $\delta = 3$. Fix a basepoint in Γ. Define an equivalence relation $\sim$ on Γ_2 by

$$x \sim y \text{ iff there is a vertex } u \in \Gamma_3 \text{ with } x, y \in \Gamma_1(u)$$

By homogeneity, any two vertices of Γ_2 at distance 2 are in the same class. If C_1, C_2 are two distinct equivalence classes in Γ_2 then each vertex of C_1 is at distance 3 from at most one vertex of C_2, and is adjacent to the remainder. In particular there are some edges between C_1 and C_2, and therefore there are none within C_1 or C_2, in other words Γ_1 is an independent set, and Γ omits K_3.

Let $a_1, a_2 \in \Gamma_3$ be distinct. We have shown that for any vertex u in $\Gamma_1(a_1)$ there is at most one vertex in $\Gamma_1(a_2)$ not adjacent to u. Taking a_3 a vertex at distance 3 from a_1, a_2, we may take $u_1 \in \Gamma_1(a_1)$, $u_2 \in \Gamma_1(a_2)$ adjacent. Then the number of vertices of $\Gamma_1(a_3)$ not adjacent to both u_1 and u_2 is at most 2, but since Γ_3 is triangle free, there are no vertices adjacent to both u_1 and u_2, so $|\Gamma_1| \le 2$. But then Γ is a cycle, and this contradicts $n > 1$.

So now suppose the diameter is 4. Take $a_1, a_2, a_3 \in \Gamma$ pairwise at distance 3, and $u_1, v_1 \in \Gamma_1(a_1)$ with $d(u_1, v_1) = 2$. Then the distances occurring in the triangle (a_3, u_1, v_1) are $3, 3, 2$.

Consider $v_2 \in \Gamma_1(a_2)$ with $d(v_1, v_2) = 4$. As $d(u_1, v_1) = 2$ and $2 = \delta/2$ we find that $d(u_1, v_2) = 2$. Hence the triangles (a_3, u_1, v_1) and (a_3, u_1, v_2) are isometric. By metric homogeneity there is an isometry carrying (a_3, u_1, v_1, a_1) to (a_3, u_1, v_2, b_1) for some b_1. Then $u_1, v_2 \in \Gamma_1(b_1)$ and $d(b_1, a_3) = 4$. But then $d(a_1, b_1), d(a_2, b_1) \le 2$ while a_1, a_2, a_3, b_1 are all in the same antipodality class, forcing $a_1 = b_1 = a_2$, a contradiction. $\square$

Now we need to extend Lemma 7.3 to some cases involving distances which may be greater than $\delta/2$.

LEMMA 7.6. *Let Γ be metrically homogeneous and antipodal, of diameter $\delta \geq 3$, and let $\Gamma_\delta \cong K_n^{(\delta)}$ with $1 < n \leq \infty$. Suppose $d(a, a') = \delta$ and $i < \delta/2$. Suppose $u \in \Gamma_i(a)$, $u' \in \Gamma_i(a')$, with $d(u, u') = \delta$. If $v \in \Gamma_i(a)$ and $d(u, v) = 2i$, then $d(u', v) = \delta - 2i$.*

PROOF. We have $d(a, u') = \delta - i$. Take $v_0 \in \Gamma_i(a)$ so that (a, v_0, u') is a geodesic, that is $d(v_0, u') = \delta - 2i$. As $u, v_0 \in \Gamma_i(a)$ we have $d(u, v_0) \leq 2i$. On the other hand $d(u, u') = \delta$ and $d(v_0, u') = \delta - 2i$, so $d(u, v_0) \geq 2i$. Thus $d(u, v_0) = 2i$.

So we have at least one triple (a, u, v_0) with $v_0 \in \Gamma_i(a)$, $d(u, v_0) = 2i$, and with $d(v_0, u') = \delta - 2i$. Let (a, u, v) be any triple isometric to (a, u, v_0). Then the quadruples (a, u, v, a') and (a, u, v_0, a') are also isometric since $u, v, v_0 \in \Gamma_i(a)$ with $i < \delta/2$. But as a, u together determine u', we then have (a, u, v, a', u') and (a, u, v_0, a', u') isometric, and in particular $d(v, u') = \delta - 2i$. □

After these preliminaries we can prove Theorem 11.

PROOF. We show that $n = 1$, after which the rest follows directly since if u determines u', then $d(u, v)$ must determine $d(u', v)$.

By Lemma 7.5 we may suppose that $\delta \geq 5$. We fix a_1, a_2, a_3 at mutual distance δ, and fix $i < \delta/2$, to be determined more precisely later.

Take $u_1, v_1 \in \Gamma_i(a_1)$ with $d(u_1, v_1) = 2i$, and then correspondingly $u_2, v_2 \in \Gamma_i(a_2)$, $u_3, v_3 \in \Gamma_i(a_3)$, with u_1, u_2, u_3 and v_1, v_2, v_3 triples of vertices at mutual distance δ.

Now $d(u_1, v_3) = \delta - 2i$, and $d(u_1, u_3) = \delta$, so as usual $d(u_3, v_3) = 2i$. We now consider the following property of the triple (u_1, v_1, a_3): *For $v_3 \in \Gamma_i(a_3)$ with $d(v_1, v_3) = \delta$, we have $d(u_1, v_3) = \delta - 2i$.* The triple (u_1, v_1, a_3) is isometric with (v_3, u_3, a_2). It follows that $d(v_3, u_2) = \delta - 2i$. So

$$\delta = d(u_1, u_2) \leq 2(\delta - 2i)$$

This shows that $i \leq \delta/4$, so for a contradiction we require

$$\delta/4 < i < \delta/2$$

and for $\delta > 4$ this is possible. □

We now give the classification of metrically homogeneous antipodal graphs of diameter 3; this is also treated in [**AMp10**].

THEOREM 15. *Let G be one of the following graphs: the pentagon (5-cycle), the line graph $E(K_{3,3})$ for the complete bipartite graph $K_{3,3}$, an independent set I_n ($n \leq \infty$), or the random graph Γ_∞. Let G^* be the graph obtained from G by adjoining an additional vertex adjacent to all vertices of G, and let Γ be the graph obtained by taking two copies H_1, H_2 of G^*, with a fixed isomorphism $u \mapsto u'$ between them, and with additional edges (u, v') or (v', u), for $u, v \in H_1$, just when (u, v) is not an edge of H_1. Then Γ is a homogeneous antipodal graph of diameter 3 with pairing the given isomorphism $u \mapsto u'$. Conversely, any connected metrically homogeneous antipodal graph of diameter 3 is of this form.*

PROOF. Let Γ be connected, metrically homogeneous, and antipodal, of diameter 3. Fix a basepoint $* \in \Gamma$ and let $G = \Gamma_1(*)$. Then G is a homogeneous graph, which may be found in the Lachlan/Woodrow catalog given in §4.

Let $H_1 = G \cup \{*\}$. Then the pairing $u \leftrightarrow u'$ on Γ gives an isomorphism of H_1 with $H_2 = \Gamma_1(*') \cup \{*'\}$. Furthermore, for $u, v \in \Gamma_1(*)$, we have $d(u, v') = \delta - d(u, v)$, so the edge rule in Γ is the one we have described. It remains to identify the set of homogeneous graphs G for which the associated graph Γ is metrically homogeneous.

We claim that for any homogeneous graph G, the associated graph Γ has the following homogeneity property: if A, B are finite subgraphs of Γ both containing the point $*$, then any isometry $A \to B$ fixing $*$ extends to an automorphism of Γ. Given such A, B, we first extend to $\hat{A}, \hat{B}$ by closing under the pairing $u \leftrightarrow u'$, then reduce to G by taking $\tilde{A} = \hat{A} \cap G$, $\tilde{B} = \hat{B} \cap G$. Then apply the homogeneity of G to get an isometry extending the given one on $\tilde{A}$ to all of H_1, fixing $*$, which then extends canonically to Γ. It is easy to see that this agrees with the given isometry on A.

This homogeneity condition implies that for such graphs Γ, the graph will be metrically homogeneous if and only if $\mathrm{Aut}(\Gamma)$ is transitive on vertices. For any of these graphs Γ, whether metrically homogeneous or not, we have the pairing $u \leftrightarrow u'$. Furthermore, we can reconstruct Γ from $\Gamma_1(v)$ for any $v \in \Gamma$. So the homogeneity reduces to this: $\Gamma_1(v) \cong G$ for $v \in G = \Gamma_1(*)$; here we use the pairing to reduce to the case $v \in G$.

Now $\Gamma_1(v)$ is the graph obtained from the vertex $*$, the graph $G_1(v)$ induced on the neighbors of v in G, and the graph $G_2(v)$ induced on the non-neighbors of v in G, by taking the neighbors of $*$ to be $G_1(v)$, and switching the edges and nonedges between $G_1(v)$ and $G_2(v)$. Another way to view this would be to replace v by $*$, and then perform the switching between $G_1(v)$ and $G_2(v)$. So it is really only the latter that concerns us.

We go through the catalog. In the degenerate cases, with G complete or independent, there is no switching, so the corresponding graph Γ is homogeneous. But when G is complete this graph is not connected, so we set that case aside.

When G is imprimitive, we switch edges and non-edges between the equivalence classes not containing the fixed vertex v, and the vertices in the equivalence class of v other than v itself. As a result, the new graph becomes connected with respect to the equivalence relation on G, so this certainly does not work.

When G is primitive, nondegenerate, and finite, we have just the two examples mentioned above for which the construction does work, by inspection.

Lastly, we consider the Henson graphs $G = \Gamma_n$, generic omitting K_n, their complements, and the random graph Γ_∞. The Henson graphs and their complements will not work here. For example, if $G = \Gamma_n$, then $G_2(v)$ contains K_{n-1}, and switching edges and nonedges with $G_1(v)$ will extend this to K_n. The complementary case is the same. So we are left with the case of the Rado graph. This is characterized by extension properties, and it suffices to check that these still hold after performing the indicated switch; and using the vertex v as an additional parameter, this is clear. □

There is a good deal more to the general analysis of [**AH06**], Proposition 2.5 through Corollary 2.10, all with some parallels in our case, but the main examples in the finite case do not satisfy our conditions, while the main examples in our case have no finite analogs, so the statements gradually diverge, and it is better for us to turn to the consideration of graphs which are exceptional in another sense, and only then come back to the bipartite case.

In particular the main result of [**AH06**] is the following.

FACT 7.7 ([**AH06**, Theorem 3.3]). *An antipodal and bipartite finite distance transitive graph of diameter* 6 *and degree at least* 3 *is isomorphic to the* 6*-cube.*

We have infinite connected metrically homogeneous graphs of any diameter $\delta \geq 3$ which are both bipartite and antipodal. In the case $\delta = 6$, the associated graph Γ_2 is the generic bipartite graph of diameter 4, and the associated graph Γ_3 is isomorphic to Γ itself. So even in the context of Smith's Theorem, the two pictures eventually diverge. But graphs of odd diameter which are both bipartite and antipodal are quite special.

THEOREM (12). *Let* Γ *be a connected metrically homogeneous graph of odd diameter* $\delta = 2\delta'+1$ *which is both antipodal and bipartite. Then* $B\Gamma$ *is connected, and* Γ *is the bipartite double cover of* $B\Gamma$*. The graph* $B\Gamma$ *is a metrically homogeneous graph with the following properties:*

(1) $B\Gamma$ *has diameter* δ'*;*
(2) *No triangle in* $B\Gamma$ *has perimeter greater than* $2\delta' + 1$*;*
(3) $B\Gamma$ *is not antipodal.*

Conversely, for any metrically homogeneous graph G *of diameter with the three stated properties, there is a unique antipodal bipartite graph of diameter* $2\delta' + 1$ *such that* $B\Gamma \cong G$*.*

We remark that conditions (1-3) on $B\Gamma$ imply that $B\Gamma_d$ is a clique of order at least 2. One exceptional case included under this theorem is that of the $(4d+2)$-gon, of diameter $2d + 1$, associated with the $(2d + 1)$-gon, of diameter d.

PROOF. Let A, B be the two halves of Γ. Then the metric on Γ is determined by the metrics on A and B and the pairing

$$a \leftrightarrow a'$$

between A and B determined by $d(a, a') = 2d + 1$, since

$$d(a_1, a_2') = 2d + 1 - d(a_1, a_2)$$

So the uniqueness is clear.

Let us next check that the conditions on $B\Gamma$ are satisfied. The first is clear. For the second, suppose we have vertices (a_1, a_2, a_3) in $B\Gamma$ forming a triangle of perimeter at least $2d+2$; we may construe these as vertices of A forming a triangle of perimeter $P \geq 4d + 4$. Then looking at the triangle (a_1, a_2, a_3'), we have

$$\begin{aligned} d(a_1, a_3') + d(a_2, a_3') &= (4d + 2) - [d(a_1, a_3) + d(a_2, a_3)] \\ &= (4d + 2) - P + d(a_1, a_2) \\ &< d(a_1, a_2), \end{aligned}$$

contradicting the triangle inequality. Finally, consider an edge (a_1, a_2) of $B\Gamma$, which we construe as a pair of vertices of A at distance 2. Then there must be a vertex a such that a' is adjacent to both, and this means that $a_1, a_2 \in \Gamma_{2d}(a)$, that is $a_1, a_2 \in B\Gamma_d(a)$.

Conversely, suppose G is a metrically homogeneous graph of diameter d, and Γ is the metric space on $G \times \{0, 1\}$ formed by doubling the metric of G on $A = G \times \{0\}$ and on $B = G \times \{1\}$, pairing A and B by $(a, \epsilon)' = (a, 1 - \epsilon)$, and defining

$$d(a_1, a_2') = 2d + 1 - d(a_1, a_2) \text{ for } a_1, a_2 \text{ in } A$$

The triangle inequality follows directly from the bound on the perimeters of triangles. We claim that Γ is a homogeneous metric space. Furthermore, the pairing $a \leftrightarrow a'$ is an isometry of Γ, and is recoverable from the metric: $y = x'$ if and only if $d(x, y) = 2d + 1$.

Let X, Y be finite subspaces of Γ, and f an isometry between them. Then f extends canonically to their closures under the antipodal pairing. So we may suppose X and Y are closed under the antipodal pairing; and composing f with the antipodal pairing if necessary, we may suppose f preserves the partition of Γ into A, B. Then restrict f to $A \cap X$, extend to A by homogeneity, and then extend back to Γ. Thus Γ is a homogeneous metric space.

Finally, we claim that the metric on Γ is the graph metric, and for this it suffices to show that vertices at distance 2 in the metric have a common neighbor in Γ. So let a_1, a_2 be two such vertices, taken for definiteness in A; write $a_i = (v_i, 0)$. Taking $v \in G$ with $v_1, v_2 \in G_d(v)$, and $a = (v, 1)$, we find that a is adjacent to a_1 and a_2. $\square$

Corollary 12.1 made this result explicit for the case in which $B\Gamma$ is already in the catalog.

Corollary (12.1). *Let Γ be an antipodal bipartite graph of diameter 5. Then $B\Gamma$ is either a pentagon, or the generic homogeneous graph omitting I_3 (the complement of the Henson graph G_3) and Γ is its bipartite double cover.*

Proof. Let $G = B\Gamma$. Then G has diameter 2, so it is a homogeneous graph, on the list of Lachlan and Woodrow.

Furthermore, by the theorem, G contains I_2 but not I_3. As $d = 2$ here, G contains a path of length 2 as well as a vertex at distance 2 from both vertices of an edge. By the Lachlan/Woodrow classification, in the finite case G is a 5-cycle and in the infinite case it must be the generic graph omitting I_3. $\square$

The bipartite double cover of C_5 is C_{10}, and the bipartite double cover of G_3^c is $\Gamma^5_{1,4;12,11;\emptyset}$, the generic antipodal bipartite graph of diameter 5.

8. Exceptional Metrically Homogeneous Graphs

We turn now from the imprimitive case to the case in which the graph Γ_1 is exceptional.

If Γ is a metrically homogeneous graph, then Γ_1 is a homogeneous graph, and must occur in the short list of such graphs described in §4. There are three possibilities for Γ_1 which are compatible with the Fraïssé constructions we have described: an infinite independent set, the Henson graphs, and the Rado graph. Thus we make the following definition.

Definition 8.1. Let Γ be a metrically homogeneous graph.
1. Γ is of *generic type* if the graph Γ_1 is of one the following three types:
  (1) an infinite independent set I_∞;
  (2) generic omitting K_n for some $n \geq 3$;
  (3) fully generic (the Rado graph).
2. If Γ is not of generic type, then we say it is of *exceptional type.*

All the bipartite graphs (indeed, all the triangle free ones) are included under "generic type" and will have to be treated under that heading. But it is useful to dispose of the classification of the exceptional metrically homogeneous graphs as a separate case.

Thus in the exceptional case we have the following possibilities for Γ_1.

- $\Gamma_1 \cong C_5$ or the line graph of $K_{3,3}$.
- $\Gamma_1 \cong m \cdot K_n$ or $K_m[I_n]$ with $1 \leq m, n \leq \infty$; and not of the form I_∞.
- Γ_1 the complement of a Henson graph

We will classify those falling under the first two headings and show that the third case does not occur. The case in which Γ_1 is finite is covered by [**Cam80, Mph82**], split under two headings: (a) Γ finite; or (b) Γ infinite and Γ_1 finite. Those proofs use considerably less than metric homogeneity. We include a treatment of those cases here, but making full use of our hypothesis.

8.1. Exceptional graphs with Γ_1 finite and primitive.

LEMMA 8.2. *Let Γ be a connected metrically homogeneous graph of diameter at least 3 and degree at least 3, and suppose that Γ_1 is one of the primitive finite homogeneous graphs containing both edges and nonedges, that is C_5 or the line graph of $K_{3,3}$. Then Γ is the antipodal graph of diameter 3 obtained from Γ_1 in the manner of Theorem 15.*

PROOF. We fix a basepoint $*$ in Γ so that Γ_i is viewed as a specific subgraph of Γ for each i. The proof proceeds in two steps.

(1) There is a $*$-definable function from Γ_1 to Γ_2.

We will show that for $v \in \Gamma_1$, the vertices of Γ_1 not adjacent to v have a unique common neighbor v' in Γ_2.

If Γ_1 is a 5-cycle then this amounts to the claim that every edge of Γ lies in two triangles, and this is clear by inspection of an edge $(*, v)$ with $*$ the basepoint and $v \in \Gamma_1$.

Now suppose Γ_1 is $E(K_{3,3})$. We claim that every induced 4-cycle $C \cong C_4$ in Γ has exactly two common neighbors.

Consider $u, v \in \Gamma_1$ lying at distance 2, and let $G_{u,v}$ be the metric space induced on their common neighbors in Γ. This is a homogeneous metric space. Since these common neighbors consist of the basepoint $*$, the two common neighbors a, b of u, v in Γ_1, and whatever common neighbors u, v may have in Γ_2, we see that pairs at distance 1 occur, and the corresponding graph has degree 2 (looking at $*$) and is connected (looking at $(a, *, b)$). So $G_{u,v}$ is a connected metrically homogeneous graph of degree 2, and furthermore embeds in $\Gamma_1(u) \cong \Gamma_1$. So $G_{u,v}$ is a 4-cycle, and therefore (u, a, v, b) has exactly 2 neighbors, as claimed. This proves (1).

Now let $f : \Gamma_1 \to \Gamma_2$ be $*$-definable. By homogeneity f is surjective, and as Γ_1 is primitive, it is bijective. It also follows from homogeneity that for $u, v \in \Gamma_1$, $d(u, v)$ determines $d(f(u), f(v))$, so f is either an isomorphism or an anti-isomorphism. Since Γ_1 is isomorphic to its complement, $\Gamma_1 \cong \Gamma_2$ in any case. Hence the vertices of Γ_2 have a common neighbor v, and $v \in \Gamma_3$. We claim that $|\Gamma_3| = 1$.

By homogeneity all pairs (u, v) in $\Gamma_2 \times \Gamma_3$ are adjacent. In particular for $v_1, v_2 \in \Gamma_3$ we have $\Gamma_1(v_1) = \Gamma_1(v_2)$ and $d(v_1, v_2) \leq 2$. Since we have pairs of vertices u_1, u_2 in Γ_1 at distance 1 or 2 for which $\Gamma_1(u_1) \neq \Gamma_1(u_2)$, we find $|\Gamma_3| = 1$.

It now follows that Γ is antipodal of diameter 3 and the previous analysis applies. □

8.2. Exceptional graphs with $\Gamma_1 \cong K_m[I_n]$. We will prove the following.

PROPOSITION 8.3. *Let Γ be a connected metrically homogeneous graph of diameter δ, and suppose $\Gamma_1 \cong K_m[I_n]$ with $1 \le m, n \le \infty$ Then one of the following occurs.*

(1) *$\delta \le 2$, Γ is homogeneous and found under the Lachlan/Woodrow classification.*
(2) *$\Gamma \cong C_n$, an n-cycle, for some n;*
(3) *$m = 1$, $\delta \ge 3$, and one of the following occurs.*
   (a) *n is finite, and Γ is the bipartite complement of a perfect matching between two sets of order $n + 1$.*
   (b) *$\delta = \infty$, Γ is a regular tree of degree n, with $2 \le n \le \infty$.*
   (c) *$n = \infty$, and any two vertices at distance 2 have infinitely many common neighbors.*

LEMMA 8.4. *Let Γ be a connected metrically homogeneous graph of diameter at least 3, and suppose that Γ_1 is a complete multipartite graph of the form $K_m[I_n]$ (the complement of $m \cdot K_n$). Then $m = 1$.*

PROOF. Suppose that $m > 1$. Fix a geodesic $(*, u, v)$ of length 2 and let $\Gamma_i = \Gamma_i(*)$. Let A be the set of neighbors of u in Γ_1. Then $A \cong K_{m-1}[I_n]$, and the neighbors of u in Γ include $*$, v, and A. Now "$d(x, y) > 1$" is an equivalence relation on $\Gamma_1(u)$, and $*$ is adjacent to A, so v is adjacent to A. Now if we replace u by $u' \in A$ and argue similarly with respect to $(*, u', v)$, we see that the rest of Γ_1 is also adjacent to v, that is $\Gamma_1 \subseteq \Gamma_1(v)$. Now switching $*$ and v, by homogeneity $\Gamma_1(v) \subseteq \Gamma_1$. But then the diameter of Γ is 2, a contradiction. Thus $m = 1$. □

We next consider the case $m = 1$, that is, Γ_1 is an independent set. We suppose in this case that the set of common neighbors of any pair of points at distance 2 is finite, which covers the case in which n is finite but also picks up the case of an infinitely branching tree. When $n = 2$ we have either a cycle or a 2-way infinite path, so we will leave this case aside in our analysis.

LEMMA 8.5. *Let Γ be a metrically homogeneous graph of diameter at least 2, with $\Gamma_1 \cong I_n$, $3 \le n \le \infty$. Suppose that for $u, v \in \Gamma$ at distance two, the number k of common neighbors of u, v is finite. Then either $k = 1$, or $n = k + 1$.*

PROOF. We consider Γ_1 and Γ_2 with respect to a fixed basepoint $* \in \Gamma$. For $u \in \Gamma_2$, let I_u be the k-set consisting of its neighbors in Γ_1. Any k-subset of Γ_1 occurs as I_u for some u. For $u, v \in \Gamma_2$ let $u \cdot v = |I_u \cap I_v|$. If $u \cdot v \ge 1$ then $d(u, v) = 2$. Now $\mathrm{Aut}(\Gamma)_*$ has a single orbit on pairs in Γ_2 at distance 2, while every value i in the range $\max(1, 2k - n) \le i \le k - 1$, will occur as $u \cdot v$ for some such u, v. Therefore $k \le 2$ or $k = n - 1$, with n finite in the latter case.

The case $k = 2 < n - 1$ is eliminated by a characteristic application of homogeneity. A set of three pairs in Γ_1 which intersect pairwise may or may not have a common element (once $n \ge 4$), so if we choose u_1, u_2, u_3 and v_1, v_2, v_3 in Γ_2 corresponding to these two possibilities for the associated I_{u_j} and I_{v_j}, we get isometric configurations $(*, u_1, u_2, u_3)$ and $(*, v_1, v_2, v_3)$ which lie in distinct orbits of $\mathrm{Aut}\,\Gamma$. So we have $k = 1$ or $n = k + 1$. □

LEMMA 8.6. *Let Γ be a connected metrically homogeneous graph of diameter at least 3, with $\Gamma_1 \cong I_n$, and $3 \le n \le \infty$. Suppose any pair of vertices at distance 2 have k common neighbors, with $k < \infty$. Then one of the following occurs.*

(1) *$n = k+1$, and Γ is the complement of a perfect matching, in other words the antipodal graph of diameter 3 obtained by doubling Γ_1.*
(2) *$k = 1$, and Γ is a k-regular tree.*

PROOF. We fix a basepoint $*$ and write Γ_i for $\Gamma_i(*)$.

Suppose first that $n = k+1$. Then any two vertices of Γ_2 lie at distance 2, and there is a $*$-definable function $f : \Gamma_2 \to \Gamma_1$ given by the nonadjacency relation. By homogeneity f is surjective, and as Γ_2 is primitive, f is bijective. In particular $\Gamma_2 \cong \Gamma_1$ and there is a vertex $v \in \Gamma_3$ adjacent to all vertices of Γ_2, hence $\Gamma_1(v) = \Gamma_2$. It follows readily that $|\Gamma_3| = 1$ and Γ is antipodal of diameter 3. The rest follows by our previous analysis.

Now suppose that $k = 1$. It suffices to show that Γ is a tree.

Suppose on the contrary that there is a cycle C in Γ, which we take to be of minimal diameter d. Then the order of C is $2d$ or $2d+1$.

Suppose the order of C is $2d$. Then for $v \in \Gamma_d$, v has at least two neighbors u_1, u_2 in Γ_{d-1}, whose distance is therefore 2. Furthermore, in Γ_{d-2} there are no edges, and each vertex of Γ_{d-2} has a unique neighbor in Γ_{d-3}, so each vertex of Γ_{d-2} has at least two neighbors in Γ_{d-1}, whose distance is therefore 2.

So for $u_1, u_2 \in \Gamma_{d-1}$, there is a common neighbor in Γ_d, and also in Γ_{d-2}. This gives a 4-cycle in Γ, contradicting $k = 1$.

So the order of C is $2d+1$. In particular, each $v \in \Gamma_d$ has a unique neighbor in Γ_{d-1}, and Γ_d contains edges.

Let G be a connected component of Γ_d. Suppose u, v in G are at distance 2 in G. As any vertex in Γ_{d-1} has at least two neighbors in Γ_d, the vertices u, v must have a common neighbor in Γ_{d-1} as well as in G, and this contradicts the hypothesis $k = 1$. So the connected components of Γ_d are simply edges.

Take $a \in \Gamma_{d-1}$, $u_1, v_1 \in \Gamma_d$ adjacent to a, and take $u_2, v_2 \in \Gamma_d$ adjacent to u_1, v_1 respectively. By homogeneity there is an automorphism fixing the basepoint $*$ and interchanging u_1 with v_2; this also interchanges u_2 and v_1. Hence $d(u_2, v_2) = 2$. It follows that u_2, v_2 have a common neighbor b in Γ_{d-1}. Now $(a, u_1, u_2, b, v_2, v_1, a)$ is a 6-cycle. Since the minimal cycle length is odd, we have $|C| = 5$ and $d = 2$.

Furthermore, the previous paragraph shows that the element b is determined by $a \in \Gamma_1$ and the basepoint $*$: we take $u \in \Gamma_2$ adjacent to a, $v \in \Gamma_2$ adjacent to u, and $b \in \Gamma_1$ adjacent to v. So the function $a \mapsto b$ is $*$-definable. However Γ_1 is an independent set of order at least 3, so this violates homogeneity. □

This completes the proof of Proposition 8.3.

8.3. Exceptional graphs with $\Gamma_1 \cong m \cdot K_n$, $n \geq 2$. We deal here with the tree-like graphs $T_{r,s}$ derived from the trees $T(r,s)$ as described in §5.3. With $r, s < \infty$ these graphs are locally finite (that is, the vertex degrees are finite). Conversely:

FACT 8.7 (Macpherson, [**Mph82**]). *Let G be an infinite locally finite distance transitive graph. Then G is $T_{r,s}$ for some finite $r, s \geq 2$.*

The proof uses a result of Dunwoody on graphs with nontrivial cuts given in [**Dun82**].

The graph $T_{r,s}$ has $\Gamma_1 \cong s \cdot K_{r-1}$. We are interested now in obtaining a characterization of these graphs, also with r or s infinite, in terms of the structure of Γ_1.

Our goal is the following.

PROPOSITION 8.8. *Let Γ be a connected metrically homogeneous graph with $\Gamma_1 \cong m \cdot K_n$, with $n \geq 2$ and $\delta \geq 3$. Then $m \geq 2$, and $\Gamma \cong T_{m,n+1}$.*

Our main goal will be to show that any two vertices at distance two have a unique common neighbor. We divide up the analysis into three cases. Observe that in a metrically homogeneous graph with $\Gamma_1 \cong m \cdot K_n$, the common neighbors of a pair $a, b \in \Gamma$ at distance 2 will be an independent set, since for u_1, u_2 adjacent to a, b, and each other, we would have the path (a, u_1, b) inside $\Gamma_1(u_2)$.

LEMMA 8.9. *Let Γ be a metrically homogeneous graph of diameter at least 3 and suppose that $\Gamma_1 \cong m \cdot K_n$ with $n \geq 3$. Then for $u, v \in \Gamma$ at distance 2, there are at most two vertices adjacent to both.*

PROOF. Supposing the contrary, every induced path of length 2 is contained in two distinct 4-cycles. Fix a basepoint $* \in \Gamma$ and let $\Gamma_i = \Gamma_i(*)$.

Fix $v_1, v_2 \in \Gamma_1$ adjacent. For $i = 1, 2$, let H_i be the set of neighbors of v_i in Γ_2. The connected components of H_1 and H_2 are cliques. We claim

$$H_1 \cap H_2 = \emptyset$$

Otherwise, consider $v \in H_1 \cap H_2$ and the path $(*, v_1, v)$ contained in $\Gamma_1(v_2)$.

We will find $u_1 \in H_1$, and $u_2, u_2' \in H_2$ distinct, so that

$$d(u_1, u_2) = d(u_1, u_2') = 1$$

Extend the edge (v_1, v_2) to a 4-cycle (v_1, v_2, u_2, u_1). Then $u_1, u_2 \notin \Gamma_1 \cup \{*\}$, so $u_1 \in H_1$ and $u_2 \in H_2$. By our hypothesis there is a second choice of u_2 with the same properties.

With the vertices u_1, u_2, u_2' fixed, let A, B, B' denote the components of H_1 and H_2, respectively, containing the specified vertices. Observe that B and B' are distinct: otherwise, the path (u_1, u_2, v_2) would lie in $\Gamma_1(u_2')$.

The relation "$d(x, y) = 1$" defines a bijection between A and B

With $u \in A$ fixed, it suffices to show the existence and uniqueness of the corresponding element of B. The uniqueness amounts to the point just made for u_1, namely that $B \neq B'$.

For the existence, we may suppose $u \neq u_1$. Then $d(u, v_2) = d(u, u_2) = 2$. We have an isometry

$$(*, v_1, u, v_2, u_2) \cong (*, v_2, u_2, v_1, u)$$

and hence the triple (u, u_1, u_2) with $u_1 \in H_1$ corresponds to an isometric triple (u_2, u', u) with $u' \in H_2$.

Thus we have a bijection between A and B definable from $(*, v_1, v_2, u_1, u_2)$ and hence we derive a bijection between B and B' definable from $(*, v_1, v_2, u_1, u_2, u_2')$. Using this, we show

$$n = 2$$

The graph induced on $B \cup B'$ is $2 \cdot K_n$, and any isometry between finite subsets of $B \cup B'$ containing u_2, u_2' which fixes u_2 and u_2' will be induced by $\mathrm{Aut}(\Gamma)$. So if there is a bijection between B and B' invariant under the corresponding automorphism group, we find $n = 2$. □

LEMMA 8.10. *Let Γ be a metrically homogeneous graph of diameter at least 3 and suppose that $\Gamma_1 \cong m \cdot K_n$ with $n \geq 2$. Let $u, v \in \Gamma$ lie at distance 2, and suppose u, v have finitely many common neighbors. Then they have a unique common neighbor.*

PROOF. We fix a basepoint $*$, and for $u \in \Gamma_2$ we let I_u be the set of neighbors of u in Γ_1. Our assumption is that $k = |I_u|$ is finite. Then any independent subset of Γ_1 of cardinality k occurs as I_u for some $u \in \Gamma_2$.

We consider the $k - 2$ possibilities:

$$|I_u \cap I_v| = i \text{ with } 1 \leq i \leq k-1$$

As $n \geq 2$, all possibilities are realized, whatever the value of m. However in all such cases, $d(u, v) \leq 2$, so we find $k - 1 \leq 2$, and $k \leq 3$.

We claim that for $u, v \in \Gamma_2$ adjacent, we have $|I_u \cap I_v| = 1$.

There is a clique v, u_1, u_2 with $v \in \Gamma_1$ and $u_1, u_2 \in \Gamma_2$. As u_1, u_2 are adjacent their common neighbors form a complete graph. On the other hand I_{u_1} and I_{u_2} are independent sets, so their intersection reduces to a single vertex. By homogeneity the same applies whenever $u_1, u_2 \in \Gamma_2$ are adjacent, proving our claim.

Now suppose $k = 3$. Then $|I_u \cap I_v|$ can have cardinality 1 or 2, and the case $|I_u \cap I_v| = 2$ must then correspond to $d(u, v) = 2$.

Now $k \leq m$, so we may take pairs (a_i, b_i) for $i = 1, 2, 3$ lying in distinct components of Γ_1. Of the eight triples t formed by choosing one of the vertices of each of these pairs, there are four in which the vertex a_i is selected an even number of times. Let a vertex $v_t \in \Gamma_2$ be taken for each such triple, adjacent to its vertices. Then the four vertices v_t form a complete graph K_4. It follows that K_3 embeds in Γ_1, that is $n \geq 3$. So we may find independent triples I_1, I_2, I_3 such that $|I_1 \cap I_2| = |I_2 \cap I_3| = 1$ while $I_1 \cap I_3 = \emptyset$. Take $u_1, u_2, u_3 \in \Gamma_2$ with $I_i = I_{u_i}$ and with $d(u_1, u_2) = d(u_2, u_3) = 1$. Then $I_{u_1} \cap I_{u_3} = \emptyset$, while $d(u_1, u_2) \leq 2$, a contradiction. Thus $k = 2$.

With $k = 2$, suppose $m > 2$. For $u \in \Gamma_2$, let $\hat{I}_u$ be the set of components of Γ_1 meeting I_u. We consider the following two properties of a pair $u, v \in \Gamma_2$:

$$|I_u \cap I_v| = 1; |\hat{I}_u \cap \hat{I}_v| = i \ (i = 1 \text{ or } 2)$$

These both occur, and must correspond in some order with the conditions $d(u, v) = 1$ or 2. But just as above we find u_1, u_2, u_3 with $d(u_1, u_2) = d(u_2, u_3) = 1$ and $I_{u_1} \cap I_{u_3} = \emptyset$, and as $d(u_1, u_3) \leq 2$ this is a contradiction.

So we come down to the case $m = k = 2$. But then for $v \in \Gamma_2$, all components of $\Gamma_1(v)$ are represented in Γ_1, and hence v has no neighbors in Γ_3, a contradiction. □

LEMMA 8.11. *Let Γ be a metrically homogeneous graph of diameter at least 3 and suppose that $\Gamma_1 \cong m \cdot K_n$ with $n \geq 2$. Let $u, v \in \Gamma$ lie at distance 2, and suppose u, v have infinitely many common neighbors. Then $n = \infty$.*

PROOF. We fix a basepoint $*$, and for $u \in \Gamma_2$ we let I_u be the set of neighbors of u in Γ_1. Our assumption is that I_u is infinite. Then any finite independent subset of Γ_1 is contained in I_u for some $u \in \Gamma_2$.

For $u \in \Gamma_2$, let $\hat{I}_u$ be the set of components of Γ_1 which meet I_u, and let $\hat{J}_u$ be the set of components of Γ_1 which do not meet I_u. We show first that

$$\hat{J}_u \text{ is infinite.}$$

Supposing the contrary, let $k = |\hat{J}_u| < \infty$ for $u \in \Gamma_2$. Any set of k components of Γ_1 will be $\hat{J}_u$ for some $u \in \Gamma_2$, and the $k+1$ relations on Γ_2 defined by

$$|\hat{J}_u \cap \hat{J}_v| = i$$

for $i = 0, 1, \ldots, k$ will be nontrivial and distinct. Furthermore, for any preassigned k components $\hat{J}$, and any vertex $a \in \Gamma_1$ not in the union of $\hat{J}$, there is a vertex u with $\hat{J}_u = \hat{J}$ and $a \in I_u$, so our $(k+1)$ relations are realized by pairs $u, v \in \Gamma_2$ with $I_u \cap I_v \neq \emptyset$, and hence $d(u,v) \leq 2$. Hence $k+1 \leq 2$, $k \leq 1$.

Suppose $k = 1$ and fix a vertex $v_0 \in \Gamma_1$. Then for $u, v \in \Gamma_2$ adjacent to v_0, the two relations $\hat{J}_u = \hat{J}_v$, $\hat{J}_u \neq \hat{J}_v$ correspond in some order to the relations $d(u,v) = 1$, $d(u,v) = 2$, and since the first relation is an equivalence relation, they correspond in order.

With $u \in \Gamma_2$, $v_0, v_1 \in I_u$ distinct, there are u_0, u_1 in Γ_2 with u_0 adjacent to u and v_0, and with u_1 adjacent to u and v_1. The neighbors of u form a graph of type $\infty \cdot K_n$, so $d(u_0, u_1) = 2$. However $d(u, u_0) = d(u, u_1) = 1$ and hence $\hat{J}_{u_0} = \hat{J}_{u_1}$, a contradiction.

So $k = 0$ and for $u \in \Gamma_2$, the set I_u meets every component of Γ_1. That is, $\Gamma_1(u)$ meets every component of Γ_1, and after switching the roles of u and the basepoint $*$, we conclude Γ_1 meets every component of $\Gamma_1(u)$, which is incompatible with the condition $\delta \geq 3$. So $\hat{J}_u$ is infinite for $u \in \Gamma_2$.

Now we claim

$$\text{For } u, v \in \Gamma_2 \text{ adjacent, } \hat{I}_u \setminus \hat{I}_v \text{ is infinite}$$

Supposing the contrary, for all adjacent pairs $u, v \in \Gamma_2$, the sets $\hat{I}_u$ and $\hat{I}_v$ coincide up to a finite difference.

Take $u \in \Gamma_2$, $v_0, v_1 \in \Gamma_1$ adjacent to u, and u_0, u_1 adjacent to u, v_0 or u, v_1 respectively. Then, as above, $d(u_0, u_1) = 2$, while $\hat{J}_{u_0} = \hat{J}_u = \hat{J}_{u_1}$. Thus for $u, v \in \Gamma_2$ with $d(u,v) \leq 2$ we have $\hat{J}_u = \hat{J}_v$. Furthermore, the size of the difference $|\hat{J}_u \setminus \hat{J}_v|$ is bounded, say by ℓ. But we can fix $u \in \Gamma_2$ and then find $v \in \Gamma_2$ so that I_u meets I_v but $\hat{I}_v$ picks up more than ℓ components of $\hat{J}_u$. Since I_u meets I_v we have $d(u,v) \leq 2$, and thus a contradiction. This proves our claim.

We make a third and last claim of this sort.

$$\text{For } u, v \in \Gamma_2 \text{ adjacent, } \hat{I}_u \cap \hat{I}_v \text{ is infinite}$$

Supposing the contrary, let k' be $|\hat{I}_u \cap \hat{I}_v|$ for $u, v \in \Gamma_2$ adjacent, fix $u, v \in \Gamma_2$ adjacent, and let I be a finite independent subset consisting of representatives for more than k' components in $\hat{I}_v \setminus \hat{I}_u$.

Take $a \in I_u \cap I_v$ and take $v' \in \Gamma_2$ adjacent to a and to u, with $J \subseteq \hat{I}_{v'}$ and $I_v \neq I_{v'}$.

Now u, v, v' are adjacent to a, and v, v' are adjacent to u, so v and v' are adjacent. But by construction $|\hat{I}_{v'} \setminus \hat{I}_v| > k'$. This proves the third claim.

Now, finally, suppose n is finite. Take $u \in \Gamma_2$, $v \in I_u$, and let A, B be components of Γ_1 which are disjoint from I_u but meet $I_{u'}$ for some u' adjacent to u, v. Then

We can then find neighbors $w_{a,b}$ of u, v in Γ_2 for which the intersection of I_w with A and B respectively is an arbitrary pair of representatives a, b. But this requires $n - 1 \geq n^2$, and gives a contradiction. $\square$

COROLLARY 8.11.1. *Let Γ be a metrically homogeneous graph of diameter at least 3 and suppose that $\Gamma_1 \cong m \cdot K_n$ with $n \geq 2$. Then for $u, v \in \Gamma$ with $d(u, v) = 2$, there is a unique vertex adjacent to both.*

PROOF. Apply the last three lemmas. If u, v have infinitely many common neighbors, then n is infinite. In particular, $n \geq 3$. But then they have at most two common neighbors. So in fact u, v have finitely many neighbors, and we apply Lemma 8.10. $\square$

After this somewhat laborious reduction, we can complete the proof of Proposition 8.8.

Proof of Proposition 8.8. If $m = 1$ then evidently Γ is complete, contradicting the hypothesis on δ. So $m \geq 2$.

By definition, the blocks of Γ are the maximal 2-connected subgraphs. Any edge of Γ is contained in a unique clique of order $n + 1$. It suffices to show that these cliques form the blocks of Γ, or in other words that any cycle in Γ is contained in a clique.

Supposing the contrary, let C be a cycle of minimal order embedding into Γ as a subgraph, and not contained in a clique. The cycle C carries two metrics: its metric d_C as a cycle, and the metric d_Γ induced by Γ. We claim these metrics coincide. In any case, $d_\Gamma \leq d_C$.

If the metrics disagree, let $u, v \in C$ be chosen at minimal distance such that $d_\Gamma(u, v) < d_C(u, v)$, let $P = (u, \ldots, v)$ be a geodesic in Γ, and let $Q = (u, \ldots, v)$ be a geodesic in C. Let v' be the first point of intersection of P with Q, after u. Then $P \cup Q$ contains a cycle C' smaller than C, and containing u, v'. Hence by hypothesis the vertices of C' form a clique in Γ, and in particular u, v' are adjacent in Γ.

If u, v' are adjacent in C, then $d_C(v, v') = d_C(u, v) - 1$, $d_\Gamma(v, v') = d_\Gamma(u, v) - 1$, so $d_\Gamma(v, v') < d_C(v, v')$, and this contradicts the choice of u, v. So they are not adjacent, and the edge (u, v') belongs to two cycles C_1, C_2 whose union contains C, both smaller than C. So the vertices of C_1 and C_2 are cliques. Since C is not a clique, there are $u_1 \in C_1$ and $u_2 \in C_2$ nonadjacent in Γ. Then $d(u_1, u_2) = 2$, and there is a unique vertex adjacent to u_1 and u_2; but u, v' are two such, a contradiction.

Thus the embedding of C into Γ respects the metric. In particular, C is an induced subgraph of Γ.

Let d be the diameter of C, so that the order of C is either $2d$ or $2d + 1$. Fix a basepoint $* \in C$, and let $\Gamma_i = \Gamma_i(*)$.

For $v \in \Gamma_i$ with $i < d$, we claim that there is a unique geodesic $[*, v]$ in Γ. Otherwise, take $i < d$ minimal such that Γ contains vertices u, v at distance i with two distinct geodesics P, Q from u to v. By the minimality, these geodesics are disjoint. Their union forms a cycle smaller than C, hence they form a clique. As they are geodesics, $i = 1$ and then in any case the geodesic is unique.

Let $v \in \Gamma_{d-1}$, and let $H = \Gamma_1(v)$, a copy of $m \cdot K_n$. Then H contains a unique component meeting Γ_{d-2}. We claim that no other component of H meets Γ_{d-1}.

Suppose on the contrary that u_1, u_2 are adjacent to v with $u_1 \in \Gamma_{d-1}$, $u_2 \in \Gamma_{d-2}$, and $d(u_1, u_2) = 2$. Then taking P_1, P_2 to be the unique geodesics from u_1 or u_2 to $*$, $P_1 \cup P_2$ contains a cycle smaller than C, and containing the path (u_1, v, u_2), and hence is not a clique. This is a contradiction.

Since Γ_d contains at least one component of H, in particular there is an edge in Γ_d whose vertices have a common neighbor in Γ_{d-1}. Using this, we eliminate the case $|C| = 2d+1$ as follows.

If $|C| = 2d+1$ then $C \cap \Gamma_d$ consists of two adjacent vertices v_1, v_2, whose other neighbors u_1, u_2 in C lie in Γ_{d-1}. Furthermore v_1, v_2 have a common neighbor u in Γ_{d-1}, and $u \neq u_1, u_2$. Let P, P_1 be the unique geodesics in Γ connecting $*$ with u, u_1 respectively. Then $P \cup P_1 \cup \{v_1\}$ contains a cycle shorter than C, which contains the path (u_1, v, v_1), and we have a contradiction.

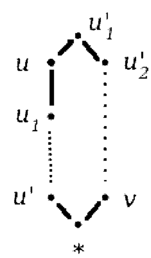
Thus $|C| = 2d$, in other words the vertices $v \in \Gamma_d$ are connected to the basepoint $*$ by at least two distinct geodesics, and any two such geodesics will be disjoint.

Take $u \in \Gamma_{d-1}$, and $u' \in \Gamma_1$, with $d(u, u') = d-2$. Take $v' \in \Gamma_1$ with $d(u', v') = 2$. We claim

$$d(u, v') = d$$

Otherwise, with P, Q the geodesics from u to u' and v' respectively, we have $|P \cup Q| \leq 2d-1 < |C|$, and hence $(u', *, v')$ is contained in a cycle smaller than C, a contradiction since u', v' are nonadjacent.

Thus $d(u, v') = d$, and the extension of the geodesic from u to u' by the path $(u', *, v')$ gives a geodesic Q from u to v'. There is a second geodesic Q' from u to v', disjoint from Q. Let u_1 be the unique neighbor of u in Γ_{d-2}; this lies on Q. Let u'_1 be the neighbor of u in Q'. As the cycle $Q \cup Q'$ satisfies the same condition as the cycle C, the metric on this cycle agrees with the metric in Γ, and in particular $d(u_1, u'_1) = 2$. By the minimality of C, we have $u'_1 \in \Gamma_d$. Let u'_2 be the following neighbor of u'_1 in Q'. Then $d(u, u'_2) = 2$, with $u, u'_2 \in \Gamma_{d-1}$.



Suppose $m \geq 3$. Then there is u^* adjacent to u and at distance 2 from u_1 and u'_1. In view of the minimality of C, the vertex u^* cannot be in Γ_{d-2} or Γ_{d-1}, hence lies in Γ_d. Hence any two vertices of Γ_d at distance 2 have a common neighbor in Γ_{d-1}. This applies to u and u'_2 and contradicts Corollary 8.11.1. We conclude that

$$m = 2$$

Fix an edge a_1, a_2 in Γ_{d-1}. For $i, j = 1, 2$ in some order, set

$$H_{ij} = \{v \in \Gamma : d(v, a_i) = 1,\ d(v, a_j) = 2\}$$

We claim $H_{ij} \subseteq \Gamma_d$.

As a_1, a_2 have the same unique neighbor in Γ_{d-2}, we have

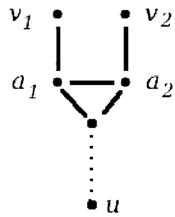
$$(H_{12} \cup H_{21}) \cap \Gamma_{d-2} = \emptyset$$

Now suppose $v \in H_{12} \cap \Gamma_{d-1}$. Then $d(v, a_2) = 2$ and $v, a_2 \in \Gamma_{d-1}$. Now since $|C| = 2d$, there is $w \in \Gamma_d$ adjacent to v and a_2. Then a_2, v have the two common neighbors w and a_1, a contradiction. So $H_{12}, H_{21} \subseteq \Gamma_d$.

Let $v_1 \in H_{12}$, $v_2 \in H_{21}$. We claim that

$$d(v_1, v_2) = 3$$

Otherwise, there is a cycle of length at most 5, not contained in a clique. As $|C|$ is even, it follows that C is a 4-cycle, so vertices at distance 2 have at least two common neighbors, a contradiction. Thus for any other choice $v_1' \in H_{12}$, $v_2' \in H_{21}$, we have $(*, a_1, a_2, v_1, v_2)$ isometric with $(*, a_1, a_2, v_1', v_2')$.



There is a unique element $u \in \Gamma_1$ at distance $d-2$ from a_1 and a_2; namely, the element at distance $d-3$ from their common neighbor in Γ_{d-2}. On the other hand, for $v \in \Gamma_d$, if I_v is the set $\{v' \in \Gamma_1 : d(v, v') = d-1\}$, then by our hypotheses, I_v is a pair of representatives for the two components of Γ_1. And if $v \in H_{12}$ or H_{21}, one of these representatives will be u. Let B be the component of Γ_1 not containing u. Then the distance from a_1 or a_2 to a vertex of B is d. It follows that all vertices of B will occur as the second vertex of I_v for some $v_1 \in H_{12}$ and for some $v_2 \in H_{21}$. Therefore, we may choose pairs (v_1, v_2) and (v_1', v_1') with $v_1, v_1' \in H_{12}$, $v_2, v_2' \in H_{21}$, and $I_{v_1} = I_{v_2}$, while $I_{v_1'} \neq I_{v_2'}$. But as $(*, a_1, a_2, v_1, v_2)$ and $(*, a_1, a_2, v_1', v_2')$ are isometric, this contradicts homogeneity. □

8.4. Γ_1 cannot be the complement of a Henson graph. The last exceptional case requiring consideration is the one in which Γ_1 is the complement of a Henson graph.

LEMMA 8.12. *Let Γ be a metrically homogeneous graph of diameter at least 3 with Γ_1 infinite and primitive. Then Γ_1 contains an infinite independent set.*

PROOF. Suppose the contrary. Evidently Γ_1 contains an independent pair. By the Lachlan/Woodrow classification, if n is minimal such that Γ_1 contains no independent set of order n, then for any finite graph G which contains no independent set of order n, G occurs as an induced subgraph of Γ_1.

We consider a certain amalgamation diagram involving subspaces of Γ. Let A be the metric space with three points a, b, x constituting a geodesic, with $d(a, b) = 2$, $d(b, x) = 1$, $d(a, x) = 3$. Let B be the metric space on the points a, b and a further set Y of order $n-1$ with the metric given by

$$\begin{aligned} d(a, y) &= d(b, y) = 1 \quad (y \in Y) \\ d(y, y') &= 2 \quad (y, y' \in Y \text{ distinct}) \end{aligned}$$

As the diameter of Γ is at least 3, the geodesic A certainly occurs as a subspace of Γ. On the other hand, the metric space B embeds into Γ_1, and hence into Γ. Therefore there is some amalgam $G = A \cup B$ embedding into Γ as well.

Now for $y \in Y$, the structure of (a, x, y) forces $d(x, y) \geq 2$. On the other hand, the element b forces $d(x, y) \leq 2$. Thus in G, the set $Y \cup \{x\}$ is an independent set of order n. Furthermore this set is contained in $\Gamma_1(b)$, so we arrive at a contradiction. □

We may sum up this section as follows.

THEOREM (10). *Let Γ be a connected metrically homogeneous graph and suppose that Γ_1 is either imprimitive or contains no infinite independent set. Then Γ_1 is found in our catalog under case I or case II.*

PROOF. By the Lachlan-Woodrow classification of homogeneous graphs and Lemma 8.12, Γ_1 must be finite, imprimitive, or the complement of a Henson graph, and we have just ruled out this last possibility.

Suppose Γ_1 is finite and primitive. Then Lemma 8.2 applies if Γ_1 is neither complete nor an independent set. If Γ_1 is complete, then Γ is complete. Proposition 8.3 covers the case in which Γ_1 is a finite independent set.

Lastly, suppose Γ_1 is imprimitive. If Γ_1 is of the form $K_m[I_n]$ with $m, n \geq 2$ then Proposition 8.3 applies. Alternatively, Γ_1 may have the form $m \cdot K_n$ with $m, n \geq 2$. This is covered by Proposition 8.8. □

It is now appropriate to return to the bipartite case and look at metrically homogeneous bipartite graphs Γ for which the associated graph $B\Gamma$ falls on the exceptional side.

9. Exceptional Bipartite Metrically Homogeneous Graphs

If Γ is a connected bipartite metrically homogeneous graph, then we consider the graph $B\Gamma$ induced on each half of the bipartition by the edge relation "$d(x, y) = 2$", as described in §7. In addition to the exceptional graphs considered in the previous section, we wish to consider those for which $B\Gamma$ is itself exceptional in the sense of the previous section. The result in that case will be as follows.

THEOREM (13). *Let Γ be a connected, bipartite, and metrically homogeneous graph, of diameter at least 3, and degree at least 3, and with Γ_1 infinite. Then either $B\Gamma_1$ is isomorphic to the Rado graph, or $B\Gamma$ and Γ are in the catalog under one of the following headings.*

(1) *$B\Gamma \cong T_{\infty,\infty}$, and Γ is an infinitely branching tree $T_{2,\infty}$.*
(2) *$B\Gamma \cong K_\infty$, and Γ has diameter 3, with Γ either the complement of a perfect matching, or the generic bipartite graph $\Gamma^3_{\infty,0;10,7;\emptyset}$.*
(3) *$B\Gamma \cong K_\infty[I_2]$, Γ has diameter 4, and $\Gamma \cong \Gamma^4_{\infty,0;10,9;\emptyset}$ is the generic antipodal bipartite graph of diameter 4.*
(4) *$B\Gamma \cong G_n^c$, the complement of a Henson graph, for some $n \geq 3$, and $\Gamma \cong \Gamma^4_{\infty,0;14,9;\{I_n^{(4)}\}}$ the generic bipartite graph in which there is no set of n vertices which are pairwise at distance 4.*
(5) *$B\Gamma \cong G_3^c$, $\Gamma \cong \Gamma^5_{\infty,0;12,11;\emptyset}$ antipodal bipartite of diameter 5 as in Corollary 12.1.*

If Γ_1 is finite then this was already covered in §8. So we may suppose $\Gamma_1 \cong I_\infty$. This means that $B\Gamma$ contains an infinite clique and thus $B\Gamma_1$ also contains an infinite clique. Furthermore, it follows from the connectedness and homogeneity of Γ that $B\Gamma$ is connected.

By the Lachlan/Woodrow classification this already reduces the possibilities to the following.

(1) $B\Gamma_1$ is imprimitive of the form $m \cdot K_\infty$ or $K_\infty[I_m]$, with $1 \leq m \leq \infty$;
(2) $B\Gamma_1$ is generic omitting I_n, for some finite $n \geq 2$;

(3) $B\Gamma_1$ is the Rado graph.

The possibilities for the exceptional graph $B\Gamma$ when $B\Gamma_1$ falls under (1) or (2) above are then known, and will be treated in the following order.

(1) $B\Gamma$ is a clique;
(2) $B\Gamma \cong K_\infty[I_m]$;
(3) $B\Gamma \cong T_{r,s}$ with $2 \leq r, \infty \leq \infty$;
(4) $B\Gamma$ is the complement of a Henson graph.

9.1. Bipartite graphs with $B\Gamma$ a clique.

PROPOSITION 9.1. *Let Γ be a connected, bipartite, and metrically homogeneous graph, of diameter $\delta \geq 3$. Suppose that $B\Gamma$ is an infinite clique. Then $\delta = 3$ and Γ is either the complement of a perfect matching, or a generic bipartite graph.*

These possibilities occur in the catalog as $\Gamma^3_{a,3}$ and $\Gamma^3_{\infty,0;10,7}$, respectively.

PROOF. As $B\Gamma$ is complete, $\delta \leq 3$, and so $\delta = 3$.

Let Γ' be the graph obtained from Γ by switching edges and non-edges between the two halves of Γ. If Γ' is disconnected, then by homogeneity of Γ, the relation "$d(x, y) = 3$" defines a bijection between the halves, and Γ is the complement of a perfect matching. So suppose that Γ' is connected. Then the graph metric on Γ' results from the metric on Γ by interchanging distance 1 and 3.

Any isometry $A \cong B$ in Γ' corresponds to an isometry in Γ and is therefore induced by an automorphism of Γ. These automorphisms preserve the partition of Γ, possibly switching the two sides. So they act as automorphisms of Γ' as well. It follows that Γ' is a metrically homogeneous graph.

If Γ or Γ' has bounded degree, we contradict Theorem 10. Thus if A, B are the two halves of the partition of Γ, for $u \in A$ the set B_u of neighbors of u in B is infinite, with infinite complement. In particular for any two disjoint finite subsets B_1, B_2 of B, there is $u \in A$ such that u is adjacent to all vertices of B_1 and no vertices of B_2. This is the extension property which characterizes the Fraïssé limit of the class of finite bipartite graphs, so Γ is generic bipartite. □

9.2. Bipartite graphs with $B\Gamma \cong K_\infty[I_m]$.

PROPOSITION 9.2. *Let Γ be a connected, bipartite, and metrically homogeneous graph. Suppose that $B\Gamma \cong K_\infty[I_m]$ with $2 \leq m \leq \infty$. Then $m = 2$, and $\Gamma \cong \Gamma^4_{\infty,0;10,9;\emptyset}$, the generic antipodal bipartite graph of diameter 4.*

PROOF. The relation $d(x, y) \in \{0, 4\}$ is an equivalence relation on Γ. It follows that Γ is antipodal of diameter 4 and in particular that $m = 2$. We claim that Γ is the generic antipodal graph of diameter 4.

We call a metric subspace of Γ antipodal if it is closed under the pairing $d(x, y) = 4$ in Γ.

It suffices to show that for any antipodal subspace G of Γ and any extension $G \cup \{v\}$ of G to a graph in which all perimeters of triangles are even and bounded by 8, the extension $G \cup \{v\}$ embeds into Γ over G.

Let $X = G \cap A$, $Y = G \cap B$. We may suppose both are nonempty, and that v is on the B side in the sense that its distances from vertices of X are odd.

Our hypotheses imply that $d(v, y) = 2$ for all $y \in Y$, and that the neighbors of v in X form a set of representatives X_0 for the components of X. Thus to realize this extension in Γ, it suffices to see that there are infinitely many vertices $v \in \Gamma$ adjacent to all vertices of X_0.

As X is finite, there is an infinite subset of B consisting of vertices whose neighbors in X are identical. Furthermore, the set of neighbors in question is a set of representatives for the components of X, hence isometric with X_0. By homogeneity, X_0 has infinitely many common neighbors in B. $\square$

9.3. Bipartite graphs with $B\Gamma$ treelike.

PROPOSITION 9.3. *Let Γ be a connected, bipartite, and metrically homogeneous graph. Suppose that $B\Gamma \cong T_{r,\infty}$ with $2 \leq r \leq \infty$. Then $r = \infty$ and Γ is an infinitely branching tree.*

PROOF. Let A, B be the two halves of Γ, and identify A with $B\Gamma$. In particular, for $u \in B$, the neighbors of u in A form a clique in the sense of $B\Gamma$.

Suppose that vertices u, u' in B are adjacent to two points v_1, v_2 of A. These points are contained in a unique clique of A, so all the neighbors of u, u' lie in a common clique. Therefore this gives us an equivalence relation on B, with u, u' equivalent just in case they have two common neighbors in A. But the graph structure on B is also that of $B\Gamma$, which is primitive, so this relation is trivial, and distinct vertices of B correspond to distinct cliques in A. In particular, $r = \infty$.

At the same time, every edge in A lies in the clique associated with some vertex in B, so the neighbors of the vertices of B are exactly the maximal cliques of A. Evidently the edge relation in B corresponds to intersection of cliques in A. Thus B is identified with the "dual" of A with vertices corresponding to maximal cliques, and maximal cliques corresponding to vertices. At this point the structure of Γ has been recovered uniquely from the structure of $T_{\infty,\infty}$, and must therefore be the infinitely branching tree. $\square$

9.4. Bipartite graphs with $B\Gamma$ the complement of a Henson Graph. In this case, the diameter of Γ is 4 or 5. We take up the case of diameter 4 first.

PROPOSITION 9.4. *If $B\Gamma$ is the generic homogeneous graph omitting I_n and the diameter of Γ is 4, then $\Gamma \cong \Gamma^4_{\infty,0;14,9;\{I_n^{(4)}\}}$.*

PROOF. Let Δ, Δ' be the two halves of Γ, each isomorphic to $2B\Gamma$ (that is, isomorphic to $B\Gamma$ after rescaling the metric). We know the finite subspaces of $2B\Gamma$. Our claim is that for any finite $A, B \subseteq 2B\Gamma$ and any metric on the disjoint union $A \cup B$ in which all cross-distances between A and B are equal to either 1 or 3, there is an embedding of A into Δ and B into Δ' such that the metric of Γ induces the specified metric on $A \cup B$. Note that $A \cup B$ and Γ are both considered as metric spaces here rather than as graphs.

We will prove this by induction on the order of B. The case $k = 0$ is known so we treat the inductive step. With k fixed, we proceed by induction on the number a_4 of pairs u_1, u_2 in A with

$$d(u_1, u_2) = 4$$

So we consider an appropriate metric space $A \cup B$ with $|B| = k$. We have to treat the cases $a_4 = 0$, and the induction step for a_4. We begin with the latter.

Suppose $a_4 > 0$ and fix $u_1, u_2 \in A$ with $d(u_1, u_2) = 4$. Take $v \in B$. We may suppose $d(v, u_1) = 3$. We adjoin a vertex a to A as follows:

$$\begin{array}{ll} d(a, u_1) = 4, & d(a, u') = 2 \text{ for } u' \in A,\, u' \neq u_1 \\ d(a, v) = 1, & d(a, v') = 3 \text{ for } v' \in A,\, v' \neq v \end{array}$$

Now the configuration $(A \cup \{a\}, B \setminus \{v\})$ embeds into Γ by induction on $k = |B|$, and the configuration $(A \setminus \{u_1\} \cup \{a\}, B)$ embeds into Γ by induction on a_4 (with k fixed). So there is an amalgam $(A \cup \{a\}, B)$ of these two configurations embedding into Γ, and the metric on this amalgam agrees with the given metric on (A, B) except possibly at the pair (u_1, v). But we have $d(u_1, a) = 4$, $d(v, a) = 1$, so $d(u_1, v) \geq 3$, $d(u_1, v)$ is odd, and $d(u_1, v) \leq 4$. Thus $d(u_1, v) = 3$ also in the amalgam.

There remains the case in which $a_4 = 0$, so that in Γ, A is an independent set of the form $I_m^{(2)}$.

Suppose first that $k > 1$. Then we extend (A, B) to a finite configuration (A, B_1) with the following properties.

(1) Some vertex $b \in B_1$ is adjacent to all vertices of A.
(2) No two vertices of A have the same neighbors in B_1.

Now consider the configurations $(\{a\}, B_1)$ for $a \in A$. If they all embed into Γ, then some amalgam does as well, and this amalgam must be isomorphic to (A, B_1) since the vertices of A must remain distinct and the metric is then determined. So it suffices to check that these configurations $(\{a\}, B_1)$ embed into Γ.

As there is an automorphism of Γ switching the two halves of its bipartition, it suffices to deal with with the configuration $(B_1, \{a\})$ instead. In this configuration, the value of $k = 1$, so we conclude by induction on k.

Now suppose

$$k = 1$$

Thus $|B| = 1$; and $a_4 = 0$ by our case hypothesis.

Take a basepoint $*$ in Γ and a vertex u in Γ_2. It suffices to show that the set I_u of neighbors of u in Γ_1 is an infinite and coinfinite subset of Γ_1. By Theorem 10, the set I_u is infinite. Furthermore there are by assumption vertices u_1, u_2 in Γ with $d(u_1, u_2) = 4$, and we may suppose that the basepoint $*$ lies at distance 2 from both. Then $u_1, u_2 \in \Gamma_2$ and I_{u_1}, I_{u_2} are disjoint. Thus they are coinfinite.

This completes the identification of Γ. □

Now we turn to diameter 5. The claim in this case is as follows.

PROPOSITION 9.5. *If $B\Gamma$ is the generic homogeneous graph omitting I_n and the diameter of Γ is 5, then $n = 3$ and Γ is the generic antipodal bipartite graph of diameter 5, $\Gamma^5_{\infty,0;12,11;\emptyset}$.*

We treat the cases $n = 3$ and $n > 3$ separately. Once we have Γ antipodal, Corollary 12.1 completes the analysis.

Since we are now taking the diameter to be 5, the following will allow us to simplify the statements of some results.

LEMMA 9.6. *Let Γ be a connected bipartite metrically homogeneous graph of diameter 5. Then one of the following holds.*

(1) *Γ is the cycle C_{10}.*
(2) *$B\Gamma$ is the complement of a Henson graph G_n.*
(3) *$B\Gamma$ is the Rado graph.*

Proof. Theorem 10 takes care of the case in which Γ_1 is finite. So we may suppose $B\Gamma$ contains an infinite clique. Proposition 9.1 eliminates the possibility that $B\Gamma$ is an infinite clique. As $B\Gamma$ is connected, Proposition 9.2 disposes of the case in which $B\Gamma$ is imprimitive. By the Lachlan/Woodrow classification, this leaves the cases in which $B\Gamma$ is the complement of a Henson graph, or the Rado graph. $\square$

Lemma 9.7. *Let Γ be a metrically homogeneous bipartite graph of diameter 5 with $B\Gamma \cong G_3^c$. Then Γ is antipodal.*

Proof. Suppose $|\Gamma_5| \geq 2$. We will show first that there is a triple (a, b, c) in Γ with $d(a, b) = 4$, $d(a, c) = 5$, and $d(b, c) > 1$.

Take a pair $u_1, u_2 \in \Gamma_5$. Then $d(u_1, u_2)$ is 2 or 4, and if it is 4 then our triple (a, b, c) can be $(u_1, u_2, *)$ with $*$ the chosen basepoint. If $d(u_1, u_2) = 2$ then extend u_1, u_2 to a geodesic (u_1, u_2, u_3) with $d(u_2, u_3) = 1$, $d(u_1, u_3) = 3$. As $d(u_2, u_3) = 1$ we find $u_3 \in \Gamma_4$ and therefore the triple $(*, u_3, u_1)$ will do.

Now fix a triple (a, b, c) with $d(a, b) = 4$, $d(a, c) = 5$, and $d(b, c) = 3$ or 5. Take a triple (b, c, d) with $d(c, d) = 1$ and $d(b, d) = 4$; this will be a geodesic of length 4 or 5, and therefore exists in Γ by homogeneity.

Now $d(a, b) = d(b, d) = 4$, and consideration of the path (a, c, d) shows that $d(a, d) \geq 4$, and $d(a, b)$ is even, so $d(a, d) = 4$ as well, and we have $I_3^{(4)}$ in Γ, a contradiction. $\square$

It remains to show that in an infinite metrically homogeneous bipartite graph Γ of diameter 5 for which $B\Gamma$ contains an independent set of order 3, $B\Gamma$ contains arbitrarily large independent sets. We will subdivide this case further according to the structure of Γ_5. Since in this case Γ is not antipodal, it follows that Γ_5 is infinite. Note that if we rescale the metric on Γ_5 by 1/2 we get a homogeneous graph contained in $B\Gamma$. Thus the possible structure on Γ_5 is quite limited. The first case to be considered is that of a clique.

Lemma 9.8. *Suppose that Γ is bipartite of diameter 5, that $I_3^{(4)}$ embeds into Γ, and that $\Gamma_5 = I_\infty^{(2)}$. Then $I_\infty^{(4)}$ embeds in $B\Gamma$ and $B\Gamma$ is the Rado graph.*

Proof. Our claim is that $I_n^{(4)}$ embeds into Γ for all n; then since $\delta = 5$, $B\Gamma$ is a connected homogeneous graph containing an infinite clique and an infinite independent set. Proposition 9.2 implies that $B\Gamma$ is not of the form $K_\infty[I_\infty]$ and then the Lachlan/Woodrow classification leaves only the Rado graph.

We proceed by induction, with the case $n = 3$ assumed.

Suppose $I_n^{(4)}$ embeds into Γ, with $n \geq 3$. Let $I \cong I_{n-1}^{(4)}$ be a metric subspace of Γ. We aim to embed subspaces $A = I \cup \{a, u\}$ and $B = I \cup \{b, u\}$ into Γ, with $I \cup \{a\} \cong I \cup \{b\} \cong I_n^{(4)}$, and with u chosen so that

$$\begin{aligned} d(u, a) &= 1 \quad d(u, b) = 5 \\ d(u, x) &= 3 \qquad (x \in I) \end{aligned}$$

Supposing we have this, considering (a, u, b) we see that $d(a, b) \geq 4$ and hence $I \cup \{a, b\} \cong I_{n+1}^{(4)}$.

We treat the second factor $I \cup \{b, u\}$ first. Consider the metric space $I \cup \{b, b'\}$ in which b' lies at distance 2 from each point of $I \cup \{b\}$. The corresponding configuration in $B\Gamma$ is a point b' adjacent to an independent set of order n, and this we have in $B\Gamma$. Thus the space $I \cup \{b, b'\}$ embeds into Γ.

By hypothesis, there is also a triple (u, b, b') with $b, b' \in \Gamma_5(u)$. Amalgamate $I \cup \{b, b'\}$ with (u, b, b') over b, b'. For $x \in I$, considering (u, b', x), we see that $d(u, x) \geq 3$, and that $d(u, x)$ is odd. Our hypothesis on Γ_5 implies that $d(u, x) \neq 5$, so $d(u, x) = 3$ for all $x \in I$. Thus $I \cup \{b, u\}$ is as desired.

The construction of the factor $I \cup \{a, u\}$ is more elaborate. Consider the metric spaces $A = I \cup \{a\} \cup J$ and $B = J \cup \{a, u\}$, where $J \cong I_n^{(2)}$, and J may be labeled as $\{v^* : v \in I\}$ in such a way that

$$\begin{array}{cc} d(a, v^*) = d(v, v^*) = 4 & (v \in I) \\ d(v, w^*) = 2 & v, w \in I \text{ distinct} \\ d(u, v) = 5 & (v \in J) \end{array}$$

Supposing that A and B embed into Γ, take their amalgam over $J \cup \{a\}$. Then for $v \in I$ the triple (u, a, v) shows that $d(u, v) \geq 3$, and the distance is odd, while the triple (u, v', v) and the hypothesis on Γ_5 shows that this distance is not 5. Thus the space $I \cup \{a, u\}$ will have the desired metric. It remains to construct A and B.

Consider $B = J \cup \{a, u\}$. The graph Γ contains an edge (a, u) as well as a copy of $J \cup \{u\}$, the latter by the hypothesis on Γ_5. Furthermore, in any amalgam of (a, u) with $J \cup \{u\}$, the only possible value for the distance $d(a, v)$, for $v \in J$, is 4. So this disposes of B.

Now consider $A = I \cup J \cup \{a\}$, in which all distances are even. So we need to look for the rescaled graph $(1/2)A$ in $B\Gamma$. It suffices to check that the maximal independent sets of vertices in $(1/2)A$ have order at most n. This is the case for $I \cup \{a\}$, and any independent set meeting J would have order at most 3. Since $n \geq 3$, we are done. □

LEMMA 9.9. *Suppose that Γ is metrically homogeneous, bipartite, infinite, not antipodal, and of diameter 5. Then Γ_5 contains a subspace of the form $I_\infty^{(2)}$; in other words, $(1/2)\Gamma_5$ contains an infinite clique.*

PROOF. Supposing the contrary, for each $u \in \Gamma_4$, the set I_u of neighbors of u in Γ_5 is finite and nonempty, of fixed order k. Since any subset of Γ_5 isomorphic to $I_{k+1}^{(2)}$ would have a common neighbor $u \in \Gamma_4$, it follows that the I_u represent maximal cliques of $(1/2)\Gamma_5$.

As $B\Gamma$ is either generic omitting I_n for some $n \geq 3$, or the Rado graph, it follows that $(1/2)\Gamma_4$ is primitive and contains both edges and nonedges. Now the map $u \to I_u$ induces an equivalence relation on Γ_4 which can only be equality, that is the map is a bijection. Since $(1/2)\Gamma_4$ contains both edges and non-edges, it follows that Γ_5 is primitive. As each vertex $v \in \Gamma_5$ has infinitely many neighbors in Γ_4, we have $|I_u| > 1$ for $u \in \Gamma_4$. On the other hand if $|I_u| \geq 3$ then for $u, u' \in \Gamma_4$ we have the possibilities $|I_u \cap I_{u'}| = 0, 1, 2$ while there are only two distances occurring in Γ_4. So $|I_u| = 2$. That is, $(1/2)\Gamma_5$ is generic triangle-free, and the vertices of Γ_4 correspond to edges of $(1/2)\Gamma_5$. It follows that vertices of Γ_4 lie at distance two iff the corresponding edges meet, that is $(1/2)\Gamma_4$ is the line graph of $(1/2)\Gamma_5$. But there are pairs of vertices in the latter graph at distance greater than 2, so $(1/2)\Gamma_4$ is not homogeneous, and we have a contradiction. □

LEMMA 9.10. *Suppose that Γ is bipartite, metrically homogeneous, and of diameter 5, and Γ_5 contains a pair of vertices at distance 4. Then the relation*

$$\text{“}d(x, y) = 0 \text{ or } 4\text{”}$$

is not an equivalence relation on Γ_5*.*

PROOF. Supposing the contrary, we have

$$\Gamma_5 \cong I_\infty^{(2)}[I_k^{(4)}]$$

for some k with $2 \le k \le \infty$.

Suppose first $k \ge 3$. Fix two equivalence classes C, C' in Γ_5, and choose a triple u_1, u_2, u_3 in C_1 and a vertex u_1' in C'. Choose $v \in \Gamma$ with $d(v, u_1) = d(v, u_1') = 1$, and let $d_i = d(v, u_i)$ for $i = 2, 3$. We may then choose u_2', u_3' in C' so that $d(v, u_i') = d_i$ for $i = 2, 3$.

Now the permutation of the u_i, u_i' which switches u_1' and u_2' and fixes the other elements is an isometry, so there is an element v' with $d(v', u) = d(v, u)$ for $u = u_1, u_2, u_3, u_3'$, but with $d(v', u_1') = d_2$, $d(v', u_2') = 1$.

As u_1 is adjacent to v, v' we have $d(v, v') = 2$. Now u_3, v, v' is isometric with u_3', v, v', and the equivalence class of u_3 contains a common neighbor of v, v'; therefore the equivalence class of u_3' contains a common neighbor of v, v'. But v can have at most one neighbor in an equivalence class, so this contradicts the choice of v'.

So we are left with the case $k = 2$:

$$\Gamma_5 \cong I_\infty^{(2)}[I_2^{(4)}]$$

In this case we will consider a specific amalgamation.

Let $\gamma = (u, v, w)$ be a geodesic with

$$d(u, v) = 1; d(v, w) = 4; d(u, w) = 5$$

Let $A = \gamma \cup \{a\}$, $B = \gamma \cup \{b\}$, with the metrics given by

	u	v	w
a	4	3	5
b	4	5	5

It is easy to see that A, B embed into Γ. In an amalgama of the two, the relation of a and b to v prevents them from being identified in the amalgam. However $a, b, u \in \Gamma_5(w)$ and $d(a, u) = d(b, u) = 4$. So $d(a, b) = 4$ by our assumption, and this contradicts $k = 2$. □

LEMMA 9.11. *Suppose that* Γ *is bipartite, metrically homogeneous, and of diameter* 5*. Then* Γ_i *is connected with respect to the edge relation given by* $d(x, y) = 2$*, for* $1 \le i \le 5$*.*

PROOF. This is true for Γ_1 automatically. It is true for $i = 2$ or 4 in view of the structure of $B\Gamma$. It remains to prove it for $i = 3$ or 5.

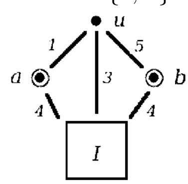
If Γ_i is disconnected with respect to this relation, then for $u \in \Gamma_{i-1}$, the set I_u of neighbors of u in Γ_i is contained in one of the equivalence classes of Γ_i, and there is more than one such class. Thus we have a function from Γ_{i-1} to the quotient of Γ_i. As $i - 1$ is even, in view of the structure of $B\Gamma$ we know Γ_{i-1} is primitive, so as Γ_i contains more than one equivalence class, this function is $1-1$. Then the sets I_u for $u \in \Gamma_{i-1}$ must be exactly the equivalence classes of Γ_i, and Γ_{i-1} is in bijection with the quotient. In particular, only one distance occurs in Γ_{i-1}. But in view of the structure of $B\Gamma$, this is not the case. □

COROLLARY 9.11.1. *Suppose that* Γ *is bipartite, metrically homogeneous, and of diameter* 5, *and* Γ_5 *contains a pair of vertices at distance* 4. *Then* Γ_5 *is primitive, infinite, and contains a copy of* $I_\infty^{(2)}$.

LEMMA 9.12. *Suppose that* Γ *is bipartite, metrically homogeneous, and of diameter* 5, *that* Γ_5 *contains a pair of vertices at distance* 4, *and that* $B\Gamma$ *is generic omitting* I_{n+1}. *Then* $I_{n-1}^{(4)}$ *embeds into* Γ_5.

PROOF. Let k be maximal so that $I_k^{(4)}$ embeds into Γ_5, and suppose $k \le n-2$.

Let $I \cong I_{n-1}^{(4)}$, and suppose a, b, u are additional vertices with $I \cup \{a\} \cong I \cup \{b\} \cong I_n^{(4)}$, and with $d(u,a) = 1$, $d(u,b) = 5$, $d(u,v) = 3$ for $v \in I$. If $I \cup \{a,u\}$ and $I \cup \{b,u\}$ embed into Γ, then so does an amalgam $I \cup \{a,b,u\}$, and the auxiliary vertex u forces $d(a,b) = 4$, and $I \cup \{a,b\} \cong I_{n+1}^{(4)}$, a contradiction as $B\Gamma$ omits I_{n+1}. So it suffices to embed $I \cup \{a,u\}$ and $I \cup \{b,u\}$ into Γ.



Construction of $I \cup \{a,u\}$.

Introduce a metric space $J = \bigcup_{v\in I} J_v$ with $J_v \cong I_k^{(4)}$ and with $d(x,y) = 2$ for $x \in J_v$, $y \in J \setminus J_v$. Extend to a metric on $I \cup J$ by taking

$$d(v,x) = \begin{cases} 4 & \text{if } x \in J_v \\ 2 & \text{if } x \in J \setminus J_v \end{cases}$$

for $v \in I$.

Give $J \cup \{a,u\}$ the metric with $d(a,x) = 4$, $d(u,x) = 5$ for $x \in J$. We claim that $I \cup J \cup \{a\}$ and $J \cup \{a,u\}$ embed into Γ. Now $B\Gamma$ is generic omitting I_{n+1}, and $(1/2)\Gamma_5$ is generic omitting I_{k+1}. Since the space $I \cup J \cup \{a\}$ does not contain $I_{n+1}^{(4)}$, and all its distances are even, it embeds into Γ. Since J does not contain $I_{k+1}^{(4)}$, it embeds into Γ_5, so $J \cup \{u\}$ embeds into Γ. In any amalgam of $J \cup \{u\}$ with $\{a,u\}$ we have $d(a,x) = 4$ for $x \in J$, so $J \cup \{a,u\}$ embeds into Γ as well.

Thus an amalgam of $I \cup J \cup \{a\}$ and $J \cup \{a,u\}$ embeds into Γ. For $v \in I$, consideration of (u,a,v) shows that $d(u,v)$ is 3 or 5, and consideration of $J_v \cup \{u,v\}$ shows that $d(u,v)$ is not 5. Thus we have $d(u,v) = 3$ for all $v \in I$ in our amalgam, and thus $I \cup \{a,u\}$ embeds isometrically into Γ.

Construction of $I \cup \{b,u\}$.

Let $J' = \bigcup_{v\in I} J'_v$ with $J'_v \cong I_{k-1}^{(4)}$. Put a metric on $I \cup J' \cup \{b,u\}$ by taking $d(u,x) = 5$, $d(b,x) = 4$ for $x \in J'$, while for $v \in I$ we take $d(v,x) = 4$ for $x \in J'_v$, and $d(v,x) = 2$ for $x \in J \setminus J'_v$.

Introduce an auxiliary vertex b' with $d(b',u) = 5$, $d(b',x) = 2$ for $x \in I \cup J' \cup \{b\}$.

We claim that $I \cup J' \cup \{b,b'\}$ and $J' \cup \{u,b,b'\}$ embed isometrically in Γ. For $I \cup J' \cup \{b,b'\}$ we use the structure of $B\Gamma$, together with the condition $k+1 < n$, and for $J' \cup \{u,b,b'\}$ we use the structure of Γ_5 to check that $J \cup \{b,b'\}$ embeds into Γ_5.

Therefore some amalgam $I \cup J' \cup \{u, b, b'\}$ embeds into Γ. Let $v \in I$. In the amalgam, the auxiliary vertex b' ensures that $d(u, v)$ is 3 or 5. Consideration of $J'_v \cup \{u, b, v\}$ shows that $d(u, v)$ is not 5. Thus $d(u, v) = 3$ for $v \in I$, and $I \cup \{b, u\}$ embeds isometrically in Γ. □

We will need some additional amalgamation arguments to complete our analysis, beginning with the following preparatory lemma.

LEMMA 9.13. *Let Γ be bipartite of diameter 5, and not antipodal. Suppose $B\Gamma$ is generic omitting $I^{(4)}_{n+1}$, with $n \geq 3$, and Γ_5 contains a pair of vertices at distance 4. Then the following hold.*

(1) *$I^{(4)}_n$ embeds in Γ_3;*
(2) *Γ_3 is primitive.*

PROOF.
1. $I^{(4)}_n$ embeds in Γ_3:
We show inductively that $I^{(4)}_m$ embeds into Γ_3 for $m \leq n$.

Let $I \cong I^{(4)}_{m-1}$. Form extensions $I \cup \{u\}$ and $I \cup \{v\}$ with $d(u, x) = 5$, $d(v, x) = 3$ for $x \in I$. Then $I \cup \{u\}$ embeds into Γ since $m - 1 \leq n - 1$, while $I \cup \{v\}$ embeds into Γ by induction on m. So some amalgam $I \cup \{u, v\}$ embeds into Γ with $d(u, v)$ either 2 or 4.

Consider a geodesic $\{u, v, w\}$ with $d(u, w) = 1$, $d(v, w) = 3$, and $d(u, v)$ as specified. There is an amalgam $I \cup \{u, v, w\}$ of $I \cup \{u, v\}$ with $\{u, v, w\}$ over u, v, and consideration of (w, u, x) for $x \in I$ show that $I \cup \{w\} \cong I^{(4)}_m$. As $I \cup \{w\} \subseteq \Gamma_3(v)$, the induction is complete.

2. Γ_3 is primitive:
By Lemma 9.11 we have $(1/2)\Gamma_3$ connected. Suppose now that Γ_3 is disconnected with respect to the edge relation "$d(x, y) = 4$."

Fix two connected components C, C' with respect to this relation. By (1) these have order n, and by assumption $n \geq 3$. Fix $u \in C$ and $u'_1, u'_2 \in C'$, and $v_1 \in \Gamma_2$ adjacent to u, u'_1. With $*$ the chosen basepoint for Γ, consider the isometry of $C \cup C' \cup \{*\}$ which interchanges u'_1 and u'_2 and fixes the remaining vertices. Then this extends to an isometry $C \cup C' \cup \{*, v\} \cong C \cup C' \cup \{*, v'\}$ for some vertex v'. Take $u'_3 \in C'$, distinct from u_1, u_2. Then the map $(*, u_3, u'_3, v, v') \mapsto (*, u'_3, u_3, v, v')$ is an isometry and therefore extends to Γ; its extension interchanges C and C' and fixes v, v'. However $d(v, x) = d(v', x)$ for $x \in C$, so the same applies to C'. But $d(v, u'_1) = 1$, $d(v', u'_2) = 1$, and $d(u'_1, u'_2) = 4$, so this is impossible. □

Now we can assemble these ingredients.

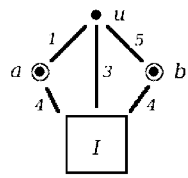
LEMMA 9.14. *If Γ is bipartite of diameter 5 and not antipodal, then $B\Gamma$ is the universal homogeneous graph (Rado's graph).*

PROOF. The alternative is that $B\Gamma$ is generic omitting I_{n+1} for some $n \geq 3$. By Lemma 9.8, we may suppose that Γ_5 contains a pair of vertices at distance 4, and hence $I^{(4)}_{n-1}$ embeds in Γ_5 by Lemma 9.12.

To get a contradiction, we will aim at an amalgamation of the following form. Let $I \cong I^{(4)}_{n-1}$, let $I \cup \{a\} \cong I \cup \{b\} \cong I^{(4)}_n$, and adjoin a vertex u such that

$$d(u, a) = 1;\ d(u, b) = 5;\ d(u, x) = 3 \text{ for } x \in I$$

We will embed $I \cup \{a,u\}$ and $I \cup \{b,u\}$ in Γ, and then in their amalgam we will have $I \cup \{a,b\} \cong I^{(4)}_{n+1}$, a contradiction. Each of the factors $I \cup \{a,u\}$ and $I \cup \{b,u\}$ will require its own construction.



Construction of the first factor, $I \cup \{a,u\}$.

Let $I = I_0 \cup \{c\}$ and introduce a vertex v with

$$\begin{array}{lll} d(v,a) = 1 & d(v,c) = 5 & d(v,u) = 2 \\ d(v,x) = 4 & (x \in I_0 \cup \{a\}) & \end{array}$$

If $I_0 \cup \{a,u,v\}$ and $I_0 \cup \{c,u,v\}$ embed into Γ, then in their amalgam $I_0 \cup \{a,c,u,v\}$ we have $d(a,c) = 4$ and thus the desired metric space $I \cup \{a,u\}$ is embedded into Γ.

Construction of $I_0 \cup \{a,u,v\}$.

We first embed $I_0 \cup \{a,u\}$ into Γ. Introduce a vertex a' with

$$d(a',u) = 1;\ d(a',a) = 2;\ d(a',x) = 2 \text{ for } x \in I_0$$

On the one hand, the geodesic (a,u,a') embeds into Γ; on the other hand, the metric space $I_0 \cup \{a,a'\}$ embeds into $B\Gamma$. So some amalgam $I_0 \cup \{a,a',u\}$ embeds into Γ, and for $x \in I_0$, consideration of the paths (u,a,x) and (u,a',x) shows that $d(u,x) = 3$, as required. Thus $I_0 \cup \{a,u\}$ embeds into Γ.

Take a as basepoint. Then $u \in \Gamma_1$, $I_0 \subseteq \Gamma_4$, and $d(u,x) = 3$ for $x \in I_0$. Let $I_1 \subseteq I_0$ be obtained by removing one vertex, so $I_1 \cong I^{(4)}_{n-3}$. Consider the sets

$$\begin{array}{rcl} A & = & \{u \in \Gamma_1 : d(u,x) = 3 \text{ for } x \in I_1\} \\ B & = & \{u \in \Gamma_4 : d(u,x) = 4 \text{ for } x \in I_1\} \end{array}$$

The partitioned metric space (A,B) is homogeneous with respect to the metric plus the partition. We consider the structure of (A,B).

We show first that A is infinite. Assuming the contrary, consider a configuration $I_1 \cup I_2$ in $I_2 \cong I^{(2)}_\infty$, and $I_1 \cup \{x\} \cong I^{(4)}_{n-2}$ for each $x \in I_2$. This configuration embeds into Γ_4. There is a pair $x,y \in I_2$ such that $I_1 \cup \{x\}$ and $I_1 \cup \{y\}$ have the same vertices at distance 3 in Γ_1. Now Γ_4 is generic omitting $I^{(4)}_n$. By homogeneity it follows easily that any two subsets of Γ_4 isomorphic to $I^{(4)}_{n-2}$ have the same vertices at distance 3 in Γ_1. This yields a nonempty subset of Γ_1 definable without parameters, and a contradiction. So A is infinite.

Now B is generic omitting $I^{(4)}_3$. In particular B is primitive. Furthermore, each vertex of B lies at distance 3 from some vertex of A. By primitivity, this vertex cannot be unique. Take $c \in B$ and $u,v \in A$ so that $d(c,u) = d(c,v) = 3$. Then $I_0 \cup \{c,a,u,v\}$ has the desired structure.

Construction of $I_0 \cup \{c,u,v\}$.

We introduce a vertex d at distance 1 from u and v, and distance 4 from c, with the relation of d to I_0 will be determined below.

In any amalgam of $I_0 \cup \{c, d, u\}$ with $I_0 \cup \{c, d, v\}$ over $I_0 \cup \{c, d\}$ we have $d(u, v) = 2$. It remains to construct $I_0 \cup \{c, d, u\}$ and $I_0 \cup \{c, d, v\}$.

We claim first that $I_0 \cup \{c, u\}$ embeds into Γ, in other words that $I_0 \cup \{c\}$ embeds into $\Gamma_3(u)$. This holds by Lemma 9.13. Now we may form $I_0 \cup \{c, d, u\}$ by amalgamating $I_0 \cup \{c, u\}$ with $\{c, d, u\}$ (a geodesic) to determine the metric on $I_0 \cup \{d\}$; all distances $d(x, d)$ will be even for $x \in I_0$. This amalgamation determines the structure of $I_0 \cup \{d\}$ and thereby completes the determination of the second factor $I \cup \{c, d, v\}$ as well.

We claim that $I \cup \{c, d, v\}$ embeds into Γ. Since the distance $d(c, d) = 4$ is forced in any amalgam of $I_0 \cup \{v, c\}$ with $I_0 \cup \{v, d\}$, we consider these two metric spaces separately.

Now $I_0 \cup \{v, d\} \cong I_0 \cup \{u, d\}$, so this is not at issue, and we are left only with $I_0 \cup \{c, v\}$. This last embeds into the second factor $I \cup \{b, u\}$, so we may turn finally to a consideration of this second factor.

Construction of the second factor, $I \cup \{b, u\}$.

We introduce another vertex v satisfying

$$d(v, b) = 1;\ d(v, u) = 4;\ d(v, x) = 5 \text{ for } x \in I$$

This will force $d(b, x) = 4$ for $x \in I$. So it will suffice to embed $I \cup \{u, v\}$ and $\{b, u, v\}$ separately into Γ. Since $\{b, u, v\}$ is a geodesic, we are concerned with $I \cup \{u, v\}$.

Introduce a vertex d with

$$d(u, d) = 1;\ d(v, d) = 5;\ d(x, d) = 2 \text{ for } x \in I$$

Then amalgamation of $I \cup \{d, u\}$ with $I \cup \{d, v\}$ forces $d(u, v) = 4$. It remains to embed $I \cup \{d, u\}$ and $I \cup \{d, v\}$ into Γ.

The second of these, $I \cup \{d, v\}$, has a simple structure with $I \cup \{d\} \subseteq \Gamma_5(v)$, and since $I \cup \{d\}$ has order n, with all distances even, it embeds into Γ_5 by Lemma 9.12. So we need only construct $I \cup \{d, u\}$.

Taking d as base point, and I contained in Γ_2, we are looking for a vertex $u \in \Gamma_1$ at distance 3 from all elements of I. For $v \in I$, let I_v be the set of neighbors of v in Γ_1. Any vertex $u \in \Gamma_1$ which is not in $\bigcup_{v \in I} I_v$ will do. So it remains to be checked that $\bigcup_{v \in I} I_v \neq \Gamma_1$.

The sets I_v for $v \in I$ are pairwise disjoint. Suppose they partition Γ_1. We may take a second set $J \cong I^{(4)}_{n-1}$ in Γ_2 overlapping with I so that $|I \cap J| = n - 2$, and then the I_v for $v \in J$ will also partition Γ_1; so the vertices $v_1 \in I \setminus J$ and $v_2 \in J \setminus I$ have the same neighbors in Γ_1. As Γ_2 is primitive, it follows that all vertices of Γ_2 have the same neighbors in Γ_1, a contradiction. □

At this point, the proof of Proposition 9.5, and also Theorem 13, is complete. We review the analysis.

Proof of Theorem 13. $B\Gamma_1$ falls under the Lachlan/Woodrow classification.

If Γ_1 is finite, then Theorem 10 applies, and as we assume diameter at least 3 and degree at least 3, we arrive at either the complement of a perfect matching or a tree in this case.

With Γ_1 infinite, $B\Gamma$ contains an infinite clique, and hence so does $B\Gamma_1$. As noted at the outset, $B\Gamma$ is connected. So by the Lachlan/Woodrow classification, $B\Gamma$ is either imprimitive of the form $K_\infty[I_m]$ or $m \cdot K_\infty$, $2 \leq n \leq \infty$, or generic omitting I_n for some finite $n \geq 2$, or universal homogeneous (Rado's graph).

When $B\Gamma_1$ is imprimitive, the classification in Theorem 10 applies to $B\Gamma$, and as $B\Gamma$ contains an infinite clique, the result is that $B\Gamma$ is either an imprimitive homogeneous graph of the form $K_\infty[I_m]$ ($2 \le m \le \infty$), or one of the tree-like graphs of Macpherson, $T_{r,\infty}$ with $2 \le r \le \infty$.

When $B\Gamma$ is of the form $K_\infty[I_m]$ with $m \ge 2$, Lemma 9.2 applies. When $B\Gamma$ is tree-like, Lemma 9.3 applies, and Γ is a tree.

Thus we may suppose that $B\Gamma_1$ is primitive. We have set aside the case in which $B\Gamma_1$ is universal homogeneous as a distinct (and typical) case. So we are left with the possibility that $B\Gamma_1$ is generic omitting I_n with $2 \le n < \infty$. In view of Theorem 10, $B\Gamma$ must have diameter at most 2, and be homogeneous as a graph. Then our hypothesis on $B\Gamma_1$ implies that $B\Gamma$ is also generic omitting I_n.

In case $n = 2$, Lemma 9.1 applies. If $n > 2$ then the diameter of Γ is 4 or 5. If the diameter is 4, then Lemma 9.4 applies, and leads to case 4 of the theorem.

This leaves us with the case taken up in Proposition 9.5: Γ has diameter 5, and $B\Gamma$ is generic omitting I_n. As the diameter is 5, we have $n \ge 3$. As $B\Gamma_1$ does not contain I_∞, Γ is antipodal by Lemma 9.14. So Corollary 12.2 applies. □

10. Graphs of small diameter

In the Appendix to [**Che98**], we gave an exhaustive list of certain amalgamation classes for highly restricted languages. The languages considered were given by a certain number of irreflexive binary relations, symmetric or asymmetric, with the proviso that every pair of distinct elements satisfies one and only one of the given relations. The cases of interest here are the languages with either 3 or 4 symmetric irreflexive binary relations. The amalgamation classes $\mathcal{A}$ listed were those satisfying the following three conditions:

(1) The class $\mathcal{A}$ is determined by a set of forbidden triangles.
(2) The Fraïssé limit of the class is primitive.
(3) The class in question is not a free amalgamation class.

This last point means that there is no single relation $R(x, y)$ such that every amalgamation problem $A_0 \subseteq A_1, A_2$ can be completed by taking R to hold between $A_1 \setminus A_0$ and $A_2 \setminus A_0$. This excludes some readily identified metrically homogeneous graphs of diameter 3, but none of greater diameter.

For example, in the case of three symmetric relations $A(x, y)$, $B(x, y)$, $C(x, y)$, the only such class (up to a permutation of the language) is the one given by the following constraints:

$$(AAB), (ACC), (AAA)$$

In this notation, AAB represents a triple x, y, z with $A(x, y)$, $A(x, z)$, and $B(y, z)$.

Now for the Fraïssé limit of this class to correspond to a metric space of diameter 3, one of the forbidden configurations must correspond to the triangle (113). Since the configuration corresponding to (113) may be either (AAB) or (ACC), there are two distinct homogeneous metric spaces of this type, with the following constraints:

$$(113), (122), (111) \text{ or } (233), (113), (333)$$

The first of these has no triangle of odd perimeter 5 or less, that is $K_1 = K_2 = 3$, $C_0 = 10$, $C_1 = 11$. The second has no triangle of perimeter 8 or more, that is $K_1 = 1$, $K_2 = 3$, $C_0 = 8$, $C_1 = 9$. We recognize these as falling within our previous classification with $C > 2\delta + K_1$ and $C' = C + 1$.

Of course, any of our examples of type $\mathcal{A}^3_{K_1,K_2;C_0,C_1}$ will be a free amalgamation class unless one of the forbidden triangles involves the distance 2, which means:

$$K_1 = 3 \text{ or } C \leq 8$$

And as our list was confined to the primitive case it omits bipartite and antipodal examples.

We view the classification of the amalgamation classes determined by triangles as a natural ingredient of a full catalog of "known" types, and a natural point of departure for an attempt at a full classification. In [**AMp10**] (a working draft) the problem is taken from the other end, in the case of diameter 3, and it is shown that in the triangle-free case (i.e., K_3-free), the classification does indeed reduce to identifying the combinations of triangle constraints and $(1, 2)$-space constraints which (jointly) define amalgamation classes.

10.1. Diameter 4. The explicit classification of the metrically homogeneous graphs of diameter 4 whose minimal forbidden configurations are triangles is more complex. Once the diameter δ exceeds 3, the possibility of "free amalgamation" falls by the wayside, as one can use an amalgamation to force any particular distance strictly between 1 and δ, using the triangle inequality. So in diameter 4, the table in [**Che98**] covers all primitive homogeneous structures with 4 symmetric 2-types which can be given in terms of forbidden triangles.

There are 27 such (up to a permutation of the language) of which 17 correspond to homogeneous metric spaces, some in more than one way (permuting the distances matters to us, if not to the theory). We will exhibit those classes in a number of formats. Table 4, at the end of the present paper, gives all 27 classes in the order they were originally given, using the symbols A, B, C, D for the binary relations involved. The numbering of cases used in the next two tables conforms to the numbering given in that table. Note that the entry "1" means the given configuration is included in the constraint set, and hence is *omitted* by the structures under consideration.

In Table 1 we have converted A, B, C, D to distances, usually in the order $1, 2, 3, 4$, or $4, 2, 3, 1$, and in one case in the order $2, 4, 3, 1$, for those cases in which the result is a metric space. In checking the possibilities, begin by identifying the forbidden triangle $(1, 2, 4)$, involving three distinct distances; in all 27 cases this can only be the triple (ABD), by inspection. Thus C corresponds to distance 3. After that, look for the forbidden triangle (113): by inspection, this is either CDD or AAC in each case. Thus A or D corresponds to distance 1, after which there is at most one assignment of distances that produces the constraint (114).

All primitive metrically homogeneous graphs of diameter 4 whose constraints are all of order 3 are listed in the resulting table. We omit the columns corresponding to the non-geodesic triangles of types $(1, 2, 4)$, $(1, 1, 3)$, and $(1, 1, 4)$, which are of course present as constraints in all cases.

In Table 2 we list these metric spaces together with their defining parameters K_1, K_2, C, C'.

Let us compare the outcome to the statement of Theorem 9.

The table contains no examples with $C \leq 2\delta + K_1$. The case $C' > C + 1$ and $3K_2 = 2\delta - 1$ is impossible with $\delta = 4$, while the case $K_1 = \infty$ is imprimitive and omitted. On the other hand, when $C' = C + 1$, if $C = 2\delta + 1$ then again the graph is imprimitive, while if $C \geq 2\delta + 2$ then $K_1 \geq 2$, and the condition

#	111	122	133	144	223	244	334	444	344	A,B,C,D
23′	0	0	0	1	0	0	0	1	1	1,2,3,4
7	0	0	0	1	0	0	0	1	0	4,2,3,1
3	0	0	0	1	0	0	0	0	0	4,2,3,1
25	1	0	0	0	0	0	0	1	1	1,2,3,4
25′	1	0	0	0	0	0	0	1	1	4,2,3,1
8	1	0	0	1	0	0	0	1	0	4,2,3,1
5	1	0	0	1	0	0	0	0	0	4,2,3,1
18	1	1	0	1	0	0	0	1	0	4,2,3,1
16	1	1	0	1	0	0	0	0	0	4,2,3,1
26′	0	0	0	0	0	1	1	1	1	1,2,3,4
21′	0	0	0	0	0	0	0	1	1	1,2,3,4
4	0	0	0	0	0	0	0	1	0	4,2,3,1
1	0	0	0	0	0	0	0	0	0	4,2,3,1
24	1	0	0	0	0	0	0	1	1	4,2,3,1
6	1	0	0	0	0	0	0	1	0	4,2,3,1
2	1	0	0	0	0	0	0	0	0	4,2,3,1
17	1	1	0	0	0	0	0	1	0	4,2,3,1
15	1	1	0	0	0	0	0	0	0	4,2,3,1
26	1	1	1	0	1	0	0	0	0	2,4,3,1
22	0	0	0	1	0	0	0	0	1	4,2,3,1
22′	0	0	0	1	0	0	0	0	1	1,2,3,4
23	1	0	0	1	0	0	0	0	1	4,2,3,1

TABLE 1. 22 metric spaces, with duplication

$K_1 + 2K_2 \leq 2\delta - 1 = 7$ gives $K_2 = K_1 = 2$, and hence $C = 2(K_1 + K_2) + 1 = 2\delta + 1$ after all.

So what we see here is the range of possibilities illustrating the third case under Theorem 9: $K_1 < \infty$, $C > 8 + K_1$, $K_1 + 2K_2 \geq 7$, $3K_2 \geq 8$, and if $K_1 + 2K_2 = 7$ then $C \geq 10 + K_1$, while if $C' > C + 1$ then $C \geq 8 + K_2$.

If $C' = C + 1$ these constraints amount to $1 \leq K_1 \leq K_2$, $K_2 = 3$ or 4, $C \geq 9 + K_1$, and if $K_1 = 1$ and $K_2 = 3$ then $C \geq 11$. This corresponds to the first two sections of the table, arranged according to increasing K_2.

If $C' > C + 1$ then there is the added constraint $C \geq 8 + K_2$ and this implies $C = 11$, $C' = 14$, $K_2 = 3$. Since $C > 2\delta + K_1$ we find $K_1 \leq 2$. This corresponds to the last two lines of the table.

Part III. Extension Properties of Finite Triangle Free Graphs

11. Extension properties

A structure has the *finite model property* if every first order sentence true in the structure is true in some finite structure. A slightly stronger property is the *finite submodel property*, where the finite approximation should be taken to lie within the original structure. There is little difference between the notions in the cases of most immediate concern here. We focus mainly on the finite model property for the generic triangle free graph, and as "triangle free" is part of the first order

#	K_1	K_2	C	C'
$23'$	1	3	11	12
7	1	3	12	13
3	1	3	13	14
$25, 25'$	2	3	11	12
8	2	3	12	13
5	2	3	13	14
18	3	3	12	13
16	3	3	13	14
$26'$	1	4	10	11
$21'$	1	4	11	12
4	1	4	12	13
1	1	4	13	14
24	2	4	11	12
6	2	4	12	13
2	2	4	13	14
17	3	4	12	13
15	3	4	13	14
26	4	4	13	14
$22,22'$	1	3	11	14
23	2	3	11	14

TABLE 2. 20 metric spaces, sorted

theory of this structure, the finite model property and the finite submodel property are equivalent here. We do not necessarily expect the finite submodel property to be true. The evidence in either direction is meager. However that may be, the problem is a concrete one connected with problems in finite geometries, and we will take pains to put it in an explicit form.

While the finite model problem has attracted the attention of a number of combinatorialists and probabilists, the main conclusion to date is that it is elusive, and there is hardly any literature to be found on the subject. Here I assemble and document the main information that has come my way about the problem.

There is some literature on lower bounds for finite approximations to Fraïssé limits, in cases where probabilistic considerations guarantee their existence, e.g. [**Sz²65, Bon09, Bon10**]. Such explorations may also shed light on cases where the existence of any such approximation remains in question. Below we will discuss some explorations of that sort in the case of triangle free graphs, where we see clear possibilities for some meaningful analysis.

11.1. Extension Properties E_n, E'_n, Adj_n. In general, the theory of a homogeneous locally finite structure is axiomatized by two kinds of axioms: negative axioms defining the "forbidden substructures," e.g. triangles in the case that most concerns us, and extension properties stating that for each k, any k-generated subset of k elements may be extended to a $(k+1)$-generated subset in any way not explicitly ruled out by the negative constraints. We will drop all further mention of k-generated structures here, and suppose that the language is purely relational, so that the issue is one of extending k given elements by one more element.

In the context of triangle free graphs, the natural extension properties are the following three.

(E_n): for any set A of at most n vertices, and any subset B of A consisting of independent vertices, there is a vertex v adjacent to all vertices of B, and to none of $A \setminus B$.

(E'_n): (a) any maximal independent set contains at least n vertices; and (b) for any set A of n independent vertices, and any subset B of A, there is a vertex v adjacent to all vertices of B, and to none of $A \setminus B$.

(Adj_n): Any set of at most n independent vertices has a common neighbor.

Here the properties E_n, together with the axiom stating that triangles do not occur, gives the full axiomatization of the generic triangle free graph. Therefore the problem of the finite model property for the generic triangle free graph is simply the question, whether for each n some finite triangle free graph has the extension property E_n.

The equivalence of E_n and E'_n in triangle free graphs is straightforward, with the former condition more easily applied, and the latter more readily checked. The mutual adjacency condition Adj_n is manifestly weaker, but only because it allows some relatively degenerate examples: the complete bipartite graph, and some less obvious ones—these can all be satisfactorily classified. As a result, we can replace the property E'_n by a mild strengthening of Adj_n, giving us the simplest version to check.

Our first order of business will be to sort out the force of these extension properties. Along the way we will find it useful to classify explicitly all the triangle free graphs which satisfy the condition Adj_n and E_2 but not E_n. For this, a description of triangle free graphs satisfying Adj_3 in terms of combinatorial geometries (or hypergraphs) is useful. These geometries can be given in two ways, either as a bipartite graph with one set called "points" and the other set called "blocks," calling the edge relation between points and blocks "incidence," or more concretely by taking the geometry to consist of a set of points together with a collection of sets of points called blocks (or hyperedges). The two points of view are not identical: in the bipartite setting, distinct blocks may be incident with the same set of points; in the more concrete setting where the blocks are taken to be sets of points, each such set can occur only once.

Given a graph G and a vertex $v \in \Gamma$ we form the geometry G_v whose points are the neighbors of v in G and whose blocks are the nonneighbors distinct from v, with the edge relation between points and blocks given by the edge relation in G. Edges between pairs of points and pairs of blocks are ignored. This is of interest to us only if the graph can be reconstructed from the geometry. In the triangle free case there are no edges between points, and there are no edges between blocks which intersect. The graphs that interest us will not only be triangle free but maximal triangle free (i.e., adding an edge produces a triangle). In such a case the edge relation between blocks is determined by the geometry: two blocks are joined by an edge if and only if they are disjoint. This point of view corresponds to the usual construction of the Higman-Sims graph, and while none of our geometries will have a geometry with the elegance of that graph's, the same point of view will still be useful. In general, the geometry obtained will depend both on the graph and on the choice of a base vertex v in the graph.

We refer to triangle free graphs satisfying condition E_n as *n-e.c.*, which stands for "n-existentially complete" (for the category of triangle free graphs). This terminology comes from model theory.

We will see that there are a number of infinite families of geometries all corresponding to 3-e.c. graphs, but that the families known to date are neither varied nor robust. In the M_{22} geometry which is associated with the Higman-Sims graph, every block has 6 points; but in the infinite families of 3-e.c. graphs known to us, there is always at least one block of order 2 (and possibly just one). At the other extreme, in a geometry associated with a 4-e.c. graph, the minimal block size is at least 19. Barring some breakthrough taking us to finite 4-e.c. triangle free graphs and beyond, it would be interesting to make the acquaintance of more robust 3-e.c. graphs. We will give explicit descriptions of some infinite families, in the hope that this may stimulate someone to find better constructions.

Of the six known non-trivial triangle free *strongly regular* graphs, two provide interesting examples of 3-e.c. graphs. The question as to whether there are more such to be found, and possibly an infinite family, seems to be tied up with fundamental problems in that subject. However following a suggestion of Peter Cameron, we can eliminate the possibility that a strongly regular triangle free graph could be 4-e.c.

The following general principle is immediate.

REMARK 11.1. *If M is a structure with the finite model property, and M' is a structure which can be interpreted in M, then M' inherits the finite model property.*

In particular if the generic K_n-free graph has the finite model property for one value of n, then so does the generic K_m-free graph for $m \leq n$, interpreting the latter as the graph induced on the set of vertices adjacent to $(n-m)$ vertices of the former. Thus the finite model property for the generic triangle free graph is the weakest instance of the problem still open in the case of homogeneous graphs.

One would of course like to have general methods for settling the finite model property in homogeneous structures. The finite model property for the Rado graph and for similarly unconstrained homogeneous structures holds by a simple probabilistic argument. Just as a random (countable) graph will be isomorphic to the Rado graph with probability one, a large finite random graph will have one of the appropriate extension properties with asymptotic probability 1 [**Fag76**]. But probabilistic constructions behave very poorly in the presence of constraints.

For example, if we use counting measure on the set of triangle free graphs of a given size, a random one will be bipartite with high probability [**EKR76**] (cf. [**KPR87**] for the K_n-free case), so that the theory of the random finite triangle free graph does not approximate the theory of the generic triangle free graph at all well.

Some time ago Vershik raised the question of a Borel measure invariant under the full infinite symmetric group and concentrating on the generic triangle free graph, a question answered positively in [**PV08**] (with a classification of the measures in question). But this does not seem to help with the finite model property.

11.2. Equivalence of E_n and $E_{n'}$. We examine the relationships among the natural extension properties. The property E_2' merits separate consideration.

LEMMA 11.2. *For triangle free graphs G the following properties are jointly equivalent to E_2':*

(*i*) *G is maximal triangle free;*
(*ii*) *G is indecomposable;*
(*iii*) *G contains an independent set of order 3.*

Maximality means that the adjunction of any additional edge would create a triangle, which is the same as the mutual adjacency condition: any two independent vertices have a common neighbor.

A graph is *decomposable* if it carries a nontrivial congruence, that is, an equivalence relation such that for any two classes C_1, C_2 either all pairs in $C_1 \times C_2$ are edges, or none are.

Proof of 11.2. The implication from E_2' to $(i-iii)$ is immediate, so we deal with the converse direction.

Given (ii, iii) it is easy to see that every maximal independent set has more than one vertex, which is clause (a) of E_2'. So we consider clause (b): for any independent pair of vertices $A = \{u, v\}$ and any subset B of A, we have a vertex adjacent to all vertices of B and no vertices of $A \setminus B$. There are four cases here, all of them relevant.

Suppose conditions (i, ii) hold, and that $|G| \geq 3$. It is easy to see that if condition $E_2'(b)$ fails, then G is a 5-cycle. We give the details.

Fix an independent pair $A = \{u, v\}$ in G. By the indecomposability of G, there is a vertex u' in G adjacent to exactly one of u and v; suppose u' is adjacent to u, but not to v. The pair $\{u', v\}$ also has a common neighbor v', which is therefore adjacent to v but not u. Thus for $B = \{u\}$ or $B = \{v\}$, clause (b) of the property E_2' holds.

Thus under the hypotheses (i, ii), with $|G| \geq 3$, any violation of E_2' consists of a pair $A = \{u, v\}$ of independent vertices which is a maximal independent set. Then $G \setminus A$ divides into three subsets over A: the set G_u of vertices adjacent to u but not v, the set G_v of vertices adjacent to v but not u, and the residual set $G' = G \setminus (A \cup G_u \cup G_v)$. At this point the edge relation on G is completely determined by conditions (i, ii): apart from the edges involving u or v, the remaining edges must connect G_u to G_v; conversely, by maximality, the induced graph on $G_u \cup G_v$ is bipartite. Thus by indecomposability $|G_u| = |G_v| = 1$ and $|G'| \leq 1$. The case $G' = \emptyset$ gives a 4-cycle, which is decomposable, so that is excluded, and we are left with the case $|G'| = 1$, which gives a 5-cycle.

If we assume (iii) as well, we have a contradiction. □

Next we check that the properties E_n' and E_n are equivalent.

LEMMA 11.3. *For G triangle free, and n arbitrary, properties E_n and E_n' are equivalent.*

PROOF. We must show that $E_n' \implies E_n$.

Let A be a vertex set of order at most n, and B an independent subset of A. We must show that there is a vertex adjacent to all vertices of B and no vertices of $A \setminus B$. We proceed by induction on n, and on $|A \setminus B|$.

If there is an edge (a, b) with $a \in A \setminus B$ and $b \in B$, then let $A_0 = A \setminus \{a\}$, and apply induction to n to get v adjacent to all vertices of B, and no vertices of $A_0 \setminus B$, and hence to no vertex of $A \setminus B$.

So we may suppose there is no edge connecting $A \setminus B$ and B, but that there is an edge in A. In particular $|A \setminus B| \geq 2$. We claim then: *There is a vertex $u \notin A$ which is adjacent to some vertex $a \in A$ and to no vertex of B.*

Let $A_0 \subseteq A$ be a maximal independent subset of A containing B, and u a vertex not adjacent to any vertex of A_0. Then u is not in A, and if u is adjacent to some vertex of A we have our claim. So suppose u is adjacent to no vertex of A. Take $a \in A \setminus B$. Then $B \cup \{a, u\}$ is an independent set. Take a vertex v adjacent to u and a, and to no vertex in B. Then $v \notin A$, and v meets the conditions of our claim.

Now applying the claim with u adjacent to $a \in A$ and to no vertex of B, let $B_1 = B \cup \{u\}$. Let $A_1 = (A \setminus \{a\}) \cup \{u\}$. By induction on $|A \setminus B|$ we find v adjacent to all vertices of B_1 and no vertex of $A_1 \setminus B_1$. Then the set of neighbors of v in A is B, as required. □

11.3. The strength of Adj_3**.** We will show that with few exceptions, graphs having the properties E_2 and Adj_n satisfy the full n-e.c. property E_n. The delicate case arises when $n = 3$, and we first dispose of the others.

LEMMA 11.4. *If G is a 3-e.c. triangle free graph with the mutual adjacency property* Adj_n*, then G is n-e.c.*

PROOF. It suffices to verify the condition E'_n. We proceed by induction on n.

Fix A an independent set of order at most n. We may assume $|A| = n$ or conclude by induction. Our objective is to show that every subset of A occurs as the set of neighbors in A of some vertex of G.

For $v \in G$, let us write A_v for the set of neighbors in A of the vertex v. Then for v' chosen adjacent to v and to all vertices of $A \setminus A_v$, we find $A_{v'} = A \setminus A_v$. So the collection $\{A_v : v \in G \setminus A\}$ is closed under complementation in A.

Fix $X \subseteq A$. We will show that X is A_v for some vertex v. Taking complements if necessary, suppose $|X| \leq n/2$. Take $a \in A \setminus X$ and u a vertex whose neighbors in $A \setminus \{a\}$ are the vertices in X. We may suppose that $A_u = X \cup \{a\}$.

Let Y be the complement $A \setminus (X \cup \{a\})$ and let v be a vertex with $A_v = Y$. If $|Y| > 1$ then taking u' (inductively) whose set of neighbors in $(A \setminus Y) \cup \{v\}$ is $X \cup \{v\}$, we finish.

We conclude that $|X| = n - 2 \leq n/2$, so $n \leq 4$. As G is 3-e.c., $n \geq 4$. Thus $n = 4$ and $|X| = 2$. In particular any singleton occurs as A_v for some v.

Write $A = \{a_1, a_2, b_1, b_2\}$ and let $X = \{a_1, a_2\}$. Let b'_1 be a vertex with $A_{b'_1} = \{b_1\}$. Let b_2 be a vertex whose unique neighbor in $\{a_1, a_2, b'_1, b_2\}$ is b_2. Let u be a vertex adjacent to a_1, a_2, b'_1, b'_2. Then $A_u = X$. □

Now we take up the graphs satisfying E_2 and Adj_3, but not E_3, and we begin with a construction.

11.4. The Linear Order Geometries. Given any triangle free graph G with the mutual adjacency property Adj_3 we associate a combinatorial geometry to each vertex v of G by taking as the set of points P the neighbors of v, and as the blocks of the geometry its non-neighbors. A point p lies on a block b if the pair (p, b) is an edge. The geometry obtained may depend on the vertex chosen, but given one such geometry, the graph G may be entirely reconstructed as follows. As the vertex set for G we take $P \cup B \cup \{v_0\}$ where v_0 is an additional vertex. We take as edges all pairs (v_0, p) with p in P, all pairs (p, b) with p on b, and all pairs (b_1, b_2) with b_1, b_2 disjoint when viewed as subsets of P; and we symmetrize. This agrees with the original graph G: in particular, if b_1, b_2 are not adjacent in G, then by the property Adj_3 there is a point p lying on both, and hence they are not adjacent in the reconstructed version of G.

The following uses extremely weak assumptions, but then we intend to apply it to an extremely weak geometry.

LEMMA 11.5. *Let (P, B) be a combinatorial geometry on at least 3 points satisfying the following conditions.*

(1) *No three blocks are pairwise disjoint.*
(2) *No pair of distinct blocks correspond to the same subset of P.*
(3) *For every block b and every point p not in b, there is a block containing p and disjoint from b.*
(4) *For every pair of points there is a block containing one but not the other.*
(5) *No block is incident with every point of P.*

Then the associated graph G is triangle free and 2-e.c.

PROOF. One checks that G is maximal triangle free, indecomposable, with an independent set of size at least three.

One point that requires checking is that no block is empty; this is part of the verification that G is maximal triangle free. For this, use the assumptions to get two nonempty blocks which are disjoint, and observe that the empty block would extend this to a pairwise disjoint triple. □

Let L be a linear order. Let B be a set of proper initial segments of L and proper terminal segments of L satisfying the conditions:

(1) For all $a < b$ in L, there is an initial segment in B containing a and not b, and a terminal segment in B containing b and not a.
(2) If I is an initial segment and $a \in B \setminus I$ is a lub for I, then the terminal segment $[a, \infty)$ is in B; and dually.

One way to meet these conditions is to let B consist of the proper segments of the form $(-\infty, a]$ and $[a, \infty)$, and this is the only way to achieve even the first of them if L is finite. Let us call such a geometry a *linear geometry.*

LEMMA 11.6. *Let (P, B) be a linear geometry on at least 3 points. Then the associated graph satisfies the conditions E_2 and Adj_n for all n (if P is finite, this is vacuous for n larger than $|P|$).*

PROOF. The conditions on the geometry have been written to ensure that our criterion for the property E_2 applies.

Now suppose A is any independent subset of the associated graph. If A contains no points, then A consists of some blocks, and possibly the base point, and as the blocks all meet pairwise, it suffices to take a point common to the minimal initial segment in A, and the minimal terminal segment in A. That point is then a vertex adjacent to all vertices of A.

If A contains a point p, then it cannot contain both initial segments and terminal segments, as they would be separated by p. So suppose for example A contains initial segments, and let I be the greatest among them. Let $a \in A$ be the least point. Then $a \notin I$. One of our two conditions on B then applies to give a terminal segment disjoint from I, and containing a.

Of course if A consists exclusively of points, the base point will suffice as a common neighbor. □

If L is finite of size n, then the resulting graph G has order $3n-1$ and can be construed as follows. The elements of G are the integers $0, 1, \ldots, 3n-2$; the edge relation is defined by $|i-j| \equiv 1 \mod 3$. The value $n=2$, which is not permitted as we require 3 points, corresponds to the pentagon, and the first legitimate example is a graph of order 8. The maximal independent sets have size n and are the points of the geometry, with respect to a basepoint which is their common neighbor. The geometry obtained is independent of the base point.

Our next point is that the converse holds: a graph with property E_2 and property Adj_n which does not have property E_n must be obtained in this way from a linear geometry, and in particular if it is finite then it isomorphic to the graph just described explicitly. It will suffice to treat the case $n=3$, since once E_3 is satisfied, Adj_n implies E_n.

11.5. Graphs with E_2, Adj_3, and not E_3.

DEFINITION 11.7. An independent set of vertices I in a graph G will be said to be *shattered* if every subset of I occurs as $\{a \in I : a, v \text{ are adjacent}\}$ for some vertex v.

We will be concerned for the present with graphs which are 2-e.c. but not 3-e.c., and therefore contain independent triples which are not shattered. We want to show that within the independent set of all neighbors of any fixed vertex of G, if one triple is shattered, then all are. We arrive at this gradually by considering various special cases.

LEMMA 11.8. *Let G be a triangle free graph with properties E_2 and Adj_3, and suppose that a, b, c is a shattered independent triple. If a', b, c is another independent triple with a, a' adjacent, then a', b, c is also shattered.*

PROOF. Since we have property Adj_3, and the collection of subsets of the set $A = \{a', b, c\}$ which occur as the set of neighbors in A of vertices of G is closed under complementation, it suffices to consider a single vertex $u \in \{a', b, c\}$ and to show that u occurs as the unique neighbor of some vertex among a', b, c.

For $u = a'$ we take a as the witnessing vertex. For $u = b$ we take u' adjacent to a, b and not c. For $u = c$ proceed similarly. □

LEMMA 11.9. *Let G be a triangle free graph with properties E_2 and Adj_3, and suppose that the triple (a, b, c) is independent and not shattered. Then there is a unique vertex u in $\{a, b, c\}$ which does not occur as the unique neighbor among a, b, c of a vertex in G.*

PROOF. There must be at least one such vertex, say b. Taking u adjacent to b and not to a, the neighbors of u among a, b, c will be b and c; and then a common neighbor of u and a will have a as its unique neighbor among a, b, c; similarly c will occur as the unique neighbor of some vertex among a, b, c. □

LEMMA 11.10. *Let G be a triangle free graph with properties E_2 and Adj_3, and suppose that $I = \{a, b, c, d\}$ is an independent quadruple with a, b, c shattered, while some vertex u has a and d as its only neighbors in I. Then (b, c, d) is shattered.*

PROOF. By Lemma 11.8 the triple u, b, c is shattered, and by another application of the lemma, b, c, d is shattered. □

LEMMA 11.11. *Let G be a triangle free graph with properties E_2 and* Adj_3*, and suppose that $I = \{a, b, c, d\}$ is an independent quadruple with a, b, c shattered, while no vertex u has precisely two neighbors in I. Then the triple b, c, d is shattered.*

PROOF. Fix vertices a', b', c' having as their unique neighbors in a, b, c the vertices a, b, or c respectively. By our hypothesis, none of the vertices a', b', c' is adjacent to d.

Now at least two of the vertices a', b', c' are nonadjacent; let u_1, u_2 be two such. Take v adjacent to u_1, u_2, and d. Then v is adjacent to at most two vertices of a, b, c, d and hence, by our hypothesis, to at most one; that is, v is adjacent to d and not to a, b, c.

Suppose now that $u_1 = a'$. Then $I' = \{a', b, c, d\}$ is an independent quadruple with a', b, c shattered, and v has only a' and d as its neighbors in I'. By the previous lemma, b, c, d is shattered.

So we may suppose that $u_1 = b'$ and $u_2 = c'$. Then b', c', and v have as their unique neighbors in b, c, d the vertices b, c, and d respectively, and so b, c, d is shattered. ☐

LEMMA 11.12. *Let G be a triangle free graph with properties E_2 and* Adj_3*, and suppose that $I = \{a, b, c, d\}$ is an independent quadruple with a, b, c shattered. Suppose there are vertices b', c' in G having as their neighbors in I the pairs a, b and a, c respectively. Then b, c, d is shattered.*

PROOF. The triple b', c', d is independent. Let d' be adjacent to b', c', d. Then the vertices b', c', d' show that the triple b, c, d is shattered. ☐

LEMMA 11.13. *Let G be a triangle free graph with properties E_2 and* Adj_3*, and suppose that $I = \{a, b, c, d\}$ is an independent quadruple with a, b, c shattered. Suppose there are vertices u, u' in G having as their neighbors in I the pairs a, b and c, d, and no other pairs from a, b, c, d occur in this fashion. Then b, c, d is shattered.*

PROOF. Applying Lemma 11.10 to the quadruple (c, a, b, d), it follows that the triple a, b, d is shattered.

Now if a, c, d is shattered, we argue similarly that b, c, d is shattered by looking at the quadruple (a, c, d, b).

Take a vertex a' adjacent to a and not adjacent to b, c. By our hypothesis, a' is not adjacent to d either. Now take a vertex v adjacent to a', b, d. Then v is not adjacent to a, so by our hypothesis v is adjacent to c.

We now consider two cases. First, if there is a vertex d' adjacent to d but not to a, b, c, then we apply Lemma 11.8 to the series of independent triples a, b, d, a, b, d', a, v, d', a, c, d', a, c, d to conclude.

Now suppose that there is no such vertex d'. Then there is no vertex whose unique neighbor in b, c, d is d. By Lemma 11.9, there is a vertex c' whose unique neighbor among b, c, d is c. Hence, by our hypothesis, c is the only neighbor of c' among a, b, c, d. If (a', c') is an edge, we conclude by applying Lemma 11.8 to the sequence of independent triples

$$(a, b, d); (a', b, d); (c', b, d); (c, b, d)$$

So we may suppose (a', c') is not an edge. Take a vertex w adjacent to a', b, c'. Then w is adjacent to b but not a or c, and hence not d either. Now apply Lemma 11.8 to the independent triples (a, b, d), (a, w, d), (a, c', d), (a, c, d) to conclude. ☐

LEMMA 11.14. *Let G be a triangle free graph with properties E_2 and Adj_3, and let I be an independent set containing some shattered triple. Then all triples of vertices from I are shattered.*

PROOF. By the foregoing lemmas, if (a, b, c, d) is any independent quadruple containing a shattered triple, then all of its triples are shattered. The general case follows. □

PROPOSITION 11.15. *Let G be a triangle free graph with properties E_2 and Adj_3, but not E_3. Let v be a vertex of G. Then the geometry (P, B) associated to the vertex v in G is a linear geometry, and G is the associated graph.*

PROOF. G is certainly the associated graph, so everything comes down to recognizing the geometry on (P, B), with P the set of neighbors of v and B the set of non-neighbors.

We first choose v to be a vertex of G having a triple of neighbors which is not shattered, and let (L, B) be the associated geometry. Once we verify that this is a linear geometry, the structure of G is determined, and it follows that the same applies to the geometry at any vertex of G.

We next look for a *linear betweenness relation* on L. This is a ternary relation $\beta(x, y, z)$, irreflexive in the sense that it requires x, y, z to be distinct, which picks out for each triple x, y, z a unique element which is between the other two, i.e., $\beta(x, y, z)$ implies $\beta(z, y, x)$ and not $\beta(y, z, x)$ or $\beta(z, x, y)$. In addition to these basic properties we have the axiom:

$$\text{For } x, y, z, t \text{ distinct, } \beta(x, y, z) \text{ implies } \beta(x, y, t) \text{ or } \beta(t, y, z)$$

Any linear order gives rise to a linear betweenness relation, and the reverse order gives the same betweenness relation; conversely, a betweenness relation determines a unique pair of linear orders which give rise to it (assuming there are at least two points).

We define $\beta(x, y, z)$ on L as follows: $\beta(x, y, z)$ holds if any vertex adjacent to y is adjacent to x or z; equivalently (taking complements) any vertex adjacent to x and z is adjacent to y. This is symmetric in x and z, and by Lemma 11.9, the relation picks out of each independent triple (x, y, z) which is not shattered, a unique y satisfying $\beta(x, y, z)$ and $\beta(z, y, x)$. Furthermore by our choice of v and Lemma 11.14, none of the triples in L are shattered. So we only have to check the critical axiom: assuming $\beta(x, y, z)$, with t a fourth vertex in L, we claim that $\beta(x, y, t)$ or $\beta(t, y, z)$ holds.

Suppose $\beta(x, y, z)$ holds and $\beta(x, y, t)$ fails. We will show that $\beta(t, y, z)$ holds. As $\beta(x, y, t)$ fails, we have $\beta(x, t, y)$ or $\beta(t, x, y)$.

Suppose $\beta(x, t, y)$ holds. Take a vertex u adjacent to y and not t. Then by $\beta(x, t, y)$, u is not adjacent to x. By $\beta(x, y, z)$, u is adjacent to z. This proves $\beta(t, y, z)$, as claimed.

Now suppose $\beta(t, x, y)$ holds, and take a vertex u adjacent to y, but not adjacent to z. Then by $\beta(x, y, z)$ we have u adjacent to x, and by $\beta(x, t, y)$ we have u adjacent to t. Thus $\beta(t, y, z)$ holds.

Accordingly, β is a linear betweenness relation on L and we may fix a linear ordering giving rise to this relation. By the definition of β, the blocks of B are convex, and are not bounded both above and below. Therefore they are initial and terminal segments of L (reversing the order will of course interchange these two notions). It remains to check that the blocks are proper, pairwise distinct, and sufficiently dense in L to satisfy our axioms for a linear geometry. This all follows from the assumption that G satisfies E_2, now that the general shape of the geometry has been established. □

With this result in hand we can give a reasonably efficient axiomatization of the geometries associated with 3-e.c. graphs.

11.6. Geometries associated with 3-e.c. graphs.

DEFINITION 11.16. An E_3-geometry is a combinatorial geometry (P, B) satisfying the following axioms.

- I There are no three disjoint blocks.
- II No block is contained in any other.
- III There are at least two points. For any two distinct points, there is a block containing exactly one of them, and a block containing neither.
- IVa If b is a block and p, q are points not in b, then there is a block disjoint from b containing p and q.
- IVb If b, b' are blocks which intersect, and p is a point outside their union, then some block containing p is disjoint from b and b'.
- V If three blocks intersect pairwise, then they either have a point in common, or some block is disjoint from their union.

The E_3-geometries are just the geometries associated with 3-e.c. triangle free graphs, as we shall show. We avoid the seemingly natural term "3-e.c. geometry", as the natural interpretation for that term would be a considerably stronger set of conditions in which, notably, in axiom V we would require both a point in common and a disjoint block, that is, we would apply the 3-e.c. condition directly to the geometry, specifying the type of the element of $P \cup B$ realizing the giving condition. The thrust of this is considerably more like 4-e.c. in the corresponding graph; in fact, it is 4-e.c. restricted to quadruples including the base point. No doubt this is an interesting class of geometries in its own right, and more tractable than those associated with 4-e.c. graphs, but still a good deal beyond anything we can construct, or analyze, at present.

LEMMA 11.17. *The geometry associated to any vertex of a 3-e.c. triangle free graph is an E_3-geometry, and conversely the graph constructed in the usual way from an E_3-geometry is a 3-e.c. triangle free graph.*

PROOF. One can read off all these axioms directly from the 3-e.c. property (with the triangle free condition accounting for the first of them). The point is to check that conditions (I–V) are strong enough. For that, we use the analysis of the previous subsection. By Axiom II, we exclude the linear geometries of section §11.4, and therefore it suffices to check that the associated graph is triangle free, 2-e.c., and satisfies the adjacency condition Adj_3, which is more or less what the axioms assert (with IV(a), IV(b), and V corresponding to different instances of Adj_3). □

We derive some further consequences of the axioms. Observe that Axioms I and III imply that all blocks are nonempty, and that a further application of Axiom III implies that there are at least 3 points.

LEMMA 11.18. *Let (P, B) be an E_3-geometry. Then the union of two blocks is never P.*

Proof. Suppose first that b_1, b_2 are two blocks which meet, and let $p \in b_2 \setminus b_1$. There is a block b' containing p and disjoint from b_1, and as b' cannot be contained in b_2, it follows that $b_1 \cup b_2 \neq P$.

Now suppose that b_1, b_2 are disjoint and their union is P. We may suppose that $|b_1| \geq 2$. Take two points of b_1 and a block b containing just one of them. As b is not contained in b_1 and $b_1 \cup b_2$ is P, b meets b_2. By construction $b \cup b_2$ is not P, so there is a block b'' disjoint from $b \cup b_2$. Then b'' is a proper subset of b_1 and we have a contradiction. □

Lemma 11.19. *Let (P, B) be an E_3-geometry. Then any two points belong to some block.*

Proof. Call two points of P *collinear* if they lie in a common block. We claim first that this is an equivalence relation on P.

Suppose that $p, q \in b_1$, and $q, r \in b_2$. Take a point $a \notin b_1 \cup b_2$, and a block b disjoint from $b_1 \cup b_2$ containing a. As p, r lie outside b, Axiom IVa applies, and p, r are collinear.

By Axiom I, there are at most two equivalence classes for the collinearity relation, and we claim there is only one.

Suppose there are two collinearity classes P_1, P_2. As $|P| \geq 3$, we may suppose $|P_1| \geq 2$. Take a block b meeting (and hence contained in) P_1. As there are no inclusions between blocks, it follows from Axiom III that b is a proper subset of P_1. Therefore by Axiom IVa we have a block meeting P_1 and P_2, a contradiction. □

Corollary 11.19.1. *Let (P, B) be an E_3-geometry. Then every block contains at least two points.*

At this point we are through sorting through the basic axioms and we can begin to look more closely at examples of 3-e.c. triangle free graphs and their associated geometries.

We first take up the strongly regular case, then look into the minimal size of an E_4-geometry.

12. Strongly regular graphs and E_4-geometries

We will discuss the extension properties of the known strongly regular triangle free graphs and then show that such a graph cannot have property E_4, following up on an old suggestion of Peter Cameron. We also look at the minimum degree of a graph with property E_4, or in other words, the minimal size of an E_4-geometry.

12.1. The Higman-Sims Graph. This is constructed from the M_{22} geometry, defined as follows. Let P_0 be the projective plane over the field of order 4, with 21 points. Adjoin an additional point ∞ to get $P = P_0 \cup \{\infty\}$. The associated blocks will be of two kinds. The first kind are obtained by extending an arbitrary line ℓ of P_0 by the point ∞: $\ell^* = \ell \cup \{\infty\}$. The second kind are called *hyperovals.* A hyperoval is a set of 6 points in P_0 which meets any line of P_0 in an even number of points. There are 168 hyperovals, and on this set the relation "$|O_1 \cap O_2|$ is even" is an equivalence relation, with three classes of 56 hyperovals each, permuted among themselves by the automorphism group of the base field. Any one class of 56 hyperovals may be taken, together with the extended lines, as the set of blocks B for the M_{22} geometry, on 22 points. Thus there are 77 blocks, each with 6 points, and 100 vertices in the associated graph, the Higman-Sims graph. Its automorphism group is vertex transitive and edge transitive, so we get the same geometry from any base point, and any adjacent point can play the role of the "new" point ∞.

We check that this graph is a 3-e.c. triangle free graph. Given three independent vertices, one may be taken to be the base point v, and the other two will then represent two intersecting blocks of the geometry. We may suppose that they both contain the point ∞ in the associated geometry, and therefore they represent two extended lines, whose intersection has order 2. So we have the condition Adj_3 with multiplicity two, that is there are two points meeting the adjacency conditions in every case. Since the graph visibly satisfies the E_2 condition, and there are no containments between blocks, this completes the verification that it is 3-e.c. On the other hand, it is not 4-e.c. As we know, this comes down to the 4-adjacency property Adj_4. Taking a triple of points lying on a projective line ℓ in P_0, and another line meeting ℓ in a different point, a common neighbor of the four vertices involved would be a block containing the given three points and disjoint from the given line; but there is only one block containing three given points, so this is impossible.

The Higman-Sims graph is an example of a *strongly regular* triangle free graph. In general, a graph on n vertices is strongly regular with parameters (n, k, λ, μ) if it is regular of degree k, and any pair of vertices v, v' has λ common neighbors if v, v' are adjacent, and μ common vertices otherwise. In the case of triangle free graphs ($\lambda = 0$), leaving aside the complete bipartite graphs and the pentagon, there are six known examples, which go by the names of the Petersen, Clebsch, Hoffman-Singleton, Gewirtz, M_{22}, and Higman-Sims graphs [**Br11, vLW92, CvL91, BvL84**]. Two of these graphs are 3-e.c., the Clebsch graph and the Higman-Sims graph.

While there are no other known strongly regular triangle free graphs, there are many "feasible" sets of parameters, that is combinations of parameters which are compatible with all known constraints on such graphs. Following a suggestion of Peter Cameron, we will use that theory to show that there are no 4-e.c. strongly regular triangle free graphs, leaving entirely open the problem whether there are any more, or infinitely many more, 3-e.c. strongly regular triangle free graphs.

We will also take a closer look at the known strongly regular triangle free graphs, notably the Clebsch graph, which serves as the basis for the simplest construction of an infinite family of 3-e.c. graphs, first proposed by Michael Albert.

12.2. Strongly regular graphs and properties E_2, E_3. Leaving aside the complete bipartite graphs and the pentagon, the known strongly regular triangle free graphs have the following parameters.

(1) Petersen: $(10, 3, 0, 1)$;
(2) Clebsch: $(16, 5, 0, 2)$;
(3) Hoffman-Singleton: $(50, 7, 0, 1)$;
(4) Gewirtz: $(56, 10, 0, 2)$;
(5) M_{22}: $(77, 16, 0, 4)$;
(6) Higman-Sims: $(100, 22, 0, 6)$

The entry "0" here simply says that the graph is triangle free. The Gewirtz graphs and the M_{22} graph can be seen naturally inside the Higman-Sims graph (the Hoffman-Singleton graph, less naturally): the Gewirtz graph is the graph on the hyperovals of the M_{22} geometry, which could be viewed as the set of vertices in Higman-Sims nonadjacent to two vertices lying on an edge. The M_{22} graph is the graph on the blocks of the M_{22} geometry, and appears as the constituent of Higman-Sims on the non-neighbors of a fixed vertex.

In the Higman-Sims graph, any independent triple of vertices has exactly two common neighbors, as noted previously. In particular if the vertices represent hyperovals with two common points, then their common neighbors are both represented by points, and hence do not lie in the M_{22} graph. Thus neither the Gewirtz graph nor the M_{22} graph can satisfy the condition Adj_3.

The Hoffman-Singleton and Petersen graphs, with the fourth parameter $\mu = 1$, do not fit into this framework at all: there is no useful geometry induced on the set of neighbors of a fixed vertex, as its blocks would consist of single points. The Petersen graph does play a respectable role as the set of blocks in the geometry associated to the Clebsch graph, and the latter is indeed a 3-e.c. graph. In the case of the Clebsch graph, the geometry is extremely degenerate: it consists of all pairs from a set of order 5. Nonetheless this geometry is an E_3-geometry.

Thus the Clebsch graph and Higman-Sims graph both are 3-e.c., and the Petersen, Gewirtz, and M_{22} graphs are naturally represented as descriptions of part or all of the associated geometries.

Any strongly regular graph other than the complete bipartite graphs and the pentagon graph will satisfy the condition E_2, and have no proper inclusion between blocks, so in all other cases the condition E_n will be equivalent to the adjacency condition Adj_n.

However the condition Adj_4 is already incompatible with strong regularity for triangle free graphs, as we now show, following the notation of [**Big09**], which relies on the eigenvalue theory for strongly regular graphs, expressing everything in terms of the minimal eigenvalue for the adjacency matrix of the graph. So we begin by reviewing that material.

12.3. Eigenvalues and E_4 in the strongly regular case. Let G be strongly regular with parameters (n, k, λ, μ). Let A be the $n \times n$ adjacency matrix for G, with 0 entries for non-adjacent pairs of vertices, and 1 for adjacent pairs. With J the $n \times n$ matrix consisting entirely of 1's, the condition of strong regularity, with the specified parameters, translates into the matrix condition

$$A^2 + (\mu - \lambda)A - (k - \mu)I = \mu J$$

and as J has the eigenvalues n with multiplicity 1 and 0 with multiplicity $n-1$, A has three eigenvalues: k with multiplicity 1, and two eigenvalues α, β which are roots of the quadratic equation

$$x^2 + (\mu - \lambda)x - (k - \mu) = 0$$

with multiplicities m_α, m_β satisfying

$$m_\alpha + m_\beta = n - 1; m_\alpha \cdot \alpha + m_\beta \cdot \beta = 0$$

since the trace of A is zero. Specializing to the case $\lambda = 0$ this gives the following formulas in terms of the parameter $s = \sqrt{\Delta}$, Δ being the discriminant $\mu^2 + 4(k-\mu)$, an integer in the nontrivial cases (leaving aside the pentagon and complete bipartite graph):

$$\alpha = \frac{s - \mu}{2}, \ \beta = \frac{-s - \mu}{2}$$

Following [**Big09**], we write q for the eigenvalue of minimal absolute value (α, above) and express everything in terms of q and μ as follows.

$$\begin{aligned} k &= (q+1)\mu + q^2 \\ n &= (q^2+3q+2)\mu + (2q^3+3q^2-q) + (q^4-q^2)/\mu \\ &= (q+2)k + (q^3+q^2-q) + (q^4-q^2)/\mu \\ \mu &\le q(q+1) \end{aligned}$$

The inequality on μ is not obvious but is derived rapidly from elementary considerations of linear algebra in [**Big09**].

We note that the extremal values $\mu = q(q+1)$, $k = q^3+3q^2+q$, $n = q^2(q+3)^2$ satisfy all known feasibility constraints and give $k \approx (q+1)^3$, $n \approx (q+1\frac{1}{2})^4$, so that k^4 and n^3 are fairly close, and as we will see in a moment this makes the refutation of condition E_4 a little delicate. Namely, given E_4, or what amounts to the same thing, Adj_4, our condition is that the collection of all independent 4-tuples of vertices (u_1, u_2, u_3, u_4) should be covered by the subset consisting of the independent 4-tuples lying in neighborhoods of the vertices, with the former being slightly less than n^4 (at least $n(n-(k+1))(n-2(k+1))(n-3(k+1))$) and the latter approximately $n \cdot k^4$, leading to an estimate of roughly the form $k^4 > n^3$. At least for large values of n it is clear that this will not be satisfied at the extreme values and is less likely to hold lower down. But we will work through this more precisely to get the following.

PROPOSITION 12.1. *There is no strongly regular triangle free graph with the property E_4.*

PROOF. Begin with the estimates

$$\begin{aligned} n &= (q+2)k + (q^3+q^2-q) + (q^4-q^2)/\mu \\ &\ge (q+2)k + (q^3+q^2-q) + (q^2-q) \\ &= (q+2)k + (q^3+2q^2-2q) \\ &\ge (q+3)k - (\mu+2q) \\ &\ge (q+3)k - (q^2+3q) \end{aligned}$$

Now the number of independent quadruples of vertices in our graph G is at least $n[n-(k+1)][n-2(k+1)+\mu][n-3(k+1)+\mu]$ and assuming E_4, they all occur in neighborhoods of individual vertices, so the number is at most $nk(k-1)(k-2)(k-3)$. So we have

$$[n-(k+1)][n-2(k+1)+\mu][n-3(k+1)+\mu] \le k(k-1)(k-2)(k-3)$$

and we show this is impossible.

We have

$$\begin{aligned} [n-(k+1)] &\ge (q+2)k - (q^2+3q+1) \\ &\ge (q+2)(k-(q+1)) \end{aligned}$$

and

$$\begin{aligned} (n-2(k+1)+\mu)(n-3(k+1)+\mu) &\ge [(q+1)k-(2q+2)][qk-(2q+3)] \\ &= q(q+1)(k-2)(k-(2+3/q)) \end{aligned}$$

Furthermore $q(q+1)(q+2) \geq k+q$, so

$$[n-(k+1)][n-2(k+1)+\mu][n-3(k+1)+\mu] \geq (k+q)(k-(q+1)](k-2)[k-(2+3/q)]$$

Suppose $\mu \geq q+1$. Then $k \geq (q+1)^2 + q^2 \geq 2q(q+1)$, so

$$(k+q)(k-(q+1)) = k^2 - k - q(q+1) \geq k(k-1\frac{1}{2})$$

If we take $q \geq 6$ as well then this yields

$$\begin{aligned}[n-(k+1)][n-2(k+1)+\mu][n-3(k+1)+\mu] &\geq k(k-1\frac{1}{2})(k-2)(k-2\frac{1}{2}) \\ &> k(k-1)(k-2)(k-3)\end{aligned}$$

ruling out this case.

For $q < 6$ we may consult the tables in [**Big09**]. The two cases in which the necessary inequality holds are one for $q = 2$, namely the Higman-Sims graph, already ruled out, and one for $q = 3$, with parameters $n = 324$, $k = 57$, $\mu = 12$, where oddly enough the two sides are exactly equal. One way to eliminate this is to show that the block size is too small: later we will give a lower bound of 19 for the minimal block size in a geometry associated with a 4-e.c. graph.

There remains the marginal case $\mu \leq q$. In this case as $q \leq 2q^2 + q$ we have $n \geq (q+3)k + q^3 - 3q \geq (q+3)k + 3$ for $q \geq 3$ and thus we can use the crude estimate

$$(n-(k+1))(n-2(k+1))(n-3(k+1)) \geq [(q+2)k][(q+1)k][qk] \geq k^4$$

to reach a contradiction. □

Since we have quoted a lower bound for the block sizes in geometries associated with 4-e.c. triangle free graphs, we will give that next.

12.4. E_4-geometries: Block size. We will refer to a geometry associated with a 4-e.c. triangle free graph as an E_4-geometry. We will not try to write out the axioms explicitly. These would consist of conditions encoding the 2-e.c. property as was done in the case of E_3-geometries, the condition that no block is contained in another, to eliminate the degenerate case of a linear geometry, and finally the main axioms which correspond to the adjacency condition Adj_4 which takes on various forms in the geometrical context depending on how the various vertices are interpreted in the geometry. We may use any instance of the 4-e.c. condition, and are not confined to the special cases corresponding directly to our reduced set of axioms.

We omit the elementary proofs of the next three lemmas.

LEMMA 12.2. *Let (P, B) be an E_4-geometry, let b, b_1 be intersecting blocks, and let b_2 be any other block. Then*

$$|(b \cap b_1) \setminus b_2| \geq 2$$

LEMMA 12.3. *Let (P, B) be an E_4-geometry, let b, b_1 be intersecting blocks, Then $|b \cap b_1| \geq 5$.*

LEMMA 12.4. *Let (P, B) be an E_4-geometry, and let b, b_1, b_2 be blocks with a point in common. Let b_3 be a block meeting b but disjoint from b_1 and b_2. Then $|b \setminus (b_1 \cup b_2 \cup b_3)| \geq 5$.*

PROPOSITION 12.5. *Let (P, B) be an E_4-geometry. Then any block contains at least 19 points.*

PROOF. Let b be a block. We claim first that there are b_1, b_2 with $b \cap b_1 \cap b_2 \neq \emptyset$, and with $|(b \cap b_2) \setminus b_1| \geq 4$.

Begin with $b \cap b_1 \cap b_2$ nonempty and with the three blocks distinct, and suppose $|(b \cap b_2) \setminus b_1| \leq 3$. Take a block b_3 so that:

$$|b_3 \cap [(b \cap b_2) \setminus b_1]| = 1; b_3 \cap b_1 = \emptyset$$

Then $|b \cap b_2 \cap b_3| = 1$ and $(b \cap b_3) \setminus b_2| \geq 4$.

So fix blocks b_1, b_2 with $b \cap b_1 \cap b_2 \neq \emptyset$, and with $|(b \cap b_2) \setminus b_1| \geq 4$. Then $|b \cap (b_1 \cup b_2)| \geq 9$. Now take b_3 disjoint from b_1, b_2 and meeting b. Then we have $|b \cap (b_1 \cup b_2 \cup b_3)| \geq 14$. And then by the previous lemma $|b| \geq 19$. □

It would be good to have a more sophisticated lower bound here. We can convert this bound into a crude but decent lower bound for the number of points in such a geometry.

LEMMA 12.6. *Let (P, B) be an E_4-geometry. Then there are intersecting blocks b_1, b_2 with $|b_1 \cup b_2| \geq 33$.*

PROOF. Let $m = \min(|b_1 \cap b_2| : b_1 \cap b_2 \neq \emptyset)$. If $m = 5$ we are done.

Suppose $m \geq 6$. Take any two intersecting blocks b_1, b_2. Take a block b meeting b_1 and disjoint from b_2. Take a point p in $b_1 \setminus (b_2 \cup b)$, and a block b' disjoint from b_2 containing p.

Then b_1, b, b' meet pairwise and hence by the 4-e.c. condition have a point in common. Furthermore b_1 meets b_2 and b, b' are disjoint from b_2. So by Lemma 12.4, $|b_1 \setminus (b \cup b' \cup b_2)| \geq 5$. Furthermore $|(b_1 \cap b') \setminus b| \geq 2$. So $|b_1 \setminus (b \cup b_2)| \geq 7$ and $|b_1 \setminus b_2| \geq m + 7 \geq 13$.

Also, in the proof of the previous lemma, the lower bound obtained on the block size is actually $m + 4 + m + 5$ which with $m \geq 6$ would give a block size of at least 21 and a lower bound for $|b_1 \cup b_2|$ of at least 34 in this case. □

LEMMA 12.7. *Let (P, B) be an E_4-geometry, $n = |P|$. Then $n \geq 66$.*

PROOF. Take intersecting blocks b_1, b_2 with $|b_1 \cup b_2| \geq 33$. Let

$$m = \min(|b_3 \cap b_4| : b_3 \cap b_4 \neq \emptyset, (b_3 \cup b_4) \cap (b_1 \cup b_2) = \emptyset)$$

If $m = 5$ the result is immediate, so take $m = 6$ and argue as in the previous lemma. □

Another way of stating all of this is as follows.

COROLLARY 12.7.1. *Let G be a 4-e.c. graph. Then every vertex has degree at least* 66*, every pair of independent vertices has at least* 19 *common neighbors, and every triple of independent vertices has at least 5 common neighbors.*

In the Higman-Sims graph the corresponding numbers are 22, 6, and 2.

It might be of interest to explore the known feasible parameter sets for strongly regular triangle free graphs to see which seem compatible with the E_3-condition. At the extreme value $\mu = q(q + 1)$, the ratio of $k(k - 1)(k - 2)$ to

$$[n - (k + 1)][n - 2(k + 1) + \mu]$$

is q, which in the two known cases of the Clebsch and Higman-Sims graphs actually corresponds to the condition Adj_3 with multiplicity q. There are many other cases where the necessary inequality is satisfied with smaller values of μ. In fact, the majority of the cases listed in the appendix to [**Big09**] meet this condition. The only case consistent with this inequality in which the Adj_3 condition is known to fail is the case of the M_{22} graph.

In a similar vein, dropping the E_4 condition, we add some comments on the relationship between the multiplicity with which Adj_3 is satisfied, and the multiplicity with which Adj_2 is satisfied.

Let G be a 3-e.c. triangle free graph. Define $\mu_n(G)$ as the minimum over all independent sets $I \subseteq G$ of order n of the cardinality of the set of common neighbors of I. Thus for example in the Higman-Sims graph, $\mu_2(G) = 6$ and $\mu_3(G) = 2$.

LEMMA 12.8. *Let G be a 3-e.c. triangle free graph. If $\mu_3(G) \geq 2$ then $\mu_2(G) \geq 5$.*

PROOF. Fix two vertices $u_1, u_2 \in G$, and v adjacent to both. With v as base point work in the associated geometry. We look for 4 blocks containing the points u_1, u_2.

Let b_1 be a block containing u_1, u_2, and $u_3 \in P \setminus b_1$. Let b_2 be a block containing u_1, u_2, u_3 ($\mu_3(G) = 2$), and $u_4 \in P \setminus b_1 \cup b_2$. Let b_3 be a block containing u_1, u_2, u_4. It suffices to show that $b_1 \cup b_2 \cup b_3 \neq P$.

Let b be a block containing u_4 and disjoint from b_1, b_2. Then $b \setminus b_3 \subseteq P \setminus (b_1 \cup b_2 \cup b_3)$. $\square$

So we have what appears to be a sharply descending series of successive weakenings of the E_4 condition, with no known examples of even the weakest condition other than subgraphs of the Higman-Sims graph.

(1) G is 4-e.c.
(2) $\mu_3(G) \geq 5$
(3) $\mu_3(G) \geq 2$
(4) $\mu_2(G) \geq 5$
(5) $\mu_2(G) \geq 3$
(6) In the geometry associated to some base point, every block contains at least 3 points.

What we can do, as mentioned, is produce an infinite family of E_3-geometries with a *unique* block of order 2, but even in this case we get no bound on the number of blocks of order 2 for *other* geometries associated with the same graph, at different basepoints.

We return now to the case of E_3-geometries. The first examples of an infinite family of 3-e.c. triangle free graphs was given by Michael Albert. An examination of the corresponding geometries leads naturally to the consideration of a more general class of geometries with rather special properties. From one point of view these examples are degenerate; on the other hand, examples can be constructed naturally from projective geometries.

13. Some E_3-geometries

13.1. Albert geometries. We first present Michael Albert's original construction. Observe that the Clebsch graph can be represented as a collection of 4 copies of a 4-cycle, related systematically to one another: a vertex v in one copy will be connected only to one vertex in any other copy, namely the vertex corresponding to the one diagonally opposed to v. Evidently the same recipe can be extended to any number of copies of a 4-cycle, and any triple of vertices in one of these extended graphs embeds into a copy of the Clebsch graph; so the "stretched" Clebsch graph inherits the 3-e.c. property from the Clebsch graph.

In terms of the associated combinatorial geometry, the Clebsch graph corresponds to the geometry on 5 points in which every pair is a block. The stretched Clebsch graphs correspond to a geometry on n points, $n \geq 5$, in which the blocks are of two sorts:

(i) all the pairs containing either of two fixed points;

(ii) all the sets of points of order $n-3$ not containing either of those two points.

DEFINITION 13.1.

1. A point p in a combinatorial geometry (P, B) will be said to be *isolated* if every pair of points containing p is a block.

2. An E_3-geometry will be called an *Albert geometry* if it has at least one isolated point.

LEMMA 13.2. *For $n \geq 5$, there is a unique Albert geometry on n points having two isolated points.*

PROOF. If p, q are two isolated points, and b is a block not containing p or q, then $|b| \geq n-3$ as otherwise there will be three pairwise disjoint blocks.

Let $a \in P$, $a \neq p, q$. There is a block b disjoint from the blocks $\{a, p\}$ and $\{a, q\}$, and as $|b| \geq n-3$, $b = P \setminus \{a, p, q\}$. So the identification of the geometry is complete. □

We will look at some examples of Albert geometries with a unique isolated point. We do not expect that one can classify these without some further restrictions. In general, the geometry in the associated graph will depend on the base point selected, so it is noteworthy that if one of these geometries is an Albert geometry, then they all are. We note that even the number of points in the geometry may depend on the base point, in other words the corresponding graphs are not regular in general. For the specific case of the geometry with two isolated points just described, the corresponding graph is vertex transitive (as is clear from Albert's original description of it), so the same geometry is obtained from any base point.

If we remove an isolated point from an Albert geometry and look at the geometry induced on the remaining points, we get a reasonable class of geometries. This point of view is useful for the construction of examples.

DEFINITION 13.3. Let (P, B) be an Albert geometry, a an isolated point. The *derived geometry* (P_0, B_0) with respect to a has point set $P_0 = P \setminus \{a\}$, and blocks $B_0 = \{b \in B : a \notin b\}$.

The reconstruction of (P, B) from (P_0, B_0) is immediate. We can phrase the E_3-conditions directly in terms of the derived geometry as follows.

I-D There are at least two points.

II-D For any three points p, q, r there is a block containing p, q, and not r.

III-D There is no inclusion between distinct blocks.
IV-D For any block, its complement is a block.
V-D If b_1, b_2, b_3 are blocks intersecting pairwise, but with no common point, then $b_1 \cup b_2 \cup b_3 \neq P_0$.

Any geometry satisfying these axioms will be called a *derived geometry.* This terminology is justified by the following lemma.

LEMMA 13.4. *If (P, B) is an Albert geometry with a an isolated point then the derived geometry with respect to a satisfies Axioms I-D through V-D. Conversely, if (P_0, B_0) is a derived geometry, then the combinatorial geometry on $P = P_0 \cup \{a\}$ whose blocks are the blocks of B_0 together with the pairs containing a is an Albert geometry.*

We omit the verification. Using this result, we can give examples in a very convenient form.

EXAMPLES 13.1.

(1) Let (P_0, H) be a projective geometry with H the set of hyperplanes. Let (P_0, B_0) have as its blocks the elements of H and their complements. This is a derived geometry.
(2) Let (P_0, L) be a projective plane and let L' be a set of lines satisfying one of the following conditions:
    (a) Every point lies on at least 3 lines of L';
    (b) L' contains all the lines of L not passing through some fixed point p, and two of the lines passing through p.

    Let L^* consist of the lines in L' and their complements. Then (P_0, L^*) is a derived geometry.
(3) Let (P_0, L) be a projective plane and $\ell \in L$ a fixed line. Let L_ℓ consist of the line ℓ together with the sets $\ell \cup \ell_1 \setminus \ell \cap \ell_1$ as ℓ_1 varies over the remaining lines. Taking these sets and their complements as blocks, we get an Albert geometry, to which we return below.
(4) Let (P_0, H) be as in (1) and let ∞ be an additional point. Extend (P_0, H) to (P_1, H_1) with $P_1 = P_0 \cup \{\infty\}$, $H_1 = \{h \cup \{\infty\} : h \in H\}$, and let H_1^* consist of the elements of H_1 and their complements. Then (P_1, H_1^*) is a derived geometry.
(5) Let (P_0, B_0) be a derived geometry, and $a \in P_0$. Let

$$P_a = \{b \in B_0 : a \in b\} \cup \{\infty\}$$

with ∞ a new point. Let $B_a = P_0 \setminus \{a\}$. For $p \in B_a$ let $b_p = \{b \in B_0 : a \in b\} \cup \{\infty\}$. Let B_a^* consist of $\{b_p : p \in B_a\}$ and the complements $\{P_a \setminus b_p : p \in B_1\}$. Then (P_a, B_a^*) is a derived geometry; indeed, it is just the derived geometry obtained by passing from (P_0, B_0) to the corresponding graph, and then using the point p as a base point in place of the original base point. If we repeat this construction to form the geometry $(P_{a\infty}, B_{a\infty})$, we recover (P_0, B_0).

If (P_0, B_0) is a derived geometry with $|P_0| = n$ and $|B_0| = 2n'$, we will say that (n, n') are the *parameters* of the derived geometry. In Example 13.1 (5) above, if (P_0, B_0) has parameters (n, n'), then (P_a, B_a) has parameters $(n' + 1, n - 1)$. For example: the derived geometry associated with a projective geometry has type (n, n) and the new geometry (P_a, B_a) thus has parameters $(n + 1, n - 1)$.

As we are ultimately interested in graphs, we will want to consider how the geometry varies as we change the basepoint. We may suppose that any change of basepoint takes place in a series of steps, replacing a given basepoint by one adjacent to it, thus iterating the construction in Example 13.1 (5).

PROPOSITION 13.5. *Let (P, B) be an Albert geometry, G the associated graph with base point v and vertex set $\{v\} \cup P \cup B$, and $u \in G$ any vertex. Let (P_u, B_u) be the geometry associated to the graph G with respect to the base point u. Then (P_u, B_u) is again an Albert geometry.*

PROOF. Since G is connected, it suffices to prove the claim when u is adjacent to the base point v of G.

The case in which u is an isolated point of (P, B) must be handled separately. In this case, there is an involution $i \in \mathrm{Aut}(G)$ defined by

$$v \leftrightarrow u;\ a \leftrightarrow \{a, u\}\ (a \in P \setminus \{u\});\ b \leftrightarrow P \setminus b \text{ on blocks}$$

Thus in this case (P_u, B_u) is isomorphic via this automorphism to the original geometry (P, B).

Suppose now that $p_0 \in P$ is an isolated point, and $u \neq p_0$. Let $p_1 = \{p_0, u\} \in P_u$. We claim that p_1 is an isolated point of P_u.

The vertex $p_0 \in B_u$ is incident with $\{v, p_1\}$ in P_u. For any other $b \in P_u \setminus \{v, p_1\}$, we have $b \in B_0$, $u \in b$, and $b' = P_0 \setminus b$ is in B_u, with b' incident with p_1 and b. Thus p_1 is an isolated point. □

LEMMA 13.6. *Let (P, B) be an Albert geometry, $p_0 \in P$ an isolated point, and (P_0, B_0) the corresponding derived geometry with parameters (n, n') where $n = |P_0|$ and $n' = |B_0|/2$. Let G be the graph associated with (P, B), with base point v. Then for any vertex $u \in B_0 \cup \{p_0, v\}$, the associated geometry with base point u has the same parameters (n, n'), while for $u \in P_0 \cup (B \setminus B_0)$, the associated geometry has parameters $(n' + 1, n - 1)$. Thus at most two vertex degrees occur in the associated graph.*

PROOF. Let g be the order of the graph G. We have $g = 1 + (n+1) + (2n'+n) = 2n + 2n' + 2$. For $p \in P_0$, the geometry $(P_, B_p)$ has parameters (n_p, n'_p) where $n_p = |P_p| - 1$ with P_p the set of neighbors of p in G, namely v, $\{p, p_0\}$, and the blocks in B_0 which contain p: so $n_p = n' + 1$, and therefore $n'_p = n - 1$.

Making use of the involution $v \leftrightarrow p_0$ which interchanges $B \setminus B_0$ and P_0, the same parameters are associated with $u \in B \setminus B_0$.

On the other hand, for a block $b \in B_0$, the set P_b of neighbors of b consists of the points p belonging to b, the pairs $\{p_0, p\}$ with $p \in P_0$ not belonging to b, and the complement b' of b in P_0, leading to $n_b = |P_b| - 1 = |P_0| = n$ and thus $n'_b = n'$. □

This raises the question as to what sorts of geometries are associated on the one hand with regular graphs, and on the other hand with graphs having just two vertex degrees. These conditions do not seem to be very restrictive, and it may be of interest to impose similar conditions generalizing strong regularity, perhaps allowing some further use of algebraic methods. As an example, if we begin with the Albert geometry based on a projective geometry with a single hyperplane removed, we get a regular graph.

EXAMPLE 13.2. Let G be the graph associated to the Albert geometry whose derived geometry comes from the projective plane. Let v be the base point, $p_0 \in P$ the isolated point, and $\ell \in B$ a line. The associated geometry has points P_ℓ consisting of:

$$\begin{cases} \text{points } p \in P_0 \text{ lying on } \ell \\ \text{co-points } \{p, p_0\} \text{ with } p \notin \ell \end{cases}$$

Then for $\ell_1 \neq \ell$ a line, the points of P_ℓ incident with ℓ as a block in B_ℓ are

$$\{p \in \ell \setminus \ell_1\} \cup \{\hat{p} : p \in \ell_1 \setminus \ell\}$$

In other words, this corresponds to the union of the fixed line ℓ with ℓ_1, with their common point removed. The block associated with the base point v is $\{p : p \in \ell\}$.

In the specific case of the projective plane of order 2, the resulting geometry again comes from the projective plane of order 2. Otherwise, the geometry is a different one, and the automorphism group of the graph leaves the pair $\{v, p_0\}$ invariant, and is $\mathrm{Aut}(P_0, B_0) \times \mathbb{Z}_2$. But for $q = 2$ the group is transitive on $B_0 \cup \{p_0, v\}$.

13.2. Another series of E_3-geometries. Moving away from Albert geometries, what we would like to see next is an infinite family of 3-e.c. graphs G with $\mu_2(G) \to \infty$, ideally even $\mu_3(G) \to \infty$. But we are far from this. Leaving aside the M_{22} geometry with its remarkably good properties ($\mu_2(G) = 6$, $\mu_3(G) = 2$), in *infinite families* the best we have done to date is to reduce the number of blocks of order 2 to a single one, while all other blocks can be made arbitrarily large. But in this construction we deal with a single geometry, rather than the set of geometries associated with a given graph. So there is much to be improved on even at this weak level.

EXAMPLE 13.3. With $m_1, m_2, m_3 \geq 2$, let the geometry $A(m_1, m_2, m_3)$ be defined as follows. Our pointset is the union of three disjoint sets P_1, P_2, P_3 with $|P_i| = m_i$, together with two distinguished points p', p^-, and the following blocks.

(1) $b_0 = \{p', p^-\}$ (size 2);
(2) For $a \in P_i$: $a' = P_i \setminus \{a\} \cup \{p'\}$ (size m_i);
(3) For $a \in P_i$: $a^+ = \{a\} \cup P_{i+1}$ (addition modulo 3), of size $m_{i+1} + 1$;
(4) For $a \in P_i$: $a^- = \{a\} \cup P_{i-1} \cup \{p^-\}$ (subtraction modulo 3), of size $m_{i-1} + 2$.

If $m_1, m_2, m_3 \geq 2$ then $A(m_1, m_2, m_3)$ is an E_3-geometry, and if $m_1, m_2, m_3 \geq 3$ then it has a unique block of order 2, with the other blocks of order at least $\min(m_1, m_2, m_3)$. Let $G(m_1, m_2, m_3)$ be the associated graph.

LEMMA 13.7. *The dihedral group of order* 8 *acts on* $G = G(m_1, m_2, m_3)$ *as a group of automorphisms, extending the natural action on the* 4*-cycle* (v, p', b_0, p^-)*, with* v *the base point,* $b_0 = \{p', p^-\}$*. In this action the classes* $P_0 = P_1 \cup P_2 \cup P_3$*,* $B' = \{a' : a \in P_0\}$*,* $B^+ = \{a^+ : a \in P_0\}$*, and* $B^- = \{a^- : a \in P_0\}$ *are permuted.*

PROOF. We define two involutions in $\mathrm{Aut}(G)$ by:

$$\begin{array}{lllll} \iota' : & v \leftrightarrow p' & b_0 \leftrightarrow p^- & a \leftrightarrow a' & a^+ \leftrightarrow a^- \\ \iota^- : v \leftrightarrow p^- & b_0 \leftrightarrow p' & a \leftrightarrow a^- & a^+ \leftrightarrow a' & \end{array}$$

□

LEMMA 13.8. *The graph $G(m_1, m_2, m_3)$ is regular.*

PROOF. It suffices to check the degree of a vertex $a \in P_i$. The neighbors of a are v, a^+, a^- and a'_1 for $a_1 \in A_i \setminus \{a\}$, b^- for $\in A_{i+1}$, c^+ for $c \in A_{i-1}$, for a total of $3 + (m_i - 1) + m_{i+1} + m_{i-1}$ which is the degree of the base point. □

The geometries associated with $G(m_1, m_2, m_3)$ (i.e., those giving an isomorphic graph) are not well behaved.

LEMMA 13.9. *Let $G = G(m_1, m_2, m_3)$ corresponding to $A(m_1, m_2, m_3)$ with point set $P_1 \cup P_2 \cup P_3 \cup \{p', p^-\}$, and take $a \in P_i$. Then the geometry (P_a, B_a) associated with the base point a has at least $m_i + 1$ blocks of order 2, and exactly $2(m_i - 1)$ blocks of order $m_1 + m_2 + m_3 - 1 = n - 2$ with $n = |P|$ the degree of G. On the set $P_i^* = \{a'_1 : a_1 \neq a, a_i \in P_i\} \cup \{a^+, a^-\}$ of order $m_i + 1$, the induced geometry is the Albert geometry with two isolated points a^+, a^-, in the sense that all pairs $\{a^+, a'_1\}$ and $\{a^-, a'_1\}$ occur as blocks of the associated geometry (P_a, B_a), while all subsets of order $n - 3$ contained in P_a which are disjoint from $\{a^+, a^-\}$ and which contain $P_a \setminus P_i^*$ also occur as blocks.*

Note that in the "restricted" geometry on P_i^* we are taking as blocks, those which lie within P_i^*, and those which contain its complement.

PROOF.

$$P_a = \{v, a^+, a^-, b'(b \in P_i, b \neq a), c^-(c \in P_{i+1}), d^+(d \in P_{i-1})\}$$

The block associated with a' is $\{a^+, a^-\}$. The block associated with b^+ for $b \in A_i$, $b \neq a$ is $\{a^-, b'\}$. The block associated with b^- for $b \in P_i$ is $\{a^+, b'\}$.

Finally, the block associated with $b \in P_i$ $(b \neq a)$ is

$$\{v\} \cup P_{i+1}^- \cup A_{i-1}^+ \cup \{b'_1 : b_1 \in P_i, b_1 \neq a, b\}$$

□

We observe that the smallest geometry for which we are able to get a single block of order 2 has order 11, namely $A(3, 3, 3)$, and this is sharp. We will give some additional information concerning small geometries.

13.3. E_3-Geometries of order at most 7.

LEMMA 13.10. *Let (P, B) be an E_3-geometry, and $n = |P|$.*

1. *The maximal block size is at most $n - 3$.*
2. *If there is a block of order $n - 3$ then all pairs lying in its complement are blocks.*
3. *$n \geq 5$.*
4. *If p is a point which is contained in at least $n - 3$ blocks of order 2 then p is isolated.*
5. *If $n = 5$ or 6 then (P, B) is the Albert geometry with two isolated points.*

PROOF. The first three points are immediate. For the fourth, let q, q' be the two points not known to occur together with p as a block of order 2. Take a block containing p, q and not q'; it must be $\{p, q\}$. Similarly $\{p, q'\}$ is block.

For the last point, the case $n = 5$ is immediate, so take $n = 6$. Then there is a pair of points p, q which do not constitute a block, and therefore they are contained in two blocks of order 3, say $\{p, q, r\}$ and $\{p, q, s\}$. Let $t \notin \{p, q, r, s\}$. Then the pairs containing t and neither of p, q are blocks, by our second point. So t is an isolated point. So the geometry has two isolated points, and we are done. □

We will also carry through the analysis for the case $n = 7$, finding in this case that the geometry is necessarily an Albert geometry, and that there are only two possibilities: the Albert geometry with two isolated points, and one other.

EXAMPLE 13.4. We construct a derived geometry $B'(n_1, n_2)$ as follows. Let A_1, A_2 be sets of orders n_1, n_2 respectively, and c an additional point. Take as blocks:

$$a_1A_2(a_1 \in A_1); a_2A_1(a_2 \in A_2; cA_i'(A_i' \subseteq A_i, |A_i \setminus A_i'| = 1)$$

Let $B(n_1, n_2)$ be the corresponding Albert geometry, with $n_1 + n_2 + 2$ points.

If $n_1 = n_2 = 2$ then this is the Albert geometry with two isolated points on 6 vertices. Otherwise, it is an Albert geometry with one isolated vertex. In particular we have the case $n_1 = 2$, $n_2 = 3$ of order 7.

LEMMA 13.11. *Let (P_0, B_0) be the derived geometry associated with an Albert geometry, with $n_0 = |P_0| > 4$. Call the blocks of order 2 in B_0 edges, and view P_0 as a graph with respect to these edges. Then all edges in P_0 have a common vertex.*

PROOF. As $n_0 > 4$ there can be no disjoint edges. So we need only eliminate the possibility that there is a triangle.

Suppose p, q, r form a triangle: any pair is a block. Take further points s, t, and a block b containing r, s but not t. Then b is disjoint from p, q and hence the complement of b is p, q. But then t is in b, a contradiction. □

LEMMA 13.12. *An Albert geometry on 7 points with one isolated point must be isomorphic with $B(2, 3)$.*

PROOF. We work in the derived geometry (P_0, B_0) on 6 points, and consider the graph on P_0 whose edges are the pairs occurring as blocks in B_0. If there are three or more edges then their common vertex is a second isolated point, a contradiction.

If there are exactly two edges we identify the geometry $B(2, 3)$ as follows. Let the vertex common to the edges be called c, and let the other vertices on the edges be $A = \{a_1, a_2\}$, while the remaining vertices are $B = \{b_1, b_2, b_3\}$. As there are just the two edges, the other blocks containing c are triples of the form $\{c, b, b'\}$ with $b, b' \in B$, and as we may exclude any element of B, all such triples occur. Thus we know all the blocks containing c and taking complements, we have all the blocks.

Suppose there is at most one edge, and take four points $A = \{a_1, a_2, a_3, a_4\}$ containing no edge, and let c_1, c_2 be the other points. Then every block containing c_1 and meeting A is a triple, thus the blocks containing c_1 meet A in a certain set of pairs E_1, and for any two points $a, a' \in A$ there is a pair in E_1 containing a and not a'. If E_1 contains no disjoint pairs, it follows that the edge set E_1 forms a triangle in A; and if we define E_2 similarly, and E_2 contains no disjoint pairs, then E_2 forms a triangle in A. However $E_1 \cup E_2$ covers all pairs in A. So we may suppose that E_1 contains two disjoint pairs, say $\{a_1, a_2\}$ and $\{a_3, a_4\}$. We may suppose then that E_2 contains the pair $\{a_1, a_3\}$, and hence B_0 contains the blocks $\{c_1, a_1, a_2\}, \{c_1, a_3, a_4\}, \{c_2, a_1, a_3\}$ which meet pairwise but have no point in common. But as the union of these blocks is P_0, we contradict our axioms. □

And lastly we claim that up to this point no non-Albert geometry occurs.

LEMMA 13.13. *Let (P, B) be an E_3-geometry on $n \leq 7$ points. Then (P, B) is an Albert geometry.*

PROOF. We have dealt with the cases $n < 7$ and we suppose $n = 7$.

We call a block of order 2 an *edge* in P.

Suppose first that there is some block A of order 4. We claim that any $a \in A$ lies on an edge.

Let $A = \{a, a_1, a_2, a_3\}$ and take blocks b_1, b_2 with $a_i, a_3 \in b_i$, $a \notin b_i$ for $i = 1, 2$. Take b disjoint from $b_1 \cup b_2$ with $a \in b$. Since any pair disjoint from A is an edge, b must also be an edge.

Now as $|A| = 4$ we can find two points a', a'' in A for which there are edges $\{p, a'\}, \{p, a''\}$ with a common neighbor $p \in P \setminus A$. Then p lies on 4 edges and is therefore an isolated point.

¿From now on suppose that *there is no block of order* 4*,* and we will arrive at a contradiction.

We show first that every point lies on an edge. Suppose the point p lies on no edge, so that every block containing p has order 3. Take two blocks b_1, b_2 whose intersection is $\{p\}$. Then the complement of $b_1 \cup b_2$ is an edge e.

Take $q \in b_1, r \in b_2$ with $q, r \neq p$, and with q, r not an edge, using the fact that there are no three disjoint edges. Take a block b_3 containing q, r, and not p. Then b_1, b_2, b_3 meet pairwise but have no common point, so they are disjoint from a block, which must be e. Thus $b_3 \subseteq b_1 \cup b_2$. There is a point $s \notin b_3 \cup \{p\} \cup e$. Form a block b containing q, s and not p, and take a block b' disjoint from $b \cup b_3$ and containing p. Then $p \in b' \subseteq \{p\} \cup e$ and thus b' is an edge containing p.

Now consider the graph on P formed by the edges. Every vertex lies on an edge, there are no three disjoint edges, and furthermore no vertex has degree greater than 3, as it would then be an isolated point of the geometry.

By the first two conditions, some vertex p must have degree at least 3, and hence exactly 3. Then it follows by inspection that there is some vertex q not adjacent to p such that every point of P other than p, q is adjacent to one of the two points p, q. Consider a block b containing p, q. Since b cannot contain any edge at p or q, b is $\{p, q\}$, so these points are adjacent and we have a contradiction. $\square$

13.4. A small non-Albert geometry. We record some further information about small E_3-geometries. The smallest non-Albert geometry lives on a set with 8 points, and is unique up to isomorphism. This geometry has 7 blocks of order 2.

The smallest geometry in which one has a unique block of order 2 is the geometry $A(3, 3, 3)$ on 11 points. To get an E_3 geometry with no block of order 2 one may take the geometry of lines and hyperovals in the projective plane over a field of order 4, with 21 points. On the other hand, we have checked that such a geometry must have at least 13 points. We would like to know the minimum size of such a geometry, and what the geometry is. We observe at this point some distinct and possibly very substantial gap between the degenerate cases we have discussed and the next level.

By brute force search, all of the E_3-geometries of order 8 may be identified. There are 11 such geometries, corresponding to 7 graphs. Four of these graphs correspond to two geometries of order 8, two correspond to one geometry of order 8 apiece, while the last graph corresponds to one geometry of order 8 and one of order 9. These geometries are Albert geometries except for one pair of geometries corresponding to a single graph. We list the geometries as follows, including the block counts (the number of blocks of each size, from the minimum size 2 up to the maximum size). We will refer to one geometry as a "variant" of another if it defines an isomorphic graph.

E_3-geometries of order 8:

(1) The Albert geometry with two isolated points. Block count $(13, 0, 0, 6)$.
(2) Albert geometries with unique isolated points:
    (a) The geometry whose derived geometry comes from a projective plane minus a line, with a vertex transitive automorphism group. Block count $(7, 6, 6)$.
    (b) The geometry whose derived geometry comes from a projective plane. In the associated graph, there is also a geometry on 9 points. Block count $(7, 7, 7)$.
    (c) The geometry $B(2, 4)$ with block count $(9, 4, 4, 2)$ and a variant with block count $(7, 6, 6)$.
    (d) The geometry $B(3, 3)$ with block count $(7, 6, 6)$ and a variant with block count $(9, 4, 4, 2)$.
    (e) An Albert geometry with block count (8,5,5,1) and a variant with block count $(7, 6, 6)$.
(3) A pair of non-Albert geometries with block counts $(6, 10, 1, 2)$ and $(7, 6, 6)$.

14. Appendix: Amalgamation Classes (Tables)

The following table shows the full list of 27 amalgamation classes determined by constraints on triangles, not allowing free amalgamation, and with the associated Fraïssé limit primitive. The structures are assumed to have four nontrivial 2-types, all of them symmetric (e.g., one may think of these structures as complete graphs, with 4 colors of edges). The data are taken from [**Che98**], with slightly different notation but the same numbering. Those which can be interpreted as homogeneous metric spaces were put into a clearer form in §10.

#	ABD	CDD	AAC	ADD	AAD	BBD	CCA	CCD	BDD	BAA	AAA	DDD
1	1	1	0	1	0	0	0	0	0	0	0	0
2	1	1	0	1	0	0	0	0	0	0	0	1
3	1	1	0	1	1	0	0	0	0	0	0	0
4	1	1	0	1	0	0	0	0	0	0	1	0
5	1	1	0	1	1	0	0	0	0	0	0	1
6	1	1	0	1	0	0	0	0	0	0	1	1
7	1	1	0	1	1	0	0	0	0	0	1	0
8	1	1	0	1	1	0	0	0	0	0	1	1
9	1	1	0	0	1	0	0	0	0	0	0	1
10	1	0	0	1	0	0	0	0	0	0	1	1
11	0	1	0	1	0	1	0	0	0	0	0	1
12	0	1	0	1	1	1	0	0	0	0	0	1
13	0	1	0	1	0	1	0	0	0	0	1	1
14	0	1	0	1	1	1	0	0	0	0	1	1
15	1	1	0	1	0	1	0	0	0	0	0	1
16	1	1	0	1	1	1	0	0	0	0	0	1
17	1	1	0	1	0	1	0	0	0	0	1	1
18	1	1	0	1	1	1	0	0	0	0	1	1
19	1	1	0	0	1	1	0	0	0	0	0	1
20	1	1	0	0	1	1	0	0	0	0	1	1
21	1	1	1	0	1	0	0	0	0	0	0	1
22	1	1	1	1	1	0	0	0	0	0	0	0
23	1	1	1	1	1	0	0	0	0	0	0	1
24	1	1	1	1	0	0	0	0	0	0	1	1
25	1	1	1	1	1	0	0	0	0	0	1	1
26	1	1	1	0	1	0	0	1	1	0	0	1
27	1	0	0	1	1	0	1	1	1	1	1	1

TABLE 3. 27 amalgamation classes [**Che98**]. See §10

References

[AH78] F. Abramson and L. Harrington, "Models without indiscernibles" *J. Symbolic Logic* **43** (1978), 572–600.

[AL95] R. Akhtar and A. Lachlan, "On countable homogeneous 3-hypergraphs," *Arch. Math. Logic* **34** (1995), 331–344.

[AH06] M. Alfuraidan and J. Hall, "Smith's theorem and a characterization of the 6-cube as distance-transitive graph," *J. Algebraic Combin.* **24** (2006), 195–207.

[AMp10] D. Amato and H. D. Macpherson, "Infinite metrically homogeneous graphs of diameter 3," Preprint, April 2010.

[Ash71] C. Ash, "Undecidable $\aleph_0$-categorical theories," *Notices Amer. Math. Soc.* **18** (1971), 423.

[Big09] N. Biggs, "Strongly regular graphs with no triangles," *Research Report, September 2009.* arXiv:0911.2160v1.

[Bog02] S. A. Bogatyĭ, "Metrically homogeneous spaces," *Uspekhi Mat. Nauk* **57** (2002), 3–22 (Russian); English translation in *Russian Mathematical Surveys*, bf 57 (2002), 221-240.

[Bon09] A. Bonato, "The search for n-e.c. graphs" *Contrib. Discrete Math.* **4** (2009), 40–53.

[Bon10] A. Bonato, "Bounds and constructions for n-e.c. tournaments," *Contributions to Discrete Mathematics* **5** (2010), 52–66.

[BM08] J. A. Bondy and U. R. S. Murty, *Graph Theory*, *Graduate Texts in Mathematics* **244**, Springer, 2008, xii+651 pp.

[BDM07] C. Bridges, A. Day, and S. Manber, "k-Saturated graphs," REU report, Clemson, 2007, http://www.math.clemson.edu/ kevja/REU/2007/.

[Br11] A. Brouwer, "Graph descriptions," `http://www.win.tue.nl/~aeb/graphs/`.

[BCN89] A. E. Brouwer, A. M. Cohen, and A. Neumaier, *Distance-Regular Graphs,* Springer-Verlag, New York, 1989, 485 pp.,

[BH95] A. E. Brouwer and W. Haemers, "Association Schemes," in *Handbook of Combinatorics, Vol. 1, 2*, 747–771. Elsevier, Amsterdam 1995.

[BvL84] A. E. Brouwer and J. H. van Lint, *Strongly regular graphs and partial geometries*, in *Enumeration and design (Waterloo, Ont. , 1982*, 85–122, Academic Press, Toronto, ON, 1984.

[Cam80] P. Cameron, "6-transitive graphs," *J. Combinatorial Theory, Series B* **28** (1980), '68–179.

[Cam90] P. Cameron, *Oligomorphic permutation groups, London Mathematical Society Lecture Note Series,* **152.** Cambridge University Press, Cambridge, 1990. viii+160 pp.

[Cam96] P. Cameron, "The random graph," in J. Nešetřil, R. L. Graham (Eds.), *The Mathematics of Paul Erdős,* Springer, Berlin, 1996, pp. 331–351.

[Cam97] P. Cameron, "Oligomorphic groups and homogeneous graphs," *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, 497, Kluwer Acad. Publ., Dordrecht, 1997.

[Cam98] P. Cameron, "A census of infinite distance transitive graphs" *Discrete Math.* **192** (1998), 11–26.

[Cam03] P. Cameron, "Homogeneous permutations," in *Permutation patterns (Otago, 2003), Electron. J. Combin.* **9** (2002/03), no. 2, Research paper 2, 9 pp.

[Cam05] P. Cameron, "The random graph has the strong small index property," *Discrete Math.* **291** (2005), 41–43.

[CvL91] P. Cameron and J. H. van Lint, *Designs, graphs, codes and their links*, *London Mathematical Society Student Texts* **22**. Cambridge University Press, Cambridge, 1991. x+240 pp.

[Che88] G. Cherlin, "Homogeneous tournaments revisited," *Geom. Ded.* **26** (1988), 231-240.

[Che93] G. Cherlin, "Combinatorial Problems connected with finite homogeneity," *Proceedings of the International Conference on Algebra, Part 3 (Novosibirsk, 1989)*, 3–30, *Contemp. Math.* **131**, Part 3, Amer. Math. Soc., Providence, RI, 1992.

[Che98] G. Cherlin, "The classification of countable homogeneous directed graphs and countable homogeneous n-tournaments." *Mem. Amer. Math. Soc.* **131** (1998), no. 621, xiv+161 pp.

[Che99] G. Cherlin, "Infinite imprimitive homogeneous 3-edge-colored complete graphs," *J. Symbolic Logic* **64** (1999), 159–179.

[ChH03] G. Cherlin and E. Hrushovski, *Finite Structures with Few Types,* Annals of Math. Studies **152**, 2003.

[ChL00] G. Cherlin and B. Latka, "Minimal antichains in well-founded quasi-orders with an application to tournaments," *J. Combin. Theory Ser. B*, **80** (2000), 258–276.

[CSW93] G. Cherlin, D. Saracino, and C. Wood, "On homogeneous nilpotent groups and rings" *Proc. Amer. Math. Soc.* **119** (1993), 1289–1306.

[CT02] G. Cherlin and S. Thomas, "Two cardinal properties of homogeneous graphs," *J. Symbolic Logic* **67** (2002)

[DNT86] J. Dixon, P. Neumann, and S. Thomas, "Subgroups of small index in infinite symmetric groups," *Bull. London Math. Soc.* **18** (1986), 580-586.

[Dun82] M. J. Dunwoody, "Cutting up graphs," *Combinatorica* **2** (1982), 15–23.

[Ehr72] A. Ehrenfeucht, "There are continuum $\aleph_0$-categorical theories," *Bull. Acad. Polon. Sci.* **20** (1972), 425-427.

[EKR76] P. Erdős, D. Kleitman, B. Rothschild, "Asymptotic enumeration of K_n-free graphs," in *Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), Tomo II*, 19–27. *Atti dei Convegni Lincei* **17** (1976), Accad. Naz. Lincei, Rome.

[Fag76] R. Fagin, "Probabilities on finite models," *J. Symb. Logic* **41** (1976), 50-58.

[Fra54] R. Fraïssé, "Sur l'extension aux relations de quelques propriétés des ordres," *Ann. Ecole Normale Sup.* **7** (1954), 361–388.

[Gar76] A. Gardiner, "Homogeneous graphs," *J. Comb. Th.* **20** (1976), 94-102.

[Gla71] W. Glassmire, Jr., "There are $2^{\aleph_0}$ countably categorical theories," *Bull. Acad. Polon. Sci.* **19** (1971), 185–190.

[GK09] L. Graham and J. M. Kantor, *Naming Infinity: A True Story of Religious Mysticism and Mathematical Creativity,* Belknap Press of Harvard University Press, 2009.

[Hen71] C. W. Henson, "A family of countable homogeneous graphs," *Pacific J. Math.* **38** (1971), 69–83.

[Hen72] C. W. Henson, "Countable homogeneous relational systems and categorical theories," *J. Symb. Logic* **37** (1972), 494-500.

[HHLS93] W. Hodges, I. Hodkinson, D. Lascar, and S. Shelah, "The small index property for ω-stable ω-categorical structures and for the random graph," *J. London Math. Soc.* **48** (1993) 204–218.

[HN05] J. Hubička and J. Nešetřil, "Finite presentation of homogeneous graphs, posets and Ramsey classes," *Probability in mathematics, Israel J. Math.* **149** (2005), 21–44.

[Huš08] M. Hušek, "Urysohn space, its development and Hausdorff's approach" *Topology and its Applications* **155** (2008), 1493–1501. Special number devoted to the proceedings of a Workshop on the Urysohn space, Beer-Sheva, 2006.

[KLM89] W. Kantor, M. Liebeck, and H. D. Macpherson, "$\aleph_0$-categorical structures smoothly approximated by finite substructures," *Proc. London Math. Soc.* **59** (1989), 439–463.

[KPT05] A. Kechris, V. Pestov, and S. Todorcevic, "Fraïssé limits, Ramsey theory, and topological dynamics of automorphism groups," *Geom. Funct. Anal.* **15** (2005), 106–189.

[GGK96] M. Goldstern, R. Grossberg, and M. Kojman, "Infinite homogeneous bipartite graphs with unequal sides," *Disc. Math.* **149** (1996), 69–82

[GK78] Ya. Gol'fand and Yu. Klin, "On k-homogeneous graphs," (Russian), *Algorithmic studies in combinatorics (Russian)* **186**, 76–85, (errata insert), "Nauka," Moscow, 1978.

[KPR87] Ph. Kolaitis, H. Prömel, and B. Rothschild, "$K_{\ell+1}$-free graphs: asymptotic structure and a 0–1 law," *Trans. Amer. Math. Soc.* **303** (1987), 637-671.

[Kom99] P. Komjáth, "Some remarks on universal graphs," *Discrete Math.* **199** (1999), 259–265.

[KMP88] P. Komjáth, A. Mekler and J. Pach, "Some universal graphs," *Israel J. Math*, **64** (1988), 158–168.

[KT01] D. Kuske and J. Truss, "Generic automorphisms of the universal partial order," *Proc. Amer. Math. Soc.* **129** (2001), 1939–1948.

[Lac82] A. Lachlan, "Finite homogeneous simple digraphs," *Proceedings of the Herbrand symposium (Marseilles, 1981),* 189–208, *Stud. Logic Found. Math.*, **107**, North-Holland, Amsterdam, 1982.

[Lac84] A. Lachlan, "Countable homogeneous tournaments," *Trans. Amer. Math. Soc.* **284** (1984), 431-461.

[Lac86a] A. Lachlan, "Binary Homogeneous Structures II," *Proc. London Math. Soc.* **52** (1986), 412–426.

[Lac86b] A. Lachlan, "Homogeneous structures," *in* **Proceedings of the ICM 1986**, American Mathematical Society, 1987, 314-321.

[LT95] A. Lachlan and A. Tripp, "Finite homogeneous 3-graphs," *Math. Logic Quart.* **41** (1995), 287–306.

[Lac96] A. Lachlan, "Stable finitely homogeneous structures: a survey," *Algebraic model theory (Toronto, ON, 1996)* 145–159, *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, **496**, Kluwer Acad. Publ., Dordrecht, 1997.

[LW80] A. Lachlan and R. Woodrow, "Countable ultrahomogeneous undirected graphs," *Trans. Amer. Math. Soc.* **262** (1980), 51-94.

[Lat94] B. J. Latka, "Finitely constrained classes of homogeneous directed graphs," *J. Symb. Logic*, 59:124–139, 1994.

[Lat02] B. J. Latka, "A structure theorem for tournaments omitting $IC_3(I, I, L_3)$," Preprint, 1994, revised 2002.

[Lat03] B. J. Latka, "A structure theorem for tournaments omitting N_5," *J. Graph Theory* **42** (2003), 165–192.

[Mph82] H. D. Macpherson, "Infinite distance transitive graphs of finite valency," *Combinatorica* **2** (1982), 63–69.

[Mph11] H. D. Macpherson, "A survey of homogeneous structures," Preprint 2011. To appear, *Discrete Mathematics*.

[Neš07] J. Nešetril, "Metric spaces are Ramsey," *European J. Combin.* **28** (2007), 457–468.

[NR75a] J. Nešetril and V. Rödl, "A Ramsey graph without triangles exists for any graph without triangles," *Infinite and Finite Sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday, Vol. III*, 1127–1132. *Colloq. Math. Soc. Janos Bolyai* **10**, North-Holland, Amsterdam, 1975.

[NR75b] J. Nešetril and V. Rödl, "Type theory of partition properties of graphs," *Recent Advances in Graph Theory (Proc. Second Czechoslovak Sympos., Prague, 1974* 405–412, Academia, Prague, 1975.

[NR76] J. Nešetril and V. Rödl, "The Ramsey property for graphs with forbidden complete subgraphs," *J. Combin. Theory Ser. A* **20** (1976), 243–249.

[NR77a] J. Nešetril and V. Rödl, "A structural generalization of the Ramsey Theorem," *Bull. Amer. Math. Soc.* **83** (1977), 127–128.

[NR77b] J. Nešetril and V. Rödl, "Partition of relational and set systems," *J. Combin. Theory Ser. A* **22** (1977), 289–312.

[NR83] J. Nešetril and V. Rödl, "Ramsey classes of set systems," *J. Combin. Theory Ser. A* **34** (1983), 183–201.

[NR90] J. Nešetril and V. Rödl, eds., *Mathematics of Ramsey theory,* Algorithms and Combinatorics, 5. Springer-Verlag, Berlin, 1990. xiv+269

[Pes02] V. Pestov, "Ramsey-Milman phenomenon, Urysohn metric spaces, and extremely amenable groups" *Israel J. Math.* **127** (2002), 317–357.

[PV08] F. Petrov and A. Vershik, "Invariant measures on the set of graphs and homogeneous uncountable universal graphs," Preprint, 2008.

[Rad64] R. Rado, "Universal graphs and universal functions," *Acta Arith.* **9** (1964), 331–340.

[Rub94] M. Rubin, "On the reconstruction of $\aleph_0$ categorical structures from their automorphism groups," *Proc. London Math. Soc.* **69** (1994), 225–249.

[SW86] D. Saracino and C. Wood, "Finite homogeneous rings of odd characteristic," *Logic colloquium '84 (Manchester, 1984)*, 207–224, *Stud. Logic Found. Math.*, **120**, North-Holland, Amsterdam, 1986.

[SWS93] N. Sauer, R. E. Woodrow, B. Sands, *Finite and Infinite combinatorics in Sets and Logic (Proc. NATO Advanced Study Inst., Banff, Alberta, Canada, May 4, 1991),* Kluwer-Academic Publishers, 1993.

[Sch79] J. Schmerl, "Countable homogeneous partially ordered sets," *Alg. Univ.* **9** (1979), 317-321.

[She74] J. Sheehan, "Smoothly embeddable subgraphs," *J. London Math. Soc.* **9** (1974), 212-218.

[Smi71] D. Smith, "Primitive and imprimitive graphs," *Quart. J. Math. Oxford* **22**, (1971), 551–557.

[Sz^{2}65] E. Szekeres and G. Szekeres, "On a problem of Schütte and Erdős," *Math. Gazette* **49** (1965), 290–293.

[Tho91] S. Thomas, "Reducts of the random graph," *J. Symbolic Logic* **56** (1991), 176–181.

[Tho96] S. Thomas, "Reducts of random hypergraphs," *Ann. Pure Appl. Logic* **80** (1996), 165–193.

[Tru85] J. Truss, "The group of the countable universal graph," *Math. Proc. Cambridge Phil. Soc.* **98** (1985), 213-245.

[Tru92] J. Truss, "Generic automorphisms of homogeneous structures," *Proc. London Math. Soc.* **65** (1992), 121–141.

[Tru09] J. Truss, "On the automorphism group of the countable dense circular order," *Fund. Math.* **204** (2009), 97–111.

[Tru03] J. Truss, "The automorphism group of the random graph: four conjugates good, three conjugates better," *Discrete Math.* **268** (2003), 257–271.

[TrT08] J. Truss and S. Torrezão de Sousa, "Countable homogeneous coloured partial orders," *Dissertationes Math. (Rozprawy Mat.)* **455** (2008), 48 pp.

[Ury25] P. Urysohn, "Sur un espace métrique universel," *CRAS* **180** (1925), 803–806.

[Ury27] P. Urysohn, "Sur un espace métrique universel," *Bull. Sci. Math.* **51** (1927), 43–64 and 74–90.

[vLW92] J. H. van Lint and R. Wilson, *A course in combinatorics*, Cambridge University Press, Cambridge, 1992.

[Wat07] S. Waton, "On Permutation Classes Defined by Token Passing Networks, Gridding Matrices and Pictures: Three Flavours of Involvement," Ph.D. thesis, St. Andrews, 2007.

Department of Mathematics, Rutgers University, 110 Frelinghuysen Rd., Piscataway, NJ 08854

This page intentionally left blank

Contemporary Mathematics
Volume **558**, 2011

On symmetric indivisibility of countable structures

Assaf Hasson, Menachem Kojman, and Alf Onshuus

Abstract. A structure $\mathcal{M} \leq \mathcal{N}$ is symmetrically embedded in $\mathcal{N}$ if any $\sigma \in \mathrm{Aut}(\mathcal{M})$ extends to an automorphism of $\mathcal{N}$. A countable structure $\mathcal{M}$ is *symmetrically indivisible* if for any coloring of M by two colors there exists a monochromatic $\mathcal{M}' \leq \mathcal{M}$ such that $\mathcal{M}' \cong \mathcal{M}$ and $\mathcal{M}'$ is symmetrically embedded in $\mathcal{M}$.

We give a model-theoretic proof of the symmetric indivisibility of Rado's countable random graph [**9**] and use these new techniques to prove that $\mathbb{Q}$ and the generic countable triangle-free graph $\Gamma_\triangle$ are symmetrically indivisible. Symmetric indivisibility of $\mathbb{Q}$ follows from a stronger result, that symmetrically embedded elementary submodels of $(\mathbb{Q}, \leq)$ are dense. As shown by an anonymous referee, the symmetrically embedded submodels of the random graph are not dense.

1. Introduction

Indivisibility of countable structures is a broad phenomenon. The ordered rationals $(\mathbb{Q}, \leq)$, Rado's countable random graph, Γ, and the generic countable K_n-free graph, Γ_n, for $n \geq 3$, known also as Henson graphs, are examples of countable structures which are indivisible under finite partitions: when partitioned to finitely many parts, one of the parts contains an isomorphic copy of the whole structure.

Indivisibility of $(\mathbb{Q}, \leq)$ is a triviality. That of Γ is well known, easy to prove and in fact true in a stronger form: whenever Γ is partitioned into finitely many parts, one of the parts itself is isomorphic to Γ [**1**]. Indivisibility of the generic triangle-free graph is somewhat harder and was proved in [**10**]. Indivisibility of all Henson graphs was proved in [**4**].

In this paper we investigate indivisibility with respect to a stronger condition on the desired copy of the whole structure, which we now define. A substructure $\mathcal{M}$ of a structure $\mathcal{N}$ is *symmetrically embedded* in $\mathcal{N}$ if every automorphism of $\mathcal{M}$ extends

1991 *Mathematics Subject Classification.* Primary 05C63, 11U09, 05D10; Secondary 03C64, 22F50, 20B27.

Key words and phrases. Rado's random graph, ordered rationals, generic triangle-free graph, automorphism, $\aleph_0$-categorical structure, stable embedding, indivisibility, symmetric indivisibility.

The first author was partially supported by an Israel Science Foundation grant number 1156/10.

The second author thanks the Institute for Advanced Study, Princeton, NJ, for its hospitality during his work on this paper.

to an automorphism of $\mathcal{N}$. Let us call a countable structure $\mathcal{N}$ *symmetrically indivisible* if for every partition of $\mathcal{N}$ into finitely many parts, one of the parts contains an isomorphic copy of $\mathcal{N}$ which is symmetrically embedded in $\mathcal{N}$.

The symmetric indivisibility of Rado's random graph was established recently in [**9**]. In this paper we investigate symmetric indivisibility in the ordered rationals and in Henson's generic triangle-free graphs. The main results are:

Theorem (Theorem 3.11 below). The symmetrically embedded copies of $\mathbb{Q}$ are *dense* in all copies of $\mathbb{Q}$ in $\mathbb{Q}$, that is, for every $\mathcal{P} \leq \mathbb{Q}$ which is isomorphic to $\mathbb{Q}$ there exists $\mathcal{Q} \leq \mathcal{P}$ isomorphic to $\mathbb{Q}$ and symmetrically embedded in $\mathbb{Q}$.

This of course implies the symmetric indivisibility of $\mathbb{Q}$.

Theorem (Theorem 5.30 below). The generic triangle-free graph is symmetrically indivisible.

The next major open problem (addressed in a sequel paper) concerning symmetric indivisibility in countable structures is, then:

Problem. Does symmetric indivisibility hold in all Henson graphs?

The symmetric indivisibility of Rado's graph Γ was established in [**9**] along the following scheme. (1) The composition of Γ with itself, denoted $\Gamma[\Gamma]$ is symmetrically embedded in Γ. (2) For every partition of $\Gamma[\Gamma]$ to two parts, one of the parts contains a symmetrically embedded copy of Γ.

In Section 2 we investigate further the scheme of using composition of structures, use it to prove the symmetric indivisibility of $(\mathbb{Q}, \leq)$, and indicate its limitations.

In Section 3 we develop a different strategy for proving symmetric indivisibility, by proving the density of symmetrically embedded copies of an indivisible structure. This suffices for a second proof of the symmetric indivisibility of $\mathbb{Q}$, and of $\mathbb{Q}$-trees, which are handled in Section 4.

However, this method too has limitations: most notably, density of symmetrically embedded copies is false in Rado's graph, as proved by an anonymous referee of this paper (see Theorem 5.3 below).

Finally, to obtain the symmetric indivisibility of the generic triangle-free graph, a combination of both approaches with a new ingredient is employed in Section 5. The new ingredient is the notion of a *stably embedded* sub-structure. A substructure $\mathcal{M} \leq \mathcal{N}$ is stably embedded in $\mathcal{N}$ if every subset $A \subseteq \mathcal{M}$ which is definable with parameters in $\mathcal{N}$ is also definable in $\mathcal{M}$ with parameters in $\mathcal{M}$. We prove that (1) the stably embedded copies of $\Gamma_\triangle$ are indestructible by finite partitions, and, (2) the symmetrically indivisible copies of the $\Gamma_\triangle$ are dense in the stably embedded copies.

Symmetric indivisibility of all Henson graphs can be proved along similar lines, but the proof of indestructibility of stably embedded copies, which follows the proof in [**4**], is quite involved, and will be presented elsewhere. We conclude the paper with the discussion of elementary symmetric indivisibility in Section 6.

Thus, the main progress on the methods of [**9**] is the improved understanding of the distribution of symmetrically embedded copies of a structure. The common feature to both density of symmetric copies of $\mathbb{Q}$ and the existence of symmetrically embedded copies of, say, Γ in stably embedded ones, is the following: it is possible to decrease the space of types realized over an infinite subset of a structure by thinning out the set. Since thinning out a set puts more elements in its complement — which may realize more types over the set — this phenomenon runs somewhat agains one's intuition.

In the case of the rationals a copy of $\mathbb{Q}$ can be thinned out so that up to conjugation only two types are realized over the remaining set. In the case of the random graph a stably embedded copy can be thinned out so that every vertex outside the remaining copy has only a finite set of neighbors inside it.

The main results in this paper are combinatorial, with the model-theoretic methods serving to obtain them. We did, though, include some results which are model-theoretic. Such are Proposition 2.14, Theorem 3.12, which looks ahead to Section 5, and is not needed in the sequel, and all of Section 6. These results are not needed for the rest of the paper and may be skipped without any harm to the comprehension of the main combinatorial results.

We tried to minimize the model theoretic jargon in the parts needed for the combinatorics. The concepts of definability and of stable embedding in the particular context of the random graph and of the generic triangle-free graphs are simpler than in their most general form, so (at least in our sincere opinion) a reader who is interested in symmetric indivisibility but is less interested in model theory can still follow all proofs without learning model theory thoroughly.

Acknowledgment. The authors are grateful to two anonymous referees for many valuable comments that improved this paper. We are especially grateful to one of the referees for proving Theorem 5.3 below and allowing us to include it here.

2. The composition of two structures

In [**9**] the symmetric indivisibility of the countable random graph, Γ, is proved in three steps:

(1) It is observed that the composition[1] $\Gamma[\Gamma]$ of Γ with itself embeds symmetrically into Γ.
(2) Any fibre and any section of $\Gamma[\Gamma]$ are isomorphic to Γ and symmetrically embedded in $\Gamma[\Gamma]$.
(3) Any coloring $c : \Gamma[\Gamma] \to \{0, 1\}$ contains either a monochromatic section or a monochromatic fibre.

This argument translates readily to show that $(\mathbb{Q}, \leq)$ is symmetrically indivisible, with a significant strengthening of (1): $\mathbb{Q}[\mathbb{Q}]$ does not only embed symmetrically into $\mathbb{Q}$, it is actually isomorphic to $\mathbb{Q}$. This stronger form of (1) is false for the random graph. As we shall see below, this difference between Γ and $\mathbb{Q}$ runs deeper, as in $\mathbb{Q}$ the symmetrically embedded submodels are dense (Theorem 3.11 below), while in Γ they are not dense (Theorem 5.3 below by an anonymous referee).

We give a now more general treatment of composition:

Definition 2.1. Let $\mathcal{L}$ be a relational first order language and $\mathcal{M}$ a structure for $\mathcal{M}$. The *composition* $\mathcal{M}[\mathcal{M}]$, of $\mathcal{M}$ with itself is the $\mathcal{L}$-structure whose universe

[1] Kojman and Geschke refer to the composition of two graphs as the wreath product. At the suggestion of one of the referees we replaced the terminology to one which seems somewhat more accurate.

is $M \times M$ and such that for $R(x_1, \ldots, x_k) \in \mathcal{L}$ we set

$$\begin{aligned}
&R(M[M]) := \\
&\{((a, a_1), \ldots, (a, a_k)) : a \in M, M \models R(a_1, \ldots, a_k)\} \cup \\
&\{((a_{1,1}, a_{1,2}), \ldots, (a_{k,1}, a_{k,2})) : \bigwedge_{i \neq j} (a_{i,1} \neq a_{j,1}), M \models R(a_{i,1}, \ldots, a_{j,1})\}
\end{aligned}$$

Recall that:

Definition 2.2. A structure $\mathcal{M}$ is *transitive* if $\mathrm{Aut}(\mathcal{M})$ acts transitively on M.

As we will see, transitivity is a natural condition when studying (symmetric) indivisibility. We need the following technical definition:

Definition 2.3. Call a theory (or a structure) in a relational language Δ-free if every relation is either binary or does not intersect any diagonal.

Note that the notion of Δ-freeness is a harmless technicality, since any n-ary relation, R, can be written as the union of an n-ary Δ-free relation with Δ-free relations of smaller arities (interpreted as the intersection of R with the various diagonals). At all events, the following is now obvious:

Lemma 2.4. *If $\mathcal{M}$ is a transitive structure in a Δ-free relational language then every section and every fibre of $\mathcal{M}[\mathcal{M}]$ are isomorphic, with the induced structure, to $\mathcal{M}$.*

Where, naturally, a *fibre* of the composition $\mathcal{M}[\mathcal{M}]$ is a set of the form $\{(a, b) : b \in M\}$ for some fixed $a \in M$ and a *section* is a set of the form $\{(a, f(a)) : a \in M\}$ for some function $f : M \to M$.

Observe also that if $c : \mathcal{M}[\mathcal{M}] \to \{0, 1\}$ is any coloring then there exists either a 0-monochromatic fibre or a 1-monochromatic section. For if there is no 0-monochromatic fibre, we can always find a function $f : M \to M$ such that $c(a, f(a)) = 1$. Thus we get:

Lemma 2.5. *Let $\mathcal{M}$ be a transitive structure in a Δ-free relational language $\mathcal{L}$. Let $W \subseteq \mathcal{M}$ be isomorphic to $\mathcal{M}[\mathcal{M}]$. Then $\mathcal{M}$ is indivisible. Moreover, under any coloring $c : M \to \{0, 1\}$ the monochromatic sub-structure of $\mathcal{M}$ isomorphic to $\mathcal{M}$ can be chosen either a section or a fibre of W.*

The above suggests the following definition:

Definition 2.6. Let $\mathcal{M}$ be a structure in a Δ-free relational first order language. Say that $\mathcal{M}$ is *(symmetrically) self similar* if $\mathcal{M}[\mathcal{M}]$ embeds (symmetrically) into $\mathcal{M}$.

Example 2.7. The structure $\mathcal{M} := (\mathbb{Q}, \leq)$ is symmetrically self-similar. The structure $\mathcal{M}[\mathcal{M}]$ is simply $(\mathbb{Q}^2, \leq_{\mathrm{lex}})$, which is a countable dense linear order with no endpoints. Since all such structures are isomorphic to each other we get that $\mathcal{M}[\mathcal{M}] \cong \mathcal{M}$.

The importance of the composition $\mathcal{M}[\mathcal{M}]$ to the study of symmetric indivisibility is that it has many copies of $\mathcal{M}$ which are symmetrically embedded:

LEMMA 2.8. *Let $\mathcal{M}$ be a transitive structure in a relational language and $\mathcal{M}_0 \leq \mathcal{M}[\mathcal{M}]$ a fibre or a section. Then $\mathcal{M}_0$ is symmetrically embedded in $\mathcal{M}$.*

PROOF. Let $\sigma \in \mathrm{Aut}(\mathcal{M}_0)$. Assume, first, that $\mathcal{M}_0$ is a fibre. Define $\tau : \mathcal{M}[\mathcal{M}]$ by

$$\tau(a) = \begin{cases} a & \text{if } a \notin \mathcal{M}_0, \\ \sigma(a) & \text{if } a \in \mathcal{M}_0. \end{cases}$$

Then, by the definition of $\mathcal{M}[\mathcal{M}]$ it follows that τ preserves all relations, and being bijective, we are done. So it remains to treat the case where $\mathcal{M}_0$ is a section.

In that case, let $f : \mathcal{M} \to \mathcal{M}$ be the function defining the section. We define $\tilde{\sigma} : \mathcal{M} \to \mathcal{M}$ by $\tilde{\sigma}(a) = \pi_1(\sigma(a, f(a)))$, where π_1 is the projection onto the first coordinate. Let $(a, f(a)) \in \mathcal{M}_0$, then — by transitivity — there exists $\sigma_a \in \mathrm{Aut}(\mathcal{M})$ such that $\sigma_a(f(a)) = \pi_2(\sigma(a, f(a))$, where π_2 is the projection onto the second coordinate. Now we define:

$$\tau(a, b) = (\tilde{\sigma}(a), \sigma_a(b)).$$

We claim that τ satisfies the requirements. First, by definition, for $(a, f(a)) \in \mathcal{M}_0$ we have $\pi_1\tau(a, f(a)) = \tilde{\sigma}(a) = \pi_1(\sigma(a, f(a)))$ and $\pi_2\tau(a, f(a)) = \sigma_a(f(a)) = \pi_2(\sigma(a, f(a)))$, so τ extends σ. To see that $\tau \in \mathrm{Aut}(\mathcal{M}[\mathcal{M}])$ let R be an n-ary relation and $\bar{a}_i := (a_i, b_i)$, $1 \leq i \leq n$ such that $\mathcal{M}[\mathcal{M}] \models R(\bar{a}_1, \ldots, \bar{a}_n)$. Observe that either $a_i = a_j := a$ or $a_i \neq a_j$ for all $i \neq j \leq n$ (and by transitivity $n > 1$). In the former case $\tau(\bar{a}_i) = (\tilde{\sigma}(a), \sigma_a(b_i)))$, and as $\sigma_a \in \mathrm{Aut}(\mathcal{M})$ the definition of the composition gives $\mathcal{M}[\mathcal{M}] \models R(\tau(\bar{a}_1, \ldots, \bar{a}_n))$. In the latter case, $\pi_1\tau(\bar{a}_i) \neq \pi_1\tau(\bar{a}_j)$ for all $i \neq j$ and, again — since $\tilde{\sigma} \in \mathrm{Aut}(\mathcal{M})$ the definition of the composition implies $\mathcal{M}[\mathcal{M}] \models R(\tau(\bar{a}_1), \ldots, \tau(\bar{a}_2))$. □

Summing up all of the above we get:

COROLLARY 2.9. *Let $\mathcal{M}$ be a transitive symmetrically self similar structure, then $\mathcal{M}$ is symmetrically indivisible.*

PROOF. Apply Lemma 2.5 with a symmetrically embedded copy W of $\mathcal{M}[\mathcal{M}]$. Then apply the last lemma and the transitivity of the property of being symmetrically embedded. □

As an application we get, using Example 2.7:

COROLLARY 2.10. *The structure $(\mathbb{Q}, \leq)$ is symmetrically self-similar.*

The same argument can be used to show that:

COROLLARY 2.11. *The countable universal n-uniform hypergraph, $\mathcal{G}_n$, is symmetrically indivisible for all $n \geq 2$.*

The case $n = 2$ is done in **[9]** and the general case follows the same lines. Since any countable n-hypergraph embeds in $\mathcal{G}_n$ we readily get that $\mathcal{G}_n$ is self-similar. It will suffice, therefore, to show:

LEMMA 2.12. *Any countable n-uniform hypergraph embeds symmetrically into $\mathcal{G}_n$.*

PROOF. For a set A let us denote $F_k(A) := [A]^k$. Let G_n^0 be any countable n-hypergraph and G_n^{m+1} the n-hypergraph obtained by adjoining to G_n^m vertices $\{a_I : I \in [F_{n-1}(G_n^m)]^{<\omega}\}$ such that $G(b_1, \ldots, b_{n-1}, b_n)$ holds if and only if one of the two following options hold:

(1) $b_i \in G_n^m$ for all $1 \leq i \leq n$ and $G(b_1, \dots, b_{n-1}, b_n)$ is satisfied in G_n^m, or
(2) $b_n \notin G_n^m$, in which case $b = a_I$ for some $I \in [F_{n-1}(G_n^m)]^{<\omega}$ and $\{b_1, \dots, b_{n-1}\} \in I$.

Finally set $\mathcal{G} := \bigcup_{m \in \omega} G_n^m$. We claim that $\mathcal{G} \cong \mathcal{G}_n$ and that G_n^0 is symmetrically embedded in $\mathcal{G}$. To show that $\mathcal{G} \cong \mathcal{G}_n$ it suffices, of course to show that $\mathcal{G} \equiv \mathcal{G}_n$, since the latter is $\aleph_0$-categorical. But $\mathrm{Th}(\mathcal{G}_n)$ is axiomatized by stating that it is an n-hypergraph and that for all k_1, k_2 and distinct $\bar{a}_1, \dots \bar{a}_{k_1}$, $\bar{b}_1, \dots, \bar{b}_{k_2}$ all in $F_{n-1}(\mathcal{G}_n)$ there is an element d such that

$$(*) \qquad \bigwedge_{i=1}^{k_1} G(\bar{a}_i, d) \wedge \bigwedge_{i=1}^{k_2} \neg G(\bar{b}_i, d).$$

So assume that $\bar{a}_1, \dots \bar{a}_{k_1}$, $\bar{b}_1, \dots, \bar{b}_{k_2}$ all in $F_{n-1}(\Gamma)$ are distinct elements. Let $m \in \mathbb{N}$ be such that $\bar{a}_i, \bar{b}_j \subseteq G_n^m$ for all $1 \leq i \leq k_1$ and $1 \leq j \leq k_2$. Let $I = \{\bar{a}_1, \dots, \bar{a}_{k_1}\}$ and let $d := a_I \in G_n^{m+1}$. Then a_I witnesses $(*)$. To see that G_n^0 is symmetrically embedded in $\mathcal{G}$ it is enough to check that, in fact, given any isomorphism $\tau : G_n^m \to G_n^m$, τ can be extended to an automorphism of G_n^{m+1}. But this is obvious, since τ induces an automorphism of $(F_{n-1}(G_n^m), \subseteq)$ (which, for simplicity of notation, we will also denote τ), and it is clear by definition that extending τ by $a_I \mapsto a_{\tau(I)}$ for all $I \in F_{n-1}(G_n^m)$ we get an automorphism of G_n^{m+1}. □

We conclude this section by pointing out that the composition of structures allows us to construct new (symmetrically) indivisible structures from old ones. First, we slightly generalize the notion of composition:

Definition 2.13. Let $\mathcal{M}$, $\mathcal{N}$ be structures in a relational language, $\mathcal{L}$. The composition of $\mathcal{N}$ with $\mathcal{M}$, denoted $\mathcal{M}[\mathcal{N}]$ is the $\mathcal{L}$-structure whose universe is $M \times N$ where for an n-ary relation $R \in \mathcal{L}$ we set

$$\begin{aligned}
&R(M[N]) := \\
&\{((a, a_1), \dots, (a, a_k)) : a \in M, M \models R(a_1, \dots, a_k)\} \cup \\
&\{((a_{1,1}, a_{1,2}), \dots, (a_{k,1}, a_{k,2})) : \bigwedge_{i \neq j} (a_{i,1} \neq a_{j,1}), N \models R(a_{i,1}, \dots, a_{j,1})\}
\end{aligned}$$

Let $\mathcal{M}[\mathcal{N}]^s$ be $\mathcal{M}[\mathcal{N}]$ expanded by a new equivalence relation E_s interpreted as $\{((a, b), (a, c)) : a \in M, b, c \in N\}$.We call E_s the *standard equivalence relation* on the composition.

The following is now easy:

Proposition 2.14. *Let $\mathcal{N}$, $\mathcal{M}$ be countable structures in a Δ-free relational language $\mathcal{L}$. If $\mathcal{N}$ and $\mathcal{M}$ are both (symmetrically) indivisible (and the number of symmetrically embedded copies of $\mathcal{N}$ in $\mathcal{N}$ is finite up to conjugacy in* $\mathrm{Aut}(\mathcal{N})$*), then $\mathcal{M}[\mathcal{N}]^s$ is (symmetrically) indivisible.*

We remark that in the next Section it will be shown that the number of symmetrically embedded copies of $\mathbb{Q}$ in $\mathbb{Q}$ is finite. Thus, by this theorem, $\Gamma[\mathbb{Q}]^s$ is symmetrically indivisible, for example.

PROOF. Let $c : M \times N \to \{0,1\}$ be any coloring. For each $a \in M$ denote $N_a = \{(a,b) : b \in N\}$ and $c_a : N \to \{0,1\}$ the coloring induced by c (i.e. $c_a(b) = c(a,b)$). Since N_a, with the induced structure, is isomorphic to $\mathcal{N}$ (through the natural projection $(a,b) \mapsto b$), we identify N_a with $\mathcal{N}$. We define a coloring $\tilde{c}$ of M as follows: For $a \in M$ define $\tilde{c}(a) = (0,i)$ if there is a (symmetric) $N'_a \subseteq N_a$ such that $c_a(N'_a) = \{0\}$ (and the conjugacy type of N'_a is the i-th in the finite list of conjugacy types of symmetrically embedded copies of $\mathcal{N}_n$ in itself). N'_a with the induced structure is isomorphic to $\mathcal{N}$ (and symmetrically embedded in N_a). Otherwise, let $c(s) = (1,j)$, when there is a 1-monochromatic copy of $\mathcal{N}$ (with the j-th conjugacy type in the finite list of types).

Because $\mathcal{M}$ is (symmetrically) indivisible we can find a $\tilde{c}$-monochromatic $M_0 \subseteq M$ such that, with the induced structure, $M_0 \cong \mathcal{M}$ (and M_0 is symmetrically embedded in $\mathcal{M}$). If $\tilde{c}(M_0) = \{(0,i)\}$ then, by definition, for all $a \in M_0$ we can find $N'_a \subseteq N_a$ such that $c(N_a) = \{0\}$, $N'_a \cong \mathcal{N}$ (and symmetrically embedded in N_a of conjugacy type i). Otherwise, because $\mathcal{N}$ is (symmetrically) indivisible we can find for each a some $N'_a \subseteq N_a$ such that $N'_a \cong \mathcal{N}$, $c(N_a) = \{1\}$ (and N_a is symmetrically embedded of conjugacy type j).

In any case we can define $W^c = \bigcup\{N'_a : a \in M_0\}$. Obviously W^c is c-monochromatic, and by construction $W^c \cong \mathcal{M}_0[\mathcal{N}] \cong \mathcal{M}[\mathcal{N}]$. So it remains only to verify that if $\mathcal{M}$ and $\mathcal{N}$ are both symmetrically indivisible every automorphism of W^c extends to $\mathcal{M}[\mathcal{N}]$.

Let $\alpha \in \mathrm{Aut}(W^c)$. So α induces, through the projection on the first coordinate, an automorphism α_1 on $\mathcal{M}_0$, which by symmetry of $\mathcal{M}_0$ in $\mathcal{M}$ we can extend, call the extension α_1, to the whole of $\mathcal{M}$. On the other hand, since α preserves the E_s structure, each fibre N'_a of W^c maps through α to the fibre $N'_{\alpha_1(a)}$. By construction, there is an automorphism, α_2^a of $\mathcal{N}$ such that $N'_a = N'_{\alpha_1(a)}$. So we can finally define $(\alpha_1 \times \alpha_2) : \mathcal{M}[\mathcal{N}]^s \to \mathcal{M}[\mathcal{N}]^s$ by:

$$(\alpha_1 \times \alpha_2)(a,b) = \begin{cases} (\alpha_1(a), b) & \text{if } a \notin \mathcal{M}_0, \\ (\alpha_1(a), \alpha_2^a(b)) & \text{if } a \in \mathcal{M}_0. \end{cases}$$

Then $(\alpha_1 \times \alpha_2) \in \mathrm{Aut}(\mathcal{M}[\mathcal{N}])$, and since $\alpha_1 \times \alpha_2$ extends α, this is what we needed.
. $\square$

Observe that in the last proof the definability of the standard equivalence relation was only needed to prove the symmetric indivisibility of $\mathcal{M}[\mathcal{N}]$. It may be worth pointing out that in many cases, the standard equivalence relation E_s is definable (without parameters) in the composition. Observe, for example, that if we expand $\mathcal{M}$ by a binary relation $R(x,y)$ interpreted as equality then for the resulting structure $\mathcal{M}_=$ we have that $\mathcal{M}_=[\mathcal{N}]$ is $\mathcal{M}[\mathcal{N}]^s$.

REMARK 2.15. Formally, we have only defined the composition $\mathcal{M}[\mathcal{N}]$ for structures in a common relational language. The general case (for relational languages) can be obtained by interpreting every relation in $\mathcal{L}(\mathcal{N}) \setminus \mathcal{L}(\mathcal{M})$ as the empty relation in $\mathcal{M}$, and the other way around.

Finally, we point out that the converse of Proposition 2.14 is also true. I.e., if $\mathcal{M}$, $\mathcal{N}$ are structures in a relational language $\mathcal{L}$ and $\mathcal{M}[\mathcal{N}]^s$ is symmetrically indivisible then so are $\mathcal{M}$ and $\mathcal{N}$.

3. A new approach to symmetric indivisibility: density of symmetrically embedded structures

As we have seen in the previous section, $(\mathbb{Q}, \leq)$ is symmetrically indivisible because the composition of $\mathbb{Q}$ with itself is isomorphic to $\mathbb{Q}$. That proof, however, does not reveal the full nature of the situation. In fact, we will show that in $(\mathbb{Q}, \leq)$ symmetry and indivisibility can be viewed as independent phenomena.

The main result of this section is the *density of symmetrically embedded copies of* $\mathbb{Q}$ *in* $\mathbb{Q}$. Symmetric indivisibility follows directly from indivisibility and density of symmetrically embedded copies.

Definition 3.1. Let $\mathcal{M}$ be a (countable) structure. Say that the symmetrically embedded copies of $\mathcal{M}$ in itself are *dense* if for any $\mathcal{N} \leq \mathcal{M}$ which is isomorphic to $\mathcal{M}$ there exists $\mathcal{N}' \leq \mathcal{N}$ such that $\mathcal{N}'$ is isomorphic to $\mathcal{M}$ and $\mathcal{N}'$ is symmetrically embedded in $\mathcal{M}$.

In fact we obtain a stronger density result: we prove that there are 36 types of symmetrically embedded copies of $(\mathbb{Q}, \leq)$ in $\mathbb{Q}$ and that four of these types are dense.

We conclude this section with a study of the theory of pairs $(\mathbb{Q}, P)$, where P is a symmetrically embedded submodel of $\mathbb{Q}$ of a certain type.

Our proof of the density of the symmetrically embedded copies of $\mathbb{Q}$ is a construction. In order to get a picture of the direction our construction has to take we begin by analyzing the possible symmetrically embedded copies of $\mathbb{Q}$ up to conjugacy with respect to $\mathrm{Aut}(\mathbb{Q}, \leq)$. The first important structural result is Corollary 3.7. This part of the work culminates in Theorem 3.9, giving the full list of all 36 possible types of symmetrically embedded copies of $\mathbb{Q}$ in $\mathbb{Q}$.

We start with some terminology. Suppose $\mathcal{N} \leq \mathbb{Q}$ is a copy of $\mathbb{Q}$. Every real number $r \in \mathbb{R} \setminus \mathcal{N}$ satisfies over $\mathcal{N}$ one of the following *types*:

(1) The lower bound $LB(x) = \{x < q : q \in \mathcal{N}\}$.
(2) The upper bound $UB(x) = \{q < x : q \in \mathcal{N}$.
(3) The left infinitesimal of a point $q \in \mathcal{N}$, $L(q, x) = \{q < x\} \cup \{q' < x : q' \in \mathcal{M} \wedge q < q'\}$.
(4) The right infinitesimal of a point $q \in \mathcal{N}$, $R(q, x) = \{q < x\} \cup \{x < q' : q' \in \mathcal{M} \wedge q < q'\}$.
(5) The cut $(L, R, x) = \{q < x \wedge x < p : q \in L, p \in R\}$ over $\mathcal{N}$, where $L \cup R = \mathcal{N}$, $L \cap R = \emptyset$, $L \neq \emptyset \neq R$ and $p < q$ for all $p \in L$, $q \in R$.

If $\sigma \in \mathrm{Aut}(\mathcal{N})$ and t is a type, $\sigma(t)$ is gotten from t by replacing every parameter p in the type by $\sigma(p)$. Two types t and t' over $\mathcal{N}$ are *conjugate* if there is $\sigma \in \mathrm{Aut}(\mathcal{N})$ such that $\sigma(t) = t'$.

Claim 3.2. *Suppose $\mathcal{N} \leq \mathbb{Q}$ is a copy of $\mathbb{Q}$. Two types over $\mathcal{N}$ are conjugate if and only if they are of the same kind in the list (1)–(5) above.*

Proof. Direct inspection shows that two types of different kinds are not conjugate. If $p, q \in \mathcal{N}$ then any automorphism of $\mathcal{N}$ which takes p to q takes $L(p, x)$ to $L(q, x)$ and $R(p, x)$ to $R(q, x)$. Given two cuts (L, R, x) and (L', R', x), since the order types of all four sets is $\mathbb{Q}$, there is an automorphism of $\mathcal{N}$ carrying L to L' and R to R'. $\square$

A type $t(x)$ is *realized* by a real $r \in \mathbb{R}$ iff r satisfies all formulas in t when substituted for x. For a copy $\mathcal{N} \leq \mathbb{Q}$ of $\mathbb{Q}$ it is possible for each of the types to be realized by a member of $\mathbb{Q}$. The type $LB(x)$ may be realized by a greatest rational number or there may not exist a greatest rational realization of it, and similarly for $UB(x)$.

Corollary 3.3. *If $\mathcal{M} \leq \mathbb{Q}$ is symmetrically embedded in $\mathbb{Q}$ and a type t over $\mathcal{M}$ is realized by some element of $\mathbb{Q}$, then all types of the same kind of t are realized in $\mathbb{Q}$.*

Proof. If t, t' are types of the same kind over a symmetrically embedded copy $\mathcal{M} \leq \mathbb{Q}$ and $q \in \mathbb{Q} \setminus \mathcal{M}$ realizes t, fix $\sigma \in \operatorname{Aut}(\mathcal{M})$ which carries t to t' and an extension $\sigma \subseteq \bar{\sigma} \in \operatorname{Aut}(\mathbb{Q})$. Now $\bar{\sigma}(q)$ satisfies t'. □

A cut (L, R) over a copy $\mathcal{N} \leq \mathbb{Q}$ may be realized by a single real number. If r realizes a cut over $\mathcal{N}$ and is the unique real number which realizes this cut, we call it a two-sided limit of $\mathcal{N}$. So r is a two-sided limit of $\mathcal{N}$ if $r \notin \mathcal{N}$ and $r = \inf\{q \in \mathcal{N} : q > r\} = \sup\{q \in N : q < r\}$.

Claim 3.4. *Suppose $\mathcal{N} \leq \mathbb{Q}$ is a copy of $\mathbb{Q}$. If $U \subseteq \mathbb{R}$ is an open interval and $|N \cap U| > 1$, there are continuum many irrationals $r \in U$ which are two-sided limits of N.*

Proof. Since $N \cap U$ contains a copy of $\mathbb{Q}$ there are continuum many cuts over $N \cap U$. If a cut over $N \cap U$ is realized by more than one real number, it is realized by a rational. Therefore, except for a countable set of cuts, every cut over $N \cap U$ is realized by a single irrational number. □

By this, almost all cuts over any copy $\mathcal{M} \leq \mathbb{Q}$ are not realized in $\mathbb{Q}$. Since realizing one cut over a symmetrically embedded copy $\mathcal{M} \leq \mathbb{Q}$ implies, by Claim 3.3, realizing all, we conclude:

Corollary 3.5. *If $\mathcal{N} \leq \mathbb{Q}$ is a symmetrically embedded copy of $\mathbb{Q}$ then no cuts over $\mathcal{M}$ are realized in $\mathbb{Q}$.*

Except for cuts, however, types over $\mathcal{N}$ which are realized by reals, are also realized by rational numbers. A convex subset of the rationals is a subset A that contains $\mathbb{Q} \cap (q, p)$ whenever $q < p$ belong to A.

Claim 3.6. *Suppose $\mathcal{N} \leq \mathbb{Q}$ is a copy of $\mathbb{Q}$.*

(1) *If the left [right] infinitesimal type $L(q, x)$ [$R(q, x)$] of some $q \in \mathcal{N}$ is realized by some real, then the set of rational numbers which realize it is a convex copy of $\mathbb{Q}$.*
(2) *If the type $LB(x)$ is realized by some real, then the set of rationals which realize it is either a convex copy of $\mathbb{Q}$, in case this set has no maximal element, or a convex copy of $\mathbb{Q} \cap (-\infty, 0]$, in case this set contains a maximal element. Similarly for $UB(x)$.*

Proof. Obvious. □

Corollary 3.7. *A copy $\mathcal{M} \leq \mathbb{Q}$ is symmetrically embedded if an only if no cuts over $\mathcal{M}$ are realized in $\mathbb{Q}$ and exactly one of the following conditions holds:*

- *For all $q \in \mathcal{M}$ neither $L(q)$ nor $R(q)$ are realized in $\mathbb{Q}$ (in which case $\mathcal{M}$ is an interval in $\mathbb{Q}$).*

- *For all $q \in \mathcal{M}$ both $L(q)$ and $R(q)$ are realized in $\mathbb{Q}$.*
- *For all $q \in \mathcal{M}$ the type $L(q)$ is realized in $\mathbb{Q}$ but $R(q)$ is not realized in $\mathbb{Q}$.*
- *For all $q \in \mathcal{M}$ the type $L(q)$ is not realized in $\mathbb{Q}$ but $R(q)$ is realized in $\mathbb{Q}$.*

In particular, a copy $\mathcal{M} \leq \mathbb{Q}$ of $\mathbb{Q}$ is symmetrically embedded if and only if for every $a < b$ in $\mathcal{M}$ the interval $(a, b) \cap M$ is symmetrically embedded in $\mathbb{Q}$.

Proof. The fact that no cuts are realized over a symmetrically embedded copy has been proved above. Suppose $\mathcal{M} \leq \mathbb{Q}$ is a copy with no cuts realized in $\mathbb{Q}$ and one of the four possibilities in the claim holds. In the first case, any automorphism of $\mathcal{M}$ can be extended as the identity on $\mathbb{Q} \setminus \mathcal{M}$. In the other three, if $\sigma \in \mathrm{Aut}(M)$ and $\sigma(q) = p$ the set of left [right] infinitesimals of q can be carried to the set of left [right] infinitesimals of p by an order isomorphism. $\square$

Let us remark that something stronger holds for symmetrically embedded copies $\mathcal{M} \leq \mathbb{Q}$. We can find a *homomorphism* $\varphi : \mathrm{Aut}(\mathcal{M}) \to \mathrm{Aut}(\mathbb{Q})$ so that $\sigma \subseteq \varphi(\sigma)$ for all $\sigma \in \mathrm{Aut}(\mathcal{M})$. Extend via the identity above $\sup \mathcal{M}$ and below $\inf \mathcal{M}$. If right rational infinitesimals exist for members of $\mathcal{M}$, fix some $q_0 \in \mathcal{M}$ and fix an automorphisms φ_p^r from the set right rational infinitesimals of q to the set of right rational infinitesimal of p for each $p \in \mathcal{M}$ (with $\varphi_{q_0}^r = \mathrm{id}$) and extend each σ by moving the right infinitesimals of p to the right infinitesimals of $\sigma(p)$ via $\varphi_{p'}^r \circ (\varphi_p^r)^{-1}$. Similarly for left rational infinitesimals. It is easy to check that $\varphi(\sigma \circ \tau) = \varphi(\sigma) \circ \varphi(\tau)$ for $\sigma, \tau \in \mathrm{Aut}(M)$.

We get, then:

Corollary 3.8. *A copy $\mathcal{M} \leq \mathbb{Q}$ is symmetrically embedded in $\mathbb{Q}$ if an only if there is a homomorphism $\varphi : \mathrm{Aut}(\mathcal{M}) \to \mathrm{Aut}(\mathbb{Q})$ satisfying $\sigma \subseteq \varphi(\sigma)$ for all $\sigma \in \mathrm{Aut}(\mathcal{M})$.*

We now get:

Theorem 3.9. *Up to conjugacy in $\mathrm{Aut}(\mathbb{Q}, \leq)$ there are exactly 36 symmetrically embedded copies of $\mathbb{Q}$.*

Proof. There are 3 possibilities concerning $LB(x)$ over a copy $\mathcal{M} \leq \mathbb{Q}$: not realized, realized with a rational greatest realization and realized with no greatest rational realization, and similarly 3 independent possibilities for $UB(x)$. There are 4 independent possibilities concerning the realization of left and right infinitesimals. Any two symmetrically embedded copies are conjugate if and only if they satisfy the same possibility out of the potential 36 described.

In fact, all 36 possibilities do occur. $\mathbb{Q}$ itself is a copy of $\mathbb{Q}$ in $\mathbb{Q}$ with no upper or lower bounds and with neither left nor right rational infinitesimals. Upper and lower bounds can be realized with or without least upper/ greatest lower one.

So it only remains to check that only left, only right and both sides of rational infinitesimals can occur over symmetrically embedded copies. Let $\mathbb{Q}$ be presented as $\mathbb{Q}[\mathbb{Q}]$ and let $N_0 = \{(q, 0) : q \in \mathbb{Q}\}$, $N_1 = \{(q, p) : p \in \mathbb{Q} \wedge p \geq 0\}$ and $N_2 = \{(q, p) : p \in \mathbb{Q} \wedge p \leq 0\}$. These copies of $\mathbb{Q}$ are symmetrically embedded in $\mathbb{Q}$ and of the three required conjugacy types. $\square$

We proceed to proving the density result claimed at the beginning of this section. Our main technical tool is:

THEOREM 3.10. *For every substructure* $\mathcal{N} \leq \mathbb{Q}$ *isomorphic to* $\mathbb{Q}$ *there is an isomorphism* $\varphi : \operatorname{conv}(\mathcal{N}) \to \mathbb{Q}[\mathbb{Q}]$ *such that* $\varphi^{-1}[\mathcal{N}] \cap \{(q,p) : p \in \mathbb{Q}\} \cong \mathbb{Q}$ *for every* $q \in \mathbb{Q}$.

PROOF. Suppose that $\mathcal{N} \subseteq \mathbb{Q}$ is order isomorphic to $\mathbb{Q}$. By replacing $\mathbb{Q}$ with the convex hull $\operatorname{conv}(\mathcal{N})$ we may assume that $\operatorname{conv}(\mathcal{N}) = \mathbb{Q}$, for simplicity of notation.

Abusing notation, let us write $I < J$ for disjoint nonempty intervals of $\mathbb{R}$ if every point in I is smaller than any point in J, that is, we shall use $<$ for both the real ordering and the *interval ordering.* Let us fix an enumeration $\{q_n : n \in \mathbb{N}\}$ of $\mathbb{Q}$.

It will suffice to present $\mathbb{Q}$ as a disjoint union of intervals $\bigcup_n I_n$ such that:

(i) The mapping $q_n \mapsto I_n$ is an order isomorphism between the rational order and the interval order on $\{I_n : n \in \omega\}$;

(ii) $I_n \cap \mathcal{N}$ is isomorphic to $\mathbb{Q}$ for each n.

We define I_n by induction so that:

(1) $I_n = (a_n, b_n)$, $a_n < b_n$ are irrational numbers each of which is a two-sided limit of N, hence $\mathcal{N} \cap (a_n, b_n)$ is isomorphic to $\mathbb{Q}$.

(2) $m < n$ implies $I_n \cap I_m = \emptyset$.

(3) The type of I_{n+1} over $\{I_m : m \leq n\}$ with respect to the interval ordering is equal to the type of q_{n+1} over $\{q_m : m \leq n\}$ with respect to the rational ordering.

(4) If q is the first rational number (in some fixed enumeration of $\mathbb{Q}$) outside $\bigcup_{m \leq n} I_m$ with the type of $\{q\}$ over $\{I_m : m \leq n\}$ equal to the type of q_{n+1} over $\{q_m : m \leq n\}$ then $q \in I_{n+1}$.

To carry the induction let $I_0 = (a_0, b_0)$ be an arbitrary interval which satisfies (1). At step $n+1$ let q be the least rational in the enumeration which satisfies the hypothesis in (4). Find $a < a_{n+1} < q < b_{n+1} < b$, such that both a_{n+1} and b_{n+1} are two-sided limits of N satisfying the same type as q over $\{I_m : m \leq n\}$.

To prove that $\bigcup_n I_n = \mathbb{Q}$ suppose this is not so and let q be the least in the enumeration of $\mathbb{Q}$ which is outside this union. At some stage n_0 it holds that q is the least in the enumeration outside of $\bigcup_{m \leq n_0} I_m$. For some $n \geq m_0$ the type of $\{q\}$ over $\{I_m : m \leq n\}$ coincides with that of q_{n+1} over $\{q_m : m \leq n\}$ hence $q \in I_{n+1}$ contrary to our assumption. □

THEOREM 3.11. *For every copy* $\mathcal{M} \leq \mathbb{Q}$ *of* $\mathbb{Q}$ *there exists a copy* $\mathcal{N} \leq \mathcal{M}$ *of* $\mathbb{Q}$ *which is symmetrically embedded in* $\mathbb{Q}$*, bounded from both sides and with both right and left rational infinitesimals for every* $q \in \mathcal{N}$*. The existence of a greatest upper bound and of least upper bound of* $\mathcal{N}$ *in* $\mathbb{Q}$ *can be chosen at will.*

PROOF. Given $\mathcal{M} \leq \mathbb{Q}$ fix, by Theorem 3.10, a presentation $\operatorname{conv}(\mathcal{M}) = \bigcup I_q$ where $I_p < I_q \iff p < 1$ and $I_q \cap \mathcal{M}$ is isomorphic to $\mathbb{Q}$. Choose $p(q) \in I_q \cap \mathcal{M}$ for each q and let $N = \{p(q) : q \in \mathbb{Q}\}$.

$\mathcal{N}$ can be intersected with an open, closed or half-closed half-open interval with end-points in $\mathbb{Q}$ to arrange the existence of a greatest lower bound and of a least upper bound. □

Recalling Definition 3.1, the previous theorem can be described as the fact that the symmetrically embedded copies of $\mathbb{Q}$ of 4 of the 36 types of symmetrically embedded copies are dense.

For the next Theorem see below Definition 5.5 of a "stably embedded substructure" and Corollary 5.6 immediately following it.

THEOREM 3.12. *For any coloring $c : \mathbb{Q} \to \{0,1\}$ there exists a monochromatic submodel $\mathcal{P} \leq \mathbb{Q}$ such that the pair $(\mathbb{Q}, \mathcal{P})$ is $\aleph_0$-categorical and $\mathcal{P}$ is stably embedded in $\mathbb{Q}$.*

PROOF. Let $c : \mathbb{Q} \to \{0,1\}$ be any coloring. Since $\mathbb{Q}$ is indivisible, we can find $\mathcal{P}_0 \leq \mathbb{Q}$ monochromatic and bounded from both sides. By Theorem 3.10 we can find $\mathcal{P} \leq \mathcal{P}_0$ symmetrically embedded in $\mathbb{Q}$ and a family of disjoint open intervals $\{I_p\}_{p \in \mathcal{P}}$ whose union is $\operatorname{conv} \mathcal{P}_0$ and such that $I_p \cap P = \{p\}$ for all $p \in P$. By Corollary 5.6 $\mathcal{P}$ is stably embedded in $\mathbb{Q}$.

It remains to check that $(\mathbb{Q}, \mathcal{P})$ is $\aleph_0$-categorical. Let Q denote the convex hull of P in $\mathbb{Q}$.

Claim The structure $(\mathcal{Q}, \mathcal{P})$ is $\aleph_0$-categorical.

PROOF. To define an isomorphism between any two countable models $(\mathcal{Q}, \mathcal{S})$ and $(\mathcal{Q}', \mathcal{S}')$ first choose an isomorphism of $\mathcal{Q}$ with $\mathcal{Q}'$. Since the conjugacy class of S in $\mathbb{Q}$ is given (and equals that of S' in $\mathcal{Q}'$), we can now find an automorphism of $\mathcal{Q}'$ taking the image of S to S', which is all we need. $\square_{claim}$

Now to construct an isomorphism between any two models of $(\mathbb{Q}, \mathcal{N})$ first use the claim to construct an isomorphism between the convex hull of the smaller model. Then observe that each of the types at $\pm\infty$ (over the predicate) is realized in one model if and only if it is realized in all models. Moreover, the set of realizations of each such type is isomorphic to $\mathbb{Q}$, if it is non-empty. So the partial isomorphism can be extended. $\square$

4. A symmetrically indivisible $\mathbb{Q}$-tree

Traditionally, induced Ramsey theory is concerned with relational languages. But structures in non-relational languages can still be indivisible. In this section we give a new example of such a structure. The main point of this new example is that once the proof of indivisibility is obtained, symmetric indivisibility can be inferred from the "density" results which concluded the previous section.

Consider the language $\mathcal{L} := \{\leq, \wedge\}$ and the theory T (see Remark 4.3 below) stating that:

- $\leq$ is a partial order whose lower cones are linear (i.e. the universe is a tree).
- $\leq$ is dense with no endpoints.
- $\wedge$ is a binary function such that $z = x \wedge y$ implies $z \leq x, z \leq y$ and if $z < w \leq x$ then $w \not\leq y$.
- For any element z there are infinitely many elements $\{x_i\}$ pairwise $\leq$-incomparable such that $z = x_i \wedge x_j$ for all $i \neq j$.

We will sometimes refer to the relation $a \geq b$ as "a is over b" or "a is above b". We will denote $a \perp_b c$ if a is incomparable to c and $b = a \wedge c$, we will say that a, c are *orthogonal* over b. We will also call a *branch* over b a maximal chain of elements lying over b. The following observation will be useful:

REMARK 4.1. For any natural number $n > 0$ let $\psi_n(z, x_0, \ldots, x_n)$ be the $\mathcal{L}$-formula stating that if $x_i \perp_z x_j$ for all $i \neq j$ then there exists $y > z$ such that $y \wedge x_i = z$ for all $i \leq n$. Then $T \models (\forall z \forall x_1, \ldots x_n)\psi_n$ for all n.

PROOF. Let $\mathcal{M} \models T$ be any model. Let $\{x_1, \ldots, x_n\}$ and z, be elements in M as in the statement. Let $\{y_i\}_{i=0}^{n+1}$ in M be such that $y_i \wedge y_j = z$ for all $i \leq n+1$ (use the last axiom). Observe that for all $0 \leq i \leq n$ there exists at most one $0 \leq j \leq n+1$ such that $x_i \wedge y_j > z$. But then there exists $j \leq n+1$ such that $y_j \wedge x_i = z$ for all $0 \leq i \leq n$, as required. □

REMARK 4.2. Let $\{a_i\}_{i \in I}$ be a set of elements pairwise orthogonal over b and let $\{c_i\}_{i \in I}$ be such that $c_i \geq a_i$ for all $i \in I$. Then $\{c_i\}_{i \in I}$ is a set of pairwise orthogonal elements over b.

REMARK 4.3. The theory T is consistent.

PROOF. We will construct a model for T in $\mathbb{Q}^{\mathbb{Q}}$. For a function $f \in \mathbb{Q}^{\mathbb{Q}}$ and $q \in \mathbb{Q}$ denote $f^{\downarrow q} := f|(-\infty, q]$. Let $\mathcal{F} := \{f^{\downarrow q} : f \in \mathbb{Q}^{\mathbb{Q}}, q \in \mathbb{Q}\}$. Let $\leq$ be the natural partial order on $\mathcal{F}$ given by $f \leq g$ if there exists $q \in \mathbb{Q}$ such that $f = g^{\downarrow q}$. Then $(\mathcal{F}, \leq)$ is a dense tree (or, rather, forest). Let $\mathcal{F}_0 := \{f \in \mathcal{F} : (\exists q) 0^{\downarrow q} \leq f\}$, where 0 is the constant function. Then $(\mathcal{F}_0, \leq)$ is a sub-tree. Say that $f \in \mathcal{F}_0$ is locally constant, if for all $r \in \mathbb{R}$ there exist rational numbers $a < r \leq b$ such that $f|(a, b]$ is constant. Let $\mathcal{F}_1 := \{f \in \mathcal{F}_0 : f \text{ is locally constant}\}$. The structure $(\mathcal{F}_1, \leq)$ will be our model (because the function $\wedge$ is definable in T from $\leq$, this suffices to determine the $\mathcal{L}$-structure). We will now verify that $\wedge$ is well defined in $(\mathcal{F}_1, \leq)$. Let $f, g \in \mathcal{F}_1$ and let $q(f, g) := \sup\{q \in \mathbb{Q} : f(q) = g(q)\}$. Note, first, that since $f, g \in \mathcal{F}_0$ we know that $q(f, g) \in \mathbb{R}$ (i.e. is not $-\infty$). Because f, g are locally constant, it follows that $q(f, g) \in \mathbb{Q}$. So we define $f \wedge g = f^{\downarrow q(f,g)}$. Verifying that $(\mathcal{F}_1, \leq, \wedge) \models T$ is now a triviality. □

REMARK 4.4. The theory T is $\aleph_0$-categorical and has quantifier elimination in the language $\mathcal{L}$.

PROOF. To see this, observe first that if $\mathcal{M} \models T$ and $A \subseteq M$ is finite then $\langle A \rangle$ — the substructure generated by A is finite. The proof is by induction on (n, m) — the number and length of the longest anti-chain in A. For let A be a set with n anti-chains of maximal length m. Let a be any member of an anti-chain of maximal length, and $a_1 < a$ be maximal in (the linearly ordered set) $\{a \wedge b : b \in A \setminus \{a\}\}$. Observe that $\langle A \rangle = \langle A_1 \rangle \cup \{a\}$ where $A_1 = A \cup \{a_1\} \setminus \{a\}$. But A_1 has fewer anti-chains of length m than A did. So the conclusion follows from the inductive hypothesis.

Now, let $\mathcal{M}_1, \mathcal{M}_2 \models T$ be countable models and $f : A_1 \to A_2$ a partial isomorphism between finite substructures of $\mathcal{M}_1$ and $\mathcal{M}_2$ respectively. Let $a \in M_1$ be any element. We will show that f can be extended to a. If $a_1 > a$ for some $a_1 \in A_1$ let c_a be the cut of a in the branch below a_1. Observe that if $d \in A_1$ and $d \wedge a_1 < a$ then $d \wedge a = d \wedge a_1$ and if $d \wedge a_1 \geq a$ then $d \wedge a = a$. In any case, if $b \in \mathcal{M}$ realizes $f(c_a)$ then $\{(a, b)\} \cup f$ is the desired extension of f (note that such b exists because the branch below $f(a_1)$ is dense with no minimum). If $a \geq a_1$ for some $a_1 \in A_1$, a_1 is maximal such and a is not bounded above in A, let $A_0 := \{b \in A : a \wedge b = a_1\}$. A similar argument as before will show that if we can find $b \in M_2$ such that $b \perp_{f(a_1)} f(a_0)$ for all $a_0 \in A_0$ then $\langle A_1, a \rangle = A_1 \cup \{a\}$ and $f \cup \{(a, b)\}$ is an extension of f. But such an element b exists in M_2 by the last claim. So it remains to extend the function f in the case where a is incomparable to A. In that case, let $a' := \max\{a \wedge a_1 : a_1 \in A\}$. Extend the function first to a' and then to a using the two previous cases. □

Let $\mathcal{M} \models T$, a countable model. We will first show that $\mathcal{M}$ is indivisible. We split the proof in to several claims:

LEMMA 4.5. *Let $\mathcal{M} \models T$ and $c : M \to \{0, 1\}$ any coloring. Assume that for some point $a \in M$ there is no densely linearly ordered set $\{a_i\}_{i\in\omega}$ above a such that $c(a_i) = 1$ for all i, then for all $b > a$ there is a dense set $\{b_i\}_{i\in\omega}$ above b such that $c(b_i) = 0$ for all i.*

PROOF. The above is true for $(\mathbb{Q}, \leq)$ and is therefore true for M using the fact that it is a dense tree. □

LEMMA 4.6. *Let $\mathcal{M} \models T$. For any $b_2 > b_1$, $b_i \in M$ let*

$$G_{b_1,b_2} := \bigcup\{b \in M : (\exists c)(b_1 < c < b_2 \text{ and } b \geq c)\}.$$

Then G_{b_1,b_2} is a submodel of $\mathcal{M}$ (and by quantifier elimination and $\aleph_0$-categoricity isomorphic to $\mathcal{M}$).

PROOF. It is obvious that G_{b_1,b_2} is a dense tree. Moreover, for all $b \in G_{b_1,b_2}$ if $c > b$ then $c \in G_{b_1,b_2}$. So every $b \in G_{b_1,b_2}$ has infinite sets $\{c_i\}_{i=1}^{\infty}$ pairwise orthogonal over b. So, by quantifier elimination, it remains to check that G_{b_1,b_2} is a substructure: if $a, b \in G_{b_1,b_2}$ there are $c_a, c_b \in (b_1, b_2)$ such that $c_a \leq a$ and $c_b \leq b$. Then $a \wedge b \geq \min\{c_a, c_b\}$. And since the right hand side is in G_{b_1,b_2} so is the left hand side. □

The next lemma provides, given a countable model $\mathcal{M} \models T$ and a coloring of M in two colors, sufficient conditions for the existence of a monochromatic elementary substructure of $\mathcal{M}$:

LEMMA 4.7. *Let $\mathcal{M} \models T$ and $c : M \to \{0, 1\}$ any coloring. Assume that for every $a \in M$ there are $\{a_i\}_{i\in\omega}$ such that:*

(1) *$\{a_i\}_{i\in\omega}$ is a dense linearly ordered set above a.*
(2) *$c(a_i) = 1$ for all i.*

Then there exists $\mathcal{N} \leq \mathcal{M}$ such that $c(N) = \{1\}$.

PROOF. We construct $\mathcal{N}$ inductively as follows. Fix $f : \mathbb{N} \to \mathbb{N}$ such that $f^{-1}(n)$ is infinite for all $n \in \mathbb{N}$. This function will serve us in keeping track, at each stage of the construction, which element should be attended to. The requirement that $f^{-1}(n)$ is infinite for all n will assure that every element will be taken care of infinitely many times. Having said that, and in order to keep the notation cleaner, the function f will be suppressed from our notation. Fix $a_0 \in M$ such that $c(a_0) = 1$ and such that there is a dense linear order of elements $b < a_0$ with $c(b) = 1$. Now assume that we have constructed a finite substructure N_i such that:

(1) For $a_1 < a_2 \in N_i$ the points $b \in M$ such that $a_1 < b < a_2$ and $c(b) = 1$ form a dense linear order.
(2) $c(N_i) = \{1\}$.

We will now construct N_{i+1}. Let d be the first element of N_i (with respect to some fixed enumeration of M) which needs to be attended to (according to the function f) We proceed in two steps:

First, we fix some $a < d$ with $c(a) = 1$ as follows: if $d = \min\{x : x \in N_i\}$ we set $a' = -\infty$, otherwise let $a' < d$ be maximal in $\{x \in N_i : x < d\}$. It follows from (1) of the hypothesis that we can choose $a \in (a', d)$. so that $c(a) = 1$ and such that (1) remains true when extending N_i with a. Observe that if y is any element in N_i such that $y \wedge d \leq a'$, then $y \wedge a = y \wedge d$ and if $y \wedge d \geq d$ then $y \wedge a = a$. So setting $N^0_{i+1} = N_i \cup \{a\}$ we satisfy the inductive hypothesis.

Second, fix some $b \in M$ such that $b > d$. By Remark 4.1 we can choose b such that $b \wedge x \leq d$ for all $x \in N^0_{i+1}$ (and in particular $b \wedge a = a$). Moreover, if $b \wedge x < d$ for some $x \in N^0_{i+1}$ then $b \wedge x = b \wedge d$. By Lemma 4.2 and our assumption, we can choose b in such a way that, in addition, $c(b) = 1$. Observe that if $y \in N_i$ is an immediate successor (in N_i) of a (i.e., $N_i \cap (a, y) = \emptyset$) then $y \perp_a b$. So setting $N_{i+1} = N^0_{i+1} \cup \{b\}$, the inductive hypothesis is preserved for N_{i+1}.

Now let $N := \bigcup_{i\in\omega} N_i$. Then $\mathcal{N}$ (N with the induced structure) is an elementary substructure. To see this note, first, that $\mathcal{N}$ is indeed a substructure because all the N_i are, and it is obviously a tree. We have to show that it is a dense tree with no end points, and that the last axiom of T holds in $\mathcal{N}$.

That $\mathcal{N}$ is dense is easy. For let $a < b \in N$ and let i be such that $a, b \in N_i$. By the choice of the function f there will be some $i' \geq i$ such that b is the first element of $N_{i'}$ to be attended to. Therefore, in $N_{i'+1}$ there exists some $a < b' < b$, as provided by the first step of our construction. That $\mathcal{N}$ has no end points follows in a similar way.

To see that for all $a \in N$ there are $\{a_i\}_{i\in\omega}$ pairwise orthogonal over a we use the second step in the construction, combined with the fact that, as the N_i are substructures of M, if $a_1 \perp_a a_2$ in the sense of N_i (some i) then they are orthogonal in M, and therefore also in N_j for all $j \geq i$. ☐

The following is an immediate corollary of the lemma.

THEOREM 4.8. *Let $\mathcal{M} \models T$ be countable, then $\mathcal{M}$ is indivisible.*

PROOF. Let $c : M \to \{0, 1\}$ be any coloring. Because T is $\aleph_0$-categorical it will suffice to show that there exists a monochromatic $\mathcal{N} \leq \mathcal{M}$. By the last lemma, if for all $a \in M$ there exists a dense linear order of points greater than a colored 1, then we are done. So we may assume that there is a point $a \in M$ for which this is not the case. So by Lemma 4.5 for all $b > a$ there exists a dense linear order of points greater than b all colored 0. Using Lemma 4.6 with $b_2 > b_1 > a$ the desired conclusion follows, again, using the previous lemma. ☐

We will now start the analysis showing that $\mathcal{M} \models T$ countable is symmetrically indivisible. It will turn out that, in fact, the symmetric substructures of $\mathcal{M}$ are dense. The following claim is our main technical tool. It will provide criteria under which we can extend isomorphisms from sub-trees of $\mathcal{M}$ all the way to an automorphism of $\mathcal{M}$.

LEMMA 4.9. *Let $\mathcal{N}_1, \mathcal{N}_2 \leq \mathcal{M}$. And assume that the following hold:*

1. *For every $a_i \in N_i$ there are $c < a_i < d$ such that $N_i \cap (c, d) = \{a_i\}$.*
2. *For every $a < b \in N_i$ the interval $(a, b) \cap N_i$ is symmetrically embedded in (a, b).*
3. *There are $r_i \in M$ $(i = 1, 2)$ which are upper lower bounds for N_i (so that $r_i < N_i$ and given any $x < N_i$ we have $x \leq r_i$).*
4. *All branches in $\mathcal{N}_i$ are bounded in $\mathcal{M}$.*

(5) *If $C \subseteq \mathcal{M}$ is a maximal set of elements pairwise orthogonal over some $c \in N_i$ then $C \setminus N_i$ is infinite.*

Then any isomorphism $\sigma : \mathcal{N}_1 \to \mathcal{N}_2$ extends to an automorphism of $\mathcal{M}$.

Proof. Assume $\mathcal{N}_1, \mathcal{N}_2$ satisfy the assumptions. Observe that $r_1 < N_1$ as provided in the assumptions is unique, so given $\sigma : \mathcal{N}_1 \to \mathcal{N}_2$ a partial isomorphism, we first extend σ by $r_1 \mapsto r_2$. This is clearly a partial isomorphism, since $x \wedge r_i = r_i$ for all $x \in N_i$.

Claim 4.10. *Let $G_i := \{x \in M : r_i \leq x\}$ and assume that we can find $\tilde{\sigma} : G_1 \to G_2$, a partial isomorphism extending σ. Then we can extend σ to* $\mathrm{Aut}(M)$.

Proof. Since $(\mathcal{M}, r_1) \cong (\mathcal{M}, r_2)$ there is an isomorphism $\tau : (\mathcal{M}, r_1) \to (\mathcal{M}, r_2)$ with $\tau(r_1) = r_2$. Let $\mathcal{M}_i$ denote the induced structure on $M \setminus G_i$ expanded by a unary predicate P interpreted as $L_i := \{x \in M : x < r_i\}$, so that $\tau|_{\mathcal{M}_1}$ is an isomorphism from $\mathcal{M}_1$ to $\mathcal{M}_2$.

Now, notice that if $a \in G_i$ and $b \in M_i$ then a, b are $\leq$-comparable if and only if $b < a$ and $b \in L_i$, so clearly $\tau|_{\mathcal{M}_1} \cup \tilde{\sigma}$ preserves the order, which implies it must be a homomorphism since the join operator is definable from it. So $\tau|_{\mathcal{M}_1} \cup \tilde{\sigma} \in \mathrm{Aut}(M)$, as required. □

Thus, it will suffice to extend σ to G_i (as above). We will show that given $d \in G_1$ we can find $d' \in G_2$ such that $\langle N_1, d\rangle \cong \langle N_2, d'\rangle$ and both structures satisfy the assumptions satisfied by $\mathcal{N}_i$. This will suffice, as a standard back and forth will then complete the argument.

So let $d \in G_1$ be any element. We now distinguish some cases:

Case I: d is bounded above and below in N_1. Let $a, b \in N_1$ be such that $a < d < b$. Since $(a, b) \cap N_1$ is symmetric in (a, b) it follows, by Remark 3.7, that d realizes a non-cut over $(a, b) \cap N_1$. So there exists (a unique element) $c(d) \in (a, b) \cap N_1$ such that either $d \models c(d)^+$ or $d \models c(d)^-$; we will assume without loss of generality that $d \models c(d)^+$. By the assumption (1) on $\mathcal{N}_2$ we can find $d^+ \in N_2$ such that $d^+ \models \sigma(c(d))^+$ (recall the notation used in Remark 3.7). By definition, extending σ by sending d to d^+ is an order preserving map from $N_1 \cup \{d\}$ to $N_2 \cup \{d\}$.

Now, let $x \neq c(d)$ in N_1, and consider $x \wedge b$. If $x \wedge b \leq c(d) < d$ then $x \wedge d = x \wedge b$; otherwise $x > d$ and $x \wedge d = d$. In either case we have $\langle N_1, d\rangle = N \cup \{d\}$, so we are done.

Case II: d is bounded from below but not bounded from above in N_1. In this case $B_d := (r_1, d) \cap N_1$ is a branch in N_1 which implies that so is $\sigma(B_d)$ in N_2. Since no branch in N_2 is unbounded in $\mathcal{M}$ we can find d' such that $d' > \sigma(B_d)$.

Now, given any $x \in N_1$ either $x < d$ and $x \wedge d = x$ or (since d is unbounded from above in N_1) there is some $x' \in (r_1, d)$ incomparable with x. In that case $d \wedge x = x' \wedge x$. Thus, once again, $\langle N_1, d\rangle = N_1 \cup \{d\}$ and setting extending σ by sending d to d' is an extension of our isomorphism.

Case III: d is bounded from above but not bounded from below in N_1. We will prove that under our hypothesis this case is impossible. Let $b > d$. Since the elements under b form a linear order, it follows that either $d > r_1$ or $d \leq r_1$.

The latter case implies that $d \notin G_1$ contradicting our hypothesis. In the former case, because r_1 is an upper lower bound there is some $a \in N_1$ such that $a \nleq d$. In this case the $a \wedge b$ is an element in N_1 comparable to d (both are smaller than b) which by transitivity cannot be greater than d so it must be smaller than d, contradicting the non lower boundedness of d in N_1.

Case IV: d is not bounded from below or above in N_1. Let $C \subseteq N_1$ be a maximal set of elements pairwise independent (over r_1). Since $\sigma(C)$ is such a set in N_2 and $\sigma(C)$ is not maximal in M, we can find $d' \in N_2$ such that $d' \notin \sigma(C)$ and $\sigma(C) \cup \{d'\}$ is pairwise independent in N_2. It is easy to check that in this case extending σ by $d \mapsto d'$ is an isomorphism.

In all of the above cases, it follows easily that $\langle N_1, d\rangle$ satisfies all assumptions that $\mathcal{N}_1$ satisfied. So the proof is complete. □

So we are now ready to prove:

PROPOSITION 4.11. *Let $\mathcal{M} \models T$ be countable. Then the symmetric submodels of $\mathcal{M}$ are dense.*

PROOF. It will suffice to show that if $\mathcal{N} \leq \mathcal{M}$ there exists $\mathcal{N}' \leq \mathcal{N}$ such that the pair $(\mathcal{N}', \mathcal{M})$ satisfies the assumptions of Lemma 4.9 (with $\mathcal{N}_1 = \mathcal{N}_2 = \mathcal{N}'$). Notice that except for (3), all the assumptions of Lemma 4.9 are hereditary under the passage to sub-trees and given any $\mathcal{N}' \leq \mathcal{N}$ satisfying (1), (2), (4), and (5), the set $\{x \in N' \mid x > a\}$ is an elementary sub-tree of $\mathcal{N}'$ satisfying all the conditions of Lemma 4.9 for any $a \in \mathcal{N}'$. Therefore, it will suffice to show that for any given $\mathcal{N} \leq \mathcal{M}$ and one of the assumptions (1), (2), (4), or (5), of Lemma 4.9, there exists $\mathcal{N}' \leq \mathcal{N}$ satisfying that specific assumption.

To make notation easier, we will refer to the assumptions of Lemma 4.9 by their number without making explicit the lemma every time.

By the remark above we will assume that (3) holds for every tree that we refer to.

Our first goal is to find, for any given tree, a tree satisfying condition (5). Let $\{a_i\}_{i\in\omega}$ be an enumeration of N. Let $C_0 = \{a_0\}$ and $b_0 > a_0$ be any element. Let $F_0 := \{x \in N : x \geq b_0\}$. Observe that $N \setminus F_0 \cong N$. Assume that we have constructed

- A set C_i isomorphic to an initial segment of $\mathcal{N}$ (with respect to the above enumeration).
- A set F_i of *forbidden* elements such that $C_i \cap F_i = \emptyset$ and $\mathcal{N} \setminus F_i \leq \mathcal{N}$.

Let $p_{i+1} := \mathrm{tp}(a_0, \ldots, a_{i+1})$. Since $C_i \subseteq N \setminus F_i \cong \mathcal{N}$ we can find $c_{n+1} \in N \setminus F_i$ such that $\mathrm{tp}(C_i, c_{i+1}) = p_{i+1}$ and set $C_{i+1} = \langle C_i \cup \{c_{i+1}\}\rangle$. Since $N \setminus F_i$ is a substructure, $C_{i+1} \cap F_i = \emptyset$. If $c_{i+1} = \min C_{i+1}$ we set $F_{i+1} = F_i$. Otherwise, let $c \in C_{i+1}$ be such that $c < c_{i+1}$ and $(c, c_{i+1}) \cap C_{i+1} = \emptyset$ (i.e., c is the immediate predecessor in C_i of c_{i+1}). Let $b_{i+1} \in N \setminus F_i$ be any element such that $b_{i+1} \perp_c c_{i+1}$ and $b_{i+1} \wedge d \leq c$ for every $d \in C_{i+1}$. Let $F_{i+1} := F_i \cup G_{c,b_{i+1}}$. In any case $F_{i+1} \cap C_{i+1} = \emptyset$ and $N \setminus F_{i+1} \leq \mathcal{N}$.

Let $\mathcal{N}' := \bigcup_{i\in\omega} C_i$. By construction $\mathcal{N}' \leq \mathcal{N}$. Let $c \in \mathcal{N}'$ be any element, and $C \subseteq N'$ a maximal set of elements pairwise orthogonal over c. Let $i \in \omega$ be such that $c \in C_i$ and such that $c < c_{i+1}$ and $(c, c_{i+1}) \cap C_i = \emptyset$. Then $b_{i+1} \perp_c c'$ for any $c' \in C$. For assume that $c < d = c' \wedge b_{i+1}$ for some $c' \in C$. Then $d \in G_{d,b_{i+1}}$, implying that $c' \in F_{i+1}$, a contradiction. Since for every $c \in \mathcal{N}'$ there are infinitely many i with the above property, it follows that $\mathcal{N}'$ satisfies (5).

To find a sub-tree which satisfies (4) we pursue a similar (though slightly simpler) construction. We only modify the construction of the sets F_i at stage i as follows: for every maximal $c \in C_i$, we choose some $b(c) \in N \setminus F_{i-1}$ with $b(c) > c$ and set

$$F_i := F_{i-1} \cup \{x : (\exists c \in \max\{C_i\})(x > b(c))\}$$

and the rest is precisely as above.

So it remains to take care of (1) and (2). We proceed as follows. Fix some $b_0 \in N$ and consider the interval $(r, b_0) \cap N$ (where $r < N$ is the upper lower bound of N). By Theorem 3.11 we can find an elementary substructure of $(r, b_0) \cap N$ symmetric in $\mathcal{M}$. Moreover, the construction in the proof of Theorem 3.11 assures that this can be done so that (1) is satisfied. So we may assume that $(r, b_0) \cap N$ is symmetrically embedded in $\mathcal{M}$. For every $b \in (r, b_0) \cap N$ let $\{c_b^i\}_{i\in\omega} \subseteq N$ an infinite set of elements pairwise orthogonal over b. Repeat the process for every interval of the form (b, c_b^i). Proceeding inductively in this way and using the "in particular" clause of Corollary 3.7, we get the desired conclusion. $\square$

Summing up all of the above we get:

THEOREM 4.12. *Let $\mathcal{M} \models T$ be a countable model. Then $\mathcal{M}$ is symmetrically indivisible.*

5. Symmetric indivisibility of the generic triangle-free graph

As we have seen, the symmetric indivisibility of $(\mathbb{Q}, \leq)$ can be viewed as the result of two separate phenomena — indivisibility and the density of symmetric substructures. This proof turned out to be more revealing than the one using composition construction. Moreover, the latter version is harder to generalize, and would not — for example — give any information on the symmetric indivisibility of the $\mathbb{Q}$-tree considered in the previous section.

In a similar way, the proof using the composition construction to show that Rado's random graph is symmetric does not give us any information, on, say, the situation with Henson graphs Γ_n, where the Henson graph for $n \geq 3$ is the generic countable K_n-free graphs (see **[5]** for the construction and basic properties of the Henson graphs).

In this section we give two alternative proofs of the symmetric indivisibility of the random graph. As before, the more revealing proof will be the one showing that symmetry and indivisibility seem not to be very closely related. It turns out that this strategy will allow us to prove, in addition, the symmetric indivisibility of the countable universal triangle free graph. In fact, the same methods give the analogous result for all Henson graphs, but there are more technicalities, and the details are deferred to a subsequent paper.

Before we embark on the proof, we set some standard terminology. A *neighborhood* in a graph G is a set of the form $R(a) := \{b : G \models R(b, a)\}$ and for a set $X \subseteq V$, the set of neighbors of a in X is $R(a, X) := \{b : b \in X \text{ and } G \models R(a, b)\}$. Throughout this section all graphs considered will have quantifier elimination in the pure language of graphs (i.e., the language containing a unique binary predicate).

In that situation, a *definable subset* of such a graph G is a finite boolean combination of neighborhoods and of singelton sets. We will introduce more notation and terminology as we proceed. We first return to Rado's random graph.

5.1. The strategy. Recall that Rado's random graph is the unique (up to isomorphism) countable graph Γ with the property that for any finite, disjoint $U, V \subseteq \Gamma$ there is a vertex v joined to every element in U and disjoint from every element in V. As in the case of $(\mathbb{Q}, \leq)$ it is well known (and very easy to show) that:

LEMMA 5.1. *The random graph,* Γ*, is indivisible.*

PROOF. It follows immediately from the axiomatisation of Γ that if $S \subseteq \Gamma$ is an infinite definable set then $S \cong \Gamma$. Fix some $c : \Gamma \to \{0, 1\}$. Then by the last remark we may assume that whenever $S \subseteq \Gamma$ is definable and infinite S is not monochromatic. It follows that for every such $S \subseteq \Gamma$ we can find a point $a_S \in S$ such that $c(a_S) = 1$. It is now obvious that the set

$$C := \{a_S : S \subseteq \Gamma, S \text{ infinite and definable}\} \cong \Gamma.$$

For assume that $a_1, \ldots, a_n, b_1, \ldots b_k \in C$ are distinct. Let

$$S := \{a \in \Gamma : \bigwedge_{i=1}^{n} R(a, a_i) \wedge \bigwedge_{i=1}^{k} \neg R(a, b_i)\}.$$

Then S is definable, and therefore $a_S \in C$ realizes S. Since $a_1, \ldots, a_m$ and $b_1, \ldots, b_k$ were arbitrary, this proves the claim. □

REMARK 5.2. An even simpler proof of a stronger version of the above lemma — namely that when partitioning Γ to two parts one of the parts is actually isomorphic to Γ — can be found in P. Cameron's blog, [**1**]: If $\mathcal{B} \cup \mathcal{R}$ is a partition of the random graph, Γ, and neither $\mathcal{B}$ nor $\mathcal{R}$ are isomorphic to Γ, there are sets $U_1, V_1 \subseteq \mathcal{B}$ and $U_2, V_2 \subseteq \mathcal{R}$ witnessing this. But in Γ there is some v connected to every vertex in $U_1 \cup U_2$ and not joined by an edge to any vertex in $V_1 \cup V_2$. However, our assumption implies that v cannot lie in neither $\mathcal{B}$ nor in $\mathcal{R}$ — a contradiction.

We chose to keep the proof of the weaker property, as the ideas appearing in it will serve us later on.

To prove that the random graph is symmetrically indivisible, one could now hope to show, as in the case of $(\mathbb{Q}, \leq)$, that the symmetrically embedded copies of Γ are dense. We were unable to prove this, apparently for a good reason... The following Theorem was proved and communicated to us by one of the two anonymous referees:

THEOREM 5.3 (Anonymous Referee). *Let* Γ *denote the countable random graph. Then the symmetrically embedded copies of* Γ *are not dense.*

PROOF. Let $\Gamma_0 \cong \Gamma$. Adjoin to Γ_0 two sets of vertices,

$$I := \{v_F : F \subseteq \Gamma_0 \text{ finite}\},\ J := \{w_F : F \subseteq \Gamma_0 \cup I \text{ finite}\}.$$

Extend the edge relation, R, from Γ_0 to $\Gamma := \Gamma_0 \cup I \cup J$, subject to the following conditions:

(1) I is an independent set of vertices.

(2) Edges connecting vertices in Γ_0 and I encode the membership relation, i.e., if $v \in \Gamma_0$ and $v_F \in I$ then $R(v, v_F)$ if and only if $v \in F$.
(3) Edges connecting vertices in $\Gamma_0 \cup I$ and J encode the non-membership relation, i.e., if $w \in \Gamma_0 \cup I$ and $w_F \in J$ then $R(w, w_F)$ if and only if $w \notin F$.
(4) Order $\Gamma_0 \cup I$ and let edges between vertices in J encode this ordering in the following way: for $a, b \in \Gamma_0 \cup I$ let $R(w_{\{a\}}, w_{\{a,b\}})$ if and only if $a < b$.
(5) Γ is the random graph, namely, whenever $F \subseteq \Gamma_0 \cup I \cup J$ is finite and p is a boolean combination of neighborhoods of elements in F, p is realized in Γ.

Conditions (1)–(4) are easily met. Starting with a minimal edge relation on Γ satisfying these conditions, we can satisfy condition (5) as follows: let $J_0 := \{w_F \in J : |F| < 3\}$ and $J_n := \{w_F \in J : |F| = n + 2\}$ for $n > 0$. For every finite $F \subseteq \Gamma_0 \cup I \cup \bigcup_{i=0}^{n} J_n$ and q as in condition (5) fix a realization of q in J_{n+1}.

We will show that no copy of Γ inside Γ_0 is symmetrically embedded in Γ. To this end define, for an arbitrary infinite set $X \subseteq \Gamma_0$,

$$S(X) := \{v \in \Gamma \setminus X : \forall F \subseteq X \text{ finite } \exists u(\neg R(u, v) \wedge F = R(u, X))\}.$$

We claim that $S(X) = (\Gamma_0 \setminus X) \cup I$. Given $w_F \in J$ let $\mathcal{F} := \{R(u, X) : u \in F\}$ and fix some finite $F' \subseteq X$ with $F' \notin \mathcal{F}$. Assume towards a contradiction that $w_F \in S(X)$. Then there would be some u such that $\neg R(u, w_F)$ and $F' = R(u, X)$. Since $F' \notin \mathcal{F}$ we know that $u \notin F$. By condition (3) above, as $\Gamma \models \neg R(u, w_F)$ it follows that $u \in J$. But $R(u, \Gamma_0)$ is co-finite, and therefore so is $R(u, X)$, a contradiction. On the other hand, if $v \in (\Gamma_0 \setminus X) \cup I$ and $F \subseteq X$ is any finite set then taking $u = v_F$ we know (conditions (1) and (2) above) that $\Gamma \models \neg R(u, v)$ and $\Gamma(u, X) = F$, implying that $v \in S(X)$. This shows that, indeed, $S(X) = (\Gamma_0 \setminus X) \cup I$, as claimed.

It follows that $S(X) \cup X = \Gamma_0 \cup I$, and thus if $\sigma \in \mathrm{Aut}(\Gamma)$ preserves X setwise it also preserves $\Gamma_0 \cup I$, and J. In particular, if $F \subseteq \Gamma_0 \cup I$ is finite, then $\sigma(w_F) = w_{\sigma(F)}$. Taking $a, b \in \Gamma_0 \cup I$ such that $a < b$ (with respect to the ordering fixed in condition (4) above) we see that

$$a < b \iff R(w_{\{a\}}, w_{\{a,b\}}) \iff R(\sigma(w_{\{a\}}), \sigma(w_{\{a,b\}})) \iff \sigma(a) < \sigma(b).$$

So any automorphism of Γ fixing an infinite subset of Γ_0 setwise preserves the order on $\Gamma_0 \cup I$. In particular, such automorphisms can only have trivial finite orbits. But the countable random graph does have automorphisms with non-trivial finite orbits. Thus, if $\Gamma_1 \subseteq \Gamma_0$ is isomorphic to Γ no automorphism of Γ_1 with non-trivial finite orbits can be extended to Γ. $\square$

Thus, we have to modify the density argument in order for it to work in the present context. In $(\mathbb{Q}, \leq)$ the density of the symmetric copies of $\mathbb{Q}$ allowed us to give a proof where indivisibility and symmetry had nothing to do with each other. It turns out that allowing a little more interaction between these two requirements, essentially the same argument still works. Instead of proving the density of the symmetric copies of the random graph, Γ, among all sub-models we will restrict ourselves to a smaller class of sub-models, $\mathcal{C}$. If we can show a stronger version of indivisibility, namely one where the monochromatic sub-model can be found in $\mathcal{C}$, and then prove that the symmetric sub-models are dense in $\mathcal{C}$, our work will be done.

So it remains to choose the class $\mathcal{C}$. Looking under the lamppost is sometimes a good idea: Given any necessary condition, $\mathcal{P}$, for a sub-model to be symmetrically embedded we may as well consider the density of symmetric indivisibility in the class of models satisfying $\mathcal{P}$. It turns out that the model theoretic notion of *stable embeddedness* is a good such condition (though we do not know whether the symmetric sub-models of the random graph are, indeed, dense in the class of stably embedded sub-models, we do know it for the generic triangle free graph, Γ_Δ, and in fact for all Henson graphs Γ_n but this latter fact is not proved in the present paper).

5.2. Stable embeddedness. As we have seen, if $S \subseteq \Gamma$ is infinite and definable then $S \cong \Gamma$. However, such a set S cannot be symmetrically embedded in Γ. The main observation required in order to prove this is:

PROPOSITION 5.4. *Let $\mathcal{M}$ be a countable $\aleph_0$-categorical structure, $D \subseteq M^n$. Then the orbit of D under* $\mathrm{Aut}(\mathcal{M})$ *is countable if and only if D is definable.*

To see how this proposition is related to the symmetric embedding of S in Γ consider any point $a \in \Gamma \setminus S$. By the characterization of the random graph $R(a, S)$ is not an S-definable subset of S (i.e., $R(a, S)$ cannot be obtained as a finite boolean combination of neighborhoods of elements in S). By Proposition 5.4 $R(a, S)$ has an uncountable orbit in S under the action of $\mathrm{Aut}(S)$. But given $\sigma \in \mathrm{Aut}(S)$ such that $\sigma(R(a, S)) \neq R(a, S)$ and $\tilde{\sigma} \in \mathrm{Aut}(\Gamma)$ extending σ it must be that $\tilde{\sigma}(a) \neq a$. But a can only have a countable orbit in Γ (under the action of $\mathrm{Aut}(\Gamma)$), so it cannot be that every $\sigma \in \mathrm{Aut}(S)$ extends to an automorphism of Γ, i.e., S is not symmetrically embedded in Γ.

In model theoretic terminology, we would say that the set S of the previous paragraph is not *stably embedded* in Γ. Thus Proposition 5.4 suggests that, at least in the context of $\aleph_0$-categorical structures, stable embeddedness is a necessary condition for symmetric embeddedness (see Corollary 5.6). In this sub-section we will develop what little we need of stable embeddedness. Our treatment will be more general than necessary for the purposes of the present paper. Readers feeling uncomfortable with model theoretic terminology are referred to Proposition 5.8 for a proof of a special case of Proposition 5.4 sufficient for our purposes. A concrete treatment of stable embeddedness in the case of the random graph or the Henson graphs is given in the paragraphs concluding this sub-section, starting with Fact 5.12.

In the proof that an infinite co-infinite definable set S is not symmetrically embedded in Γ the only special role the random graph had to play was in showing that the neighborhood of some element in Γ, when intersected with S, was not definable in S. In order to generalize this argument we need the model theoretic notion of stable embeddedness:

DEFINITION 5.5. Let $\mathcal{M}, \mathcal{N}$ be structures (not necessarily in a common language) with $\mathcal{N} \subseteq \mathcal{M}^n$ for some $n \geq 1$. Then $\mathcal{N}$ is *stably embedded* in $\mathcal{M}$ if every $\mathcal{M}$-definable (with parameters from $\mathcal{M}$) relation on $\mathcal{N}$ is $\mathcal{N}$-definable (with parameters form $\mathcal{N}$).

In this terminology, the argument showing that infinite definable subsets of the random graph are not symmetrically embedded translates word for word to give:

COROLLARY 5.6. *Let $\mathcal{M}$ be an $\aleph_0$-categorical structure symmetrically embedded in a countable structure $\mathcal{N}$. Then $\mathcal{M}$ is stably embedded in $\mathcal{N}$.*

Before we proceed to discuss in more detail the notion of stable embeddedness, we return to the proof of Proposition 5.4. Since the proof is somewhat technical, we first give a proof in a special case that covers all cases of interest for us in this paper. Towards that end we recall the following:

DEFINITION 5.7. A countable structure $\mathcal{M}$ has the *small index property* (SIP) if whenever a subgroup $G \leq \mathrm{Aut}(\mathcal{M})$ has a countable index in $\mathrm{Aut}(\mathcal{M})$ there exists a finite set A such that $\mathrm{Aut}(\mathcal{M}/A) \leq G$.

Thus we have:

PROPOSITION 5.8. *Let $\mathcal{M}$ be a countable $\aleph_0$-categorical structure with SIP. Then $D \subseteq M^n$ has countable orbit under* $\mathrm{Aut}(\mathcal{M})$ *if and only if D is definable.*

PROOF. Because $\mathcal{M}$ is $\aleph_0$-categorical, a set $D \subseteq M^n$ is definable if and only if there exists a finite set A such that $\sigma(D) = D$ for all $\sigma \in \mathrm{Aut}(\mathcal{M}/A)$. So assume that $G := \{\sigma : \sigma(D) = D\}$ has countable index in $\mathrm{Aut}(\mathcal{M})$. By SIP there is a finite set A such that $G \geq \mathrm{Aut}(\mathcal{M}/A)$, hence D is definable over A. ☐

By **[7]** the random graph and the Henson graphs (as well as many other $\aleph_0$-categorical structures) have the small index property.

We will now give an outline of the proof of Proposition 5.4:

PROOF OF PROPOSITION 5.4. Of course, we only have to prove that if D is not definable its orbit is uncountable. The proof below, which we included for completeness of presentation's sake, is of a particular instance of Kueker-Reyes Theorem, which holds in larger generality (see Theorem 4.2.4 in **[6]**).

Given a non-definable subset D of $\mathcal{M}^n$, we construct a binary tree of elementary maps, whose branches correspond to automorphism of $\mathcal{M}$ with distinct images of D. For simplicity, we will assume that $D \subseteq \mathcal{M}$ (i.e., $n = 1$) — the general case being similar, with just a little extra care in the bookkeeping. We assume also that $\mathcal{M}$ is given in a language with quantifier elimination (i.e., that partial isomorphisms between finite substructures extend to automorphisms of $\mathcal{M}$).

Let σ be a finite partial isomorphism with domain A. We will show how to find σ_i for $i < 2$, such that:

(1) $\sigma_i \supseteq \sigma$ is a finite partial isomorphisms.
(2) For every choice of $\tilde{\sigma}_i$ such that $\sigma_i \subseteq \tilde{\sigma}_i \in \mathrm{Aut}(\mathcal{M})$ it holds that $\tilde{\sigma}_0(D) \neq \tilde{\sigma}_1(D)$.

Because D is not definable, there exists a type p over $A \cup \sigma(A)$ such that p has infinitely many realizations in D and and infinitely many realizations outside D.

Fix $a_0 \in D \setminus A$ and $a_1 \notin D \cup A$ both realizations of p and fix some $b \notin \sigma(A)$ such that $b \models \sigma(p)$. Let $\sigma_0 = \sigma \cup \{(a_0, b)\}$ and $\sigma_1 = \sigma \cup \{(a_1, b)\}$. Each σ_i, for $i < 2$, is a finite partial automorphisms which extends σ, and whenever $\sigma_i \subseteq \tilde{\sigma}_i \in \mathrm{Aut}(\mathcal{M})$ it holds that $b \in \tilde{\sigma}_0(D) \setminus \tilde{\sigma}_1(D)$.

The rest is standard, so we will be brief: construct inductively for $\eta \in \{0,1\}^n$ partial isomorphisms σ_η such that:

(1) $\sigma_{\langle\rangle}$ is the empty isomorphism, and the domain of σ_η is finite for all η.
(2) If $\eta \in \{0,1\}^n$ then, with respect to some fixed enumeration of M, both the domain and range of η contain the first n elements of M.

(3) $\sigma_{\eta^\frown i} \supseteq \sigma_\eta$.
(4) If $\tilde{\sigma}_{\eta^\frown 0}, \tilde{\sigma}_{\eta^\frown 1} \in \mathrm{Aut}(\mathcal{M})$ are such that $\sigma_\eta \subseteq \tilde{\sigma}_{\eta^\frown i}$ then $\tilde{\sigma}_{\eta^\frown 0}(D) \neq \tilde{\sigma}_{\eta^\frown 1}(D)$.

By the argument of the previous paragraph this induction can be achieved. By (2) and (3) of the construction for any $\eta \in \{0,1\}^\omega$ the function $\sigma_\eta := \bigcup_{n\in\omega} \sigma_{\eta|_{0,\dots,n}}$ is an automorphism of $\mathcal{M}$, and by (4) of the constructing if $\eta_1, \eta_2 \in \{0,1\}^\omega$ are distinct then $\sigma_{\eta_1}(D) \neq \sigma_{\eta_2}(D)$, as required. □

We can now go back to the discussion of stable embeddedness. We start with a few examples:

EXAMPLE 5.9.

(1) Let $\mathbb{N}^*$ be an $\aleph_1$-saturated elementary extension of $(\mathbb{N}, +, \cdot, 0, 1)$. Then $\mathbb{N}$ is symmetrically embedded in $\mathbb{N}^*$ (because $\mathbb{N}$ is rigid), but $\mathbb{N}$ is not stably embedded because by compactness and saturation for any set of primes $S \subseteq \mathbb{N}$ there exists in $\mathbb{N}^*$ an element n_S divisible precisely by those (finitary) primes in S. Thus $S := \{n \in \mathbb{N} : n|n_S\}$. Cardinality considerations now prevent $\mathbb{N}$ from being stably embedded.
(2) On the other hand, by quantifier elimination, it is easy to check that $(\mathbb{N}, \leq)$ is stably embedded in any elementary extension.
(3) If $\mathcal{M}$ is saturated, $P \subseteq \mathcal{M}$ is definable, then it is not hard to verify that P (with all the $\mathcal{M}$-induced structure) is stably embedded if and only if it is symmetrically embedded (see the Appendix of **[3]** for the proof).
(4) Stable embeddedness with respect to subsets is weaker than stable embeddedness with respect to relations: consider the structure $\mathcal{M}$ with two infinite disjoint predicates A, B and a unique generic ternary relation $R(x_1, x_2, y)$ holding only if x_1, x_2 are in A and y is in B. Then A is vacuously stably embedded with respect to subsets, but not stably embedded with respect to binary relations.

In the present paper, however, stable embeddedness will be needed only in contexts where we have quantifier elimination in a language containing only binary relations. In such circumstances the distinction between stable embeddedness for subsets and stable embeddedness for relations does not exist. We will also need a more convenient test for stable embeddedness, one that is easier to keep track of in inductive constructions. For that purpose we recall:

DEFINITION 5.10. Let $\mathcal{M}$ be a structure and $A \subseteq M$ any set. The *type over A of an element $b \in M$ is definable* (over $B \subseteq A$) if for every formula $\varphi(x, \bar{y})$ (with $|\bar{y}| = n$, some $n \in \mathbb{N}$) the set $\varphi(b, A^n) := \{\bar{a} \in A^n : \mathcal{M} \models \varphi(b, \bar{a})\}$ is definable (over B).

It is now merely a question of unravelling the definitions to verify:

REMARK 5.11. Let $\mathcal{M}, \mathcal{N}$ be structures with $N \subseteq M$. Then $\mathcal{N}$ is stably embedded in $\mathcal{M}$ if and only if every type over N is definable.

In the context of the random graph or the Henson graphs, the discussion of the present sub-section can be given in more concrete terms for those not at ease with the model theoretic terminology. From this point to the end of this subsection Γ will denote either the countable random graph or a Henson graph.

Remark 5.12. Let $b \in \Gamma$ and $A \subseteq \Gamma$. The *type of b over A* is the set of formulas

$$\{R(x,a) : \Gamma \models R(b,a), a \in A\} \cup \{\neg R(x,a) : \Gamma \models \neg R(b,a), a \in A\}$$

Remark 5.13.

(1) If the set A in the above definition is finite, then by quantifier elimination (or $\aleph_0$-categoricity, if you prefer) the type of every element over A is merely a definable set with parameters in A, i.e., it is a finite boolean combination of neighborhoods of elements in A.
(2) If A is finite as above, it is convenient to identify the type of an element b over A with the orbit of b under the action of $\mathrm{Aut}(\mathcal{M}/A)$. Viewed from this angle, $\aleph_0$-categoricity implies that $A \subseteq \Gamma$ is definable if an only if it is a finite union of types.
(3) In the present paper all types will be types of elements in Γ (over various sets of parameters).

In this context we have

Fact 5.14. *A set $A \subseteq \Gamma$ is stably embedded if and only if for every $b \in \Gamma$ there is a finite set $A_0 \subseteq A$ and types $p_1, \dots, p_n$ over A_0, such that $\{a \in A : \Gamma \models R(a,b)\}$ is the union of all realizations of $p_1, \dots, p_n$ in A.*

We can now return to our main question, the symmetric indivisibility of the random graph.

5.3. Extending partial isomorphisms. In this subsection we carry out one part of the strategy outlined in Subsection 5.1. We introduce the machinery for constructing, given $S \leq \Gamma$ — satisfying appropriate assumptions — symmetric submodels of Γ inside S. As we will see (in the case of the generic triangle free graph) stable embeddedness is an appropriate assumption for this machinery to work. For the random graph we have a short cut. We need some technical definitions and a lemma:

Definition 5.15. (1) Let $C_i, C'_i \subseteq \Gamma$ (for $i = 1, 2, \dots, k$). By $\langle C_i\rangle_{i=1}^k \equiv \langle C'_i\rangle_{i=1}^k$ we mean that there exists a partial isomorphism $\sigma : \bigcup_{i=1}^k C_i \to \bigcup_{i=1}^k C'_i$ such that $\sigma(C_i) = C'_i$ for all i.
(2) Let $\mathcal{A} := \{\langle A_i, B_i\rangle\}_{i=0}^k$ be a sequence (of ordered pairs) of subsets of Γ, and let $A := \bigcup_{i=1}^k A_i$. Say that $\mathcal{A}$ extends partial automorphisms if:
   - $A_i \cap B_i = \emptyset$ for all $i \leq k$ and $i < j$ implies $A_i \subseteq A_j, B_i \subseteq B_j$.
   - $R(b,A) \subseteq A_i$ if $b \in B_i$ and $0 < i$. For $b \in B_0$, either $R(b,A) \subseteq A_0$ or $R(b,A) = A$.
   - Let $i < k$, $C, C' \subseteq B_i \cup A_i$ be such that $\langle C \cap A_i, C \cap B_i\rangle \equiv \langle A_i \cap C', B_i \cap C'\rangle$, σ a partial isomorphism witnessing this, and $b \in B_i$ such that $R(b,A) \subseteq C$. Then there exists $b' \in B_{i+1}$ such that $\sigma \cup \langle b, b'\rangle$ is a partial isomorphism.

As we will see below, the idea underlying the above definition is to allow an inductive construction of an increasing sequence of finite graphs, culminating in a pair of countable structures with the smaller symmetrically embedded in the larger. The definition is tailor made to suite our requirements in dealing with the random graph and, later on, with the Henson graphs, Γ_n.

LEMMA 5.16. *Assume that* Γ *is a graph in the language* $\mathcal{L} = \{R\}$, *and* $\mathcal{A} = \{\langle A_i, B_i\rangle\}_{i\in\omega}$ *a sequence of pairs of subsets of* Γ. *If* $\{\langle A_i, B_i\rangle\}_{i<k}$ *extends partial isomorphisms for each* $k \in \omega$ *and* $B_0 = \emptyset$, *then* $A := \bigcup_{i\in\omega} A_i$ *is symmetrically embedded in* $\Gamma' := A \cup \bigcup_{i\in\omega} B_i$.

PROOF. By a standard back and forth, it will suffice to show that if $C_1, C_2 \supseteq A$, $|C_i \setminus A| < \aleph_0$ and $\sigma : C_1 \to C_2$ is a partial isomorphism extending an automorphism of A then σ can be extended to any $c \in \Gamma$. Let $n \in \omega$ be such that $c \in B_n$. Without loss of generality $(C_1 \cup C_2) \setminus A \subseteq B_n$. So $R(c, A) \subseteq A_n$ (because $B_0 = \emptyset$ and therefore $c \notin B_0$). We may also assume (by increasing n if necessary) that $\sigma(R(c, A)) \subseteq A_n$. Denote $C_A = R(c, A)$, $C'_A := \sigma(R(c, A))$, $C_B := C_1 \setminus A$ and $C'_B := C_2 \setminus A$. Then, $\langle C_A, C_B\rangle \equiv \langle C'_A, C'_B\rangle$, as witnessed by σ. But $\langle A_i, B_i\rangle_{i=1}^{n+1}$ extends partial isomorphisms and applying this property to $C := C_A \cup C_B$, $C' := C'_A \cup C'_B$ and c we can find $c' \in B_{n+1}$ such that $(\sigma \restriction C) \cup \langle c, c'\rangle$ is a partial isomorphism.

It remains to show that $\sigma \cup \{\langle c, c'\rangle\}$ is a partial isomorphism. Indeed, we only have to show that if $b \in C_1$ then $\models R(c, b)$ if and only if $\models R(\sigma(b), c')$. If $b \in C$ this is obvious by the choice of c'. But, by assumption, if $b \in \Gamma \setminus C$ then $\models \neg R(c, b)$. Since $R(c', A) \subseteq C'_A$ and $\sigma(C_A) = C'_A$ the conclusion follows. □

Observe that the above proof gives a little bit more: it shows that if $\mathcal{A}$ extends partial isomorphisms then for any finite set C, if $\sigma : A \cup C \to A \cup C'$ is a partial isomorphism extending an automorphism of A, then σ can be extended to an automorphism. This suffices to prove:

LEMMA 5.17. *Let* Γ *be the random graph or a Henson graph. Let* $D \subseteq \Gamma$ *be a finite set and* $S \subseteq \Gamma$ *be such that for all* $n \in \mathbb{N}$, $b_1, \dots, b_n \in \Gamma \setminus D$ *and any formula* $\varphi(x)$ *with parameters in* S, *the formula* $\varphi(x) \wedge \bigwedge_{i=1}^{n} \neg R(x, b_i)$ *has a solution in* S *if and only if it has a solution in* Γ. *Then there exists* $\Gamma_0 \subseteq S$ *such that* $\Gamma_0 \cong \Gamma$ *and* Γ_0 *is symmetric in* Γ.

PROOF. Recall that the theory of Γ is $\aleph_0$-categorical and has quantifier elimination.

Let $D := \{d_1, \dots, d_n\} \subseteq \Gamma$ be as in the statement of the lemma. Assume that D is minimal with this property. Then for all $d_i \in D$ there are a formula $\varphi_i(x)$ and some $b_1, \dots b_k \in \Gamma \setminus D$ such that

$$\varphi_i(x) \cap \neg R(d_i, S) \cap \bigcap_{j=1}^{k} \neg R(b_j, S) = \emptyset, \quad \varphi_i(x) \cap \bigcap_{j=1}^{k} \neg R(b_j, S) \neq \emptyset.$$

Let $S_1 = \varphi_1(S) \cap \bigcap_{i=1}^{k} \neg R(b_i, S)$. Observe, first, that $S_1 \subseteq R(\Gamma, d_1)$ and that, moreover, S_1 satisfies the assumptions of the lemma (with respect to the same finite set D). Hence, replacing S with S_1 we may assume that $S \subseteq R(\Gamma, d_1)$. By induction we may now assume that S is such that $S \subseteq R(\Gamma, d)$ for all $d \in D$. For simplicity we may also assume that $S \cap D = \emptyset$.

We can now proceed with the proof under the above assumptions. Let $\{c_i\}_{i\in\omega}$ be some fixed enumeration of Γ and denote $C_n := \{c_0, \dots, c_{n-1}\}$. Let $a_0 \in S$ be any element, and set $A_0 = \{a_0\}$, $B_0 = D$. Assume that we have constructed an increasing sequence of sets A_n, B_n such that:

(1) $A_n \subseteq S$.
(2) $\langle A_i, B_i\rangle_{i=0}^{n}$ extends partial isomorphisms.

(3) $A_n \cap B_n \subseteq B_0$.
(4) $c_n \in B_n \cup A_n$.
(5) $A_n \cong C_m$, where $m = |A_n|$, witnessed by some partial isomorphism σ_n such that $\sigma_{n'} = \sigma_n | A_{n'}$ for $n' < n$.

We will now construct A_{n+1}, B_{n+1} such that the above inductive assumptions remain valid. Let $p = \sigma_n^{-1}(\mathrm{tp}(c_n/C_n)) \cup \{\neg R(x,b) : b \in B_n \setminus B_0\}$. By $\aleph_0$-categoricity of Γ and the finiteness of C_n we know that $p(x)$ is isolated, say by a formula $\varphi_p(x)$. Notice that by universality of the graphs we are considering (and the fact that in any of these graphs, given any consistent formula $\varphi(x)$ with parameters in A the formula $\varphi(x) \wedge \neg R(x,b)$ is consistent for $b \notin A$) $\varphi_p(x)$ is realized in Γ. Hence, our assumptions on S assure that we can find some $a \in p(S) \setminus A_n$. We let $A_{n+1} = A_n \cup \{a\}$.

Now we construct B_{n+1}. If $c_n \notin A_n$ we let $B_{n+1,0} = B_n \cup \{c_n\}$, otherwise $B_{n+1,0} = B_n$. Now, for all $C, C' \subseteq B_n \cup A_n$, partial isomorphism $\sigma : C \to C'$ and $b \in B_n$ as in Definition 5.15, we let $t_{CC'b} \in \Gamma$ be such that $\sigma \cup \langle b, t_{CC'b} \rangle$ is a partial isomorphism. This is possible because, as B_n and A_n are finite so are C and C' therefore, the type of $t_{CC'b}$ over $CC'b$ is isolated, hence realized in Γ. Moreover, by $\aleph_0$-categoricity of Γ the number of types to realize in this way is finite. Therefore, setting $B_{n+1} = B_{n+1,0} \cup \{t_{CC'b} : C, C', b \text{ as above}\}$, the set B_{n+1} is finite and the inductive assumptions still hold of $\langle A_i, B_i \rangle_{i=0}^{n+1}$.

Now let $\Gamma_0 := \bigcup_{n \in \omega} A_n$. Then $\Gamma_0 \cong \Gamma$ by (5) of the inductive hypothesis. So we only have to check that Γ_0 is symmetrically embedded in Γ. If $D = \emptyset$ then by Lemma 5.16 and (4) of the inductive assumptions, we are done. Otherwise, by assumption $S \subseteq R(\Gamma, d)$ for all $d \in D$, so $\mathrm{tp}(D/\Gamma_0) = \mathrm{tp}(D/\sigma(\Gamma_0))$ for all $\sigma \in \mathrm{Aut}(\Gamma_0)$. So, given $\sigma \in \mathrm{Aut}(\Gamma_0)$ we know that $\sigma' := \sigma \cup \mathrm{id}(D)$ is a partial isomorphism. By the observation following Lemma 5.16 we get that σ' can be extended to an automorphism of $\mathcal{M}$.

□

Remark 5.18. The above proof would work, essentially as written, for the class of $\aleph_0$-categorical graphs with quantifier elimination in the pure language of graphs. In this setting, however, we have to assume that in the theory any consistent formula $\varphi(x)$ is consistent with $\varphi(x) \wedge \neg R(x,b)$ provided b is not a parameter in $\varphi(x)$.

Corollary 5.19. *The random graph is symmetrically indivisible.*

Proof. Let $c : \Gamma \to \{0,1\}$ and $S_1 = \{a \in \Gamma : c(a) = 1\}$. If S_1 meets every infinite definable set then set $S = S_1$. If $S_1 \cap \varphi(\Gamma) = \emptyset$ for some infinite definable set $\varphi(x, \bar{d})$ set $S = \{a \in \Gamma : c(a) = 0 \wedge \models \varphi(a, \bar{d})\}$. Either way S satisfies the assumptions of Lemma 5.17 (with $D = \emptyset$ in the former case and $D = \{\mathrm{dom}(\bar{d})\}$ in the latter). So the conclusion follows. □

Combined with Corollary 5.6 this gives:

Corollary 5.20. *The class of stably embedded submodels of the random graph is indestructible by finite partitions, namely under any coloring of the (vertices of the) random graph in two colors we can find a stably embedded monochromatic submodel.*

Remark 5.21. We do not know whether every stably embedded submodel of the random graph Γ contains a submodel symmetrically embedded in Γ. The proof of Corollary 5.19 can be viewed as a density result in a smaller class of stably embedded sub-models: those sub-models $\Gamma_0 \leq \Gamma$ such that for all but finitely many $b \in \Gamma \setminus \Gamma_0$ the set $R_b^{\Gamma_0}$ is finite in Γ_0.

5.4. The generic triangle free graph. The proof of Corollary 5.19 cannot be generalized to prove that the Henson graphs, Γ_n, are symmetrically indivisible, since it uses the fact that every infinite definable set contains an isomorphic copy of the whole universe. This is, of course, not true in Γ_n since for every $a \in \Gamma_n$ the set $R(a, \Gamma_n)$ does not admit K_{n-1}. Of course, $\Gamma_n[\Gamma_n]$ is not a K_n-free graph, so we cannot hope to use the composition construction either.

It turns out, however, that the ideas underlying our new proofs of the symmetric indivisibility of $(\mathbb{Q}, \leq)$ and the random graph can be adapted to prove the symmetric indivisibility of the generic triangle free graphs. Namely, once we show that Γ_n is indivisible in the stably embedded sub-models, we can prove that the symmetric sub-models are dense in the stably embedded ones, to get the desired conclusion.

The proof that Γ_n is indivisible is not as easy as in the previous cases. A simple generalization of the proof of Claim 5.1 gives only the following (see [**5**] for a proof along similar lines):

FACT 5.22. *Let $c : \Gamma_n \to \{0, 1\}$ be any coloring. Then there exists $i \in \{0, 1\}$ such that for any K_n-free graph G there exists $\sigma : G \hookrightarrow \Gamma_n$ such that $c(\sigma(G)) = \{i\}$.*

The indivisibility of $\Gamma_\triangle$, the generic triangle-free graph is proved by Komjáth and Rödl [**10**]. The proof that Γ_n is indivisible (for $n > 3$) is more technical, [**4**]. Inspecting the proof of Komjáth and Rödl we see that, in fact, we get a little more than indivisibility of $\Gamma_\triangle$. The proof assures that given $c : \Gamma_\triangle \to \{0, 1\}$ either there exists a monochromatic sub-model colored 0, or there is a stably embedded monochromatic sub-model colored 1. It turns out, however, that a natural modification of the proof shows the indivisibility of $\Gamma_\triangle$ in the stably embedded sub-models. The main part of the present subsection is dedicated to a self-contained proof of this fact. We then deduce the symmetric indivisibility of $\Gamma_\triangle$. Much of the work is done in the somewhat greater generality of arbitrary Henson graphs, Γ_n, the proof of whose symmetric indivisibility is postponed to a subsequent paper.

We start with some notation and conventions. We often enumerate the elements of a countable model Γ_n. In this case, we will think of the enumeration as an ordering of Γ_n and say that $a < b$ if a appears before b in this enumeration.

We also need a few technical remarks:

DEFINITION 5.23. A definable set $S \subseteq \Gamma_n$ is generic if it is of the form $\bigvee_{j=1}^{l} \psi(x, \bar{v}_j)$ and at least one of the formulas ψ_j is given by $\bigwedge_{i=1}^{k} \neg R(x, v_{j,i})$ for some $k \in \mathbb{N}$ and $v_{j,1}, \ldots, v_{j,k} \in \Gamma_n$.

Observe that a non-generic set in Γ_n may not admit K_{n-1}, so it is, in some sense "small". Generic sets, however, are nice:

REMARK 5.24. The intersection of finitely many generic sets in Γ_n is generic.

But more importantly:

REMARK 5.25. If $S \subseteq \Gamma_n$ is generic then S contains an isomorphic copy of Γ_n.

PROOF. It is enough to show that if S is of the form $\bigwedge_{i=1}^{k} \neg R(x, v_i)$ then, in fact $S \setminus \{v_1, \ldots, v_k\} \cong \Gamma_\triangle$, which is a triviality. $\square$

But, in fact, we can show more:

LEMMA 5.26. *Assume that $S \subseteq \Gamma_n$ is generic. Then there exists $\Gamma_0 \subseteq S$ such that $\Gamma_0 \leq \Gamma_n$ and Γ_0 is stably embedded. Even more, we may choose Γ_0 such that $R(v, \Gamma_0)$ is finite for all $v \in \Gamma_n \setminus \Gamma_0$.*

PROOF. We may assume without loss of generality that S is of the form $\bigwedge_{i=1}^{k} \neg R(x, v_i)$. Let $\{v_i\}_{i\in\omega}$ be an enumeration of Γ_n. For simplicity of notation we will write $v < w$ if $v = v_j$, $w = v_i$ and $j < i$. We will construct Γ_0 by induction. Let $B_0 = \{v_1, \dots, v_k\}$ and $a_0 \in S \setminus B_0$ the least such element (in our enumeration). Assume we have constructed $\{a_0, \dots, a_n\}$ and sets $B_0, \dots, B_n$ such that:

(1) $a_i \in S$ for all i and $(a_0, \dots, a_n)$ is an increasing sequence.
(2) $(a_0, \dots, a_k) \cong (v_0, \dots, v_k)$.
(3) If $j < i$ and $v \in B_j$ then $\models \neg R(a_i, v)$.
(4) $B_{i+1} := \{v : a_i < v < a_{i+1}\}$.

We claim that if we can continue the construction (keeping the inductive assumptions) then our lemma will be proved. Indeed, taking $\Gamma_0 := \{a_i\}_{i\in\omega}$ we get (by condition (2)) that $\Gamma_0 \cong \Gamma_n$; by condition (4) $\Gamma_0 \cup \bigcup_{i\in\omega} B_i = \Gamma_n$ and by (3) for every $b \in \Gamma_n$, if $b \in B_i$ then $R(b, \Gamma_0) \subseteq \{a_0, \dots, a_i\}$, so Γ_0 is stably embedded.

Thus, it remains to verify that the construction can be carried out, but this follows from the axiomatisation of Γ_n, as at stage n we only have finitely many conditions to realize, and these conditions can be expressed by a (single) formula which is merely a specialization of the universal quantifier in an explicit axiom of Γ_n. $\square$

We will not use the above lemma explicitly, but the ideas appearing in the construction will play an important role in what follows. The following lemma is key in the proof:

LEMMA 5.27. *Assume that $S \leq \Gamma_\Delta$ and there exists a finite set D such that $R(v, S)$ is finite for all $v \in \Gamma_\Delta \setminus (S \cup D)$. Then there exists $\Gamma_0 \leq S$ stably embedded in Γ_Δ. Moreover, we can choose Γ_0 such that $R(v, \Gamma_0)$ is finite for all $v \notin \Gamma_0$.*

PROOF. By induction on $|D|$, it suffices to prove the claim for $D = \{v\}$. First, we show:

CLAIM 5.28. *Up to finitely many points, $S_0 := S \setminus R(v, S)$ is a model, i.e., there exists a finite set F such that $S_0 \setminus F \cong \Gamma_\Delta$.*

PROOF. If there exist $w_1, \dots, w_k \in S$ such that $\left(\bigcup_{i=1}^{k} R(w_i, S)\right) \setminus R(v, S)$ is a finite set $\{s_1, \dots s_l\}$, we are done: given an independent set $v_1, \dots, v_n \in S_0$ and any $u_1, \dots, u_m \in S_0$, all distinct from the w_i we can realize

$$\bigwedge_{i=1}^{l} x \neq s_i \bigwedge_{i=1}^{k} x \neq w_i \bigwedge_{i=1}^{k} \neg R(x, w_i) \wedge \bigwedge_{i=1}^{m} \neg R(x, u_i) \wedge \bigwedge_{i=1}^{n} R(x, v_i)$$

in Γ_Δ, and since all parameters in the above formula are in S, also in S. Thus, $S_0 \setminus \{w_1, \dots, w_k, s_1, \dots, s_l\}$ is a model. So we may assume that this is not the case. Let $w_1, \dots, w_k$ be independent and $u_1, \dots, u_r$ any elements, all distinct from the w_i. By our assumption we can find $u \in R(v, S)$ not connected to any of the w_i and different from all the u_i. In Γ_Δ there must be some u' such that u' is connected to u, to all the w_i and to none of the u_i. But $u' \notin R(v, \Gamma_0)$, since otherwise u, u' and v would form a triangle. Since $S \leq \Gamma_0$ such a u' can already be found in S. □

We let $B_0 = \{v\}$ if there are no $w_1, \dots, w_k, s_1, \dots, s_l \in S$ such that $R_v^S \subseteq \bigcup_{i=1}^k R(w_i, S) \cup \{s_1, \dots, s_l\}$ and $B_0 := \{w_1, \dots, w_k, s_1, \dots, s_l\}$ if $w_1, \dots, w_k, s_1, \dots, s_l$ witness that this is not the case. We let $S_0 = S \setminus (R(v, S) \cup B_0)$. From the proof of the last claim we get that $S_0 \leq \Gamma_\Delta$ (and not only up to finitely many points). We can now repeat the proof of Lemma 5.26 to obtain the desired conclusion. We give the details, as they are slightly more delicate than above. We let $a_0 \in S_0 \setminus B_0$ be the first element in some (fixed) enumeration of Γ_Δ, $\{v_i\}_{i \in \omega}$. Assume we have constructed $\{a_0, \dots, a_n\}$ and sets $B_0, \dots, B_n$ such that:

(1) $a_i \in S_0$ for all i and $(a_0, \dots, a_n)$ is an increasing sequence.
(2) $(a_0, \dots, a_n) \cong (v_0, \dots, v_n)$.
(3) If $j < i$ and $w \in B_j$ then $\models \neg R(a_i, w)$.
(4) $B_{i+1} := \{x : a_i < x < a_{i+1}\}$.

As before, it will suffice to show that the construction can be continued. Since the construction is given once we define a_{n+1}, at each stage we have to realize in S_0 a formula of the following form:

$$\varphi(x, \bar{a}) \wedge \bigwedge_{i=0}^{l} \neg R(x, b_i)$$

where $\bar{a} \subseteq S_0$, none of the b_i are in S_0, and $\varphi(x, \bar{a})$ has infinitely many solutions. By Claim 5.28, S_0 is a model so there is no problem realizing $\varphi(x, \bar{a})$ in S_0. We will need to work a bit harder to find a realization satisfying also the formulas $\neg R(x, b_i)$.

Notice that $\varphi(x, \bar{a}) \wedge \bigwedge_{b_i \in B_0} \neg R(x, b_i)$ has infinitely many solutions in S (if $v \in B_0$ use Claim 5.28, otherwise use the fact that S is a model), and obviously no such solution can be in $R(v, S)$. Setting

$$I = \{i : 1 \leq i \leq l \wedge b_i \in \Gamma_\Delta \setminus S\}$$

recall that, by assumption, $R(b_i, S_0)$ is finite for all $i \in I$. Therefore, it will suffice to check that

$$\varphi(x, \bar{a}) \wedge \bigwedge_{i \notin I} \neg R(x, b_i)$$

has infinitely many solutions in S_0. By definition $i \notin I$ implies that $b_i \in R_v^S$. Assume without loss of generality, that

$$\varphi(x, \bar{a}) = \bigwedge_{i=1}^{n} R(x, a_i) \wedge \bigwedge_{i=n+1}^{k} \neg R(x, a_i).$$

If there exists $u \in R_v^S$ such that $\bigwedge_{i=1}^n \neg R(u, a_i)$ then, as in the proof of Claim 5.28 any $u' \in S$ connected to u, to all the a_i with $1 \leq i \leq n$ and to none of the a_i with $n + 1 \leq i \leq k$ is in S_0, and, as $S \leq \Gamma_\Delta$, we are done. So we may assume that there is no such u. But then

$$\bigwedge_{i=1}^{n} R(x, a_i) \wedge \bigwedge_{i=n+1}^{k} \neg R(x, a_i) \wedge \bigwedge_{b \in B_0} \neg R(x, b) \wedge \bigwedge_{i \notin I} \neg R(x, b_i)$$

is a formula in S, and it is clearly consistent. So it must have a solution in S, but such a solution must be in S_0. □

We can now show:

PROPOSITION 5.29. *Let $c : \Gamma_\triangle \to \{0,1\}$ be any coloring. Then there exists $\Gamma_0 \leq \Gamma_\triangle$ which is monochromatic and stably embedded in $\Gamma_\triangle$.*

PROOF. This is a slight variation on the proof of [**10**] showing that $\Gamma_\triangle$ is indivisible. As already mentioned above, the proof in [**10**] assures that if there is no monochromatic sub-model colored 0 then there is a stably embedded monochromatic sub-model colored 1. The idea of the proof in [**10**] is to try and construct inductively a monochromatic sub-model colored 0. If the first attempt fails, this is because we ran into a formula $\varphi_0(x, \bar{c}_0)$ all of whose realizations are colored 1. We fix such a realization, r_0, and retry the construction of the 0-colored sub-model, starting above r_0 (with respect to some enumeration of $\Gamma_\triangle$), collecting a new 1-colored vertex, r_1, if we fail again. We proceed with this construction until we either manage to construct a 0-colored sub-model or until we constructed a sequence $\{r_i\}_{i\in\omega}$ of 1-colored points. Collecting the vertices r_i with enough ingenuity we can assure that if all our attempts in constructing a 0-colored sub-model failed then $\{r_i\}_{i\in\omega}$ is a (stably embedded) 1-colored sub-model. Our main observation is that, essentially, the same proof works if instead of trying to construct any 0-colored sub-model we apply the construction of Lemma 5.27 to try and construct a stably embedded sub-model. Since the construction of the sequence $\{r_i\}_{i\in\omega}$ in case of failure is similar to that of [**10**], this assures that indeed $\Gamma_\triangle$ is indivisible in the stably embedded sub-models. As the proof is not long we give the details.

Fix an enumeration $\{v_i\}_{i\in\omega}$ of $\Gamma_\triangle$, and again we refer to this enumeration as inducing an ordering on the vertices of $\Gamma_\triangle$. We try to construct a monochromatic copy of $\Gamma_\triangle$ colored 0, as provided by Lemma 5.26, the only difference with respect to the construction there, is that — in addition — we require that all the elements in the construction of Γ_0 are colored 0. If we succeed, we are done. So we may assume that we have failed, at some finite stage j_0. Thus, at stage j_0 we have constructed a set $Y_0 = \{y_{0,0}, \ldots, y_{0,j_0}\}$ isomorphic to an initial segment of $\Gamma_\triangle$ and a set $B_0 = \bigcup_{k=0}^{j_0} B_{0,k}$ such that

(1) $Y_0 \cup B_0$ is an initial segment of $\Gamma_\triangle$ (not only isomorphic to one).
(2) For every $b \in B_{0,k}$ if $j > k$ then $\models \neg R(b, y_{0,j})$.
(3) If z is such that $\mathrm{tp}(Y_0, z) = \mathrm{tp}(v_0, \ldots, v_{j_0+1})$ and $\models \bigwedge_{b\in B_0} \neg R(b, z)$ then $c(z) = 1$.

We fix some z_0 as in (3) above and such that $z_0 > B_0 \cup Y_0$ and restart the construction anew, above z_0. That is, we do not give up after one failure, or indeed after any finite number of failures. We keep trying over and over again, keeping track of the information we gathered: each failure gives us a formula all of whose realizations are colored 1. So each failure could, if we are careful enough, advance us one more step towards the construction of a monochromatic sub-model colored 1. We will now give the technical details explaining how this is done. Assume we have constructed for all $i < n$ sets Y_i isomorphic to an initial segment of $\Gamma_\triangle$, $B_i = \bigcup_{k=0}^{j_i} B_{i,k}$ and elements z_i satisfying conditions (1)-(3) above (for the index i, rather than 0, of course), and such that, in addition:

(4) For all $j < i$, if $b \in B_j$ then $\models \neg R(b, y_{i,k})$ for all $k < j_i$.
(5) $Y_j > Y_i \cup B_i \cup \{z_i\}$ for all $i < j$.

We now fix $y_{n,0} > Y_{n-1} \cup B_{n-1} \cup \{z_{n-1}\}$, and start all over again: set $B_{n,0} := \{v : v < y_{n,0}\}$ and try to construct inductively a 0-colored sequence $y_{n,0} < y_{n,1} \ldots$ with

$(y_{n,0}, \ldots, y_{n,i}) \cong (v_0, \ldots, v_i)$ and a collection of sets $B_{n,i} = \{v : y_{n,i-1} < v < y_{n,i}\}$ for $0 < i$ satisfying (2) above. If we do not get stuck, i.e., if we manage to construct $\{y_{n,i}\}_{i\in\omega}$ as above, then setting $\Gamma_0 := \{y_{n,i}\}_{i\in\omega}$ the construction gives $\Gamma_0 \leq \Gamma_\Delta$ which is monochromatic 0, and such that for all $v \geq y_{n_0}$, if $v \notin \Gamma_0$ then $R(v, \Gamma_0)$ is finite. By the previous lemma we can find $\Gamma_1 \leq \Gamma_0$ which is stably embedded in Γ_Δ, and we are done.

So we may assume that the construction fails at some stage j_n. In that case we choose z_n satisfying the following conditions:

(1) If $\mathrm{tp}(v_n/v_0, \ldots, v_{n-1})$ is generic we choose z_n such that $\mathrm{tp}(Y_n, z_n) = \mathrm{tp}(v_0, \ldots, v_{j_n})$ and $\models \bigwedge_{i=0}^{n-1} \neg R(z_i, z_n)$.
(2) If $\models R(v_n, v_i)$ for some $i < n$ let $i_0 + 1$ be minimal such that this is true. We choose z_n such that $\mathrm{tp}(Y_{i_0}, z_n) = \mathrm{tp}(v_0, \ldots, v_{j_{i_0}})$ and $\mathrm{tp}(z_0, \ldots, z_n) = \mathrm{tp}(v_0, \ldots, v_n)$.

By construction of the Y_i, if z_n exists then $c(z_n) = 1$. Clearly, if case (1) holds, we have no problem finding z_n satisfying the requirements. So we check what happens in case (2). We will not be able to find such z_i only if we are required to connect z_n to two vertices, w_1, w_2 which are already connected. Obviously, w_1, w_2 cannot both be in Y_{i_0} or both in $\{z_0, \ldots, z_{n-1}\}$, since z_n is required to satisfy a consistent type over both sets. So $w_1 \in Y_{i_0}$ and $w_2 = z_j$ for some j. By minimality of $i_0 + 1$, since $\models R(z_j, w_1)$, it follows that $\models R(z_{i_0+1}, z_j)$. Surely, we can find z_n such that $\mathrm{tp}(z_0, \ldots, z_n) = \mathrm{tp}(v_0, \ldots, v_n)$. Then, $z_n \models R(x, z_{i_0+1}) \wedge R(x, z_j)$, creating a triangle in Γ_Δ, which is impossible.

Thus, either at some stage we manage to construct our 0-colored copy of Γ_Δ, and we can conclude using Lemma 5.27, or $\Gamma_0 := \{z_i\}_{i\in\omega} \cong \Gamma_\Delta$ is a 1-colored elementary substructure of Γ. So it remains only to check that in the latter case Γ_0 is stably embedded. So let $v \in \Gamma_\Delta \setminus \Gamma_0$. If $v \notin \bigcup Y_i$ then $R(v, \Gamma_0)$ is finite, because $v \in B_i$ for some i, and therefore $R(v, \Gamma_0) \subseteq \{z_0 \ldots, z_i\}$. So we may assume that $v \in Y_i$ for some i. In that case $z_j \in R_v^{\Gamma_0}$ if one of two cases happen:

(1) $j < i$ and $\Gamma_\Delta \models R(z_j, v)$ or
(2) $i < j$ and $\bigwedge_{k\leq i} \neg R(z_j, z_k)$ and $\mathrm{tp}(z_j, v_0, \ldots, v_{j_i}) = \mathrm{tp}(v_{j_i+1}, v_0, \ldots, v_{j_i})$

Since each of the two conditions is clearly definable (given v) with parameters in Γ_0, we are done. □

Combining this with the earlier results of this section we get:

THEOREM 5.30. *The generic countable triangle free graph is symmetrically indivisible.*

PROOF. Let $c : \Gamma_\Delta \to \{0, 1\}$ be any coloring. Let $S \leq \Gamma_\Delta$ be a monochromatic stably embedded sub-model, as provided by the last proposition. Then, if we can show that S satisfies the assumptions of Lemma 5.17 (with $D = \emptyset$), the conclusion will follow. Indeed, let $\varphi(x)$ be any formula over S and $b_1, \ldots, b_n \in \Gamma_\Delta \setminus S$. Because S is stably embedded the set $B := \bigwedge \neg R(S, b_i)$ is a definable subset of S. Observe that the set $\bigvee R(\Gamma_\Delta, b_i)$ does not contain a copy of Γ_Δ, so it is non-generic. Therefore $S \setminus B = \bigvee R(S, b_i)$ cannot contain a model, and is therefore non-generic. Thus, since B is definable and its complement (in S) does not contain a model, B itself must be generic. But the intersection of two definable sets, one of which is generic is never empty, thus, $\varphi(B) \neq \emptyset$, as required. □

6. Elementary Symmetric indivisibility

In the examples studied in the previous sections the structure of interest had quantifier elimination in a natural language, suggesting a natural interpretation of the notion of indivisibility. But in general, indivisibility would depend in a significant way on our choice of language. For example, the structure $(\mathbb{N}, \leq)$ is trivially indivisible, but after adding a binary predicate $S(x, y)$ interpreted as "y is the immediate successor of x", it is no longer indivisible. From the model theoretic view point, in order to allow a more general discussion of indivisibility, one needs to strengthen the definition:

Definition 6.1. Let $\mathcal{M}$ be a (countable) structure. Say that $\mathcal{M}$ is *elementarily indivisible* if for any coloring of M in two colors there exists a monochromatic elementary substructure $\mathcal{N} \prec \mathcal{M}$ isomorphic to $\mathcal{M}$.

This definition releases the analysis of indivisibility from the dependence on the choice of language. As was pointed out to us by the referee, in most examples of interest indivisibility is studied in structures not only admitting quantifier elimination in a natural language, but which are also $\aleph_0$-categorical. In such cases the isomorphism requirement in the definition of indivisibility is automatic. In the abstract setting of elementary indivisibility, it may be of interest to study also *weak elementary indivisibility*, where the isomorphism condition is dropped from Definition 6.1.

In this section we study the basic properties of elementary indivisibility. Unfortunately, the results of this study seem mostly negative in the sense that most of our classification attempts ran into counter-examples constituting the main bulk of this section. These examples suggest that our understanding of elementary indivisibility is still far from being satisfactory. Most annoyingly, we were unable to settle the following:

Question 6.2. Is there an elementarily indivisible structure that is not symmetrically indivisible?

Since the notions of elementary divisibility and weak elementary divisibility coincide in the context of $\aleph_0$-categorical structures, it is natural to ask whether these notions coincide. To address this question we show, first, that $\aleph_0$-categoricity is not a necessary condition for elementary indivisibility:

Example 6.3. Let $\mathcal{C}$ be the class of all complete finite graphs in ω-many edge colors (so every $G \in \mathcal{C}$ is colored in finitely many colors, but there are infinitely many non-isomorphic structures on two elements). Clearly, $\mathcal{C}$ has the amalgamation property, but for any $G_0 \leq G_i$ $(i = 1, 2)$, colored graphs in $\mathcal{C}$ there are countably many possible amalgams of G_1 with G_2 over G_0 whose universe is $G_1 \amalg_{G_0} G_2$. Let $\mathcal{G}$ be the Fraïssé limit of $\mathcal{C}$ with respect to all such amalgams. More precisely, for every finite $G_1 \leq \mathcal{G}$ and $G_2 \in \mathcal{C}$, if $G_0 \hookrightarrow G_2$ then for an amalgam $G \in \mathcal{C}$ of G_1 with G_2 over G_0 there is an embedding $g : G_2 \hookrightarrow \mathcal{G}$, such that $g \restriction G_0 = \mathrm{id}$ and $G \cong G_1 \cup g(G_2)$ (with the induced structure).

Observe that this last property implies that if $G_1 \leq \mathcal{G}$ and $G_1 \leq G_2 \in \mathcal{C}$ then there is an embedding of G_2 into $\mathcal{G}$ over G_1. Thus, every complete countable graph on ω colors embeds into $\mathcal{G}$. In particular $\mathcal{G}[\mathcal{G}] \hookrightarrow \mathcal{G}$. So $\mathcal{G}$ is self similar and thus indivisible, but clearly not $\aleph_0$-categorical.

In fact, using the same construction as in the proof of Corollary 2.11 we can easily show that $\mathcal{G}$ is symmetrically self-similar and therefore symmetrically indivisible.

We get:

COROLLARY 6.4. *Elementary indivisibility is stronger than weak elementary indivisibility.*

PROOF. Let $\mathcal{G}_0$ be the graph obtained in the previous example with the color removed from one edge $e := (v_1, v_2)$. Then $\mathcal{G}_0$ is weakly elementarily indivisible, since $\mathcal{G}_0 \equiv \mathcal{G}$. It is not elementarily indivisible because we can separate v_1 from v_2 (and no other edges are colorless). $\square$

REMARK 6.5. The example in the above corollary also shows that a theory T may have two countable models, one elementarily indivisible and the other not.

Another corollary of the above example was supplied to us by the referee:

COROLLARY 6.6. *There exists an elementarily indivisible structure in a finite relational language which is not $\aleph_0$-categorical.*

PROOF. Let $\mathcal{G}$ be the colorful graph of Example 6.3. Let $\chi : [G]^2 \to \mathbb{N}$ be the function taking an edge e to its color $n \in \mathbb{N}$. For any binary relation S on $\mathbb{N}$ we define a 4-ary relation R^S on G by setting $R^s(x_1, y_1, x_2, y_2) \iff S(\chi(x_1, y_1), \chi(x_2, y_2))$. We let $\mathcal{G}^S$ be the structure with the same universe G, and whose unique relation is R^S.

Let S be the standard order on $\mathbb{N}$. We claim that $\operatorname{Aut}(\mathcal{G}^S) = \operatorname{Aut}(\mathcal{G})$. Obviously, $\operatorname{Aut}(\mathcal{G}) \subseteq \operatorname{Aut}(\mathcal{G}^S)$. For the other inclusion, we note that if $\alpha \in \operatorname{Aut}(\mathcal{G}^S)$ then α induces an automorphism of $(\mathbb{N}, S)$. Indeed, if e_1, e_2 are any two edges with $\chi(e_1) = \chi(e_2)$ then $\mathcal{G}^S \models \neg(S(e_1, e_2) \vee S(e_2, e_1))$, implying that $\chi(\alpha(e_1)) = \chi(\alpha(e_2))$ so α induces a bijection $\tilde{\alpha} : \mathbb{N} \to \mathbb{N}$, which is, by definition, an automorphism. Because $(\mathbb{N}, S)$ is rigid, $\tilde{\alpha} = \mathrm{id}$ for all α, thus, $\alpha \in \operatorname{Aut}(\mathcal{G})$.

Since $\mathcal{G}$ is indivisible so is $\mathcal{G}^S$. Since they have the same automorphism groups and $\mathcal{G}$ is not $\aleph_0$-categorical neither is $\mathcal{G}^s$. $\square$

We were also unable to find answers to either of:

QUESTION 6.7. Is there a rigid elementarily indivisible structure? Is an elementarily indivisible structure homogeneous?

The only result we have in this direction is the observation:

REMARK 6.8. If $\mathcal{M}$ is weakly elementarily indivisible then $|S_1(\emptyset)| = 1$, i.e., there are no formulas in one variable (with no parameters), except the formula $x = x$.

PROOF. If $p_1, p_2 \in S_1(\emptyset)$ choose $\varphi(x) \in p_1$ such that $\neg\varphi(x) \in p_2$ and color $\varphi(M)$ red and $\neg\varphi(M)$ blue. Since $\mathcal{M} \models (\exists x, y)(\varphi(x) \wedge \neg\varphi(y))$, neither $\varphi(M)$ nor $\neg\varphi(M)$ can contain any elementary substructure of $\mathcal{M}$ (let alone one which is isomorphic to $\mathcal{M}$). $\square$

Another restriction on elementarily indivisible structures is:

PROPOSITION 6.9. *If $\mathcal{M}$ is elementarily indivisible, then for every $A \subseteq M$, if $|A| < \aleph_0$ then $|\operatorname{acl}(A)| < \aleph_0$.*

Proof. We will show that if $|\operatorname{acl}(A)| = \aleph_0$ for some finite $A \subseteq M$ then M is not elementarily indivisible. So let $A \subseteq M$ be as in the assumption. Let $\{A_i\}_{i\in\omega}$ enumerate realizations of $\operatorname{tp}(A/\emptyset)$ (with respect to some enumeration of the elements of A). To show that $\mathcal{M}$ is divisible it will suffice to show that there is a coloring of M with the property that $\operatorname{acl}(A_i)$ is not monochromatic for all i. This can be done using a simple diagonalisation: Color A_0 blue and choose $a_0 \in \operatorname{acl}(A_0)$. Color a_0 red. Assume that for all $i < n$ we colored A_i and a point $a_i \in \operatorname{acl}(A_i)$ in such a way that $A_i \cup a_i$ is not monochromatic, we will now color A_n. Let $A = \bigcup_{i<n}(A_i \cup a_i)$. Then A is finite. Choose $a_n \in \operatorname{acl}(A_n)\setminus A$ (this is possible because $\operatorname{acl}(A_n)$ is infinite and A is finite). Color $A_n \cup a_n$ so that it is not monochromatic. Since $\bigcup_{i<\omega} A_i = m$ this defines a coloring of M, and we are done. □

We do not know whether Proposition 6.9 remains true for weakly elementarily indivisible structures. Observe, however, that if A is a singleton, then by Proposition 6.8 the same proof can be carried out for weakly elementarily indivisible structures. The following is an immediate corollary, which remains true for weakly indivisible structures:

Corollary 6.10. *If $\mathcal{M}$ is weakly elementarily indivisible there are no non-trivial $\emptyset$-definable equivalence relations (on M) with finite classes.*

We get

Corollary 6.11. *If $\mathcal{M}$ is weakly elementarily indivisible then* $\operatorname{acl}(a) = a$ *for every $a \in M$. In particular, there are no $\emptyset$-definable non-trivial definable unary functions.*

Proof. Note that if $x \in \operatorname{acl}(y)$ then $y \in \operatorname{acl}(x)$, for $\operatorname{acl}(x) \subseteq \operatorname{acl}(y)$. But $\operatorname{tp}(x) = \operatorname{tp}(y)$ (by Proposition 6.8), so equality must hold. By Claim 6.9 (see the discussion preceding Corollary 6.10 for the weakly elementarily indivisible case), if for some (equivalently, all) $a \in M$ we would have $\operatorname{acl}(a) \supsetneq a$ then $x \sim y$ if and only if $x \in \operatorname{acl}(y)$ would be a non-trivial equivalence relation with finite classes, contradicting 6.10. □

As we have seen in Section 4 it is not true that in an elementarily (symmetrically) indivisible structure the algebraic closure operator is totally trivial. We give some more examples:

Example 6.12.

(1) Let A be an infinite affine space over $\mathbb{F}_2$. Then A is symmetrically indivisible. This is an immediate corollary of Hindman's theorem on finite sums, [**8**]: Fix a coloring c of A in two colors $\{0,1\}$ and a point $a \in A$. Let V_a be the vector space obtained by localizing A in a. By Hindman's theorem there exists an infinite set $X \subseteq V_a$ and $i \in \{0,1\}$ such that for all finite $F \subseteq X$ we have $c(\sum_{x\in F} x) = i$. Since V_a is a vector space over $\mathbb{F}_2$, this implies that X, a generate a monochromatic (except, possibly at a), infinite dimensional sub-vector space $U_a \subseteq V_a$. Let $B \subseteq A$ be any infinite dimensional affine subspace of U_a not containing a. Then B is monochromatic and symmetric in A.

(2) It is apparently well known that the above is not true (in a strong sense) for affine spaces over any other (finite) field, see Theorem 7 of [**11**].

(3) The following was suggested to us by H. D. Macpherson: Let $G = [\mathbb{N}]^2$. Define a graph on G by setting $E := \{\{a, b\}, \{a, c\} : c \neq b\}$. Let c be any coloring of G in two colors. By Ramsey's theorem there is an infinite set $I \subseteq \mathbb{N}$ such that $c|[I]^2$ is constant. Thus, the induced graph on $[I]^2$ is monochromatic. Obviously, $[I]^2$ is isomorphic to G (with any bijection $f : \mathbb{N} \to I$ inducing an isomorphism). Note that every permutation of $\mathbb{N}$ induces an automorphism of G. But the other direction is also true. Define, an equivalence relation $\sim$ on E by $(x, y) \sim (w, z)$ if the graph induced on $\{x, y, w, z\}$ is complete. Observe that if $x = \{n, m\}$ and $y = \{n, k\}$ then $(\{n', m'\}, \{n', k'\})$ is equivalent to (x, y) if and only if $n = n'$. This gives a bijection $f : E/\sim \to \mathbb{N}$, and since E is a definable equivalence relation, it follows that any automorphism σ of G induces, through f, a permutation f_σ on $\mathbb{N}$, which - in turn - induces σ. So $[I]^2$ is symmetric in G. It is not hard to check that $\mathrm{Th}(G)$ has quantifier elimination, and thus $[I]^2 \leq G$.

The following question remains open:

QUESTION 6.13. Let $\mathcal{M}$ be a symmetrically indivisible structure in a language $\mathcal{L}$. Let $\mathcal{L}_0 \subseteq \mathcal{L}$. Is $\mathcal{M} \restriction \mathcal{L}_0$ symmetrically indivisible?

It is obvious, of course, that a reduct of an indivisible structure is indivisible, but since a reduct of a symmetric sub-structure need not be symmetric (see the example below), it is not clear what one should expect as an answer to this question.

EXAMPLE 6.14. Let Γ_0 be a countable random graph in the language $\mathcal{L} = \{R\}$. Expand the language by constants $\{c_i\}_{i\in\omega}$ enumerating Γ_0. Let $\Gamma_0 \leq \Gamma$ be a countable elementary extension. Assume that in Γ there is some infinite co-infinite $S \subseteq \omega$ and an element $b \in \Gamma$ such that $\Gamma \models R(b, c_i)$ if and only if $i \in S$.

Because Γ_0 is rigid, it is symmetric in Γ. If the set $\{c_i : i \in S\}$ is not Γ_0-definable, then its orbit under $\mathrm{Aut}(\Gamma_0)|\mathcal{L}$ is uncountable (Proposition 5.4). So as pure random graphs, Γ_0 is not symmetric in Γ.

References

[1] Peter Cameron. The pigeonhole principle. Available at the URL *http://cameroncounts.wordpress.com/2010/10/11/the-pigeonhole-property/*

[2] Zoé Chatzidakis and Ehud Hrushovski. Model theory of difference fields. *Trans. Amer. Math. Soc.*, 351(8):2997–3071, 1999.

[3] Gregory Cherlin and Ehud Hrushovski. *Finite structures with few types*, volume 152 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2003.

[4] M. El-Zahar and N. Sauer. On the divisibility of homogeneous hypergraphs. *Combinatorica*, 14(2):159–165, 1994.

[5] C. Ward Henson. A family of countable homogeneous graphs. *Pacific J. Math.*, 38:69–83, 1971.

[6] W. Hodges.*Model theory.* Encyclopedia of Mathematics and its Applications, 42. Cambridge University Press, Cambridge, 1993.

[7] Bernhard Herwig. Extending partial isomorphisms for the small index property of many ω-categorical structures. *Israel J. Math.*, 107:93–123, 1998.

[8] Neil Hindman. Finite sums from sequences within cells of a partition of *N. J. Combinatorial Theory Ser. A*, 17:1–11, 1974.

[9] Menachem Kojman and Stefan Geschke. Symmetrized induced ramsey theorems. Available on http://www.math.bgu.ac.il/~kojman/SymPart.pdf, 2008.

[10] Péter Komjáth and Vojtěch Rödl. Coloring of universal graphs. *Graphs Combin.*, 2(1):55–60, 1986.

[11] Claude Laflamme, Lionel Nguyen Van Th'e, Maurice Pouzet, and Norbert Sauer. Partitions and indivisibility properties of countable dimensional vector spaces. Available on http://arxiv.org/pdf/0907.3771v1, 2009.

Department of Mathematics, Ben Gurion University of the Negev, Be'er-Sheva 84105, Israel
E-mail address: hassonas@math.bgu.ac.il

Department of Mathematics, Ben Gurion University of the Negev, Be'er-Sheva 84105, Israel and Institute for Advanced Study, Princeton, NJ, 08540
E-mail address: kojman@math.bgu.ac.il

Universidad de los Andes, Departemento de Matemáticas, Cra. 1 No 18A-10, Bogotá, Colombia
E-mail address: onshuus@gmail.com

Contemporary Mathematics
Volume **558**, 2011

Partitions and Permutation Groups

Andreas Blass

Abstract. We show that non-trivial extremely amenable topological groups are essentially the same thing as permutation models of the Boolean prime ideal theorem that do not satisfy the axiom of choice. Both are described in terms of partition properties of group actions.

1. Introduction

The purpose of this paper is to point out that two results, in apparently quite different areas, become the same when reduced to their combinatorial essence. The first is the result of Herer and Christensen [**4**] on the existence of extremely amenable groups, with some additional information from Pestov [**10**]. The second is the result of Halpern [**2**] that the Boolean prime ideal theorem does not imply the axiom of choice, in a set theory allowing atoms, with some additional information from Halpern's proof.

Both of these results are equivalent to the existence of groups and filters of subgroups with particular combinatorial properties. In the case of extremely amenable groups, the equivalence in question is (almost) contained in the work of Kechris, Pestov, and Todorčević [**7**]. In the case of the Boolean prime ideal theorem, the required equivalence is (almost) contained my paper [**1**].

In Sections 2 and 3 of the present paper, we review background material about extremely amenable groups and about the Boolean prime ideal theorem. In Section 4, we define the relevant combinatorial properties of group actions. We prove some basic facts about these properties, and we also raise an open question. Finally, in Section 5, we recall the results from [**7**] and [**1**] mentioned above, and we make the minor modifications needed to establish the claimed equivalences.

This arrangement of the material tends to obscure the connection with the topics of this volume, finite combinatorics and model theory. Nevertheless, finite combinatorics is, as will become clear from the examples in Section 4, the essence of the group-theoretic properties discussed here. There are also implicit connections with model theory, not only in the direction of models of set theory, but also in the crucial role of the compactness theorem in the results proved in [**1**] and used here in Section 5.

1991 *Mathematics Subject Classification.* Primary 05C55; Secondary 03E25, 03E35, 22F05.
Partially supported by NSF grant DMS-0653696.

Throughout this paper all topological spaces (including topological groups) are assumed to be Hausdorff spaces unless the contrary is explicitly stated.

2. Extremely Amenable Groups

An *action* of a group G on a set X is a function $\alpha : G \times X \to X$ such that, abbreviating $\alpha(g, x)$ as $g \cdot x$ and writing 1 for the identity element of G, we have $g_1 \cdot (g_2 \cdot x) = (g_1 g_2) \cdot x$ and $1 \cdot x = x$ for all $g_1, g_2 \in G$ and all $x \in X$. Equivalently, $g \mapsto (x \mapsto \alpha(g, x))$ is a homomorphism from G into the group of permutations of X. In this situation, we call X (understood as being equipped with the action) a *G-set*.

When G is a topological group and X is a topological space, we call an action α continuous if it is continuous with respect to the product topology on $G \times X$ (and the given topology on X).

Definition 2.1. A topological group is *extremely amenable* if, whenever G acts continuously on a compact Hausdorff space X, there is a fixed point in X, i.e., an $x \in X$ such that $g \cdot x = x$ for all $g \in G$.

The first question to ask about this concept was asked by Mitchell [**8**]: Are there non-trivial examples? (Of course the trivial group is extremely amenable, by the second equation in the definition of "action.") This question was answered affirmatively by Herer and Christensen [**4**] by a direct, hands-on construction. More natural examples were given by Pestov [**10**], and they had the following additional property.

Definition 2.2. A topological group has *small open subgroups* if every neighborhood of 1 includes an open subgroup.

Equivalently, the open subgroups constitute a neighborhood base at 1, so they determine the topology (because a neighborhood base at any point in a topological group can be obtained by simply translating a neighborhood base at 1).

Pestov gave, as an example of an extremely amenable group, $\mathrm{Aut}(\mathbb{Q}, <)$, the group of order-automorphisms of the rational number line. The topology is the subspace topology obtained by regarding $\mathrm{Aut}(\mathbb{Q}, <)$ as a subset of the product space $\mathbb{Q}^{\mathbb{Q}}$, which in turn has the product topology induced by the discrete topology on $\mathbb{Q}$. A neighborhood base at 1 is given by the pointwise stabilizers of finite sets, i.e., the subgroups of the form

$$\mathrm{Fix}(F) = \{\pi \in \mathrm{Aut}(\mathbb{Q}, <) : (\forall q \in F)\, \pi(q) = q\}$$

for finite subsets F of $\mathbb{Q}$. Thus, $\mathrm{Aut}(\mathbb{Q}, <)$ has small open subgroups.

We summarize the preceding results of Herer, Christensen, and Pestov as follows.

Theorem 2.3. *There exists a non-trivial extremely amenable topological group with small open subgroups.*

3. Boolean Prime Ideal Theorem

The *Boolean prime ideal theorem* (BPI) is the assertion that every non-degenerate Boolean algebra has a prime ideal. ("Non-degenerate" means that the algebra has at least two elements; equivalently $0 \neq 1$.) BPI is often stated in terms of ultrafilters rather than prime ideals; it makes no difference because ultrafilters are just the complements of prime ideals.

The usual proof of BPI uses Zorn's Lemma to produce a maximal ideal and then checks that all maximal ideals are prime. Thus, BPI is a consequence of the axiom of choice,[1] and it is natural to ask whether it is equivalent to AC in the presence of the other axioms of Zermelo-Fraenkel set theory (ZF). This question was answered negatively by Halpern [**2**] in the context of ZFA, a version of ZF that allows the existence of atoms (objects that are not sets and therefore do not have elements but can be elements of sets). Later, Halpern and Lévy [**3**] improved the result to work with ZF (without atoms), but it is the original result of Halpern that will be relevant for us here.

Halpern's proof used the Fraenkel-Mostowski-Specker technique of *permutation models* of ZFA, which we briefly review here. See [**6**, Chapter 4] for details. The method begins with a set A of atoms, a group G of permutations of A, and a normal filter $\mathcal{F}$ of subgroups of G. "Normal filter" means that $\mathcal{F}$ is a collection of subgroups of G, closed under supergroups, under finite intersections, and under conjugation by elements of G. One builds a universe $V(A)$ of sets and atoms by starting with the atoms in A and iterating transfinitely the operation of forming arbitrary sets of already formed elements. That is, one defines, by induction on the ordinal ξ,

$$V_\xi(A) = A \cup \bigcup_{\eta<\xi} \mathcal{P}(V_\eta(A))$$

where $\mathcal{P}(X)$ means the set of all subsets of X. Then $V(A)$ is defined as the union of the sets $V_\xi(A)$ over all ordinals ξ.

Any permutation π of A, in particular any element of G, naturally extends to an automorphism (preserving the membership relation $\in$) of $V(A)$ by the inductive definition

$$\pi(x) = \{\pi(y) : y \in x\} \quad \text{for all sets } x.$$

The *stabilizer* of an $x \in V(A)$ is defined as the subgroup $\{\pi \in G : \pi(x) = x\}$ of G.[2] An element $x \in V(A)$ is said to be *symmetric* if its stabilizer is in the given filter $\mathcal{F}$; it is *hereditarily symmetric* if, in addition, all elements of its transitive closure (i.e., the elements of x, their elements, their elements, etc.) are symmetric. The subuniverse $M = M(A, G, \mathcal{F}) \subseteq V(A)$ of hereditarily symmetric sets, with the membership relation $\in$ inherited from $V(A)$, satisfies ZFA. It is called a permutation model of ZFA, and, except in very special cases, it does not satisfy AC.

The method of permutation models is used to establish independence results between AC and its various weak forms by cleverly choosing A, G, and $\mathcal{F}$ so as to make certain weak axioms of choice hold in $M(A, G, \mathcal{F})$ while others fail. Halpern [**2**] showed that BPI holds while AC fails in a particular permutation model, the ordered Mostowski model [**9**] obtained by choosing the parameters as follows. A is a countable set equipped with a dense linear ordering without endpoints, i.e., an order-isomorphic copy of $(\mathbb{Q}, <)$. G is the group of order-automorphisms of A. $\mathcal{F}$ is the filter generated by the pointwise stabilizers $\mathrm{Fix}(F)$ of finite subsets F of A.

---

[1] BPI is one of the most useful consequences of AC. See for example the long list of its equivalent forms in the compendium by Howard and Rubin [**5**, Form 14].

[2] Note that, in the notation of the previous section, this would be $\mathrm{Fix}(\{x\})$, not $\mathrm{Fix}(x)$; the latter would be the smaller subgroup of those π that fix every element of x individually, whereas a π in the stabilizer of x could permute the elements of x with each other.

We summarize Halpern's result as follows.

Theorem 3.1. *There is a permutation model of ZFA that satisfies the Boolean prime ideal theorem but not the axiom of choice.*

4. Ramsey Actions

The thesis to be explained in the rest of this paper is that Theorems 2.3 and 3.1 are essentially the same theorem. A certain similarity between them is already apparent in the specific examples used to prove them. Pestov's and Halpern's arguments both use $\mathrm{Aut}(\mathbb{Q}, <)$. Our thesis, however, is much stronger: *Any* example for either of these theorems is, after some minor normalizations, also an example for the other. To justify this claim, we must look at the combinatorial core of both theorems; that core is the subject of the present section.

We begin with the notion of a Ramsey action, introduced in [**1**]. The following is not literally the definition from [**1**], but the equivalence between the definitions is included in Proposition 4.2 below.

Definition 4.1. An action of a group G on a set X is a *Ramsey action* and is said to have the *Ramsey property*, and X is a *Ramsey* G-set, if, for every finite subset F of X and every 2-coloring $c : X \to 2$, there is some $g \in G$ such that c is constant on $g \cdot F$.

We use here and below the standard convention from set theory identifying a natural number with the set of smaller natural numbers, so 2 means $\{0, 1\}$. We also use the standard notation $g \cdot F$ for $\{g \cdot x : x \in F\}$. Thus, the definition of a Ramsey action says that, given a 2-coloring of X, every finite subset of X can be translated (on the left, by an element of G) to a monochromatic set.

Before giving some examples of Ramsey actions, we point out some elementary facts about the concept. We begin with some equivalent reformulations of the definition.

Proposition 4.2. *For any action of a group G on a set X, the following are equivalent.*

(1) *The Ramsey property.*
(2) *For each finite $F \subseteq X$ there exists a finite $Y \subseteq X$ such that, whenever $c : Y \to 2$, there is $g \in G$ such that $g \cdot F \subseteq Y$ and c is constant on $g \cdot F$.*
(3) *For every finite k, the definition of Ramsey action is satisfied with k-colorings in place of 2-colorings.*
(4) *For every finite k, item* (2) *above is satisfied with k-colorings in place of 2-colorings.*

Proof. Some of the implications are trivial: (4) implies both (2) and (3), each of which implies (1).

The implication from (1) to (2) is established by the following standard compactness argument. If we had a counterexample to (2), i.e., an F for which no appropriate Y exists, then the compactness theorem for sentential logic would apply to the following set of sentences. Let the sentential variables be [x gets color i]

for each $x \in X$ and $i \in 2$, and let Σ be the set of sentences

$$\bigvee_{i\in 2}[x \text{ gets color } i] \qquad \text{for each } x \in X,$$

$$\neg([x \text{ gets color } i] \wedge [x \text{ gets color } j]) \qquad \text{for each } x \in X \text{ and distinct } i, j \in 2,$$

$$\neg \bigvee_{i\in 2} \bigwedge_{x\in F} [g \cdot x \text{ gets color } i] \qquad \text{for each } g \in G.$$

Our choice of F implies that each finite subset of Σ is satisfiable. By compactness, Σ is satisfiable, and any satisfying truth assignment describes a coloring which, together with F, witnesses the failure of (1). The same proof, with all occurrences of 2 changed to k, shows that (3) implies (4).

To complete the proof of the proposition, we show that (2) implies (4), by induction on k. Assume (2) and assume (4) for a certain $k \geq 2$. Given any finite $F \subseteq X$, use the induction hypothesis to find a finite $Y \subseteq X$ as in (4) for k. It does no harm to enlarge Y so that it includes F. Then use (2), with Y in the role of F, to find a finite $Z \subseteq X$ such that each 2-coloring of Z is constant on a set of the form $g \cdot Y$. We shall show that each $(k+1)$-coloring c of Z is constant on a set of the form $g \cdot F$.

Given $c : Z \to k+1$, form $c' : Z \to 2$ by setting

$$c'(z) = \begin{cases} 0 & \text{if } c(z) \neq k \\ 1 & \text{if } c(z) = k. \end{cases}$$

By our choice of Z, we can fix some $g \in G$ such that c' is constant on $g \cdot Y$. If the constant is 1, then c is constant with value k on $g \cdot Y$ and therefore on its subset $g \cdot F$. Assume, therefore, that the constant value of c' on $g \cdot Y$ is 0. This means that c maps $g \cdot Y$ into k (rather than $k+1$), and therefore the function c'' defined by

$$c''(y) = c(g \cdot y)$$

maps Y into k. By our choice of Y, we can fix some $h \in G$ such that c'' is constant on $h \cdot F$. Then c is constant on $(gh) \cdot F$. □

Item (2) in the preceding proposition was used as the definition of "Ramsey property" in [**1**].

Proposition 4.3. *In any Ramsey action of a group G on a set X, the action is transitive and, if X has at least two elements, then it has infinitely many.*

Proof. Recall that transitivity means that, for any two elements $x, y \in X$, there is some $g \in G$ with $g \cdot x = y$. To prove that this is necessary in any Ramsey action, let x and y be given and apply the definition of Ramsey action with F being $\{x, y\}$ and c having value 0 at all points of the form $g \cdot x$ and value 1 at all other points. We get some $h \in G$ such that

$$c(h \cdot y) = c(h \cdot x) = 0,$$

and so $h \cdot y = g \cdot x$ for some $g \in G$. Then $(h^{-1}g) \cdot x = y$; this completes the proof of transitivity.

If X were finite but had at least two points, then, in the definition of Ramsey action, we could take F to be all of X and take c to be a non-constant function $X \to 2$. The definition requires c to be constant on a set of the form $g \cdot X$, but the only such set, regardless of g, is X itself, on which c is not constant. □

In the definition of Ramsey action, if the requirements are satisfied for a particular finite set F, then they are obviously also satisfied for any subset of F. The following proposition begins with this observation and extends it in ways that will be useful in verifying some examples.

PROPOSITION 4.4. *Let G act on X, and let $\mathcal{F}$ be a family of finite subsets of X. Under any of the following hypotheses, to verify the Ramsey property, it suffices to verify the definition for $F \in \mathcal{F}$.*

(1) *Each finite subset of X is included in a member of $\mathcal{F}$.*
(2) *For each finite $F \subseteq X$, there is some $g \in G$ such that $g \cdot F$ is included in a member of $\mathcal{F}$.*
(3) *For each finite $F \subseteq X$ there is a bijection $d : X \to X$ such that $d(F)$ is included in a member of $\mathcal{F}$ and G is closed under conjugation by d, i.e., considering each $g \in G$ as a permutation of X via the action, we have $d^{-1} \circ g \circ d \in G$ for all $g \in G$.*

PROOF. The first part is obvious and the second almost as obvious. In any event, both of these parts are subsumed by the third, which we now prove.

Assume (3) and assume that the definition of Ramsey action is satisfied when $F \in \mathcal{F}$. Given now an arbitrary finite $F \subseteq X$ and an arbitrary $c : X \to 2$, let d be as in (3), and let $F' \in \mathcal{F}$ with $d(F) \subseteq F'$. Let $c' = c \circ d^{-1} : X \to 2$. By assumption, there is $g \in G$ such that c' is constant on $g(F')$ and therefore on its subset $g(d(F))$. Then c is constant on $d^{-1}(g(d(F)))$, which is, by (3), $g'(F)$ for a certain $g' \in G$. $\square$

To connect Ramsey actions with traditional Ramsey theory, we give two examples. The first is a special case of the action relevant to the constructions by Pestov and Halpern mentioned in the preceding sections. The second example will be useful in connection with an open problem to be discussed later.

EXAMPLE 4.5. Fix a natural number n. Let G be the group $\mathrm{Aut}(\mathbb{Q}, <)$. Its natural action on $\mathbb{Q}$ induces an action on the set $[\mathbb{Q}]^n$ of n-element subsets of $\mathbb{Q}$. We shall show that this is a Ramsey action. Every finite subset of $[\mathbb{Q}]^n$ is included in one of the form $[A]^n$ (the set of n-element subsets of A) for some finite $A \subseteq \mathbb{Q}$. So it suffices to check the definition of Ramsey action when F is of the form $[A]^n$. Let a be the cardinality of A, and, by the finite Ramsey theorem, let M be a natural number so large that, whenever $[M]^n$ is partitioned into two pieces, there is $H \subseteq M$ of size a with $[H]^n$ included in one of the pieces. Let Z be a subset of $\mathbb{Q}$ of size M; we claim that $Y = [Z]^n$ satisfies the requirements in item (2) of Proposition 4.2. Indeed, if $c : [Z]^n \to 2$ then, by our choice of M, there is an a-element set $H \subseteq Z$ such that c is constant on $[H]^n$. Since A and H are subsets of $\mathbb{Q}$ of the same finite size, there is an order-automorphism $g \in G$ sending A to H. (In fact, g can be chosen to be piecewise linear.) Then, in the action of G on $[\mathbb{Q}]^n$, g sends $[A]^n$ to $[H]^n$, on which c is constant.

The preceding example, taken from [**1**], can be generalized to convert many other structural Ramsey results into statements about the Ramsey property of group actions. See, for example, the material in [**7**] leading from Proposition 4.2 to Theorem 4.7. The following example, also taken from [**1**], fits the same mold, but the partition theorem it uses, van der Waerden's theorem, is usually thought of as numerical rather than structural.

EXAMPLE 4.6. Let G be the group of affine permutations of $\mathbb{Q}$, i.e., transformations of the form $x \mapsto ax + b$ with $a \in \mathbb{Q} - \{0\}$ and $b \in \mathbb{Q}$. We claim that the natural action of G on $\mathbb{Q}$ is a Ramsey action. Since every finite subset of $\mathbb{Q}$ is included in an arithmetical progression, it suffices to check the definition when F is a finite arithmetical progression. By van der Waerden's theorem, whenever c maps $\mathbb{Q}$ (or even just $\mathbb{N}$) to 2, there is an arithmetical progression P, of the same length as F, on which c is constant. There is an affine permutation mapping F onto P, and this completes the verification of the Ramsey property.

We turn next to some preservation results for the Ramsey property, constructing new Ramsey actions from old.

DEFINITION 4.7. Let $\alpha : G \times X \to X$ and $\beta : G \times Y \to Y$ be two actions of the same group G. A *homomorphism* or *G-equivariant* map from the former to the latter is a function $p : X \to Y$ such that $p(\alpha(g, x)) = \beta(g, p(x))$ (more briefly, $p(g \cdot x) = g \cdot p(x)$) for all $g \in G$ and all $x \in X$. In this situation, the *fiber* over a point $y \in Y$ is the subset $p^{-1}(\{y\})$ of X.

Notice that, with notation as in this definition and with $H = \mathrm{Fix}(\{y\})$ being the stabilizer of the point y in Y, the fiber $p^{-1}(\{y\})$ is invariant under the action of H in X, so it makes sense to speak of the induced action of H on the fiber.

PROPOSITION 4.8. *Let $p : X \to Y$ be a surjective homomorphism of G-actions. If X has the Ramsey property, then so does Y.*

PROOF. Assume the Ramsey property for X, and let a finite set $F \subseteq Y$ and a 2-coloring c of Y be given. Choose, for each $u \in F$, some v in its fiber, i.e., $p(v) = u$, and let F' be the set of these chosen v's. Let $c' = c \circ p : X \to 2$. Since X has the Ramsey property, there is $g \in G$ such that c' is constant on $g \cdot F'$. This means, in view of the definition of c', that c is constant on $p(g \cdot F') = g \cdot p(F') = g \cdot F$. □

PROPOSITION 4.9. *Let $p : X \to Y$ be a surjective homomorphism of G-actions and let $y_0 \in Y$. Let Z be the fiber over y_0, with the induced action of the stabilizer H of y_0. If X has the Ramsey property, then so does Z.*

PROOF. Notice first that, since the action of G on X is transitive (by Proposition 4.3) and since p is equivariant and surjective, the action of G on Y is also transitive. So we can fix, for each $y \in Y$, some element $m_y \in G$ with $m_y \cdot y_0 = y$.

Let a finite set $F \subseteq Z$ and a 2-coloring $c : Z \to 2$ be given. Define a 2-coloring $c' : X \to 2$ by

$$c'(x) = c({m_{p(x)}}^{-1} \cdot x).$$

This makes sense because

$$p({m_{p(x)}}^{-1} \cdot x) = {m_{p(x)}}^{-1} \cdot p(x) = y_0,$$

so ${m_{p(x)}}^{-1} \cdot x$ is in the domain Z of c. (We could have required m_{y_0} to be the identity element of G, and then c' would extend c, but this is not needed for the proof.) By the Ramsey property of X, there is $g \in G$ such that c' is constant on $g \cdot F$.

Since $F \subseteq Z = p^{-1}(\{y_0\})$, we have, for all $g \cdot z \in g \cdot F$ that $p(g \cdot z) = g \cdot p(z) = g \cdot y_0$. Thus, if we write simply m for $m_{g \cdot y_0}$, we have, for all elements $z \in F$,

$$c'(g \cdot z) = c(m^{-1} \cdot (g \cdot z)) = c((m^{-1}g) \cdot z).$$

As z varies over F, the left side of this equation remains constant; therefore so does the right side, and this means that c is constant on $(m^{-1}g) \cdot F$. Finally, note that $m^{-1}g$ is in H as required, because

$$m^{-1}g \cdot y_0 = {m_{g \cdot y_0}}^{-1} \cdot (g \cdot y_0) = y_0.$$

□

We digress briefly to mention an open problem suggested by the two preceding propositions.

QUESTION 4.10. Suppose $p : X \to Y$ is a surjective homomorphism of transitive G-actions. Suppose the action of G on Y has the Ramsey property, and suppose also that, for some (equivalently, for every) $y \in Y$, the action of its stabilizer on its fiber has the Ramsey property. Does it follow that the action of G on X has the Ramsey property?

Here is a simple special case.

QUESTION 4.11. Let G be the group of transformations of the rational plane $\mathbb{Q}^2$ of the form

$$(x, y) \mapsto (ax + b, \frac{y}{a} + c),$$

where $a \in \mathbb{Q} - \{0\}$ and $b, c \in \mathbb{Q}$. Does the natural action of this group on $\mathbb{Q}^2$ have the Ramsey property?

This is a special case of the previous question, as follows. The group maps horizontal lines $\mathbb{Q} \times \{y\}$ to horizontal lines; let Y be the set of these lines with this action of G. The map p sending each point in $\mathbb{Q}^2$ to the horizontal line through it is a surjective G-homomorphism from $\mathbb{Q}^2$ to Y. The action of G on Y is essentially the same as Example 4.6, so it has the Ramsey property. The fiber over 0 is the horizontal axis, and the stabilizer consists of the transformations which, in the notation displayed above, have $c = 0$. So the action of the stabilizer on the fiber is again that of Example 4.6. So we are in the situation of Question 4.10; a positive answer there would imply a positive answer to Question 4.11.

We can reformulate Question 4.11 by applying part (3) of Proposition 4.4 as follows. Consider dilations of $\mathbb{Q}^2$, i.e., transformations of the form $(x, y) \mapsto (rx, ry)$ with positive integer r. Conjugating an element of our group by such a dilation produces again an element of our group. Also, any finite $F \subseteq \mathbb{Q}^2$ can be moved, by such a dilation, to a subset of $\mathbb{Z}^2$ (i.e., the dilation can clear the denominators) and thus to a subset of A^2 for some interval A of integers. The images of such a set A^2 under the elements of our group are *unit-area grids*, rectangular grids of unit-area cells; more precisely, they are sets of the form $B \times C$ where B and C are arithmetical progressions in $\mathbb{Q}$ whose differences are reciprocals of each other. Thus, Question 4.11 becomes: If $\mathbb{Q}^2$ is partitioned into two pieces, must one piece contain arbitrarily large unit-area grids?

A comment is perhaps in order about the use of the reciprocal $\frac{1}{a}$ as the multiplier of y in the definition of our group. Had we used simply a instead, the final form of the question would have been not about unit-area grids but about square grids, i.e., the differences in the arithmetical progressions B and C would have been equal rather than being reciprocals. This modified question would also be a special case of Question 4.10, but it would be less interesting because a positive answer follows easily from the Hales-Jewett theorem.

In contrast to Question 4.10, there is a situation where we can infer the Ramsey property for an action from the Ramsey property for some associated smaller actions.

PROPOSITION 4.12. *Given Ramsey actions of G on X and of H on Y, let $G\times H$ act on $X \times Y$ by*

$$(g,h)\cdot(x,y) = (g\cdot x, h\cdot y).$$

This action also has the Ramsey property.

PROOF. By Proposition 4.4, we need only check the definition for F of the form $A \times B$, where A and B are finite subsets of X and Y, respectively. Let such sets be given, and let c be an arbitrary 2-coloring of $X \times Y$.

By the Ramsey property of X, or rather by part (2) of Proposition 4.2, find a finite set $M \subseteq X$ such that every 2-coloring of M is constant on $g \cdot A$ for some $g \in G$. Define a function c' from Y into the (finite) set of all functions $M \to 2$ by setting

$$c'(y)(x) = c(x,y).$$

By the Ramsey property of Y, or rather by part (3) of Proposition 4.2, there is $h \in H$ such that $c'(h\cdot b)$ is the same function, say $c'' : M \to 2$, for all $b \in B$. By our choice of M, there is $g \in G$ such that $g \cdot A$ is included in M and c'' is constant on it, say with value e. Then we have, for all $a \in A$ and all $b \in B$,

$$c\big((g,h)\cdot(a,b)\big) = c(g\cdot a, h\cdot b) = c'(h\cdot b)(g\cdot a) = c''(g\cdot a) = e.$$

So c is constant on $(g,h)\cdot(A\times B)$, as required. □

Of course, this proposition can be extended to apply to more than two factors, by routine induction.

COROLLARY 4.13. *Given Ramsey actions of groups $G_1,\dots,G_k$ on sets $X_1,\dots,X_k$, respectively, let the product group $G_1\times\cdots\times G_k$ act on $X_1\times\cdots\times X_k$ by*

$$(g_1,\dots,g_k)\cdot(x_1,\cdots,x_k) = (g_1\cdot x_1,\dots,g_k\cdot x_k).$$

This action also has the Ramsey property.

Combining this corollary with Example 4.5, we find that the natural action of $(\mathrm{Aut}(\mathbb{Q},<))^k$ on $([\mathbb{Q}]^n)^k$ or more generally on

$$[\mathbb{Q}]^{n_1}\times\cdots\times[\mathbb{Q}]^{n_k}$$

has the Ramsey property. There is a useful alternative way to look at this example. Choose a subset C of $\mathbb{Q}$ of cardinality $k-1$. It cuts $\mathbb{Q}-C$ into k intervals, each order-isomorphic to $\mathbb{Q}$. Furthermore, the pointwise stabilizer $\mathrm{Fix}(C)$ consists of order-preserving permutations that act independently on each of these intervals, so it is isomorphic to $(\mathrm{Aut}(\mathbb{Q},<))^k$. Now let F be a subset of $\mathbb{Q}$ that includes C and has exactly n_i additional elements in the i^{th} interval for each i. Thus, the elements of C occur in F at positions n_1+1, n_1+n_2+2, $\dots$, $n_1+\cdots+n_{k-1}+k-1$. Then the orbit of F under the action of $\mathrm{Fix}(C)$ consists of all sets F' having the same configuration as F relative to the elements of C. This orbit can thus be identified with $[\mathbb{Q}]^{n_1}\times\cdots\times[\mathbb{Q}]^{n_k}$ by identifying any F' with the list of its intersections with the k intervals (and identifying each interval with $\mathbb{Q}$). Thus we find that the action of $\mathrm{Fix}(C)$ on the orbit of any finite $F\subseteq\mathbb{Q}$ that includes C has the Ramsey property. This example is Halpern's main combinatorial lemma in [**2**].

The same result could be obtained as a consequence of Proposition 4.9 and Example 4.5, as follows. Given k and $n_1, \dots, n_k$ as above, let $N = n_1 + \cdots + n_k + k - 1$; this is the cardinality of F in the preceding discussion. There is an $\mathrm{Aut}(\mathbb{Q}, <)$-equivariant map $p : [\mathbb{Q}]^N \to [\mathbb{Q}]^{k-1}$ sending each $F \in [\mathbb{Q}]^N$ to the subset of its elements in positions $n_1 + 1$, $n_1 + n_2 + 2$, …,$n_1 + \cdots + n_{k-1} + k - 1$. So, in the notation above, it extracts C from F. If we fix some $C \in [\mathbb{Q}]^{k-1}$, then its fiber $p^{-1}(\{C\})$ consists of those F''s whose configuration relative to C is the same as that of F. And the stabilizer of C is $\mathrm{Fix}(C)$, because an order-preserving permutation that stabilizes C must fix it pointwise. Since the action of $\mathrm{Aut}(\mathbb{Q}, <)$ on $[\mathbb{Q}]^N$ has the Ramsey property by Example 4.5, we may apply Proposition 4.9 to it, with $[\mathbb{Q}]^{k-1}$ and C in the role of Y and y_0. The fiber and stabilizer for which we thus obtain the Ramsey property are precisely the action considered in the preceding paragraph.

In the rest of this section, we reformulate some of the preceding material in terms of subgroups of G rather than actions of G. If H is a subgroup of G, which we shall write as $H \le G$, then G acts on the set $G/H = \{kH : k \in G\}$ of left cosets by $g \cdot (kH) = (gk)H$. This is a transitive action, and it is well known that every transitive action of G is isomorphic to one of this form. Specifically, given any transitive action of G on a set X and given any element $x_0 \in X$, let H be the stabilizer $\mathrm{Fix}(\{x_0\})$ of x_0. Then there is an isomorphism $G/H \to X$ defined by $kH \mapsto k \cdot x_0$. Thus, we may usually assume that transitive actions of G are of the form G/H.

Furthermore, homomorphisms between transitive actions of G are essentially given by inclusion relations between subgroups. Specifically, suppose X and Y are transitive G-sets, $p : X \to Y$ is G-equivariant, and $x_0 \in X$. Let $y_0 = p(x_0)$, and proceed as above to identify X and Y with G/H and G/K, where H and K are the stabilizers of x_0 and y_0, respectively. Then $H \le K$ and the homomorphism $G/H \to G/K$ that corresponds to p via the identifications is simply $gH \mapsto gK$.

Definition 4.14. A subgroup H of a group G is a *Ramsey subgroup* of G if the natural action of G on G/H has the Ramsey property.

Since every Ramsey action is transitive (Proposition 4.3) and therefore isomorphic to one of the form G/H, the study of Ramsey actions is essentially the same as the study of Ramsey subgroups.

In particular, we have the following succinct reformulation of Propositions 4.8 and 4.9.

Corollary 4.15. *Suppose $K \le H \le G$ and K is a Ramsey subgroup of G. Then K is a Ramsey subgroup of H, and H is a Ramsey subgroup of G.*

Proof. Since G/K has the Ramsey property and since we have a surjective G-homomorphism $G/K \to G/H : gK \mapsto gH$, Proposition 4.8 says that G/H has the Ramsey property. Furthermore, the point $1H$ in G/H has stabilizer H and fiber H/K. So Proposition 4.9 says that H/K, as an action of H, has the Ramsey property. □

From this point of view, Question 4.10 simply asks about the converse of this corollary, namely whether "Ramsey subgroup" is a transitive relation.

Question 4.16. Is a Ramsey subgroup of a Ramsey subgroup of G necessarily a Ramsey subgroup of G?

To conclude this section, we consider a simplified yet equivalent version of a definition from [**1**].

Definition 4.17. Let G be a group and $\mathcal{F}$ a normal filter of subgroups of G. We call $\mathcal{F}$ a *Ramsey filter* if it contains a group H such that every subgroup of H in $\mathcal{F}$ is a Ramsey subgroup of H.

The corresponding definition (of "Ramsey property" for normal filters) in [**1**] required not just one H as in the present definition but many such H's, enough to form a base for $\mathcal{F}$. That is, every subgroup in $\mathcal{F}$ had to include some $H \in \mathcal{F}$ all of whose subgroups in $\mathcal{F}$ are Ramsey subgroups of H. But by Corollary 4.15, this apparently stronger requirement actually follows from the definition given here. Once we have one $H \in \mathcal{F}$ with the required property, all its subgroups in $\mathcal{F}$ inherit that property, and these subgroups form a base for $\mathcal{F}$.

We shall need the following lemma about Ramsey filters.

Lemma 4.18. *Suppose $\mathcal{F}$ is a Ramsey filter of subgroups of G, and let H be as in the definition of Ramsey filter. If the trivial subgroup $\{1\}$ is in $\mathcal{F}$, then $H = \{1\}$.*

Proof. It suffices to show that, if H is any non-trivial group, then the trivial subgroup is not a Ramsey subgroup, i.e., the action of H on itself by left translation is not a Ramsey action. Fix an element $a \in H - \{1\}$ and consider the left cosets $h\langle a\rangle$ of the cyclic subgroup $\langle a\rangle$ of H generated by a. We can color the elements of $\langle a\rangle$ with at most three colors so that no two consecutive elements a^n and a^{n+1} have the same color. (If the order of $\langle a\rangle$ is even or infinite, two colors suffice, but a third color is needed if the order is odd.) Translate this coloring to each of the other cosets $h\langle a\rangle$ along some (chosen) h. The result is a coloring c of H such that no set of the form $h \cdot \{1, a\}$ is monochromatic. This means that the Ramsey property fails. □

5. Equivalence

This final section is devoted to establishing the main thesis of the paper.

Thesis. *The following are essentially the same.*

- *A non-trivial, extremely amenable topological group with small open subgroups*
- *A permutation model that satisfies the Boolean prime ideal theorem but not the axiom of choice*
- *A group with a Ramsey filter of subgroups*

Let us first consider topological groups G with small open subgroups. In such a group, the open subgroups constitute a normal filter $\mathcal{F}$ of subgroups. (Normality comes from continuity of the group operations; any conjugate of an open subgroup is again an open subgroup.)

Conversely, if $\mathcal{F}$ is any normal filter of subgroups of a group G, then we obtain a topology on G by declaring a set $A \subseteq G$ to be open if, for each $g \in A$, there is some $H \in \mathcal{F}$ such that $gH \subseteq A$. In other words, the neighborhood filter at any $g \in G$ is obtained by translating $\mathcal{F}$ on the left by g. (Translation on the right would produce the same filter, thanks to normality.) This topology makes G a topological group with small open subgroups, and its filter of open subgroups is exactly $\mathcal{F}$.

To comply with our convention that topological spaces should be Hausdorff spaces, we should require that the intersection of all the groups in $\mathcal{F}$ contains only the identity element of G. This requirement makes little difference. If $\bigcap \mathcal{F}$ is not just $\{1\}$, then it is a non-trivial normal subgroup N of G, and we should work with G/N instead. In particular, $\mathcal{F}/N = \{H/N : H \in \mathcal{F}\}$ is a normal filter of subgroups of G/N, and the topology it defines makes G/N a Hausdorff topological group with small open subgroups. Notice also that, in any continuous action of G on a Hausdorff space, N must act trivially, so there is an induced continuous action of G/N on X. In particular, G is extremely amenable if and only if G/N is.

THEOREM 5.1. *Let G be a topological group with small open subgroups and let $\mathcal{F}$ be the filter of open subgroups of G. Then the following are equivalent.*

(1) *G is extremely amenable and non-trivial.*
(2) *Every subgroup in $\mathcal{F}$ is a Ramsey subgroup of G, and $\{1\} \notin \mathcal{F}$.*

Notice that item (2) in this theorem says slightly more than "$\mathcal{F}$ is a Ramsey filter"; it requires that the H in the definition of Ramsey filter must be G itself, not some proper subgroup. This variation is necessary because of examples like the following. Consider an extremely amenable G and the product $G \times (\mathbb{Z}/2)$. The latter is not extremely amenable, because it acts without fixed points on a two-element set. Yet the filter of open subgroups in $G \times (\mathbb{Z}/2)$ is simply the closure, under supergroups, of the filter of open subgroups of G (identified with $G \times \{0\}$). It follows immediately that, if either of these is a Ramsey filter, then so is the other. So extreme amenability cannot correspond exactly to Ramseyness of the filter of open subgroups; the latter is preserved from G to $G \times (\mathbb{Z}/2)$, and the former is not.

PROOF OF THEOREM 5.1. This theorem is almost the same as Proposition 4.2 in [**7**]. The main difference is that, in [**7**], the group G is assumed to be a subgroup of the group of permutations of $\mathbb{N}$, topologized as a subspace of the product $\mathbb{N}^{\mathbb{N}}$ (with $\mathbb{N}$ having the discrete topology). Inspection of the proof in [**7**] reveals, however, that this assumption about G was used only in order to show that G has small open subgroups. So the proof in [**7**] establishes that, for topological groups with small open subgroups, extreme amenability (i.e., our condition (1) without "non-trivial") is equivalent to our condition (2) without "$\{1\} \notin \mathcal{F}$."

That the conditions remain equivalent when we adjoin "non-trivial" to the first and "$\{1\} \notin \mathcal{F}$" to the second is immediate from Lemma 4.18 once we observe that the H there will be all of G in our present situation. $\square$

We now turn to permutation models of BPI. Recall, from Section 3, that a permutation model $M = M(A, G, \mathcal{F})$ is specified by giving a set A of atoms, a group G of permutations of A, and a normal filter $\mathcal{F}$ of subgroups of G. We can make two normalizations of these data without affecting M.

First, we can assume that each atom in A is symmetric. The reason is that, if any atom a were not symmetric then neither a nor any set with a in its transitive closure would be in the permutation model M. The model would be unchanged if we simply deleted all non-symmetric atoms from A. We intend to delete from A all the non-symmetric atoms, but, in order to do so, we must take care of some technicalities. Normality of $\mathcal{F}$ easily implies that the set A' of surviving atoms is closed under the action of G. It is possible, though, that G is no longer a group of permutations of A' because there might be non-trivial elements of G that fix all the atoms in A'. In that case, those elements form a normal subgroup N of G, and we should replace G by G/N and $\mathcal{F}$ by the filter generated by the images under the projection $G \to G/N$ of the subgroups in $\mathcal{F}$. All these changes do not affect the model M, so we may assume henceforth that all atoms are symmetric.

Second, we can assume that each group in $\mathcal{F}$ occurs as the stabilizer of some element of M. To see this, suppose it were not the case, and let $\mathcal{F}'$ be the family of those groups $H \in \mathcal{F}$ that do occur as stabilizers of elements of M. We claim that $\mathcal{F}'$ is a normal filter of subgroups of G and that the associated permutation model $M(A, G, \mathcal{F}')$ is the same as M.

To verify the claim, suppose first that $H \in \mathcal{F}'$ and $H \leq K \leq G$. So H is the stabilizer of some $x \in M$. Then M contains all the elements $g(x)$ for $g \in G$ (by an inductive proof, based on the facts that, for any y, the stabilizer of $g(y)$ is the conjugate, by g, of the stabilizer of y and that $\mathcal{F}$ is normal), and it contains the orbit $\{g(x) : g \in K\}$, whose stabilizer is K. Thus $K \in \mathcal{F}'$, and we have shown that $\mathcal{F}'$ is closed under supergroups.

The rest of the verification of the claim is even easier. $\mathcal{F}'$ is closed under finite intersections, because the stabilizer of an ordered pair (x, y) is the intersection of the stabilizers of x and of y. So $\mathcal{F}'$ is a filter of subgroups of G. It is normal because, as already mentioned, the conjugate by any $g \in G$ of the stabilizer of any $x \in M$ is the stabilizer of $g(x)$, which is also in M. Since $\mathcal{F}' \subseteq \mathcal{F}$, anything symmetric with respect to $\mathcal{F}'$ is also symmetric with respect to $\mathcal{F}$; therefore $M(A, G, \mathcal{F}') \subseteq M$. The converse inclusion is immediate from the definition of $\mathcal{F}'$.

This completes the verification of the claim and thus the justification of the second normalization. Summarizing, we have arranged, without altering the model M, that

- all atoms are symmetric, and
- every group in $\mathcal{F}$ occurs as the stabilizer of some element of M.

The first of these normalizations and the fact that G is a group of permutations of A together imply that the intersection of all the groups in $\mathcal{F}$ is only the trivial group (so that $\mathcal{F}$ induces a Hausdorff topology on G). Indeed, given any element $g \in G$ other than the identity, there is an atom a moved by g, and then the stabilizer of a is a group in $\mathcal{F}$ that doesn't contain g.

Theorem 5.2. *With the normalizations above, the following are equivalent.*

(1) *The model $M(A, G, \mathcal{F})$ satisfies the Boolean prime ideal theorem but not the axiom of choice.*
(2) *$\mathcal{F}$ is a Ramsey filter of subgroups of G not containing the trivial subgroup $\{1\}$.*

Proof. This is almost part of Theorem 2 of [**1**]. The relevant part says that $M = M(A, G, \mathcal{F})$ satisfies BPI if and only if $\mathcal{F}$ is a Ramsey filter of subgroups of G. It remains only to check under what circumstances M satisfies the axiom of choice.

It is well known (see for example the end of Section 4.1 in [**6**]) that a set $x \in M$ admits a well-ordering in M if and only if the subgroup $\mathrm{Fix}(x)$ of permutations in G that fix all elements of x is in $\mathcal{F}$. If AC holds in M, then the set A of atoms can be well-ordered in M, and $\mathrm{Fix}(A)$ is clearly $\{1\}$. So we get $\{1\} \in \mathcal{F}$. Conversely, if $\{1\} \in \mathcal{F}$ then everything in $V(A)$ is symmetric, so $M = V(A)$, and M satisfies AC. (The last step uses AC in the meta-theory.)

Thus, the requirement in item (1) of the theorem that M should not satisfy AC corresponds exactly to the requirement in part (2) that $\{1\} \notin \mathcal{F}$. □

References

[1] Andreas Blass, "Prime ideals yield almost maximal ideals," *Fund. Math.* 127 (1987) 57–66.

[2] James D. Halpern, "The independence of the axiom of choice from the Boolean prime ideal theorem," *Fund. Math.* 55 (1964) 57–66.

[3] James D. Halpern and Azriel Lévy, "The Boolean prime ideal theorem does not imply the axiom of choice," in *Axiomatic Set Theory*, Proc. Sympos. at UCLA, 1967 (Proc. Sympos. Pure Math., XIII, Part I) ed. by D. Scott, Amer. Math. Soc. (1971) 83–134

[4] Wojchiech Herer and Jens Christensen, "On the existence of pathological submeasures and the construction of exotic topological groups," *Math. Ann.* 213 (1975) 203–210.

[5] Paul Howard and Jean Rubin, *Consequences of the Axiom of Choice,* Mathematical Surveys and Monographs 59, Amer. Math. Soc. (1998).

[6] Thomas Jech *The Axiom of Choice,* Studies in Logic and Foundations of Mathematics 75, North-Holland (1973).

[7] Alexander Kechris, Vladimir Pestov, and Stevo Todorčević, "Fraïssé limits, Ramsey theory, and topological dynamics of automorphism groups," *Geom. Funct. Anal.* 15 (2005) 106–189.

[8] Theodore Mitchell, "Topological semigroups and fixed points," *Illinois J. Math.* 14 (1970) 630–641.

[9] Andrzej Mostowski, "Über die Unabhängigkeit des Wohlordnungssatzes vom Ordnungsprinzip," *Fund. Math.* 32 (1939) 201–252.

[10] Vladimir Pestov, "On free actions, minimal flows, and a problem by Ellis," *Trans. Amer. Math. Soc.* 350 (1998) 4149–4165.

Mathematics Department, University of Michigan, Ann Arbor, MI 48109–1043, U.S.A.
E-mail address: ablass@umich.edu

Contemporary Mathematics
Volume **558**, 2011

(Un)countable and (non)effective versions of Ramsey's theorem

Dietrich Kuske

Abstract. We review Ramsey's theorem and its extensions by Jockusch for computable partitions, by Sierpiński and by Erdős and Rado for uncountable homogeneous sets, by Rubin for automatic partitions, and by the author for ω-automatic (in particular uncountable) partitions.

Introduction

Every infinite graph has an infinite clique or an infinite anticlique – this is the paradigmatic formulation of Ramsey's theorem [**14**]. But this theorem is highly non-constructive since there are computable infinite graphs none of its infinite cliques and anticliques is computable (they are not even in Σ^0_2 [**7**]). On the positive side, Jockusch also showed that every infinite computable graph contains an infinite clique or anticlique from Π^0_2. Soon after Ramsey's paper from 1930, authors got interested in a quantitative analysis: how many nodes are necessary and sufficient to obtain a clique or anticlique of size $\aleph_1$. The answer was given by Sierpiński [**17**] ($(2^{\aleph_0})^+$ nodes are necessary) and Erdős & Rado [**6**] ($(2^{\aleph_0})^+$ nodes are sufficient). We unify the proofs of these positive results by the notion of a Ramsey tree. In doing so, we prove in particular that every infinite computable finitely branching tree contains an infinite path from Π^0_2.

Recall that a graph is computable if both its set of nodes and its set of edges can be decided by a Turing machine. Replacing these Turing machines by finite automata, one obtains the more restrictive notion of an *automatic graph*: the set of nodes is a regular set and whether a pair of nodes forms an edge can be decided by a synchronous two-tape automaton (this concept is known since the beginning of automata theory, a systematic study started with [**9, 3**], see [**15**] for a recent overview). Compared to computable graphs, here, the situation is much more favourable: every infinite automatic graph contains an infinite regular clique or an infinite regular anticlique [**15**].

Since automatic graphs contain at most $\aleph_0$ nodes, we need a more general notion for a recursion-theoretic analysis of uncountable graphs. For this, we use

1991 *Mathematics Subject Classification.* Primary 03D05, 05D10.
Key words and phrases. Computable and automatic graphs, uncountable graphs, Ramsey theory.

Blumensath & Grädel's [**3**] ω-automatic graphs: the names of nodes form a regular ω-language and the edge relation (on names) as well as the relation "these two names denote the same node" can be decided by a synchronous 2-tape Büchi-automaton. We show that any such graph of size $2^{\aleph_0}$ has a clique or anticlique of size $2^{\aleph_0}$, but these (anti)cliques are not necessarily described by a Büchi-automaton.

While the results by Ramsey, by Jockusch, by Erdős & Rado, and by Rubin generalise to hypergraphs, this is not the case for our result on ω-automatic graphs: We present a ternary ω-automatic hypergraph of size $2^{\aleph_0}$ that does not contain any uncountable clique or anticlique.

Finally, it should be noted that all the positive results, except those by Rubin and by Kuske are shown by an analysis of the Ramsey tree of a hypergraph. This proof technique seems not to work in the (ω-)autmatic setting where other strategies are employed.

Acknowledgment Markus Lohrey's careful and sympathetic reading of the manuscript eliminated more mistakes than I thought possible. I am very grateful for the time and energy he devoted to this task.

1. Partitions and the Ramsey tree

For a set V and a natural number $k \geq 1$, let $[V]^k$ denote the set of k-element subsets of V. A *(k, ℓ)-partition* is a tuple $P = (V, E_1, \ldots, E_\ell)$ where V is a set and $(E_1, \ldots, E_\ell)$ is a partition of $[V]^k$ into (possibly empty) sets. For $1 \leq i \leq \ell$, a set $W \subseteq V$ is *E_i-homogeneous* if $[W]^k \subseteq E_i$; it is *homogeneous* if it is E_i-homogeneous for some $1 \leq i \leq \ell$. The case $k = \ell = 2$ is special: any $(2,2)$-partition $G = (V, E_1, E_2)$ can be considered as an (undirected loop-free) graph (V, E_1). Homogeneous sets in G are then complete or discrete induced subgraphs of (V, E_1).

This paper is concerned with the following question: Does every (k, ℓ)-partition $G = (V, E_1, \ldots, E_\ell)$ with $|V| = \kappa$ have a homogeneous set of size λ (where κ and λ are cardinal numbers and $k, \ell \geq 2$ are natural numbers). If this is the case, one writes

$$\kappa \to (\lambda)^k_\ell$$

(a notation due to Erdős & Rado [**5**]).

Observation 1.1. *Let $P = (V, E_1, \ldots, E_\ell)$ be a (k, ℓ)-partition, $\leq$ a well-order on V, and $v \in V$. Then there exists a unique ordinal sequence $(v_\alpha)_{\alpha \leq \beta}$ satisfying:*

(R1) for $\alpha \leq \beta$, the node v_α is the least element of $(V, \leq)$ such that
- *$v_\gamma < v_\alpha$ for all $\gamma < \alpha$ and*
- *$A \cup \{v_\alpha\} \in E_i$ iff $A \cup \{v\} \in E_i$ for all $A \in [\{v_\gamma \mid \gamma < \alpha\}]^{k-1}$ and $1 \leq i \leq \ell$*

(R2) $v_\beta = v$

Example 1.2. We consider the Rado graph as a $(2,2)$-partition $(\mathbb{N}, E_1, E_2)$ where $\{m, n\} \in E_1$ with $m < n$ if, in the binary expansion of n, position m carries 1 (i.e., the unique presentation of n as sum of powers of 2 contains the summand 2^m).

For a word w over $\{0,1\}$, let $[w]$ denote its value at base 2, i.e., $[\varepsilon] = 0$ and $[wx] = 2 \cdot w + x$ for $x \in \{0,1\}$.

Then $0, 1, 5, 33, 37$ is the Ramsey sequence of 37: Clearly, every Ramsey sequence has to start with $v_0 = 0$. Then $[100101] = 37$, so $\{0, 37\} \in E_1$. Hence we determine the minimal number $v_1 > v_0$ whose binary expansion ends with 1 – this is obviously $v_1 = 1$. Since $\{1, 37\} \in E_2$, we now search for the minimal number

$v_2 > v_1 = 1$ whose binary expansion ends in 01 – hence $v_2 = [101] = 5$. Again, $\{5, 37\} \in E_1$, so v_3 is the minimal number above 5 whose binary expansion ends with $1abc01$ for some $a, b, c \in \{0, 1\}$ – hence $v_3 = [100001] = 33$. Now $\{33, 37\} \in E_2$, so v_4 is the minimal number whose binary expansion ends with $0w1abc01$ with $w \in \{0, 1\}^*$ of length 28 and $a, b, c \in \{0, 1\}$ – thus $v_4 = [100101] = 37$.

Let $v \in V$. Then one can construct, by transfinite induction, the sequence of nodes v_α satisfying (R1). In particular, $v_0 < v_1 < \cdots < v_{k-2}$ is the initial segment of $(V, \leq)$ of length $k - 1$ (if v is not among these elements). Eventually, v is the least element satisfying the two conditions in (R1) – at this stage β, the process terminates. We call the sequence $(v_\alpha)_{\alpha \leq \beta}$ the *Ramsey sequence of* v and denote it by $\bar{v}$; its *length* is β. Using these sequences, we can define a partial order $\preceq$ on V by setting $v \preceq w$ if and only if $\bar{v}$ is a prefix of $\bar{w}$ (which is the case if and only if v appears in the sequence $\bar{w}$). The structure $(V, \preceq)$ is the *Ramsey tree* of the (k, ℓ)-partition P with respect to the well-order $\leq$.

Some properties of the Ramsey tree are obvious:

(1) If $v \in V$, then its predecessors in the Ramsey tree (i.e., the set $\{w \in V \mid w \preceq v\}$) are well-ordered. This justifies to call $(V, \preceq)$ an order tree.

(2) Let $v \in V$ and let B be the cardinality of its Ramsey sequence. For any two distinct brothers x and y of v, there exists a $(k-1)$-elements subset A of $\bar{v}$ not containing v and $1 \leq i \leq \ell$ such that $A \cup \{x\} \in E_i$ and $A \cup \{y\} \notin E_i$. Hence the number of brothers is at most
$$\ell^{\binom{B}{k-1}} = \ell^{|[B]^{k-1}|}.$$
In particular, it is finite if $\bar{v}$ is finite, and it is at most $2^{\aleph_n}$ if $|B| = \aleph_n$.

Different versions of Ramsey's theorem will be derived from an analysis of the Ramsey tree. This is based on the following observation and result that allows to infer the existence of homogeneous sets in (k, ℓ)-partitions from their existence in $(k - 1, \ell)$-partitions.

Observation 1.3. *Let $(V, \preceq)$ be the Ramsey tree of the (k, ℓ)-partition $P = (V, E_1, \ldots, E_\ell)$ with respect to the well-order $\leq$ and let $Y \subseteq V$ be some chain in the Ramsey tree (i.e., linearly ordered subset) without maximal element.*

For $A \in [Y]^{k-1}$, set $A \in E_i^Y$ if and only if there exists $y \in Y$ with $A \prec y$ and $A \cup \{y\} \in E_i$. Then $P^Y = (Y, E_1^Y, \ldots, E_\ell^Y)$ is a $(k - 1, \ell)$-partition.

Proof. Since Y has no maximal element with respect to $\preceq$, any set $A \in [Y]^{k-1}$ belongs to some E_i^Y. Suppose $A \in E_i^Y \cap E_j^Y$. Then there are $y_i, y_j \in Y$ with $A \prec y_i, y_j$, $A \cup \{y_i\} \in E_i$, and $A \cup \{y_j\} \in E_j$. For symmetry reasons, we can assume $y_i \preceq y_j$. But then $i = j$ by (R1) since y_i appears in the Ramsey sequence of y_j. □

Proposition 1.4. *Let $k \geq 2$ and let $P = (V, E_1, \ldots, E_\ell)$ be a (k, ℓ)-partition and Y a chain without maximal element in the Ramsey tree $(V, \preceq)$. Any set homogeneous in the $(k - 1, \ell)$-partition P^Y induced by Y is homogeneous in P.*

Proof. Let $W \subseteq Y \subseteq V$ be E_i^Y-homogeneous, i.e., $[W]^{k-1} \subseteq E_i^Y$. Let $B \in [W]^k \subseteq [Y]^k$ and let $v < w$ be the maximal elements of $(B, \leq)$. Then $A := B \backslash \{w\} \in [W]^{k-1} \subseteq E_i^Y$. By the definition of E_i^Y, there exists $y \in Y$ with $A \leq v < y$ and $A \cup \{y\} \in E_i$. Since $(Y, \preceq)$ is linearly ordered, y appears in the Ramsey sequence of w or vice versa. In any case, $B = A \cup \{w\} \in E_i$ follows from (R1). □

2. The countable case

2.1. Ramsey's theorem.

THEOREM 2.1 (Ramsey [**14**]). *If $k, \ell \geq 1$ and $\kappa \geq \aleph_0$, then $\kappa \to (\aleph_0)^k_\ell$.*

PROOF. It suffices to prove the theorem for $\kappa = \aleph_0$ in which case it is shown by induction on k. The base case ($k = 1$) is the pigeonhole principle. So suppose $\aleph_0 \to (\aleph_0)^{k-1}_\ell$ and let $P = (\mathbb{N}, E_1, \dots, E_\ell)$ be a (k, ℓ)-partition. The well-order $\leq$ on the base set $\mathbb{N}$ of this partition is the natural order on the natural numbers. Then the Ramsey tree of P is infinite and, since all Ramsey sequences are finite, finitely branching. Hence, by König's lemma, it contains some infinite branch Y (that does not have a maximal element). Then P^Y is an infinite $(k-1, \ell)$-partition, so it contains by the induction hypothesis an infinite homogeneous set. But this set is, by Prop. 1.4 also homogeneous in P. $\square$

2.2. Jockusch's theorems. A partition $(\mathbb{N}, E_1, \dots, E_\ell)$ is computable if all the sets E_i are decidable. For $k, \ell \geq 2$ and classes of sets $\mathcal{C} \subseteq 2^{[\mathbb{N}]^k}$ and $\mathcal{D} \subseteq 2^{\mathbb{N}}$, write

$$(\aleph_0, \mathcal{C}) \to (\aleph_0, \mathcal{D})^k_\ell$$

if every (k, ℓ)-partition $(\mathbb{N}, E_1, \dots, E_\ell)$ with $E_1, \dots, E_\ell \in \mathcal{C}$ contains some infinite homogeneous set $H \in \mathcal{D}$. In this section, the sets $\mathcal{C}$ and $\mathcal{D}$ will be classes from the arithmetical hierarchy: A set $A \subseteq \mathbb{N}$ belongs to Σ^0_n if and only if there is a computable predicate $P \subseteq \mathbb{N}^{n+1}$ such that

$$A = \{x \in \mathbb{N} \mid \exists x_1 \forall x_2 \dots \exists/\forall x_n : (x, x_1, \dots, x_n) \in P\}.$$

The set A belongs to Π^0_n if its complement belongs to Σ^0_n, it belongs to Δ^0_n if it is both, Σ^0_n and Π^0_n, and it belongs to the arithmetical hierarchy AH if it belongs to Σ^0_n for some $n \in \mathbb{N}$. We also write REC for Δ^0_1 since these are precisely the computable sets. Identifying a finite set $A \in [\mathbb{N}]^k$ with its Gödel number, we can also say that a set $E \subseteq [\mathbb{N}]^k$ belongs to Σ^0_n etc.

We now discuss the computational content of the above proof of Ramsey's theorem. First note that clearly

$$(\aleph_0, \mathrm{REC}) \to (\aleph_0, \mathrm{REC})^1_\ell$$

since, given an infinite computable $(1, \ell)$-partition, one of the classes $E_1, \dots, E_\ell$ is infinite and they are all computable. Next, let $(\mathbb{N}, E_1, \dots, E_\ell)$ be some computable (k, ℓ)-partition and let $\leq$ be the natural order on $\mathbb{N}$. Then the Ramsey tree of P wrt. $\leq$ has a computable copy: it consists of all Ramsey sequences (that are necessarily finite) ordered by the prefix order. An isomorphism is given by mapping v to its Ramsey sequence $\bar{v}$. In order to apply Prop. 1.4, we have to find an infinite branch in this computable copy. First note that also the successor relation in this computable copy is computable. We therefore now make a short excursion into the theory of computable successor trees.

THEOREM 2.2. *Let $T = (V, \mathrm{succ})$ with $V \subseteq \mathbb{N}$ be an infinite computable finitely branching successor tree. Then T contains an infinite branch from Π^0_2.*

PROOF. On V, we define the lexicographic order: $m \leq_{\mathrm{lex}} n$ if and only if $(m, n) \in \mathrm{succ}^*$ or there exist nodes $x, m', n' \in V$ with $(x, m'), (x, n') \in \mathrm{succ}$, $(m', m), (n', n) \in \mathrm{succ}^*$, and $m' < n'$. Note that the linear order $\leq_{\mathrm{lex}}$ is decidable. Furthermore, the root of T is its minimal element and every node $v \in V$ has a successor in the linear order $(V, \leq_{\text{lex}})$ (if it is not maximal in this linear order). Hence $(V, \leq_{\text{lex}})$ is isomorphic to $\omega + (V', \leq_{\text{lex}})$ for some set $V' \subseteq V$. For $n \geq 1$ let v_n denote the n^{th} element of $(V, \leq_{\text{lex}})$ and define

$$C = \{v \in V \mid \forall n : (v \leq_{\text{lex}} v_n \Rightarrow (v, v_n) \in \text{succ}^*) \wedge \exists m : v = v_m\}.$$

This set is a chain in the tree (V, succ^*): Let $v, w \in C$. Then there exist $m, n \in \mathbb{N}$ with $v = v_m$ and $w = v_n$, we assume $m \leq n$. By the definition, this implies $v_m \leq_{\text{lex}} v_n$ and therefore (since $v = v_m \in C$) also $(v_m, v_n) \in \text{succ}^*$, i.e., $(v, w) \in \text{succ}^*$.

Our next aim is to show that C is infinite. First note that the root v_1 belongs to C. Let $c \in C$. Then there are infinitely many $m \in \mathbb{N}$ with $c \leq_{\text{lex}} v_m$ and therefore with $(c, v_m) \in \text{succ}^*$. Let c' be the lexicographically largest son of c of the form v_n. Then, for all $s \geq n$, we have $c <_{\text{lex}} c' = v_n \leq_{\text{lex}} v_s$ implying $(c, v_s) \in \text{succ}^*$ and therefore $(c', v_s) \in \text{succ}^*$. Hence $c' \in C$, i.e., the chain C does not contain a maximal element. It follows that C is even an infinite branch (i.e., downwards closed).

To show that C is Π^0_2, it suffices to argue that the relation $R = \{(n, v_n) \mid n \in \mathbb{N}\}$ is Σ^0_2. But this is obvious since it is defined by

$$\begin{aligned} \exists v_1, v_2, \ldots, v_{n-1} \forall v : \quad & \textstyle\bigwedge_{1 \leq i < n-1} v_i <_{\text{lex}} v_{i+1} \\ \wedge \quad & v \not<_{\text{lex}} v_1 \wedge \textstyle\bigwedge_{1 \leq i < n-1} \neg(v_i <_{\text{lex}} v <_{\text{lex}} v_{i+1}). \end{aligned}$$

□

LEMMA 2.3. *Let $P = (\mathbb{N}, E_1, \ldots, E_\ell)$ be a computable and infinite $(k+1, \ell)$-partition and let $C \subseteq \mathbb{N}$ be an infinite branch in its Ramsey tree from Π^0_2. Then $E_i^C \subseteq \mathbb{N}$ is Π^0_2 for all $1 \leq i \leq \ell$.*

PROOF. Let $A \in [C]^k$ with maximal element a. Let furthermore $\mathbb{A}$ denote the finite set $\{x \in V \mid x \preceq a\}$. Then $A \in E_i^C$ if and only if there exists a function $f : [\mathbb{A}]^k \to \{1, \ldots, \ell\}$ with $f(A) = i$ such that

$$\begin{aligned} &\forall y (a \prec\!\cdot\, y \wedge \bigwedge_{B \in [\mathbb{A}]^k} B \cup \{y\} \in E_{f(B)} \to y \in C) \\ &\wedge \exists y (x \prec y \wedge \bigwedge_{B \in [\mathbb{A}]^k} B \cup \{y\} \in E_{f(B)}) \end{aligned}$$

where $\prec\!\cdot$ is the immediate successor relation of the tree T. The existentially quantified formula and the premise in the universally quantified formula are computable predicates. Since the quantification over the functions f is computably bounded, the whole expression belongs to Π^0_2. □

THEOREM 2.4. *Let $P = (V, E_1, \ldots, E_\ell)$ be some infinite computable $(2, \ell)$-partition. Then P contains an infinite homogeneous set from Π^0_2.*

PROOF. Since also the successor relation of the Ramsey tree of P is computable, we can apply Theorem 2.2, i.e., there exists an infinite chain C from Π^0_2 in the Ramsey tree of P. Since C is infinite, one of the sets E_i^C is infinite. By Lemma 2.3, this set E_i^C belongs to Π^0_2 and, by Prop. 1.4, it is homogeneous in P. □

THEOREM 2.5 (attributed to Manaster in [**7**, p. 276]). *For all $k, \ell \geq 1$, we have*

$$(\aleph_0, \text{AH}) \to (\aleph_0, \text{AH})^k_\ell.$$

Proof. Note that Theorem 2.2 and Lemma 2.3 and therefore Theorem 2.4 also hold in their relativized form. Hence we can prove this theorem by induction in the same way that we proved Ramsey's theorem. □

Jockusch [**7**] proved the more precise result

$$(\aleph_0, \Delta^0_n) \to (\aleph_0, \Pi^0_{n+k-1})^k_\ell$$

for $k, \ell, n \geq 1$. In general, the proof proceeds by induction as before: from a (k, ℓ)-partition P whose partition classes all belong to Δ^0_n, one builds a $(k-1, \ell)$-partition whose homogeneous sets are also homogeneous in P. Once the inductive construction reaches a $(2, \ell)$-partition, the above Theorem 2.4 is invoked. While Manaster uses the Π^0_2-branch from Theorem 2.2 in the inductive step, Jockusch uses a branch C such that C' (the first jump of C) is Δ^0_3 – existence of such a branch is shown in [**8**].

Esthetically, this result is not as satisfactory as the previous ones since the levels of the arithmetical hierarchy on the left and on the right do not coincide. Jockusch [**7**] also showed that this cannot be avoided. His construction uses the following limit lemma twice.

Lemma 2.6 (Shoenfield [**16**]). *Let $k \geq 2$. There is a total function $F : \mathbb{N}^3 \to \{0,1\}$ computable in $\emptyset^{(k-2)}$ such that, for any $A \in \Delta^0_k$, there is $e \in \mathbb{N}$ satisfying*

(1) the sequence $(F(e,m,s))_{s\in\omega}$ is ultimately constant for all $m \in \mathbb{N}$, and
(2) $A = \{m \mid F(e,m,s) = 1$ for almost all $s\}$.

Proof. For $s \in \mathbb{N}$ and $X \subseteq \mathbb{N}$, let K^X_s denote the set of all $e \leq s$ such that the oracle Turing machine with index e stops, when started with input e and oracle X, after at most s steps.

Next, let $F : \mathbb{N}^3 \to \{0,1\}$ be defined by $F(e,m,s) = 1$ if and only if the oracle Turing machine with index e stops successfully after at most s steps, when started with input m and oracle $K^{\emptyset^{(n-2)}}_s$. Then F is computable in $\emptyset^{(k-2)}$ and total. Let $A \in \Delta^0_k$. Then A is computable in $\emptyset^{(k-1)}$, i.e., there exists $e \in \mathbb{N}$ such that the oracle Turing machine M_e with index e computes the characteristic function χ_A using the oracle $\emptyset^{(k-1)}$. Now let $m \in \mathbb{N}$. Then M_e has a unique computation C with input m and oracle $\emptyset^{(k-1)}$. Let $t \in \mathbb{N}$ be at least the length of C such that, for any oracle access o in C, we have

$$o \in \emptyset^{(k-1)} = K^{\emptyset^{(k-2)}} \iff o \in K^{\emptyset^{(k-2)}}_t .$$

Then, for any $s \geq t$, C is the computation of the oracle Turing machine M_e with oracle $K^{\emptyset^{(k-2)}}_s$ and input m. In other words, $F(e,m,s) = \chi_A(m)$ for any $s \geq t$. This proves (1) and (2). □

Jockusch's first use of the limit lemma is in the base case:

Lemma 2.7. *For any $n \geq 2$, we have $(\aleph_0, \Delta^0_{n-1}) \not\to (\aleph_0, \Delta^0_n)^2_2$*

Proof. Let F be the function from Lemma 2.6 that is computable in $\emptyset^{(n-2)}$. We set $\{m,s\} \in E_1$ (with $m < s$) if and only if $m \in H(s)$ where $H(s)$ is defined by the procedure from Fig. 1; $E_2 = [\mathbb{N}]^2 \setminus E_1$ completes the definition of the $(2,2)$-partition $(\mathbb{N}, E_1, E_2)$. Note that E_1 and E_2 are both computable in $\emptyset^{(n-2)}$ and therefore belong to Δ^0_{n-1}.

We now show by contradiction that this partition does not have any infinite homogeneous set in Δ^0_n. Let $A \in \Delta^0_n$ be infinite. Hence there is $e \in \mathbb{N}$ satisfying (1)

```
0  function H(s ∈ ℕ): subset of {1, 2, ..., s}
1  H_1 := ∅; H_2 := ∅;
2  for e := 0 to s − 1 do
3    if |{n ≤ s : F(e, n, s) = 0} \ (H_1 ∪ H_2)| ≥ 2 then
4      let a < b be minimal in {n ≤ s | F(e, n, s) = 0} \ (H_1 ∪ H_2);
5      set H_1 := H_1 ∪ {a} and H_2 := H_2 ∪ {b};
6  return H_1;
```

FIGURE 1. Computation of $H(s)$

and (2) in Lemma 2.6. Let $a_1 < a_2 < \cdots < a_{2e+2}$ be the $2e+2$ minimal elements of A. By Lemma 2.6, there is $t \in \mathbb{N}$ such that

$$\forall a \leq a_{2e+2} \forall s \geq t : a \in A \iff F(e,a,s) = 0\,. \tag{2.1}$$

Since A is infinite, there exists $s \in A$ with $s \geq a_{2e+2}$ and $s \geq t$. We analyse the set $H(s)$: Note that in any run through the loop in Fig. 1, $|H_1 \cup H_2|$ increases by at most 2. Hence, whenever we execute line 3, we get $|H_1 \cup H_2| \leq 2e$. With e satisfying (1) and (2) in Lemma 2.6, the test in line 3 succeeds since $\{n \leq s : F(e,n,s) = 0\} = \{a_1, \ldots, a_{2e+2}\}$ by (2.1). For the same reason, the elements a and b chosen in line 4 both belong to $\{a_1, \ldots, a_{2e+2}\} \subseteq A$. Since no element is ever deleted from H_1, we obtain $a \in H(s)$. On the other hand, no element from H_2 will ever be put into H_1, so $b \notin H(s)$. Hence, we found $a, b \in A$ with $\{a, s\} \in E_1$ and $\{b, s\} \in E_2$, i.e., the set A is not homogeneous. Since all we assumed about the set A was $A \in \Delta^0_n$ and A infinite, we proved that no infinite Δ^0_n-set is homogeneous in the $(2,2)$-partition $(\mathbb{N}, E_1, E_2)$. $\square$

By induction, Jockusch then increases the gap between Δ^0_{n-1} and Δ^0_n in the above lemma at the expense of partitioning $[\mathbb{N}]^k$ for $k > 2$. This inductive argument makes use of the limit lemma, again.

LEMMA 2.8. *Let $P = (\mathbb{N}, A, B)$ be a $(k,2)$-partition with $A, B \in \Delta^0_n$. Then there exists a $(k+1,2)$-partition $P' = (\mathbb{N}, A', B')$ with $A', B' \in \Delta^0_{n-1}$ such that any infinite set homogeneous in P' is also homogeneous in P.*

PROOF. Let F be the function from Lemma 2.6 that is computable in $\emptyset^{(n-2)}$. Hence there is $e \in \mathbb{N}$ satisfying (1) and (2) in Lemma 2.6 (where we identify a finite set with its Gödel number). For $a_1 < a_2 < \cdots < a_k < s$, set $\{a_1, \ldots, a_k, s\} \in A'$ if and only if $F(e, \{a_1, a_2, \ldots, a_k\}, s) = 0$ and define $B' = [\mathbb{N}]^{k+1} \setminus A'$. Clearly, A' is computable in $\emptyset^{(n-2)}$ implying $A', B' \in \Delta^0_{n-1}$.

Now let $H \subseteq \mathbb{N}$ be homogeneous in P'. It suffices to consider the case $[H]^{k+1} \subseteq A'$ since the other one is analogous. To prove $[H]^k \subseteq A$, let $h_1 < h_2 < \cdots < h_k$ be elements of H. Then, for any $t > h_k$ from H, we get $F(e, \{h_1, \ldots, h_k\}, t) = 0$. Since the sequence $(F(e, \{h_1, \ldots, h_k\}, s))_{s \in \omega}$ is ultimately constant, this implies $F(e, \{h_1, \ldots, h_k\}, s) = 0$ for almost all s, i.e., $\{h_1, \ldots, h_k\} \in A$. $\square$

THEOREM 2.9 (Jockusch [**7**]). *Let $k, \ell, n \geq 2$. Then $(\aleph_0, \Delta^0_n) \not\to (\aleph_0, \Sigma^0_{n+k-1})^k_\ell$.*

PROOF. We first prove $(\aleph_0, \Delta^0_n) \not\to (\aleph_0, \Delta^0_{n+k-1})^k_2$:

By Lemma 2.7, we find a $(2,2)$-partition $P_2 = (\mathbb{N}, E^2_1, E^2_2)$ with $E^2_1, E^2_2 \in \Delta^0_{n+k-2}$ without any homogeneous set in Δ^0_{n+k-1}. Applying Lemma 2.8 several times, we find $(m,2)$-partitions $P_m = (\mathbb{N}, E_1^m, E_2^m)$ with $E_1^m, E_2^m \in \Delta^0_{n+k-m}$ whose infinite homogeneous sets are also homogeneous in P_2 (and therefore do not belong to Δ^0_{n+k-1}). With $m = k$, the claim follows.

The general result now follows since:

- Any infinite set H from Σ^0_{n+k-1} contains an infinite set from Δ^0_{n+k-1} that is homogeneous whenever H is.
- If $\ell > 2$, extend the $(n,2)$-partition by sufficiently many empty classes.

$\square$

Let $P = (\mathbb{N}, E_1, E_2)$ be some infinite computable $(2,2)$-partition. Let $R = (\mathbb{N}, \preceq)$ be its Ramsey tree (with respect to the natural order $\leq$ on $\mathbb{N}$) and suppose $C \subseteq \mathbb{N}$ is an infinite chain in R from Σ^0_2. Then the sets E_i^C of natural numbers $c \in C$ with

$$\exists d : d \in C \wedge c < d \wedge \{c,d\} \in E_i$$

belong to Σ^0_2 and are homogeneous in P^C and therefore in P. Furthermore, one of E_1^C and E_2^C is infinite. But this contradicts $(\aleph_0, \Delta^0_1) \not\to (\aleph_0, \Sigma^0_2)^2_2$. Hence the Ramsey tree of P does not necessarily contain an infinite chain from Σ^0_2, i.e., we proved

Corollary 2.10. *There exists a computable finitely branching successor tree without any infinite chain from* Σ^0_2.

2.3. Rubin's theorem. Rubin [**15**] considered Ramsey's theorem in the context of automatic structures [**9, 3**]. While a computable structure is given by a tuple of Turing machines, an automatic structure is given by a tuple of finite automata. We therefore first sketch some basic notions from the theory of finite automata.

2.3.1. *Languages of finite words.* Let Γ be some finite alphabet. The set of all (finite) words over Γ is denoted Γ^*, the empty word is ε, and Γ^+ is the set of non-empty words. By $\leq_{\text{pref}}$, we denote the prefix order on Γ^*.

A *finite automaton* M is a tuple $M = (Q, \Gamma, \delta, I, F)$ where Q is a finite set of states, $I \subseteq Q$ is the set of initial states, $F \subseteq Q$ is the set of final states, and $\delta \subseteq Q \times \Gamma \times Q$ is the transition relation. A *run* of M on a word $x = a_1 a_2 \cdots a_n$ with $n \geq 0$ and $a_i \in \Gamma$ is a word $q_0 q_1 \ldots q_n \in Q^+$ with $q_i \in Q$ such that $(q_i, a_{i+1}, q_{i+1}) \in \delta$ for all $0 \leq i < n$. The run is *successful* if $q_0 \in I$ and $q_n \in F$. The language $L(M) \subseteq \Gamma^*$ defined by M is the set of all words that admit a successful run. A language $L \subseteq \Gamma^*$ is *regular* if there exists a finite automaton M with $L(M) = L$. By REG, we denote the class of all regular languages.

For words $x_i = a_i^0 a_i^1 a_i^2 \ldots a_i^{k_i} \in \Gamma^*$, the *convolution* $(x_1, x_2, \ldots, x_n)^\otimes \in ((\Gamma \uplus \{\#\})^n)^*$ is defined by

$$(x_1, \ldots, x_n)^\otimes = (b_1^0, \ldots, b_n^0)\,(b_1^1, \ldots, b_n^1)\,(b_1^2, \ldots, b_n^2) \cdots (b_1^k, \ldots, b_n^k)$$

where $k = \max\{k_1, \ldots, k_n\}$ and

$$b_i^\ell = \begin{cases} a_i^\ell & \text{if } \ell \leq k_i \\ \# & \text{otherwise.} \end{cases}$$

In other words, the convolution is obtained by first adding some occurrences of the new letter # to each word to make them the same length and then "glueing" the prolonged words together. An n-ary relation $R \subseteq (\Gamma^*)^n$ is called *automatic* if the language $\{(x_1, \ldots, x_n)^\otimes \mid (x_1, \ldots, x_n) \in R\}$ is regular. In order to also capture partition classes of a (k,ℓ)-partition, we will call a set $E \subseteq [\Gamma^*]^k$ automatic if the relation

$$\{(u_1,\dots,u_n) \mid \{u_1,\dots,u_n\} \in E\}$$

is automatic, i.e., if the language

$$\{(u_1,\dots,u_n)^{\otimes} \mid \{u_1,\dots,u_n\} \in E\}$$

is regular. We write A for the class of automatic subsets of $[\Gamma^*]^k$ for any $k \in \mathbb{N}$ and alphabet Γ.

2.3.2. *Ramsey's theorem for automatic partitions.* Note that any partition class from A is decidable. Hence, any (k,ℓ)-partition with automatic partition classes is computable and therefore has an infinite homogeneous set in the arithmetical hierarchy. Rubin showed that in this case, much simpler infinite homogeneous sets can be found. Namely, he proved

PROPOSITION 2.11 (Rubin [**15**]). *Let $(V, E_1,\dots,E_\ell)$ be some (k,ℓ)-partition with $E_i \in \mathsf{A}$ for all $1 \le i \le \ell$ that has some infinite E_1-homogeneous set. Then there is some infinite regular E_1-homogeneous set.*

Since every infinite partition has, by Ramsey's theorem, some homogeneous set, we get immediately:

THEOREM 2.12 (Rubin [**15**, Prop. 3.21]). *For all $k,\ell \ge 1$, we have*

$$(\aleph_0,\mathsf{A}) \to (\aleph_0,\mathsf{A})^k_\ell\,.$$

Since $L \subseteq \Gamma^*$ is automatic if and only if it is regular, we can alternatively express this as

$$(\aleph_0,\mathsf{A}) \to (\aleph_0,\mathrm{REG})^k_\ell\,.$$

The proof of Prop. 2.11 does not use the Ramsey tree of the partition for two reasons. First, the proof is simpler since it constructs the regular homogeneous set directly. Secondly, it is not clear whether the Ramsey tree of an automatic partition is automatic, i.e., can be described by finite automata. Hence, we do not know how to perform the induction in the automatic setting.

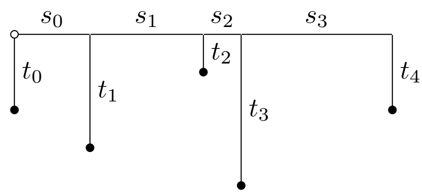
The direct construction of the regular homogeneous set in Prop. 2.11 uses the concept of a comb: a *comb* is a set of words $C = \{s_0 s_1 s_2 \dots s_{i-1} t_i \mid 0 \le i < \alpha\}$ for some $\alpha \le \omega$ where $s_i, t_i \in \Gamma^+$ and $|t_i| = |s_i|$ for all $0 \le i < \alpha$ (note that, given C, these conditions determine s_i and t_i completely). A comb C is visualized in Fig. 2: it consists of the paths from the empty node to the filled nodes since these paths are labeled t_0, s_0t_1, $s_0s_1t_2$, $\dots$, $s_0s_1s_2s_3t_4$.

Given a (k,ℓ)-partition $P = (V, E_1,\dots,E_\ell)$ with E_i automatic, our first lemma ensures that any infinite homogeneous set contains some infinite comb:

LEMMA 2.13. *Let $X \subseteq \Gamma^*$ be infinite. Then there exists an infinite comb $C \subseteq X$.*

PROOF. Let $t_0 \in X \setminus \{\varepsilon\}$ be arbitrary. Now suppose we defined $s_0,\dots,s_{j-1}$ and $t_0,\dots,t_j$ such that

(1) $s_0 s_1 \dots s_{i-1} t_i \in X$ for all $0 \le i \le j$
(2) $|s_i| = |t_i|$ for all $0 \le i < j$
(3) $X \cap s_0 s_1 \dots s_{j-1}\Gamma^+$ is infinite.



s_0	s_1	s_2	s_3	s_4
t_0	t_1	t_2	t_3	t_4
00001	00000001	01	00000000001	00001

FIGURE 2. A comb and its coding

Since there are only finitely many words of length $|t_j|$, (3) implies the existence of a word $s_j \in \Gamma^+$ with $|s_j| = |t_j|$ such that $X \cap s_0 s_1 \dots s_j \Gamma^+$ is infinite ensuring (2) and (3) for $j+1$. Choose $t_{j+1} \in \Gamma^+$ with $s_0 s_1 \dots s_j t_{j+1} \in X$ arbitrary. Since this ensures (1), we can proceed by induction. Then the set of words $s_0 s_1 \dots s_{i-1} t_i$ is an infinite comb $C \subseteq X$. □

Words from $(\Sigma^2 \times \{0,1\})^*$ encode finite combs as follows:

DEFINITION 2.14. Let $n_0, n_1, \dots, n_\alpha \in \mathbb{N}$ and $w = 0^{n_0} 1 0^{n_1} 1 \dots 0^{n_{\alpha-1}} 1 0^{n_\alpha}$. Furthermore, let $s_i, t_i \in \Sigma^*$ with $|s_i| = |t_i| = n_i + 1$ for $0 \leq i < \alpha$ and $|s_\alpha| = |t_\alpha| = n_\alpha$. The comb encoded by $(s_0 s_1 \dots s_\alpha, t_0 t_1 \dots t_\alpha, w)^\otimes$ is the set of words $\{s_0 s_1 \dots s_{i-1} t_i \mid 0 \leq i \leq \alpha\}$.

Note that any word from $(\Sigma^2 \times \{0,1\})^*$ determines α, n_i, s_i, and t_i and therefore encodes a unique comb. Conversely, every finite comb can be encoded by some worf from $(\Sigma^2 \times \{0,1\})^*$ (which is not unique since s_α is arbitrary).

LEMMA 2.15. *Let $E \subseteq [\Sigma^*]^k$ be automatic. Then the set of encodings of finite E-homogeneous combs is regular.*

PROOF. Since the class of regular languages is closed under complementation, it suffices to construct a finite automaton that accepts a word from $(\Sigma^2 \times \{0,1\})^*$ if and only if it does *not* encode an E-homogeneous comb.

So let $M = (Q, (\Sigma \cup \{\#\})^k, \delta, I, F)$ be a finite automaton that accepts the set $E^\otimes = \{(u_1, \dots, u_k)^\otimes \mid \{u_1, \dots, u_k\} \in E\}$. We can assume that M is deterministic and complete, i.e., $|I| = 1$ and for any $p \in Q$ and $a \in (\Sigma \cup \{\#\})^k$, there exists a unique state $q \in Q$ with $(p, a, q) \in \delta$. We build a finite automaton M' as follows:

- states of M' are pairs (q, f) with $q \in Q$, $f : \{1, \dots, k\} \to \{1, 2, 3\}$, and $|f^{-1}(2)| \leq 1$,
- a state (q, f) is initial if and only if $q \in I$ and $f(i) \leq 2$ for all $1 \leq i \leq k$,
- a state (q, f) is accepting (i.e., belongs to F') if and only if $q \notin F$ and $f(i) \geq 2$ for all $1 \leq i \leq k$, and
- a triple $((q, f), (a, b, x), (q', f')) \in Q' \times (\Sigma^2 \times \{0,1\}) \times Q'$ is a transition from δ' if and only if
  - $f(i) \leq f'(i) \leq f(i) + x$ for all $1 \leq i \leq k$ and

– $(q, (c_i)_{1\leq i\leq k}, q') \in \delta$ where

$$c_i = \begin{cases} a & \text{if } f(i) = 1 \\ b & \text{if } f(i) = 2 \\ \# & \text{otherwise.} \end{cases}$$

Let $(s,t,w)^{\otimes} \in (\Sigma^* \times \{0,1\})^*$. We use α, n_i, s_i, and t_i as in Definition 2.14. Then the automaton M' has a run from some initial state to (q,f) labeled $(s,t,w)^{\otimes}$ if and only if there are pairwise distinct numbers $0 \leq j_i < \alpha$ for $f(i) > 1$ and a run of M from $\iota \in I$ to q labeled $(u_1,\dots,u_k)^{\otimes}$ where

(1) $u_i = \begin{cases} s & \text{if } f(i) = 1 \\ s_0 s_1 \dots s_{j_i} t_{j_i+1} & \text{if } f(i) > 1, \end{cases}$

(2) $j_i = \alpha - 1$ implies $f(i) = 2$ or $n_\alpha = 0$, and

(3) $f(i) = 2$ implies $j_i = \alpha - 1$.

Now suppose $(s,t,w)^{\otimes} \in (\Sigma^2 \times \{0,1\})^*$. Then $(s,t,w)^{\otimes}$ is accepted by M' if and only if there is some $q \in Q \setminus F$ such that M' has a run from some initial state to (q,f) labeled $(s,t,w)^{\otimes}$ where $f(i) = 3$ for all i. This is, by the above observation, equivalent to the existence of pairwise distinct $0 \leq j_i < \alpha$ for $1 \leq i \leq k$ (where $j_i = \alpha - 1$ is only allowed if $n_\alpha = 0$) and a run of M from $\iota \in I$ to q labeled $(u_1,\dots,u_k)^{\otimes}$ where $u_i = s_0 s_1 \dots s_{j_i} t_{j_i+1}$ for $1 \leq i \leq k$. Since M is deterministic and complete, such a run exists if and only if $\{u_1,\dots,u_k\} \notin E$. Hence, indeed, the finite automaton M' accepts a word $(s,t,w)^{\otimes}$ from $(\Sigma^2 \times \{0,1\})^*$ if and only if it encodes a comb that is not E-homogeneous. ☐

Proof of Prop. 2.11. Let $(V, E_1, \dots, E_\ell)$ be an infinite (k,ℓ)-partition with E_i automatic for all $1 \leq i \leq \ell$ and let $X \subseteq V$ by E_1-homogeneous. By Lemma 2.13, there exists an infinite comb $C \subseteq X$ (that is also E_1-homogeneous). Let $C = \{s_0 s_1 s_2 \dots s_{i-1} t_i \mid 0 \leq i < \omega\}$ with $s_i, t_i \in \Gamma^+$ and $|t_i| = |s_i|$ for all $0 \leq i < \omega$. For $0 < \alpha < \omega$, let W_α encode the finite comb $\{s_0 s_1 s_2 \dots s_{i-1} t_i \mid 0 \leq i < \alpha\}$. We can assume $W_1 <_{\text{pref}} W_2 <_{\text{pref}} W_3 \dots$.

Let L consist of all words from $(\Sigma^2 \times \{0,1\})^* (\Sigma^2 \times \{1\})$ that encode some E_1-homogeneous comb. Since L is regular by Lemma 2.15, it can be accepted by a deterministic and complete finite automaton. Hence there are $m < n$ and a word W with $W_n = W_m W$ such that $W_m W^* \subseteq L$ (we write V for W_m such that $VW^* \subseteq L$).

There are words $s_V, t_V, s_W, t_W \in \Sigma^+$ and $w_V, w_W \in \{0,1\}^+$ such that $V = (s_V, t_V, w_V)^{\otimes}$ and $W = (s_W, t_W, w_W)^{\otimes}$. Let n be such that $0^n 1$ is a prefix of w_W and let t be the prefix of t_W of length $n+1$.

Since $V \in L$, the word w_V ends with 1. For $1 \leq i \leq k$, let $u_i = s_V s_W^{n_i} t$ for pairwise distinct numbers n_i and let $\alpha < \omega$ be some upper bound for the numbers n_i. Then all the words u_i belong to the comb encoded by $VW^\alpha \in L$. Since this comb is E_1-homogeneous, we get $[s_V s_W^+ t]^k \subseteq E_1$. Hence, indeed, there is some infinite regular homogeneous set. ☐

The concept of a comb was first used by Khoussainov et al. in [**10**] where they show that every automatic order tree with at least one infinite branch has a regular infinite branch. Büchi exploited the relation between automata and logic and showed, in this context, that the first-order theory of every automatic structure is decidable [**9**]. Rubin [**15**] proved, using combs again, that this holds even if first-order logic is extended by Ramsey quantifiers. This result was then extended by

Kuske & Lohrey [**12**] to a weak form of second-order quantification, namely to quantifications of the form

$$\exists R \text{ infinite} : \varphi$$

where the relational variable R occurs only negatively in φ. This decidability implies that certain graph problems that are Σ^1_1-complete for computable graphs are decidable for automatic graphs [**12**].

The original proofs of all these results used infinite combs and therefore Büchi-automata that accept infinite words. Differently, the above proof works in the realm of finite automata and finite words.

3. The uncountable case

We now deal with the existence of uncountable homogeneous sets. The first observation in this direction is due to Sierpiński who showed that the direct analogue for Ramsey's theorem $\aleph_0 \to (\aleph_0)^k_\ell$ does not hold:

THEOREM 3.1 (Sierpiński [**17**]). *If $k, \ell \geq 2$, then $2^{\aleph_0} \not\to (\aleph_1)^k_\ell$ and therefore in particular $2^{\aleph_0} \not\to (2^{\aleph_0})^k_\ell$.*

PROOF. We first sketch the proof of $2^{\aleph_0} \not\to (\aleph_1)^2_2$: Let $\sqsubseteq$ be some well-order on $\mathbb{R}$ and let $E_1 = \{\{x, y\} \in [\mathbb{R}]^2 \mid x \leq y \iff x \sqsubseteq y\}$ and $E_2 = [\mathbb{R}]^2 \setminus E_1$. If $H \subseteq \mathbb{R}$ is E_1-homogeneous, then $(H, \sqsubseteq) = (H, \leq) \hookrightarrow (\mathbb{R}, \leq)$, i.e., the well-order $(H, \sqsubseteq)$ embeds into the linear order $(\mathbb{R}, \leq)$. Hence H is countable. If H is E_2-homogeneous, then $(H, \sqsubseteq) = (H, \geq) \hookrightarrow (\mathbb{R}, \geq) \cong (\mathbb{R}, \leq)$, so again H is countable. Hence the $(2, 2)$-partition $(\mathbb{R}, E_1, E_2)$ does not contain any uncountable homogeneous set.

To also show $2^{\aleph_0} \not\to (\aleph_1)^k_\ell$ for $k > 2$, define $A \in E'_i$ if and only if $\{x_1, x_2\} \in E_i$ where $x_1 < x_2$ are the two minimal elements of A. Then $(\mathbb{R}, E'_1, E'_2, \emptyset, \ldots, \emptyset)$ is a (k, ℓ)-partition of size $2^{\aleph_0}$ without uncountable homogeneous sets. □

3.1. Erdős & Rado's theorem. This theorem is concerned with homogeneous sets of size $\aleph_1$. We will assume the generalized continuum hypothesis $2^{\aleph_k} = \aleph_{k+1}$ for all $k \in \omega$.

THEOREM 3.2 (Erdős & Rado [**6**]). *For all $k, \ell \geq 1$, we have*

$$\aleph_k \to (\aleph_1)^k_\ell\,.$$

PROOF. The proof is, as the proof of Ramsey's theorem, by induction on k where the base case $k = 1$ is trivial.

So let $P = (V, E_1, \ldots, E_\ell)$ be some $(k+1, \ell)$-partition of size $\aleph_{k+1}$ and consider the Ramsey tree of P with respect to some well-order on V. Recall that every node at level $\alpha < \omega_k$ has at most $2^{\aleph_{k-1}} = \aleph_k$ brothers. Hence, by induction, there are at most $\aleph_k$ nodes at level α. In total, there are only $\aleph_k$ nodes on levels $< \omega_k$. Since the size of the Ramsey tree equals that of the $(k + 1, \ell)$-partition, there is at least one node at level ω_k. So the Ramsey tree contains a chain Y of size at least $\aleph_k$. Let P^Y be the (k, ℓ)-partition induced by Y. Since, by induction, we have $\aleph_k \to (\aleph_1)^k_\ell$, the (k, ℓ)-partition P^Y contains some homogeneous set of size $\aleph_1$. But this set is also homogeneous in P by Prop. 1.4. □

3.2. Ramsey's theorem for ω-automatic partitions. We next want to find an effective analogue of the theorem of Erdős and Rado. For this, we use the framework of ω-automatic structures as defined formally in [**2, 3**].

3.2.1. *Languages of infinite words.* An ω-*word* over the alphabet Γ is an infinite ω-sequence $x = a_0a_1a_2\cdots$ with $a_i \in \Gamma$. The set of all ω-words over Γ is denoted by Γ^ω. For a set $V \subseteq \Gamma^+$ of finite words let $V^\omega \subseteq \Gamma^\omega$ be the set of all ω-words of the form $v_0v_1v_2\cdots$ with $v_i \in V$.

Two ω-words x and y are *eventually equal* (denoted $x \sim_e y$) if, from some position on, the two words coincide. For $\Sigma = \{0,1\}$, the support $\mathrm{supp}(x) \subseteq \mathbb{N}$ is the set of positions of the letter 1 in the word $x \in \Sigma^\omega$.

A *run* of a finite automaton $M = (Q, \Gamma, \delta, I, F)$ on an ω-word $x = a_0a_1a_2\cdots$ is an ω-word $r = p_0p_1p_2\cdots$ over the set of states Q such that $(p_i, a_i, p_{i+1}) \in \delta$ for all $i \geq 0$. The run r is *successful* if $p_0 \in I$ and there exists a final state from F that occurs infinitely often in r. The ω-language $L^\omega(M) \subseteq \Gamma^\omega$ defined by M is the set of all ω-words that admit a successful run. Note that a finite automaton M accepts a language of finite words $L(M)$ and an ω-language $L^\omega(M)$. Whenever we are more interested in the latter, we will speak of M as a Büchi-automaton. An ω-language $L \subseteq \Gamma^\omega$ is *regular* if there exists a Büchi-automaton M with $L^\omega(M) = L$.

Alternatively, regular ω-languages can be represented algebraically. To this end, one defines ω-*semigroups* to be two-sorted algebras $S = (S_+, S_\omega; \cdot, *, \pi)$ where $\cdot : S_+ \times S_+ \to S_+$ and $* : S_+ \times S_\omega \to S_\omega$ are binary operations and $\pi : (S_+)^\omega \to S_\omega$ is an ω-ary operation such that the following hold:

- $(S_+, \cdot)$ is a semigroup,
- $s * (t * u) = (s \cdot t) * u$,
- $s_0 \cdot \pi((s_i)_{i\geq 1}) = \pi((s_i)_{i\geq 0})$,
- $\pi((s_i^1 \cdot s_i^2 \cdots s_i^{k_i})_{i\geq 0}) = \pi((t_j)_{j\geq 0})$ whenever

$$(t_j)_{j\geq 0} = (s_0^1, s_0^2, \dots, s_0^{k_0}, s_1^1, \dots, s_1^{k_1}, \dots)\ .$$

The ω-semigroup S is *finite* if both, S_+ and S_ω are finite. The free ω-semigroup generated by Γ is

$$\Gamma^\infty = (\Gamma^+, \Gamma^\omega; \cdot, *, \pi)$$

where $u \cdot v$ and $u * x$ are the natural operations of prefixing a word by the finite word u, and $\pi((u_i)_{i\geq 0})$ is the ω-word $u_0u_1u_2\dots$. A homomorphism $h : \Gamma^\infty \to S$ of ω-semigroups maps finite words to elements of S_+ and ω-words to elements of S_ω and commutes with the operations $\cdot$, $*$, and π. The algebraic characterisation of regular ω-languages then reads as follows.

Proposition 3.3. *An ω-language $L \subseteq \Gamma^\omega$ is regular if and only if there exists a finite ω-semigroup S and a homomorphism $\eta : \Gamma^\infty \to S$ such that $L = \eta^{-1}(\eta(L))$.*

Hence, every Büchi-automaton M is "equivalent" to a homomorphism into some finite ω-semigroup together with a distinguished set $\eta(L^\omega(M))$ (and vice versa). Even more, this translation is effective – see [**13**] for this and more results on regular ω-languages. Note that the equation $L = \eta^{-1}(\eta(L))$ can also be read as

$$\eta(x) = \eta(y) \implies (x \in L \iff y \in L)$$

for all ω-words x and y.

For ω-words $x_i = a_i^0a_i^1a_i^2\cdots \in \Gamma^\omega$, the *convolution* $x_1 \otimes x_2 \otimes \cdots \otimes x_n \in (\Gamma^n)^\omega$ is defined by

$$(x_1, \dots, x_n)^\otimes = (a_1^0, \dots, x_n^0)\,(a_1^1, \dots, a_n^1)\,(a_1^2, \dots, a_n^2)\cdots\ .$$

An n-ary relation $R \subseteq (\Gamma^\omega)^n$ is called *ω-automatic* if the ω-language $\{(x_1, \ldots, x_n)^\otimes \mid (x_1, \ldots, x_n) \in R\}$ is regular, a partition class $E \subseteq [\Gamma^\omega]^k$ is *ω-automatic* if the relation $\{(x_1, \ldots, x_k) \mid \{x_1, \ldots, x_k\} \in E\}$ is.

To describe the complexity of ω-languages, we will use language-theoretic terms. Let ωLANG denote the class of all ω-languages (i.e., sets of ω-words over some finite set of symbols). By ωREG, we denote the regular ω-languages. A language is *context-free* if it can be accepted by a pushdown-automaton. An ω-language is *eventually regular context-free* if it is a finite union of ω-languages UV with U context-free and $V \in \omega$REG. Let co-ωerCF denote the set of complements of eventually regular context-free ω-languages.

A final, rather peculiar class of ω-languages is $\mathbf{\Lambda}$: it is the class of ω-languages L such that $(\mathbb{R}, \leq)$ embeds into $(L, \leq_{\text{lex}})$ (the name derives from the notation $\boldsymbol{\lambda}$ for the order type of $(\mathbb{R}, \leq)$).

An *ω-automatic presentation of a (k, ℓ)-partition* $(V, E_1, \ldots, E_\ell)$ is a pair (L, h) consisting of an ω-language L and a surjection $h : L \to V$ such that

$$\begin{aligned} E_i^\otimes &= \{(x_1, x_2, \ldots, x_k) \in L^k \mid \{h(x_1), h(x_2), \ldots, h(x_k)\} \in E_i\} \text{ for } 1 \leq i \leq k \text{ and} \\ R_\approx &= \{(x_1, x_2) \in L^2 \mid h(x_1) = h(x_2)\} \end{aligned}$$

are ω-automatic. An ω-automatic presentation is *injective* if h is a bijection. A (k, ℓ)-partition is *(injectively) ω-automatic* if it has an (injective) ω-automatic presentation. From [**1**], it follows that an uncountable ω-automatic (k, ℓ)-partition has $2^{\aleph_0}$ elements.

Extending notions we used before, we write

$$(\kappa, \omega\mathsf{A}) \to (\lambda, \mathcal{C})^k_\ell$$

if the following partition property holds: for every ω-automatic presentation (L, h) of a (k, ℓ)-partition P of size κ, there exists $H \subseteq L$ in $\mathcal{C}$ such that $h(H)$ is homogeneous in P and of size λ;

$$(\kappa, \omega\mathsf{iA}) \to (\lambda, \mathcal{C})^k_\ell$$

is to be understood similarly where we only consider injective ω-automatic presentations.

Remark 3.4. Let $P = (V, E_1, \ldots, E_\ell)$ be some (k, ℓ)-partition with ω-automatic presentation (L, h). Then the partition property above requires that there is a "large" homogeneous set $X \subseteq V$ and an ω-language $H \in \mathcal{C}$ such that $h(H) = X$, in particular, every element of X has at least one representative in H. Alternatively, one could require that $h^{-1}(X) \subseteq L$ is an ω-language from $\mathcal{C}$. We will only encounter classes $\mathcal{C}$ of ω-languages such that the following closure property holds: if $H \in \mathcal{C}$ and R is an ω-automatic relation, then also $R(H) = \{y \mid \exists x \in H : (x, y) \in R\} \in \mathcal{C}$. Since $h^{-1}h(H) = R_\approx(H)$, all our results also hold for this alternative requirement $h^{-1}(X) \in \mathcal{C}$.

3.2.2. *Ramsey's theorem for ω-automatic partitions.* We want to prove

$$(2^{\aleph_0}, \omega\mathsf{A}) \to (2^{\aleph_0}, \text{co-}\omega\text{erCF} \cap \mathbf{\Lambda})^2_\ell$$

for all $\ell \geq 2$. Hence we will be concerned with edge-labeled graphs. The proof will not be done via the Ramsey tree. Instead, we will identify a common substructure of all ω-automatic edge-labeled graphs (see Example 3.5) and show that this substructure gives rise to homogeneous sets as required.

EXAMPLE 3.5. Let $V = \{0,1\}^\omega$ consist of all ω-words. Two distinct such words x and y are connected by an edge (i.e., $\{x,y\}$ belongs to E_1) if and only if $\mathrm{supp}(x)\setminus\mathrm{supp}(y)$ and $\mathrm{supp}(y)\setminus\mathrm{supp}(x)$ are both infinite; $E_2 = [V]^2\setminus E_1$ denotes the set of non-edges. We construct an instructive E_1-homogeneous set of size $2^{\aleph_0}$. For this, let $u_0, u_1, u_2, \dots$ be the list of words from $\{0,1\}^*$ in length-lexicographic order (i.e., u comes before v if $|u| < |v|$ or $|u| = |v|$, $u = x0y$ and $v = x1z$). Intuitively, we list the nodes of the complete binary tree $(\{0,1\}^*, \leq_{\mathrm{pref}})$ level-wise where each level is listed from left to right. Let the ω-language N consist of all ω-words $x \in \{0,1\}^\omega$ such that

$$\{u_i \mid i \in \mathrm{supp}(x)\}$$

is an infinite branch in the complete binary tree $(\{0,1\}^*, \leq_{\mathrm{pref}})$. Then, clearly, $N \subseteq V$ has size $2^{\aleph_0}$ since its elements correspond to infinite branches. An ω-word $x \in V$ does not belong to N if and only if

- $0 \notin \mathrm{supp}(x)$ or
- $\exists i \in \mathrm{supp}(x) : |\{2i+1, 2i+2\} \cap \mathrm{supp}(x)| \neq 1$ or
- $\exists i \in \mathrm{supp}(x) \setminus \{0\} : \lfloor \frac{i-1}{2} \rfloor \notin \mathrm{supp}(x)$.

Since this condition relates positions i and $2i$, this is a typical context-free condition, hence N is the complement of an eventually regular context-free language. Finally, any two branches have a finite intersection, so $\mathrm{supp}(x) \cap \mathrm{supp}(y)$ is finite for any $x, y \in N$ distinct. Since $\mathrm{supp}(x)$ and $\mathrm{supp}(y)$ are infinite, this implies $[N]^2 \subseteq E_1$. Hence we found an E_1-homogeneous set of size $2^{\aleph_0}$ whose complement is eventually regular context-free. Let $x, y \in N$ be distinct with $X = \{u_i \mid i \in \mathrm{supp}(x)\}$ and $Y = \{u_i \mid i \in \mathrm{supp}(y)\}$. Then $x <_{\mathrm{lex}} y$ if and only if the branch X is right of the branch Y. Since $(\mathbb{R}, \leq)$ can be embedded into the set of branches (ordered naturally from right to left), there is also an embedding into $(N, \leq_{\mathrm{lex}})$. Hence $N \in \mathbf{\Lambda}$.

Proposition 3.7 below will express that this example is typical for ω-automatic $(2,\ell)$-partitions. But for its proof, we first need the following.

LEMMA 3.6. *Let $L \subseteq \Sigma^\omega$ be regular and $\approx$ an automatic equivalence relation on L with $|L/{\approx}| \geq \aleph_1$. Then there exist finite words $u_i, v_i \in \Sigma^+$ such that $u_i v_i^\omega \in L$ and $[u_i v_i^\omega]_{\sim_e} \cap [u_j v_j^\omega]_\approx = \emptyset$ for all $0 \leq i < j$.*

PROOF. Let $n \geq 0$ and suppose we have constructed pairs (u_i, v_i) with the desired properties for $0 \leq i < n$. We want to find (u_n, v_n). Suppose, for all $x \in L$, there exists $0 \leq i < n$ with $[u_i v_i^\omega]_{\sim_e} \cap [x]_\approx \neq \emptyset$. In other words, the countable set $\bigcup_{0 \leq i < n} [u_i v_i^\omega]_{\sim_e}$ intersects every equivalence class of $\approx$. But this contradicts $|L/\{\approx\}| > \aleph_0$. Hence there exists an element $x \in L$ with $[u_i v_i^\omega]_{\sim_e} \cap [x]_\approx = \emptyset$ for all $i < n$, i.e., the formula

$$\bigwedge_{0 \leq i < n} \neg \exists y : u_i v_i^\omega \sim_e y \approx x$$

is satisfiable in the structure $(L, (u_i v_i^\omega), \sim_e, \approx)$. Since the relations $\sim_e$, $\approx$, and $\{u_i v_i^\omega\}$ for $0 \leq i < n$ are automatic, also the set of ω-words x satisfying this formula is automatic and therefore a nonempty regular ω-language [**2**]. Hence it contains some ultimately periodic ω-word, i.e., there are $u_n, v_n \in \Sigma^+$ as desired. $\square$

Let u, v, w be some finite words. For $x \in \{0,1\}^\omega$, let $f_{v,w}(x)$ denote the ω-word obtained from x by replacing every occurrence of 0 by v and every occurrence of 1 by w. Furthermore, $H_{u,v,w} = u f_{v,w}(N)$ where N is the language from Example 3.5.

Proposition 3.7. *Let $P = (L, E_0, E_1, \ldots, E_\ell)$ be some $(2, 1+\ell)$-partition with injective ω-automatic presentation (L, id) such that $\{(x,y) \mid \{x,y\} \in E_0\} \cup \{(x,x) \mid x \in L\}$ is an equivalence relation (denoted $\approx$) on L with at least $\aleph_1$ equivalence classes. Then there exist nonempty words u, v, and w with v and w distinct, but of the same length, such that $H_{u,v,w}$ is E_i-homogeneous for some $1 \le i \le \ell$.*

This proposition is stated in [**11**] where I explain how to obtain it from a proof in [**1**]. Here, we give a self-contained proof that follows and simplifies the ideas from [**1, 11**].

Proof. There are finite ω-semigroups S and T and homomorphisms

$$\varphi : \Sigma^\infty \to S \text{ and } \psi : (\Sigma \times \Sigma)^\infty \to T$$

such that $\varphi^{-1}\varphi(L) = L$ and

$$\psi(x \otimes y) = \psi(x' \otimes y') \implies (\{x,y\} \in E_i \iff \{x',y'\} \in E_i)$$

for all $0 \le i \le \ell$. Since the semigroups S_+ and T_+ are finite, there exists $\alpha \ge 1$ such that x^α is idempotent for all $x \in S_+ \cup T_+$.

From Lemma 3.6, we find finite words $u_i, v_i \in \Sigma^+$ such that $u_i v_i^\omega \in L$ and

$$[u_i v_i^\omega]_{\sim_e} \cap [u_j v_j^\omega]_\approx = \emptyset$$

for all $0 \le i < j \le |S_+| \cdot |T_+|$. Let m be the maximal length of a word u_i for $0 \le i \le |S_+| \cdot |T_+|$. By extending the word u_i by the prefix of v_i^ω of length $m - |u_i|$ (and rotating the word v_i accordingly), we can assume $|u_i| = |u_j|$ for all $0 \le i < j \le |S_+| \cdot |T_+|$ without changing the words $u_i v_i^\omega$. Furthermore, we can replace v_i by its

$$\left(\alpha \cdot \prod_{\substack{0 \le j \le |S_+| \cdot |T_+| \\ i \ne j}} |v_j| \right)^{th}$$

power, i.e., we can assume that $|v_i| = |v_j|$ and that all the elements

$$\varphi(v_i) \in S_+ \text{ and } \psi(v_i \otimes v_j) \in T_+$$

are idempotent.

Now there are $0 \le i < j \le |S_+| \cdot |T_+|$ with $\varphi(v_i) = \varphi(v_j)$ and $\psi(v_i \otimes v_i) = \psi(v_j \otimes v_j)$. We may assume $i = 0$ and $j = 1$ and write, for simplicity,

$$s = \varphi(v_0) = \varphi(v_1), \qquad t = \psi(v_0 \otimes v_0) = \psi(v_1 \otimes v_1),$$

$$t_{01} = \psi((v_0v_0v_0 \otimes v_0v_1v_0)^\alpha), \text{ and } \qquad t_{10} = \psi((v_0v_1v_0 \otimes v_0v_0v_0)^\alpha).$$

We will use the following equalities repeatedly:

$$t_{01} = t_{01} \cdot t = t \cdot t_{01} = t_{01}^2 \qquad t_{10} = t_{10} \cdot t = t \cdot t_{10} = t_{10}^2 \qquad t = t^2$$

Finally, for $X \subseteq \mathbb{N}$ we define a word W_X as follows:

$$W_X = u_1 v_1 \prod_{i \ge 0} \left\{ \begin{array}{ll} (v_0v_0v_0)^\alpha & \text{if } i \in X \\ (v_0v_1v_0)^\alpha & \text{if } i \notin X \end{array} \right\}$$

Note that, with $u = u_1v_1$, $v = (v_0v_1v_0)^\alpha$, and $w = (v_0v_0v_0)^\alpha$, we get $W_X = u f_{v,w}(\chi_X)$ where $\chi_X \in \{0,1\}^\omega$ is the characteristic word of X.

We will now prove three claims concerning the words W_X.

Claim 1. *$W_X \in L$ for all $X \subseteq \mathbb{N}$.*

Proof of Claim 1. We have $\varphi(W_X) = \varphi(u_1)\varphi(v_1)s^\omega = \varphi(u_1 v_1^\omega)$ since $W_X \in u_1 v_1\{v_0, v_1\}^\omega$. Now $W_X \in L$ follows from $u_1 v_1^\omega \in L$. q.e.d.

Claim 2. *Let $X, Y \subseteq \mathbb{N}$ such that $X \setminus Y$ and $Y \setminus X$ are both infinite. Then*

$$\psi(W_X \otimes W_Y) = \psi(u_1 \otimes u_1)\begin{cases}(t_{01}t_{10})^\omega & \textit{if } \min(X\triangle Y) \in X\\ (t_{10}t_{01})^\omega & \textit{if } \min(X\triangle Y) \in Y\end{cases}$$

Proof of Claim 2.

$$\psi(W_X \otimes W_Y) = \psi(u_1 v_1 \otimes u_1 v_1)\prod_{0\le i}\left\{\begin{array}{ll}\psi(v_0v_0v_0 \otimes v_0v_0v_0)^\alpha & \text{if } i \in X\cap Y\\ \psi(v_0v_0v_0 \otimes v_0v_1v_0)^\alpha & \text{if } i \in X\setminus Y\\ \psi(v_0v_1v_0 \otimes v_0v_0v_0)^\alpha & \text{if } i \in Y\setminus X\\ \psi(v_0v_1v_0 \otimes v_0v_1v_0)^\alpha & \text{otherwise}\end{array}\right\}$$

$$= \psi(u_1 \otimes u_1)\, t\prod_{0\le i}\left\{\begin{array}{ll} t_{01} & \text{if } i \in X\setminus Y\\ t_{10} & \text{if } i\in Y\setminus X\\ t^{3\alpha} = t & \text{otherwise}\end{array}\right\}$$

Now the claim follows since there are infinitely many factors t_{01} and t_{10} (since $X \setminus Y$ and $Y \setminus X$ are both infinite), since these two factors are idempotent, and since multiplication of t with t_{10} or t_{01} absorbes t. q.e.d.

From Claim 2, we now derive that $H_{u,v,w}$ is homogeneous in the partition P: Let $A, B \in [H_{u,v,w}]^2$. Then there exist sets $X, Y, X', Y' \subseteq \mathbb{N}$ with $A = \{W_X, W_Y\}$ and $B = \{W_{X'}, W_{Y'}\}$. Furthermore, the differences $X \setminus Y$, $Y \setminus X$, $X' \setminus Y'$, and $Y' \setminus X'$ are all infinite. We assume $\min(X\triangle Y) \in X$ and $\min(X'\triangle Y') \in X'$. Then, by Claim 2, $\psi(W_X \otimes W_Y) = \psi(W_{X'} \otimes W_{Y'})$ proving $A \in E_i \iff B \in E_i$ for all $0 \le i \le \ell$. Hence $H_{u,v,w}$ is homogeneous. To show that it is not 0-homogeneous, we use the following claim.

Claim 3. $\psi(W_{2\mathbb{N}} \otimes W_{4\mathbb{N}+2}) = \psi(u_1v_1v_0^\omega \otimes u_1(v_1v_0)^\omega) = \psi(u_1(v_1v_0)^\omega \otimes u_1 v_1^\omega$

Proof of Claim 3.

$$\begin{aligned}\psi(W_{2\mathbb{N}} \otimes W_{4\mathbb{N}+2}) &= \psi(u_1v_1 \otimes u_1v_1)\\ &\quad\cdot\prod_{0\le i}\left\{\begin{array}{ll}\psi(v_0v_0v_0 \otimes v_0v_0v_0)^\alpha & \text{if } i \in 4\mathbb{N}+2\\ \psi(v_0v_0v_0 \otimes v_0v_1v_0)^\alpha & \text{if } i \in 4\mathbb{N}\\ \psi(v_0v_1v_0 \otimes v_0v_1v_0)^\alpha & \text{if } i \text{ odd}\end{array}\right\}\\ &= \psi(u_1 \otimes u_1)t\prod_{0\le i}\left\{\begin{array}{ll} t_{01} & \text{if } i\in 4\mathbb{N}\\ t^{3\alpha} & \text{otherwise}\end{array}\right\}\\ &= \psi(u_1\otimes u_1)t_{01}^\omega \text{ since } tt_{01} = t_{01}\end{aligned}$$

$$\begin{aligned}\psi(u_1v_1v_0^\omega, u_1(v_1v_0)^\omega) &= \psi(u_1v_1 \otimes u_1v_1)\psi(v_0v_0 \otimes v_0v_1)^\omega\\ &= \psi(u_1\otimes u_1)t(t\psi(v_0\otimes v_1))^\omega\\ &= \psi(u_1\otimes u_1)(t\psi(v_0\otimes v_1)t)^\omega \text{ since } t^2 = t\\ &= \psi(u_1\otimes u_1)t_{01}^\omega \text{ since } (t\psi(v_0\otimes v_1)t)^\alpha = t_{01}\end{aligned}$$

$$\begin{aligned}\psi(u_1(v_1v_0)^\omega \otimes u_1v_1^\omega) &= \psi(u_1\otimes u_1)(t\psi(v_0\otimes v_1))^\omega\\ &= \psi(u_1\otimes u_1)(t\psi(v_0\otimes v_1)t)^\omega \text{ since } t^2 = t\\ &= \psi(u_1\otimes u_1)t_{01}^\omega \text{ since } (t\psi(v_0\otimes v_1)t)^\alpha = t_{01}\end{aligned}$$

q.e.d.

Suppose $[H_{u,v,w}]^2 \subseteq E_0$. Let $A \in [H_{u,v,w}]^2$, i.e., $A = \{W_X, W_Y\}$ for some $X, Y \subseteq \mathbb{N}$ with $X \setminus Y$ and $Y \setminus X$ infinite. We assume $\min(X \triangle Y) \in X$. From Claim 2, we obtain

$$\psi(W_X \otimes W_Y) = \psi(W_{2\mathbb{N}} \otimes W_{2\mathbb{N}+1}) = \psi(W_{2\mathbb{N}+1} \otimes W_{4\mathbb{N}+2}) \,.$$

Then $A \in E_0$ implies

$$\{W_{2\mathbb{N}}, W_{2\mathbb{N}+1}\}, \{W_{2\mathbb{N}+1}, W_{4\mathbb{N}+2}\} \in E_0 \,,$$

in other words,

$$W_{2\mathbb{N}} \approx W_{2\mathbb{N}+1} \approx W_{4\mathbb{N}+2} \,.$$

Since $\approx$ is an equivalence relation, this implies $W_{2\mathbb{N}} \approx W_{4\mathbb{N}+2}$. Now from Claim 3, we obtain $u_1 v_1 v_0^\omega \approx u_1 (v_1 v_0)^\omega \approx u_1 v_1^\omega$. Since $u_1 v_1 v_0^\omega \sim_e u_0 v_0^\omega$, this contradicts our choice of the words u_i and v_i. $\square$

THEOREM 3.8 (Kuske [**11**]). *For all $\ell \geq 2$, we have*

$$(2^{\aleph_0}, \omega\mathsf{A}) \to (2^{\aleph_0}, \text{co-}\omega\text{erCF} \cap \mathbf{\Lambda})^2_\ell \,.$$

PROOF. Let $P' = (V, E'_1, \ldots, E'_\ell)$ be some $(2, \ell)$ partition with ω-automatic presentation (L, h). To apply Prop. 3.7, consider the following $(2, 1+\ell)$-partition $P = (L, E_0, \ldots, E_\ell)$:

- The underlying set is the ω-language L,
- E_0 consists of all sets $\{x, y\}$ with $h(x) = h(y)$ and $x \neq y$, and
- E_i (for $1 \leq i \leq \ell$) consists of all sets $\{x, y\}$ with $\{h(x), h(y)\} \in E'_i$.

Then (L, id) is an injective ω-automatic presentation of the $(2, 1+\ell)$-partition P. By Prop. 3.7, there exists $1 \leq i \leq \ell$ and words u, v and w such that $v \neq w$, $|v| = |w|$, and such that $H_{u,v,w}$ is E_i-homogeneous for some $1 \leq i \leq \ell$. Since $(E_0, \ldots, E_\ell)$ is a partition of $[L]^2$, we have $\{x, y\} \notin E_0$ (and therefore $h(x) \neq h(y)$) for all $x, y \in H_{u,v,w}$ distinct. Hence h is injective on $H_{u,v,w}$. Furthermore $[H_{u,v,w}]^2 \subseteq E_i$ implies $[h(H_{u,v,w})]^2 \subseteq E'_i$. Hence $h(H_{u,v,w})$ is an E'_i-homogeneous set of size $2^{\aleph_0}$. From $N \in \text{co-}\omega\text{erCF} \cap \mathbf{\Lambda}$, we get immediately $H_{u,v,w} \in \text{co-}\omega\text{erCF} \cap \mathbf{\Lambda}$. $\square$

¿From Theorem 3.8, we now derive a necessary condition for a partial order of size $2^{\aleph_0}$ to be ω-automatic. A partial order $(V, \sqsubseteq)$ is ω-automatic iff there exists a regular ω-language L and a surjection $h : L \to V$ such that the relations $R_= = \{(x, y) \in L^2 \mid h(x) = h(y)\}$ and $R_\sqsubseteq = \{(x, y) \in L^2 \mid h(x) \sqsubseteq h(y)\}$ are ω-automatic.

COROLLARY 3.9. *If $(V, \sqsubseteq)$ is an ω-automatic partial order with $|V| \geq \aleph_1$, then $(\mathbb{R}, \leq)$ or an antichain of size $2^{\aleph_0}$ embeds into $(V, \sqsubseteq)$.*

PROOF. Let $(V, \sqsubseteq)$ be a partial order, $L \subseteq \Sigma^\omega$ a regular ω-language and $h : L \to V$ a surjection such that $R_=$ and $R_\sqsubseteq$ are ω-automatic. Define an injective ω-automatic $(2, 4)$-partition $G = (L, E_0, E_1, E_2, E_3)$:

- E_0 comprises all pairs $\{x, y\} \in [L]^2$ with $h(x) = h(y)$,
- E_1 comprises all pairs $\{x, y\} \in [L]^2$ with $h(x) \sqsubset h(y)$ and $x <_{\text{lex}} y$,
- E_2 comprises all pairs $\{x, y\} \in [L]^2$ with $h(x) \sqsupset h(y)$ and $x <_{\text{lex}} y$, and
- $E_3 = [L]^2 \setminus (E_0 \cup E_1 \cup E_2)$ comprises all pairs $\{x, y\} \in [L]^2$ such that $h(x)$ and $h(y)$ are incomparable.

By Prop. 3.7, there exists $H \subseteq L$ 1-, 2- or 3-homogeneous with $(\mathbb{R}, \leq) \hookrightarrow (H, \leq_{\text{lex}})$. Since $[H]^2 \subseteq E_1 \cup E_2 \cup E_3$ and since G is a partition of L, the mapping h acts injectively on H. If $[H]^2 \subseteq E_1$ (the case $[H]^2 \subseteq E_2$ is symmetrical) then $(\mathbb{R}, \leq) \hookrightarrow (H, \leq_{\text{lex}}) \cong (h(H), \sqsubseteq)$. If $[H]^2 \subseteq E_3$, then $h(H)$ is an antichain of size $2^{\aleph_0}$. □

A linear order $(L, \sqsubseteq)$ is *scattered* if $(\mathbb{Q}, \leq)$ cannot be embedded into $(L, \sqsubseteq)$. Automatic partial orders are defined similarly to ω-automatic partial orders with the help of finite automata instead of Büchi-automata.

COROLLARY 3.10. *Any scattered ω-automatic linear order $(V, \sqsubseteq)$ is countable. Hence,*

- *a scattered linear order is ω-automatic if and only if it is automatic, and*
- *an ordinal α is ω-automatic if and only if $\alpha < \omega^\omega$.*

PROOF. If $(V, \sqsubseteq)$ is not countable, then it embeds $(\mathbb{R}, \leq)$ by the previous corollary and therefore in particular $(\mathbb{Q}, \leq)$. The remaining two claims follow immediately from [**1**] ("countable ω-automatic structures are automatic") and [**4**] ("an ordinal is automatic iff it is properly smaller than ω^ω"), resp. □

3.2.3. *Sierpiński's theorem for ω-automatic partitions.* Compare Theorem 3.8 to Sierpiński's theorem 3.1 and the theorem of Erdős and Rado: if an edge-labeled graph is ω-automatic, already $2^{\aleph_0}$ nodes guarantee the existence of a homogeneous set of size $2^{\aleph_0}$. Compared to Rubin's theorem 2.12, there are two shortcomings: We would like to extend the result to (k, ℓ)-partitions for $k > 2$, and we would like to find ω-regular homogeneous sets (and not only complements of context-free ones). The following two examples prove that this is not possible, i.e., that Theorem 3.8 is best possible.

EXAMPLE 3.11. Let $\Sigma = \{0, 1\}$, $L = \{0, 1\}^\omega$. For $H \subseteq L$, we write $\bigwedge H \in \Sigma^\infty$ for the longest common prefix of all ω-words in H, $\bigwedge\{x, y\}$ is also written $x \wedge y$. By $\leq_{\text{lex}}$, we denote the lexicographic order on the set Σ^ω (with some, implicitly assumed linear order on the letters from Σ). Then let E_1 consist of all 3-sets $\{x, y, z\} \in [L]^3$ with $x <_{\text{lex}} y <_{\text{lex}} z$ and $x \wedge y <_{\text{pref}} y \wedge z$; E_2 is the complement of E_1. This finishes the construction of the $(3, 2)$-partition (L, E_1, E_2) of size $2^{\aleph_0}$ with injective ω-automatic presentation (L, id).

Note that 1^*0^ω is a countable E_1-homogeneous set and that 0^*1^ω is a countable E_2-homogeneous set. But there is no uncountable homogeneous set: First suppose $H \subseteq L$ is infinite and $x \wedge y <_{\text{pref}} y \wedge z$ for all $x <_{\text{lex}} y <_{\text{lex}} z$ from H. Let $u \in \Sigma^*$ such that $H \cap u0\Sigma^\omega$ and $H \cap u1\Sigma^\omega$ are both nonempty and let $x, y \in H \cap u0\Sigma^\omega$ with $x \leq_{\text{lex}} y$ and $z \in H \cap u1\Sigma^\omega$. Then $x \wedge y >_{\text{pref}} u = y \wedge z$ and therefore $x = y$ (for otherwise, we would have $x <_{\text{lex}} y <_{\text{lex}} z$ in H with $x \wedge y >_{\text{pref}} y \wedge z$). Hence we showed $|H \cap u0\Sigma^\omega| = 1$. Let $u_0 = \bigwedge H$ and $H_1 = H \cap u_0 1\Sigma^\omega$. Since $H \cap u_0 0\Sigma^\omega$ is finite, the set H_1 is infinite. We proceed by induction: $u_n = \bigwedge H_n$ and $H_{n+1} = H_n \cap u_n 1\Sigma^\omega$ satisfying $|H_n \cap u_n 0\Sigma^\omega| = 1$. Then $u_0 <_{\text{pref}} u_0 1 \leq_{\text{pref}} u_1 <_{\text{pref}} u_1 1 \leq_{\text{pref}} u_2 \cdots$ with

$$H = \bigcup_{n \geq 0} (H \cap u_n 0\Sigma^\omega) \cup \bigcap_{n \geq 0} (H \cap u_n 1\Sigma^\omega) \ .$$

Then any of the sets $H \cap u_n 0\Sigma^\omega = H_n \cap u_n 0\Sigma^\omega$ and $\bigcap(H \cap u_n 1\Sigma^\omega)$ is a singleton, proving that H is countable. Thus, there cannot be an uncountable E_1-homogeneous set.

So let $H \subseteq L$ be infinite with $x \wedge y \geq_{\text{pref}} y \wedge z$ for all $x <_{\text{lex}} y <_{\text{lex}} z$. Since we have only two letters, we get $x \wedge y >_{\text{pref}} y \wedge z$ for all $x <_{\text{lex}} y <_{\text{lex}} z$ which allows to argue symmetrically to the above. Thus, indeed, there is no uncountable homogeneous set in L. This proves $(2^{\aleph_0}, \omega\mathsf{iA}) \not\to (\aleph_1, \omega\text{LANG})^3_2$; using the trick from the end of the proof of Sierpiński's theorem 3.1 (where $x_1 < x_2$ is replace by $x_1 <_{\text{lex}} x_2$), we get $(2^{\aleph_0}, \omega\mathsf{iA}) \not\to (\aleph_1, \omega\text{LANG})^k_\ell$ for all $k \geq 3$ and $\ell \geq 2$.

Example 3.12. Let L denote the regular ω-language $(1^+0^+)^\omega$. Recall that the ω-words x and y are ultimately equal, briefly $x \sim_e y$, if there exist finite words u and v of the same length and an ω-word z with $x = uz$ and $y = vz$ (i.e., from some point on, the two ω-words coincide). Let $E_1 \subseteq [L]^2$ consists of all 2-sets $\{x, y\} \subseteq L$ such that $\text{supp}(x) \cap \text{supp}(y)$ is finite or $x \sim_e y$. The set E_2 is the complement of E_1 in $[L]^2$. This completes the construction of the $(2,2)$-partition $G = (L, E_1, E_2)$. Note that (L, id_L) is an injective ω-automatic presentation of G.

Now let $n \in \mathbb{N}$ and $U_i, V_i \subseteq \{0,1\}^+$ be languages, set

$$H = \bigcup_{1 \leq i \leq n} U_i V_i^\omega\,, \tag{3.1}$$

and assume that H has size $2^{\aleph_0}$. We show that H is not homogeneous: Since H is infinite, there are $1 \leq i \leq n$ and $x, y \in U_i V_i^\omega$ distinct with $x \sim_e y$ and therefore $\{x, y\} \in E_1$.

Since $|H| > \aleph_0$, there is $1 \leq i \leq n$ with $|U_i V_i^\omega| > \aleph_0$; we set $U = U_i$ and $V = V_i$. Let $u \in U$ be arbitrary (such a word exists since $UV^\omega \neq \emptyset$). From $|U| \leq \aleph_0$, we obtain $|V^\omega| > \aleph_0$. Hence there are $v_1, v_2 \in V^+$ distinct with $|v_1| = |v_2|$. Since $uv_1^\omega \in H$ and each element of H contains infinitely many occurrences of 1, the word v_1 belongs to $\{0,1\}^*10^*$. Consider the ω-words $x' = u(v_1v_2)^\omega$ and $y' = u(v_1v_1)^\omega$ from $UV^\omega \subseteq H$. Then $x' \not\sim_e y'$ since $v_1 \neq v_2$ and $|v_1| = |v_2|$. At the same time, $\text{supp}(x') \cap \text{supp}(y')$ is infinite since v_1 contains an occurrence of 1. Hence $\{x', y'\} \in E_2$.

Thus, we found ω-words $x, y, x', y' \in H$ with $\{x, y\} \in E_1$ and $\{x', y'\} \notin E_1$ proving that H is not homogeneous.

Since all context-free and regular ω-language are of the form (3.1) (see [**18**]), this proves $(2^{\aleph_0}, \omega\mathsf{iA}) \not\to (2^{\aleph_0}, \omega\text{CF})^k_\ell$ and $(2^{\aleph_0}, \omega\mathsf{iA}) \not\to (2^{\aleph_0}, \omega\text{REG})^k_\ell$ for all $k, \ell \geq 2$.

References

[1] V. Bárány, Ł. Kaiser, and S. Rubin, *Cardinality and counting quantifiers on omega-automatic structures*, STACS'08, IFIB Schloss Dagstuhl, 2008, pp. 385–396.

[2] A. Blumensath, *Automatic structures*, Tech. report, RWTH Aachen, 1999.

[3] A. Blumensath and E. Grädel, *Finite presentations of infinite structures: Automata and interpretations*, Theory of Computing Systems **37** (2004), no. 6, 641–674.

[4] Ch. Delhommé, *Automaticité des ordinaux et des graphes homogènes*, C. R. Acad. Sci. Paris, Ser. I **339** (2004), 5–10.

[5] P. Erdős and R. Rado, *A problem on ordered sets*, Journal of the LMS **28** (1953), 426–438.

[6] ______, *A partition calculus in set theory*, Bull. AMS **62** (1956), 427–489.

[7] C.G. Jockusch, *Ramsey's theorem and recursion theory*, Journal of Symbolic Logic **37** (1972), 268–280.

[8] C.G. Jockusch and R.I. Soare, *Π^0_1-classes an degrees of theories*, Trans. Am. Math. Soc. **173** (1972), 33–56.

[9] B. Khoussainov and A. Nerode, *Automatic presentations of structures*, Logic and Computational Complexity, Lecture Notes in Comp. Science vol. 960, Springer, 1995, pp. 367–392.

[10] B. Khoussainov, S. Rubin, and F. Stephan, *On automatic partial orders*, LICS'03, IEEE Computer Society Press, 2003, pp. 168–177.

[11] D. Kuske, *Is Ramsey's theorem ω-automatic?*, STACS 2010, Leibniz International Proceedings in Informatics (LIPIcs) vol. 5, Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2010, pp. 537–548.

[12] D. Kuske and M. Lohrey, *Some natural problems in automatic graphs*, Journal of Symbolic Logic **75** (2010), no. 2, 678–710.

[13] D. Perrin and J.-E. Pin, *Infinite words*, Pure and Applied Mathematics vol. 141, Elsevier, 2004.

[14] F.P. Ramsey, *On a problem of formal logic*, Proc. London Math. Soc. **30** (1930), 264–286.

[15] S. Rubin, *Automata presenting structures: A survey of the finite string case*, Bulletin of Symbolic Logic **14** (2008), 169–209.

[16] J.R. Shoenfield, *On degrees of unsolvability*, Annals of Mathematics **69** (1959), no. 3, 644–653.

[17] W. Sierpiński, *Sur un problème de la thèorie des relations*, Ann. Scuola Norm. Sup. Pisa **2** (1933), no. 2, 285–287.

[18] L. Staiger, *ω-languages*, Handbook of Formal Languages Vol. 3 (G. Rozenberg and A. Salomaa, eds.), Springer, 1997, pp. 339–387.

Institut für Theoretische Informatik, TU Ilmenau, PF 100565, D-98684 Ilmenau, Germany

E-mail address: `dietrich.kuske@tu-ilmenau.de`

This page intentionally left blank

Contemporary Mathematics
Volume **558**, 2011

Reducts of Ramsey structures

Manuel Bodirsky and Michael Pinsker

Abstract. One way of studying a relational structure is to investigate functions which are related to that structure and which leave certain aspects of the structure invariant. Examples are the automorphism group, the self-embedding monoid, the endomorphism monoid, or the polymorphism clone of a structure. Such functions can be particularly well understood when the relational structure is countably infinite and has a first-order definition in another relational structure which has a finite language, is totally ordered and homogeneous, and has the Ramsey property. This is because in this situation, Ramsey theory provides the combinatorial tool for analyzing these functions – in a certain sense, it allows to represent such functions by functions on finite sets.

This is a survey of results in model theory and theoretical computer science obtained recently by the authors in this context. In model theory, we approach the problem of classifying the reducts of countably infinite ordered homogeneous Ramsey structures in a finite language, and certain decidability questions connected with such reducts. In theoretical computer science, we use the same combinatorial methods in order to classify the computational complexity for various classes of infinite-domain constraint satisfaction problems. While the first set of applications is obviously of an infinitary character, the second set concerns genuinely finitary problems – their unifying feature is that the same tools from Ramsey theory are used in their solution.

Contents

2010 *Mathematics Subject Classification.* 03C40; 08A35; 05C55; 03D15.

The first author has received funding from the ERC under FP7/2007-2013 grant no. 257039. The second author is supported by an APART-fellowship of the Austrian Academy of Sciences.

1. Introduction

"I prefer finite mathematics much more than infinite mathematics. I think that it is much more natural, much more appealing and the theory is much more beautiful. It is very concrete. It is something that you can touch and something you can feel and something to relate to. Infinity mathematics, to me, is something that is meaningless, because it is abstract nonsense."

(Doron Zeilberger, February 2010)

"To the person who does deny infinity and says that it doesn't exist, I feel sorry for them, I don't see how such view enriches the world. Infinity may be does not exist, but it is a beautiful subject. I can say that the stars do not exist and always look down, but then I don't see the beauty of the stars. Until one has a real reason to doubt the existence of mathematical infinity, I just don't see the point."

(Hugh Woodin, February 2010)

Sometimes, infinite mathematics is not just beautiful, but also *useful*, even when one is ultimately interested in finite mathematics. A fascinating example of this type of mathematics is the recent theorem by Kechris, Pestov, and Todorcevic [**32**], which links Ramsey classes and topological dynamics. A class of finite structures $\mathcal{C}$ closed under isomorphisms, induced substructures, and with the joint embedding property (see [**28**]) is called a *Ramsey class* [**38**, **39**] (or *has the Ramsey property*) if for all $P, H \in \mathcal{C}$ and every $k \geq 2$ there is a $S \in \mathcal{C}$ such that for every coloring of the copies of P in S with k colors there is a copy H' of H in $\mathcal{C}$ such that all copies of P in H' have the same color. This is a very strong requirement — and certainly from the finite world. Proving that a class has the Ramsey property can be difficult [**38**], and Ramsey theory rather provides a tool box than a theory to answer this question.

Kechris, Pestov, and Todorcevic [**32**] provide a characterization of such classes in topological dynamics, connecting Ramsey classes with *extreme amenability* in (infinite) group theory, a concept from the 1960s [**27**]. The result can be used in two directions. One can use it to translate deep existing Ramsey results into proofs of extreme amenability of topological groups (and this is the main focus of the already cited article [**32**]). One can also use it in the other direction to obtain a more systematic understanding of Ramsey classes. A key insight for this direction is the result of Nešetřil (see [**39**]) which says that Ramsey classes $\mathcal{C}$ have the *amalgamation property*. Hence, by Fraïssé's theorem, there exists a countably infinite homogeneous and ω-categorical structure Γ such that a finite structure is from $\mathcal{C}$ if and only if it embeds into Γ. The structure Γ is unique up to isomorphism, and is called the *Fraïssé limit* of $\mathcal{C}$. Now let $\mathcal{D}$ be any amalgamation class whose Fraïssé limit Δ is bi-interpretable with Γ. By the theorem of Ahlbrandt and Ziegler [**3**], two ω-categorical structures are first-order bi-interpretable if and only if their automorphism groups are isomorphic as *(abstract) topological groups*. In addition, the above-mentioned result from [**32**] shows that whether or not $\mathcal{D}$ is a Ramsey class only depends on the automorphism group $\mathrm{Aut}(\Delta)$ of Δ; in fact, and much more interestingly, it only depends on $\mathrm{Aut}(\Delta)$ viewed as a topological group (which has cardinality 2^ω). From this we immediately get our first example where [**32**] is used in the second direction, with massive consequences for finite structures: the Ramsey property is preserved under first-order bi-interpretations. We will see another statement of this type (Proposition 24) and more concrete applications of such statements later (in Section 5, Section 7, and Section 8).

Constraint Satisfaction. Our next example where infinite mathematics is a powerful tool comes from (finite) computer science. A *constraint satisfaction problem* is a computational problem where we are given a set of variables and a set of constraints on those variables, and where the task is to decide whether there is an assignment of values to the variables that satisfies all constraints. Computational problems of this type appear in many areas of computer science, for example in artificial intelligence, computer algebra, scheduling, computational linguistics, and computational biology.

As an example, consider the BETWEENNESS problem. The input to this problem consists of a finite set of variables V, and a finite set of triples of the form (x, y, z) where $x, y, z \in V$. The task is to find an ordering $<$ on V such that for each of the given triples (x, y, z) we have either $x < y < z$ or $z < y < x$. It is well-known that this problem is NP-complete [**24, 43**], and that we therefore cannot expect to find a polynomial-time algorithm that solves it. In contrast, when we want to find an ordering $<$ on V such that for each of the given triples (x, y, z) we have $x < y$ or $x < z$, then the corresponding problem can be solved in polynomial time.

Many constraint satisfaction problems can be modeled formally as follows. Let Γ be a structure with a finite relational signature. Then the *constraint satisfaction problem for* Γ, denoted by $\mathrm{CSP}(\Gamma)$, is the problem of deciding whether a given primitive positive sentence ϕ is true in Γ. By choosing Γ appropriately, many problems in the above mentioned application areas can be expressed as $\mathrm{CSP}(\Gamma)$. The BETWEENNESS problem, for instance, can be modeled as $\mathrm{CSP}((\mathbb{Q}; \mathit{Betw}))$ where $\mathbb{Q}$ are the rational numbers and $\mathit{Betw} = \{(x, y, z) \in \mathbb{Q}^3 \mid x < y < z \vee z < y < x\}$.

Note that even though the structure Γ might be infinite, the problem $\mathrm{CSP}(\Gamma)$ is always a well-defined and *discrete* problem. Since the signature of Γ is finite, the complexity of $\mathrm{CSP}(\Gamma)$ is independent of the representation of the relation symbols of Γ in input instances of $\mathrm{CSP}(\Gamma)$. The task is to decide whether there *exists* an assignment to the variables of a given instance, and we do not have to exhibit such a solution. Therefore, the computational problems under consideration are finitistic and concrete even when the domain of Γ is, say, the real numbers.

There are many reasons to formulate a discrete problem as $\mathrm{CSP}(\Gamma)$ for an infinite structure Γ. The advantages of such a formulation are most striking when Γ can be chosen to be ω-categorical. In this case, the computational complexity of $\mathrm{CSP}(\Gamma)$ is fully captured by the *polymorphism clone* of Γ; the polymorphism clone can be seen as a higher-dimensional generalization of the automorphism group of Γ. When studying polymorphism clones, we can apply techniques from universal algebra, and, as we will see here, from Ramsey theory to obtain results about the computational complexity of $\mathrm{CSP}(\Gamma)$.

Contributions and Outline. In this article we give a survey presentation of a technique how to apply Ramsey theory when studying automorphism groups, endomorphism monoids, and polymorphism clones of countably infinite structures with a first-order definition in an ordered homogeneous Ramsey structure in a finite language – such structures are always ω-categorical. We present applications of this technique in two fields. Let Δ be a countable structure with a first-order definition in an ordered homogeneous Ramsey structure in a finite language. In model theory, our technique can be used to classify the set of all structures Γ that are first-order definable in Δ. In constraint satisfaction, it can be used to obtain a complete complexity classification for the class of all problems $\mathrm{CSP}(\Gamma)$ where Γ is first-order definable in Δ. We demonstrate this for $\Delta = (\mathbb{Q}; <)$, and for $\Delta = (V; E)$, the countably infinite random graph.

2. Reducts

One way to classify relational structures on a fixed domain is by identifying two structures when they *define* one another. The term "define" will classically stand for "first-order define", i.e., a structure Γ_1 has a first-order definition in a structure Γ_2 on the same domain iff all relations of Γ_1 can be defined by a first-order formula over Γ_2. When Γ_1 has a first-order definition in Γ_2 and vice-versa, then two structures are considered equivalent *up to first-order interdefinability*.

Depending on the application, other notions of definability might be suitable; such notions include syntactic restrictions of first-order definability. In this paper, besides first-order definability, we will consider the notions of *existential positive definability* and *primitive positive definability*; in particular, we will explain the importance of the latter notion in theoretical computer science in Section 8.

The structures which we consider in this article will all be countably infinite, and we will henceforth assume this property without further mentioning it. A structure is called *ω-categorical* if all countable models of its first-order theory are isomorphic. We are interested in the situation where all structures to be classified are *reducts* of a single countable ω-categorical structure in the following sense (which differs from the standard definition of a reduct and morally follows e.g. [**46**]).

Definition 1. Let Δ be a structure. A *reduct* of Δ is a structure with the same domain as Δ all of whose relations can be defined by a first-order formula in Δ.

When all structures under consideration are reducts of a countably infinite base structure Δ which is ω-categorical, then there are natural ways of obtaining classifications up to first-order, existential positive, or primitive positive interdefinability by means of certain sets of functions. In this section, we explain these ways, and give some examples of classifications that have been obtained in the past. In the following sections, we then observe that these results have actually been obtained in a more specific context than ω-categoricity, namely, where the structures are reducts of an *ordered Ramsey* structure Δ which has a finite relational signature and which is *homogeneous* in the sense that every isomorphism between finite induced substructures of Δ can be extended to an automorphism of Δ. We further develop a general framework for proving such results in this context.

We start with first-order definability. Consider the assignment that sends every structure Γ with domain D to its automorphism group $\mathrm{Aut}(\Gamma)$. Automorphism groups are closed sets in the convergence topology of all permutations on D, and conversely, every closed permutation group on D is the automorphism group of a relational structure with domain D. The closed permutation groups on D form a complete lattice, where the meet of a set of groups is given by their intersection. Similarly, the set of those relational structures on D which are *first-order closed*, i.e., which contain all relations which they define by a first-order formula, forms a lattice, where the meet of a set S of such structures is the structure which has those relations that are contained in all structures in S. Now when Γ is a countable ω-categorical structure, then it follows from the proof of the theorem of Ryll-Nardzewki (see [**28**]) that its automorphism group $\mathrm{Aut}(\Gamma)$ still has the first-order theory of Γ encoded in it. And indeed we can, up to first-order interdefinability, recover Γ from its automorphism group as follows: For a set $\mathcal{F}$ of finitary functions on D, let $\mathrm{Inv}(\mathcal{F})$ be the structure on D which has those relations R which are *invariant under* $\mathcal{F}$, i.e., those relations that contain $f(r_1, \dots, r_n)$ (calculated componentwise) whenever $f \in \mathcal{F}$ and $r_1, \dots, r_n \in R$.

THEOREM 2. *Let Δ be ω-categorical. Then the mapping $\Gamma \mapsto \mathrm{Aut}(\Gamma)$ is an antiisomorphism between the lattice of first-order closed reducts of Δ and the lattice of closed permutation groups containing $\mathrm{Aut}(\Delta)$. The inverse mapping is given by $\mathcal{G} \mapsto \mathrm{Inv}(\mathcal{G})$.*

This connection between closed permutation groups and first-order definability has been exploited several times in the past in order to obtain complete classifications of reducts of ω-categorical structures. For example, let Δ be the order of the rational numbers – we write $\Delta = (\mathbb{Q}; <)$. Then it has been shown in [**20**] that there are exactly five reducts of Δ, up to first-order interdefinability, which we will define in the following.

On the permutation side, let $\leftrightarrow$ be the function that sends every $x \in \mathbb{Q}$ to $-x$. For our purposes, we can equivalently choose $\leftrightarrow$ to be any permutation that inverts the order $<$ on $\mathbb{Q}$. For any fixed irrational real number α, let $\circlearrowright$ be any permutation on $\mathbb{Q}$ with the property that $x < y < \alpha < u < v$ implies $\circlearrowright(u) < \circlearrowright(v) < \circlearrowright(x) < \circlearrowright(y)$, for all $x, y, u, v \in \mathbb{Q}$. We will consider closed groups *generated* by these permutations: For a set of permutations $\mathcal{F}$ and a closed permutation group $\mathcal{G}$, we say that $\mathcal{F}$ generates $\mathcal{G}$ iff $\mathcal{G}$ is the smallest closed group containing $\mathcal{F}$.

On the relational side, for $x_1, \dots, x_n \in \mathbb{Q}$ write $\overrightarrow{x_1 \dots x_n}$ when $x_1 < \dots < x_n$. Then we define a ternary relation *Betw* on $\mathbb{Q}$ by $\mathit{Betw} := \{(x, y, z) \in \mathbb{Q}^3 \mid \overrightarrow{xyz} \vee \overrightarrow{zyx}\}$. Define another ternary relation *Cycl* by $\mathit{Cycl} := \{(x, y, z) \in \mathbb{Q}^3 \mid \overrightarrow{xyz} \vee \overrightarrow{yzx} \vee \overrightarrow{zxy}\}$. Finally, define a 4-ary relation *Sep* by

$$\begin{aligned}\{(x_1, y_1, x_2, y_2) \in \mathbb{Q}^4 \mid\ & \overrightarrow{x_1x_2y_1y_2} \vee \overrightarrow{x_1y_2y_1x_2} \vee \overrightarrow{y_1x_2x_1y_2} \vee \overrightarrow{y_1y_2x_1x_2} \\ & \vee \overrightarrow{x_2x_1y_2y_1} \vee \overrightarrow{x_2y_1y_2x_1} \vee \overrightarrow{y_2x_1x_2y_1} \vee \overrightarrow{y_2y_1x_2x_1}\}\,.\end{aligned}$$

THEOREM 3 (Cameron [**20**]). *Let Γ be a reduct of $(\mathbb{Q}; <)$. Then exactly one of the following holds:*

- *Γ is first-order interdefinable with $(\mathbb{Q}; <)$; equivalently, $\mathrm{Aut}(\Gamma) = \mathrm{Aut}((\mathbb{Q}; <))$.*
- *Γ is first-order interdefinable with $(\mathbb{Q}; \mathit{Betw})$; equivalently, $\mathrm{Aut}(\Gamma)$ equals the closed group generated by $\mathrm{Aut}((\mathbb{Q}; <))$ and $\leftrightarrow$.*
- *Γ is first-order interdefinable with $(\mathbb{Q}; \mathit{Cycl})$; equivalently, $\mathrm{Aut}(\Gamma)$ equals the closed group generated by $\mathrm{Aut}((\mathbb{Q}; <))$ and $\circlearrowright$.*
- *Γ is first-order interdefinable with $(\mathbb{Q}; \mathit{Sep})$; equivalently, $\mathrm{Aut}(\Gamma)$ equals the closed group generated by $\mathrm{Aut}((\mathbb{Q}; <))$ and $\{\leftrightarrow, \circlearrowright\}$.*
- *Γ is first-order interdefinable with $(\mathbb{Q}; =)$; equivalently, $\mathrm{Aut}(\Gamma)$ equals the group of all permutations on $\mathbb{Q}$.*

Another instance of the application of Theorem 2 in the classification of reducts up to first-order interdefinability has been provided by Thomas [**46**]. Let $G = (V; E)$ be the random graph, i.e., the up to isomorphism unique countably infinite graph which is homogeneous and which contains all finite graphs as induced subgraphs. It turns out that up to first-order interdefinability, G has precisely five reducts, too.

On the permutation side, observe that the graph $\bar{G}$ obtained by making two distinct vertices $x, y \in V$ adjacent iff they are not adjacent in G is isomorphic to G; let $-$ be any permutation on V witnessing this isomorphism. Moreover, for any fixed vertex $0 \in V$, the graph obtained by making all vertices which are adjacent with 0 non-adjacent with 0, and all vertices different from 0 and non-adjacent with 0 adjacent with 0, is isomorphic to G. Let sw be any permutation on V witnessing this fact.

On the relational side, define for all $k \geq 2$ a k-ary relation $R^{(k)}$ on V by

$$\begin{aligned} R^{(k)} := \{(x_1, \ldots, x_k) \mid\ & \text{all } x_i \text{ are distinct,} \\ & \text{and the number of edges on } \{x_1, \ldots, x_k\} \text{ is odd}\}. \end{aligned}$$

THEOREM 4 (Thomas [**46**]). *Let Γ be a reduct of the random graph $G = (V; E)$. Then exactly one of the following holds:*

- Γ *is first-order interdefinable with G; equivalently,* $\mathrm{Aut}(\Gamma) = \mathrm{Aut}(G)$.
- Γ *is first-order interdefinable with $(V; R^{(3)})$; equivalently,* $\mathrm{Aut}(\Gamma)$ *equals the closed group generated by* $\mathrm{Aut}(G)$ *and* sw.
- Γ *is first-order interdefinable with $(V; R^{(4)})$; equivalently,* $\mathrm{Aut}(\Gamma)$ *equals the closed group generated by* $\mathrm{Aut}(G)$ *and* $-$.
- Γ *is first-order interdefinable with $(V; R^{(5)})$; equivalently,* $\mathrm{Aut}(\Gamma)$ *equals the closed group generated by* $\mathrm{Aut}(G)$ *and* $\{\mathrm{sw}, -\}$.
- Γ *is first-order interdefinable with $(V; =)$; equivalently,* $\mathrm{Aut}(\Gamma)$ *equals the group of all permuations on V.*

In a similar fashion, the reducts of several prominent ω-categorical structures Δ have been classified up to first-order interdefinability by finding all closed supergroups of $\mathrm{Aut}(\Delta)$. Examples are:

- The countable homogeneous K_n-free graph, i.e., the unique countable homogeneous graph which contains precisely those finite graphs which do not contain a clique of size n as induced subgraphs, has 2 reducts up to first-order interdefinability (Thomas [**46**]), for all $n \geq 3$.
- The countable homogeneous k-hypergraph has $2^k + 1$ reducts up to first-order interdefinability (Thomas [**47**]), for all $k \geq 2$.
- The structure $(\mathbb{Q}; <, 0)$, i.e., the order of the rationals which in addition "knows" one of its points, has 116 reducts up to first-order interdefinability (Junker and Ziegler [**31**]).

All these examples have in common that the structures have a high degree of symmetry in the sense that they are homogeneous in a finite language – intuitively, one would expect the automorphism group of such a structure to be rather large. And indeed, Thomas conjectured in [**46**]:

CONJECTURE 5 (Thomas [**46**]). *Let Δ be a countable relational structure which is homogeneous in a finite language. Then Δ has finitely many reducts up to first-order interdefinability.*

It turns out that all the examples above are not only homogeneous in a finite language; in fact, they all have a first-order definition in (in other words: are themselves reducts of) an *ordered Ramsey structure* which is homogeneous in a finite language. Functions on such structures, in particular automorphisms of reducts, can be analyzed by the means of Ramsey theory, and we will outline a general method for classifying the reducts of such structures in Sections 3 to 6.

We now turn to analogs of Theorem 2 for syntactic restrictions of first-order logic. A first-order formula is called *existential* iff it is of the form $\exists x_1 \ldots \exists x_n.\ \phi$, where ϕ is quantifier-free. It is called *existential positive* iff it is existential and does not contain any negations. Now observe that similarly to permutation groups, the *endomorphism monoid* $\mathrm{End}(\Delta)$ of a relational structure Δ with domain D is always closed in the pointwise convergence topology on the space of all functions from D to D, and that every closed transformation monoid $\mathcal{M}$ acting on D is the endomorphism monoid of the structure $\mathrm{Inv}(\mathcal{M})$, i.e., the structure with domain D which contains those relations which are invariant under all functions in $\mathcal{M}$. Note also that the set of closed transformation monoids on D, ordered by inclusion, forms a complete lattice, and that likewise the set of all existential positive closed structures forms a complete lattice. The analog to Theorem 2 for existential positive definability is an easy consequence of the homomorphism preservation theorem (see [**28**]) and goes like this:

THEOREM 6. *Let Δ be ω-categorical. Then the mapping $\Gamma \mapsto \mathrm{End}(\Gamma)$ is an antiisomorphism between the lattice of existential positive closed reducts of Δ and the lattice of closed transformation monoids containing $\mathrm{Aut}(\Delta)$. The inverse mapping is given by $\mathcal{M} \mapsto \mathrm{Inv}(\mathcal{M})$.*

All the closed monoids containing the group of all permutations on a countably infinite set D (which equals the automorphism group of the empty structure $(D; =)$) have been determined in [**8**], and their number is countably infinite. Therefore, every structure has infinitely many reducts up to existential positive interdefinability. In general, it will be impossible to determine all of them, but sometimes it is already useful to determine certain closed monoids, as in the following theorem about endomorphism monoids of reducts of the random graph from [**14**]. We need the following definitions. Since the random graph $G = (V; E)$ contains all countable graphs, it contains an infinite clique. Let e_E be any injective function from V to V whose image induces such a clique in G. Similarly, let e_N be any injection from V to V whose image induces an independent set in G.

THEOREM 7 (Bodirsky and Pinsker [**14**]). *Let Γ be a reduct of the random graph $G = (V; E)$. Then at least one of the following holds.*

- $\mathrm{End}(\Gamma)$ *contains a constant operation.*
- $\mathrm{End}(\Gamma)$ *contains e_E.*
- $\mathrm{End}(\Gamma)$ *contains e_N.*
- $\mathrm{Aut}(\Gamma)$ *is a dense subset of* $\mathrm{End}(\Gamma)$ *(equipped with the topology of pointwise convergence).*

Theorem 7 states that for reducts Γ of the random graph, either $\mathrm{End}(\Gamma)$ contains a function that destroys all structure of the random graph, or it contains basically no functions except the automorphisms. This has the following non-trivial consequence. A theory T is called *model-complete* iff every embedding between models of T is elementary, i.e., preserves all first-order formulas. A structure is said to be model-complete iff its first-order theory is model-complete.

COROLLARY 8 (Bodirsky and Pinsker [**14**]). *All reducts of the random graph are model-complete.*

PROOF. It is not hard to see (cf. [**14**]) that an ω-categorical structure Γ is model-complete if and only if $\mathrm{Aut}(\Gamma)$ is dense in the monoid of self-embeddings of Γ. Now let Γ be a reduct of G, and let $\mathcal{M}$ be the closed monoid of self-embeddings of Γ; we will show that $\mathrm{Aut}(\Gamma)$ is dense in $\mathcal{M}$. We apply Theorem 7 to $\mathcal{M}$ (which, as a closed monoid containing $\mathrm{Aut}(G)$, is also an endomorphism monoid of a reduct Γ' of G). Clearly, Γ' and Γ have the same automorphisms, namely those permutations in $\mathcal{M}$ whose inverse is also in $\mathcal{M}$. Therefore we are done if the last case of the theorem holds. Note that $\mathcal{M}$ cannot contain a constant operation as all its operations are injective. So suppose that $\mathcal{M}$ contains e_N – the argument for e_E is analogous. Let R be any relation of Γ, and ϕ_R be its defining quantifier-free formula; ϕ_R exists since G has quantifier-elimination, i.e., every first-order formula over G is equivalent to a quantifier-free formula. Let ψ_R be the formula obtained by replacing all occurrences of E by *false*; so ψ_R is a formula over the empty language. Then a tuple a satisfies ϕ_R in G iff $e_N(a)$ satisfies ϕ_R in G (because e_N is an embedding) iff $e_N(a)$ satisfies ψ_R in G (as there are no edges on $e_N(a)$) iff $e_N(a)$ satisfies ψ_R in the substructure induced by $e_N[V]$ (since ψ_R does not contain any quantifiers). Thus, Γ is isomorphic to the structure on $e_N[V]$ which has the relations defined by the formulas ψ_R; hence, Γ is isomorphic to a structure with a first-order definition over the empty language. This structure has, of course, all injections as self-embeddings, and all permutations as automorphisms, and hence is model-complete; thus, the same is true for Γ. □

It follows from [**11**, Proposition 19] that all reducts of the linear order of the rationals $(\mathbb{Q};<)$ are model-complete as well. This is remarkable, since similar structures do not have this property – for example, $(\mathbb{Q};<,0)$ is first-order interdefinable with the structure $(\mathbb{Q};<,[0,\infty))$ which is not model-complete.

We now turn to an even finer way of distinguishing reducts of an ω-categorical structure, namely up to *primitive positive interdefinability.* This is of importance in connection with the constraint satisfaction problem from the introduction, as we will describe in more detail in Section 8. We call a formula *primitive positive* iff it is existential positive and does not contain disjunctions. A *clone* on domain D is a set of finitary operations on D which contains all projections (i.e., functions of the form $(x_1,\dots,x_n)\mapsto x_i$) and which is closed under composition. A clone $\mathcal{C}$ is *closed* (also called *locally closed* or *local* in the literature) iff for each $n\geq 1$, the set of n-ary functions in $\mathcal{C}$ is a closed subset of the space D^{D^n}, where D is taken to be discrete. The closed clones on D form a complete lattice with respect to inclusion – the structure of this lattice has been studied in the universal algebra literature (see [**26**], [**44**]). Similarly, the set of relational structures with domain D which are *primitive positive closed*, i.e., which contain all relations which they define by primitive positive formulas, forms a complete lattice. For a structure Γ, we define $\mathrm{Pol}(\Gamma)$ to consist of all finitary operations on the domain of Γ which preserve all relations of Γ, i.e., an n-ary function f is an element of $\mathrm{Pol}(\Gamma)$ iff for all relations R of Γ and all tuples $r_1,\dots,r_n\in R$ the tuple $f(r_1,\dots,r_n)$ is an element of R. It is easy to see that $\mathrm{Pol}(\Gamma)$ is always a closed clone. Observe also that $\mathrm{Pol}(\Gamma)$ is a generalization of $\mathrm{End}(\Gamma)$ to higher (finite) arities.

Theorem 9 (Bodirsky and Nešetřil [**12**]). *Let Δ be ω-categorical. Then the mapping $\Gamma \mapsto \mathrm{Pol}(\Gamma)$ is an antiisomorphism between the lattice of primitive positive closed reducts of Δ and the lattice of closed clones containing $\mathrm{Aut}(\Delta)$. The inverse mapping is given by $\mathcal{C} \mapsto \mathrm{Inv}(\mathcal{C})$.*

It turns out that even for the empty structure $(X;=)$, the lattice of primitive positive closed reducts is probably too complicated to be completely described – the lattice has been thoroughly investigated in [**8**].

Theorem 10 (Bodirsky, Chen, Pinsker [**8**]). *The structure $(X;=)$ (where X is countably infinite), and therefore all countably infinite structures, have $2^{\aleph_0}$ reducts up to primitive positive interdefinability.*

Fortunately, it is sometimes sufficient in applications to understand only parts of this lattice. We will see examples of this in Section 8.

3. Ramsey Classes

While Theorems 2, 6 and 9 provide a theoretical method for determining reducts of an ω-categorical structure Δ by transforming them into sets of functions on Δ, understanding these infinite objects could turn out difficult without further tools for handling them. We will now focus on structures which have the additional property that they are reducts of an *ordered Ramsey structure* that is homogeneous in a finite relational signature; such structures are ω-categorical since homogeneous structures in a finite language are ω-categorical and since reducts of ω-categorical structures are ω-categorical. This is less restrictive than it might appear at first sight: we remark that it could be the case that all homogeneous structures with a finite relational signature are reducts of ordered homogeneous Ramsey structures with a finite relational signature (that is, we do not know of a counterexample). It turns out that in this context, certain infinite functions can be represented by finite ones, making classification projects more feasible.

Definition 11. A structure is called *ordered* iff it has a total order among its relations.

Definition 12. Let τ a relational signature. For τ-structures $\mathcal{S}, \mathcal{H}, \mathcal{P}$ and an integer $k \geq 1$, we write $\mathcal{S} \to (\mathcal{H})^{\mathcal{P}}_k$ iff for every k-coloring χ of the copies of $\mathcal{P}$ in $\mathcal{S}$ there exists a copy $\mathcal{H}'$ of $\mathcal{H}$ in $\mathcal{S}$ such that all copies of $\mathcal{P}$ in $\mathcal{H}'$ have the same color under χ.

Definition 13. A class $\mathcal{C}$ of finite τ-structures which is closed under isomorphisms, induced substructures, and with the joint embedding property (see [**28**]) is called a *Ramsey class* iff it is closed under substructures and for all $k \geq 1$ and all $\mathcal{H}, \mathcal{P} \in \mathcal{C}$ there exists $\mathcal{S}$ in $\mathcal{C}$ such that $\mathcal{S} \to (\mathcal{H})^{\mathcal{P}}_k$.

Definition 14. A relational structure is called *Ramsey* iff its *age*, i.e., the set of finite structures isomorphic to a finite induced substructure, is a Ramsey class.

Examples of Ramsey structures are the dense linear order $(\mathbb{Q};<)$ and the *ordered random graph* $(V;E,<)$, i.e., the Fraïssé limit of the class of finite ordered graphs. We remark that the random graph itself is not Ramsey, but since it is a reduct of the ordered random graph, the methods we are about to expose apply as well.

We will now see that one can find regular patterns in the behavior of any function acting on an ordered Ramsey structure which is ω-categorical.

Definition 15. Let Γ be a structure. The *type* $\mathrm{tp}(a)$ of an n-tuple $a \in \Gamma$ is the set of first-order formulas with free variables $x_1, \ldots, x_n$ that hold for a in Γ.

We recall the classical theorem of Ryll-Nardzewski about the number of types in ω-categorical structures.

Theorem 16 (Ryll-Nardzewski). *The following are equivalent for a countable structure* Γ *in a countable language.*

- Γ *is* ω*-categorical, i.e., any countable model of the theory of* Γ *is isomorphic to* Γ*.*
- Γ *has for all* $n \geq 1$ *only finitely many different types of* n*-tuples.*

We also mention that moreover, as a well-known consequence of the proof of this theorem, two tuples in a countable ω-categorical structure have the same type if and only if there is an automorphism of Γ which sends one tuple to the other.

Definition 17. A *type condition* between two structures Γ_1, Γ_2 is a pair (t_1, t_2), where each t_i is a type of an n-tuple in Γ_i. A function $f : \Gamma_1 \to \Gamma_2$ *satisfies* a type condition (t_1, t_2) if for all n-tuples $(a_1, \ldots, a_n)$ of type t_1, the n-tuple $(f(a_1), \ldots, f(a_n))$ is of type t_2.

A *behavior* is a set of type conditions between two structures. A function *has behavior* B if it satisfies all the type conditions of the behavior B. A behavior B is called *complete* iff for all types t_1 of tuples in Γ_1 there is a type t_2 of a tuple in Γ_2 such that $(t_1, t_2) \in B$.

A function $f : \Gamma_1 \to \Gamma_2$ is *canonical* iff it has a complete behavior. If $F \subseteq \Gamma_1$, then we say that f is *canonical on* F if its restriction to F is canonical.

Observe that the function $\leftrightarrow$ of Theorem 3 is canonical for the structure $(\mathbb{Q}; <)$. The function $\circlearrowright$ is not, but it is canonical on each of the intervals $(-\infty, \alpha)$ and (α, ∞). For the random graph, the function $-$ of Theorem 4 is canonical, while sw is canonical on $V \setminus \{0\}$. Also, sw is canonical as a function from $(V; E, 0)$ to $(V; E)$, where $(V; E, 0)$ denotes the structure obtained from $(V; E)$ by adding a new constant symbol for the element 0 by which we defined the function sw. Moreover, the constant function and e_E, e_N of Theorem 7 are canonical on $(V; E)$. We will now show that it is no coincidence that canonical functions are that ubiquitous.

Definition 18. Let Δ be a structure. A property P *holds for arbitrarily large finite substructures of* Δ iff for all finite substructures $F \subseteq \Delta$ there is a copy of F in Δ for which P holds.

The following observation is just an easy application of the definition of a Ramsey class, but crucial in understanding functions on ordered Ramsey structures.

Lemma 19. *Let* Δ *be ordered Ramsey and* ω*-categorical, and let* $f : \Delta \to \Delta$*. Then* f *is canonical on arbitrarily large finite substructures.*

The proof goes along the following lines: Let F be any finite substructure of Δ. Then the function f induces a mapping from the tuples in Δ to the set of types in Δ (each tuple is sent to the type of its image under f). If we restrict this mapping to tuples of length at most the size of F, then since Δ is ω-categorical, the range of this restriction is finite by Theorem 16, and thus is a k-coloring of tuples for some finite k. Now apply the Ramsey property once for every type of tuple that occurs in F – see [**16**] for details. We remark that this lemma would be false if one dropped the order assumption, which implies that coloring induced substructures and coloring tuples in Δ are one and the same thing.

The motivation for working with ordered Ramsey structures is the rough idea that all "important" functions can be assumed to be canonical. While this is simply false when stated boldly like this, it is still true for some functions when the idea is further refined, as we will show in the following. Observe that if Δ is ω-categorical, then for each $n \geq 1$ there are only finitely many possible type conditions for n-types over Δ (Theorem 16). Suppose that Δ has in addition a finite language and *quantifier elimination*, i.e., every first-order formula in the language of Δ is equivalent to a quantifier-free formula over Δ; this follows in particular from homogeneity in a finite language. Then, if $n(\Delta)$ is the largest arity of its relations, then a function $f : \Delta \to \Delta$ is canonical iff for every type t_1 of an $n(\Delta)$-tuple in Δ there is a type t_2 in Δ such that f satisfies the type condition (t_1, t_2). In other words, the complete behavior of f is already determined by its behavior on $n(\Delta)$-types. Hence, a canonical function on Δ is essentially a function on the $n(\Delta)$-types of Δ – a finite object.

Definition 20. Let $f, g : \Delta \to \Delta$. We say that f *generates* g over Δ iff g is contained in the smallest closed monoid containing f and $\mathrm{Aut}(\Delta)$. Equivalently, for every finite subset F of Δ, there exists a term $\beta \circ f \circ \alpha_1 \circ f \circ \alpha_2 \circ \cdots \circ f \circ \alpha_n$, where $\beta, \alpha_i \in \mathrm{Aut}(\Delta)$, which agrees with g on F.

Proposition 21. *Let Δ be a structure in a finite language which is ordered, Ramsey, and homogeneous. Let $f : \Delta \to \Delta$. Then f generates a canonical function $g : \Delta \to \Delta$.*

First proof. Let $(F_i)_{i \in \omega}$ be an increasing sequence of finite substructures of Δ such that $\bigcup_{i \in \omega} F_i = \Delta$. By Lemma 19, for each $i \in \omega$ we find a copy F'_i of F_i in Δ on which f is canonical. Since there are only finitely many possibilities of canonical behavior, one behavior occurs an infinite number of times; thus, by thinning out the sequence, we may assume that the behavior is the same on all F'_i. By the homogeneity of Δ, there exist automorphisms α_i of Δ sending F_i to F'_i, for all $i \in \omega$. Also, since the behavior on all the F'_i is the same, we can inductively pick automorphisms β_i of Δ such that $\beta_{i+1} \circ f \circ \alpha_{i+1}$ agrees with $\beta_i \circ f \circ \alpha_i$ on F_i, for all $i \in \omega$. The union over the functions $\beta_i \circ f \circ \alpha_i : F_i \to \Delta$ is a canonical function on Δ. □

Second proof. The identity function $\mathrm{id} : \Delta \to \Delta$ is generated by f and is canonical. □

The problem with the preceding lemma is the second proof, which makes it trivial. What we really want is that f generates a canonical function g which represents f in a certain sense – it should be possible to retain specific properties of f when passing to the canonical functions. For example, we could wish that if f violates a certain relation, then so does g; or, if f is not an automorphism of Δ, we will look for a canonical function g which is not an automorphism of Δ either.

We are now going to refine our method, and fix constants $c_1, \ldots, c_n$ such that $f \notin \mathrm{Aut}(\Delta)$ is witnessed on $\{c_1, \ldots, c_n\}$. We then consider f as a function from

$(\Delta, c_1, \ldots, c_n)$ to Δ, where $(\Delta, c_1, \ldots, c_n)$ denotes the expansion of Δ by the constants $c_1, \ldots, c_n$. It turns out that f is canonical on arbitrarily large substructures of $(\Delta, c_1, \ldots, c_n)$, and that it generates a canonical function $g : (\Delta, c_1, \ldots, c_n) \to \Delta$ which agrees with f on $c_1, \ldots, c_n$; in particular, g is not an automorphism of Δ, and the problem of triviality in Proposition 20 no longer occurs. In order to do this, we must assure that $(\Delta, c_1, \ldots, c_n)$ still has the Ramsey property. This leads us into topological dynamics.

4. Topological Dynamics

We have seen in the previous section that our approach crucially relies on the fact that when an ordered homogeneous Ramsey structure is expanded by finitely many constants, the expansion is again Ramsey (it is clear that the expansion is again ordered and homogeneous). To prove this, we use a characterization in topological dynamics of those ordered homogeneous structures which are Ramsey.

Recall that a *topological group* is an (abstract) group G together with a topology on the elements of G such that $(x, y) \mapsto xy^{-1}$ is continuous from G^2 to G. In other words, we require that the binary group operation and the inverse function are continuous.

DEFINITION 22. A topological group is *extremely amenable* iff any continuous action of the group on a compact Hausdorff space has a fixed point.

Kechris, Pestov and Todorcevic have characterized the Ramsey property of the age of an ordered homogeneous structure by means of extreme amenability in the following theorem.

THEOREM 23 (Kechris, Pestov, Todorcevic [**32**]). *Let Δ be an ordered homogeneous relational structure. Then the age of Δ has the Ramsey property iff* $\mathrm{Aut}(\Delta)$ *is extremely amenable.*

This theorem can be applied to provide a short and elegant proof of the following.

PROPOSITION 24 (Bodirsky, Pinsker and Tsankov [**16**]). *Let Δ be ordered, Ramsey, and homogeneous, and let $c_1, \ldots, c_n \in \Delta$. Then $(\Delta, c_1, \ldots, c_n)$ is Ramsey as well.*

When Δ is ordered, Ramsey, and homogeneous, then $\mathrm{Aut}(\Delta)$ is extremely amenable. Note that the automorphism group of $(\Delta, c_1, \ldots, c_n)$ is an open subgroup of $\mathrm{Aut}(\Delta)$. The proposition thus follows directly from the following fact – confer [**16**].

LEMMA 25. *Let G be an extremely amenable group, and let H be an open subgroup of G. Then H is extremely amenable.*

5. Minimal Functions

The results of the preceding section provide a tool for "climbing up" the lattice of closed monoids containing the automorphism group of an ordered Ramsey structure which is homogeneous and has a finite language.

DEFINITION 26. Let $\mathcal{C}, \mathcal{D}$ be closed clones. Then $\mathcal{D}$ is called *minimal above* $\mathcal{C}$ iff $\mathcal{D} \supseteq \mathcal{C}$ and there are no closed clones between $\mathcal{C}$ and $\mathcal{D}$.

Observe that transformation monoids can be identified with those clones which have the property that all their functions depend on only one variable. Hence, Definition 26 also provides us with a notion of a minimal closed monoid above another closed monoid.

It follows from Theorem 9 and Zorn's Lemma that if Δ is an ω-categorical structure in a finite language, then every closed clone containing $\mathrm{Pol}(\Delta)$ contains a minimal closed clone above $\mathrm{Pol}(\Delta)$. Similarly, as a consequence of Theorem 6, every closed monoid containing $\mathrm{End}(\Delta)$ contains a minimal closed monoid.

For closed permutation groups, minimality can be defined analogously. Then Theorem 2 implies that for ω-categorical structures Δ in a finite language, every closed permutation group containing $\mathrm{Aut}(\Delta)$ contains a minimal closed permutation group above $\mathrm{Aut}(\Delta)$.

Clearly, if a closed clone $\mathcal{D}$ is minimal above $\mathcal{C}$, then any function $f \in \mathcal{D} \setminus \mathcal{C}$ generates $\mathcal{D}$ with $\mathcal{C}$ (i.e., $\mathcal{D}$ is the smallest closed clone containing f and $\mathcal{C}$) – similar statements hold for monoids and groups. In the case of clones and monoids and in the setting of reducts of ordered Ramsey structures which are homogeneous in a finite language, we can standardize such generating functions. This is the contents of the coming subsections.

5.1. Minimal unary functions. Adapting the proof of Lemma 21, with the use of the Proposition 24, one can show the following.

LEMMA 27. *Let Δ be ordered, Ramsey, homogeneous, and of finite language. Let $f : \Delta \to \Delta$, and let $c_1, \ldots, c_n \in \Delta$. Then f together with* $\mathrm{Aut}(\Delta)$ *generates a function which agrees with f on $\{c_1, \ldots, c_n\}$ and which is canonical as a function from $(\Delta, c_1, \ldots, c_n)$ to Δ.*

Let Γ be a finite language reduct of a structure Δ which is ordered, Ramsey, homogeneous, and of finite language, and let $\mathcal{N}$ be a minimal closed monoid containing $\mathrm{End}(\Gamma)$. Then, setting $n(\Gamma)$ to be the largest arity of the relations of Γ, we can pick constants $c_1, \ldots, c_{n(\Gamma)} \in \Gamma$ and a function $f \in \mathcal{N} \setminus \mathrm{End}(\Gamma)$ such that $f \notin \mathrm{End}(\Gamma)$ is witnessed on $\{c_1, \ldots, c_{n(\Gamma)}\}$. By the preceding lemma, f and $\mathrm{Aut}(\Delta)$ generate a function g which behaves like f on $\{c_1, \ldots, c_{n(\Gamma)}\}$ and which is canonical as a function from $(\Delta, c_1, \ldots, c_{n(\Gamma)})$ to Δ. This function g, together with $\mathrm{End}(\Gamma)$, generates $\mathcal{N}$. Since there are only finitely many choices for the type of the tuple $(c_1, \ldots, c_{n(\Gamma)})$ and for each choice only finitely many behaviors of functions from $(\Delta, c_1, \ldots, c_{n(\Gamma)})$ to Δ, we get the following.

PROPOSITION 28 (Bodirsky, Pinsker, Tsankov [**16**]). *Let Γ be a finite language reduct of a structure Δ which is ordered, Ramsey, homogeneous, and of finite language. Then the number of minimal closed monoids above* $\mathrm{End}(\Gamma)$ *is finite, and each such monoid is generated by* $\mathrm{End}(\Gamma)$ *plus a canonical function $g : (\Delta, c_1, \ldots, c_{n(\Gamma)}) \to \Delta$, for constants $c_1, \ldots, c_{n(\Gamma)} \in \Gamma$.*

Since for every relation R of Γ we can add its negation to the language, we get the following

COROLLARY 29. *Let $\mathcal{M}$ be the monoid of self-embeddings of a finite-language structure Γ which is a reduct of a structure Δ which is ordered, Ramsey, homogeneous, and of finite language. Then the number of minimal closed monoids above $\mathcal{M}$ is finite, and each such monoid is generated by $\mathcal{M}$ and a canonical function $g : (\Delta, c_1, \ldots, c_{n(\Gamma)}) \to \Delta$.*

The following is an example for the random graph $G = (V; E)$. Since G is model-complete, its monoid of self-embeddings is just the topological closure $\langle \mathrm{Aut}(G) \rangle$ of $\mathrm{Aut}(G)$ in the space V^V. Therefore, the minimal closed monoids above the monoid of self-embeddings of G are just the minimal closed monoids above $\langle \mathrm{Aut}(G) \rangle$.

THEOREM 30 (Thomas [**47**]). *Let $G = (V; E)$ be the random graph. The minimal closed monoids containing $\langle \mathrm{Aut}(G) \rangle$ are the following:*

- *The monoid generated by a constant operation with* $\mathrm{Aut}(G)$.
- *The monoid generated by* e_E *with* $\mathrm{Aut}(G)$.
- *The monoid generated by* e_N *with* $\mathrm{Aut}(G)$.
- *The monoid generated by* $-$ *with* $\mathrm{Aut}(G)$.
- *The monoid generated by* sw *with* $\mathrm{Aut}(G)$.

5.2. Minimal higher arity functions. We now generalize the concepts from unary functions and monoids to higher arity functions and clones.

DEFINITION 31. Let Δ be a structure. For $1 \le i \le m$ and a tuple x in the power Δ^m, we write x_i for the i-th coordinate of x. The *type* of a sequence of tuples $a^1, \ldots, a^n \in \Delta^m$, denoted by $\mathrm{tp}(a^1, \ldots, a^n)$, is the cartesian product of the types of $(a_i^1, \ldots, a_i^n)$ in Δ.

With this definition, the notions of *type condition*, *behavior*, *complete behavior*, and *canonical* generalize in complete analogy from functions $f : \Gamma_1 \to \Gamma_2$ to functions $f : \Gamma_1^m \to \Gamma_2$, for structures Γ_1, Γ_2. It can be shown that for ordered structures, the Ramsey property is not lost when going to products; an example of a proof can be found in [**16**].

PROPOSITION 32. *Let Δ be ordered and Ramsey, and let $m \ge 1$. Let moreover a number $k \ge 1$, an n-tuple $(a^1, \ldots, a^n) \in \Delta^m$, and finite $F_i \subseteq \Delta$ be given for $1 \le i \le m$. Then there exist finite $S_i \subseteq \Delta$ with the property that whenever the n-tuples in $S_1 \times \cdots \times S_m$ of type $\mathrm{tp}(a^1, \ldots, a^n)$ are colored with k colors, then there are copies F_i' of F_i in S_i such that the coloring is constant on $F_1' \times \cdots \times F_m'$.*

We remark that Proposition 32 does not hold in general if Δ is not assumed to be ordered – an example for the random graph can be found in [**14**]. Similarly to the unary case (Proposition 28), one gets the following.

PROPOSITION 33 (Bodirsky, Pinsker, Tsankov [**16**]). *Let Γ be a finite language reduct of a structure Δ which is ordered, Ramsey, homogeneous and of finite language. Then every minimal closed clone above $\mathrm{Pol}(\Gamma)$ is generated by $\mathrm{Pol}(\Gamma)$ and a canonical function $g : (\Delta, c_1, \ldots, c_k)^m \to \Delta$, where $m \ge 1$, $k \ge 0$, and $c_1, \ldots, c_k \in \Delta$. Moreover, m only depends on the number of $n(\Gamma)$-types in Γ (and not on the clone), and k only depends on m and $n(\Gamma)$, and the number of minimal closed clones above $\mathrm{Pol}(\Gamma)$ is finite.*

In the case of minimal closed clones above an endomorphism monoid, the arity of the generating canonical functions can be further reduced as follows.

PROPOSITION 34 (Bodirsky, Pinsker, Tsankov [**16**]). *Let Γ be a finite language reduct of a structure Δ which is ordered, Ramsey, homogeneous and of finite language. Then every minimal closed clone above $\mathrm{End}(\Gamma)$ is generated by $\mathrm{End}(\Gamma)$ and a canonical function $g : (\Delta, c_1, \ldots, c_{n(\Gamma)}) \to \Delta$, or by $\mathrm{End}(\Gamma)$ and a canonical function $g : (\Delta, c_1, \ldots, c_m)^m \to \Delta$, where m only depends on the number of 2-types in*

Γ *(and not on the clone). In particular, the number of minimal closed clones above* $\mathrm{End}(\Gamma)$ *is finite.*

Using this technique, the minimal closed clones containing the automorphism group of the random graph $G = (V; E)$ have been determined. In the following, let $f : V^2 \to V$ be a binary operation; we now define some possible behaviors for f. We say that f is

- *of type* p_1 iff for all $x_1, x_2, y_1, y_2 \in V$ with $x_1 \neq x_2$ and $y_1 \neq y_2$ we have $E(f(x_1, y_1), f(x_2, y_2))$ if and only if $E(x_1, x_2)$;
- *of type* max iff for all $x_1, x_2, y_1, y_2 \in V$ with $x_1 \neq x_2$ and $y_1 \neq y_2$ we have $E(f(x_1, y_1), f(x_2, y_2))$ if and only if $E(x_1, x_2)$ or $E(y_1, y_2)$;
- *balanced in the first argument* iff for all $x_1, x_2, y \in V$ with $x_1 \neq x_2$ we have $E(f(x_1, y), f(x_2, y))$ if and only if $E(x_1, x_2)$;
- *balanced in the second argument* iff $(x, y) \mapsto f(y, x)$ is balanced in the first argument;
- *E-dominated in the first argument* iff for all $x_1, x_2, y \in V$ with $x_1 \neq x_2$ we have that $E(f(x_1, y), f(x_2, y))$;
- *E-dominated in the second argument* iff $(x, y) \mapsto f(y, x)$ is E-dominated in the first argument.

The *dual* of an operation $f(x_1, \ldots, x_n)$ on V is defined by $-f(-x_1, \ldots, -x_n)$.

THEOREM 35 (Bodirsky and Pinsker [**14**]). *Let $G = (V; E)$ be the random graph, and let $\mathcal{C}$ be a minimal closed clone above $\langle \mathrm{Aut}(G) \rangle$. Then $\mathcal{C}$ is generated by* $\mathrm{Aut}(G)$ *together with one of the unary functions of Theorem* 30, *or by* $\mathrm{Aut}(G)$ *and one of the following canonical operations from G^2 to G:*

- *a binary injection of type p_1 that is balanced in both arguments;*
- *a binary injection of type* max *that is balanced in both arguments;*
- *a binary injection of type* max *that is E-dominated in both arguments;*
- *a binary injection of type p_1 that is E-dominated in both arguments;*
- *a binary injection of type p_1 that is balanced in the first and E-dominated in the second argument;*
- *the dual of one of the last four operations.*

In [**15**], the technique of canonical functions was applied again to climb up further in the lattice of closed clones above $\mathrm{Aut}(G)$ – we will come back to this in Section 8.

Another example are the minimal closed clones containing all permutations of a countably infinite base set X. Observe that the set $\mathcal{S}_X$ of all permutations on X is the automorphism group of the structure $(X; =)$ which has no relations.

THEOREM 36 (Bodirsky and Kára [**10**]; cf. also [**8**]). *The minimal closed clones containing $\langle \mathcal{S}_X \rangle$ on a countably infinite set X are:*

- *The closed clone generated by $\mathcal{S}_X$ and any constant operation;*
- *The closed clone generated by $\mathcal{S}_X$ and any binary injection.*

Observe that any constant operation and any binary injection on X are canonical operations for the structure $(X; =)$.

We end this section with a last example which lists the minimal closed clones containing the self-embeddings of the dense linear order $(\mathbb{Q}; <)$. As with the random graph and the empty structure, since $(\mathbb{Q}; <)$ is model-complete it follows that the monoid of self-embeddings of $(\mathbb{Q};<)$ is just the closure of $\mathrm{Aut}((\mathbb{Q};<))$ in $\mathbb{Q}^{\mathbb{Q}}$.

Let lex be a binary operation on $\mathbb{Q}$ such that $\mathrm{lex}(a,b) < \mathrm{lex}(a',b')$ iff either $a < a'$ or $a = a'$ and $b < b'$, for all $a, a', b, b' \in \mathbb{Q}$. Observe that lex is canonical as a function from $\mathbb{Q}^2$ to $\mathbb{Q}$. Next, let pp be an arbitrary binary operation on $\mathbb{Q}$ such that for all $a, a', b, b' \in \mathbb{Q}$ we have $\mathrm{pp}(a,b) \leq \mathrm{pp}(a',b')$ iff one of the following cases applies:

- $a \leq 0$ and $a \leq a'$;
- $0 < a$, $0 < a'$, and $b \leq b'$.

The name of the operation pp stands for "projection-projection", since the operation behaves as a projection to the first argument for negative first argument, and a projection to the second argument for positive first argument. Observe that pp is canonical if we add the origin as a constant to the language. Finally, define the *dual* of an operation $f(x_1, \ldots, x_n)$ on $\mathbb{Q}$ by $\leftrightarrow(f(\leftrightarrow(x_1), \ldots, \leftrightarrow(x_n)))$.

THEOREM 37 (Bodirsky and Kára [**11**]). *Let $(\mathbb{Q};<)$ be the order of the rationals, and let $\mathcal{C}$ be a minimal closed clone above $\langle \mathrm{Aut}((\mathbb{Q};<)) \rangle$. Then $\mathcal{C}$ is generated by $\mathrm{Aut}((\mathbb{Q};<))$ together with one of the following operations:*

- *a constant operation;*
- *the operation $\leftrightarrow$;*
- *the operation $\circlearrowright$;*
- *the operation* lex*;*
- *the operation* pp*;*
- *the dual of* pp.

6. Decidability of Definability

We turn to another application of the ideas of the last sections. Consider the following computational problem for a structure Γ: Input are quantifier-free formulas $\phi_0, \ldots, \phi_n$ in the language of Γ defining relations $R_0, \ldots, R_n$ on the domain of Γ, and the question is whether R_0 can be *defined* from $R_1, \ldots, R_n$. As in Section 2, "defined" can stand for "first-order defined" or syntactic restrictions of this notion. We denote this computational problem by $\mathrm{Expr}_{ep}(\Gamma)$ and $\mathrm{Expr}_{pp}(\Gamma)$ if we consider existential positive and primitive positive definability, respectively.

For *finite* structures Γ the problem $\mathrm{Expr}_{pp}(\Gamma)$ is in co-NEXPTIME (and in particular decidable), and has recently shown to be co-NEXPTIME-hard [**49**]. For infinite structures Γ, the decidability of $\mathrm{Expr}_{pp}(\Gamma)$ is not obvious. An algorithm for primitive positive definability has theoretical and practical consequences in the study of the computational complexity of CPSs (which we will consider in Section 8). It is motivated by the fundamental fact that expansions of structures Γ by primitive positive relations do not change the complexity of $\mathrm{CSP}(\Gamma)$. On a practical side, it turns out that hardness of a CSP can usually be shown by presenting primitive positive definitions of relations for which it is known that the CSP is hard. Therefore, a procedure that decides primitive positive definability of a given relation might be a useful tool to determine the computational complexity of CSPs.

Using the methods of the last sections, one can show decidability of $\mathrm{Expr}_{ep}(\Gamma)$ and $\mathrm{Expr}_{pp}(\Gamma)$ for certain infinite structures Γ. The following uses the same terminology as in [**35**].

DEFINITION 38. We say that a class $\mathcal{C}$ of finite τ-structures (or a τ-structure with age $\mathcal{C}$) is *finitely bounded* if there exists a finite set of finite τ-structures $\mathcal{F}$ such for all finite τ-structures A we have that $A \in \mathcal{C}$ iff no structure from $\mathcal{F}$ embeds into A.

THEOREM 39 (Bodirsky, Pinsker, Tsankov [**16**]). *Let Δ be ordered, Ramsey, homogeneous, and of finite language, and let Γ be a finite language reduct of Δ. Then $\mathrm{Expr}_{ep}(\Gamma)$ and $\mathrm{Expr}_{pp}(\Gamma)$ are decidable.*

Examples of structures Δ that satisfy the assumptions of Theorem 39 are $(\mathbb{Q}; <)$, the Fraïssé limit of ordered finite graphs (or tournaments [**39**]), the Fraïssé limit of finite partial orders with a linear extension [**39**], the homogeneous universal 'naturally ordered' C-relation [**13**], just to name a few. CSPs for structures that are definable in such structures are abundant in particular for qualitative reasoning calculi in Artificial Intelligence.

We want to point out that that decidability of primitive positive definability is already non-trivial when Γ is trivial from a model-theoretic perspective: for the case that Γ is the structure $(X; =)$ (where X is countably infinite), the decidability of $\mathrm{Expr}_{pp}(\Gamma)$ has been posed as an open problem in [**8**]. Theorem 39 solves this problem, since $(X; =)$ is isomorphic to a reduct of the structure $(\mathbb{Q}; <)$, which is clearly finitely bounded, homogeneous, ordered, and Ramsey.

The proof of Theorem 39 goes along the following lines, and is based on the results of the last sections. We outline the algorithm for $\mathrm{Expr}_{pp}(\Gamma)$; the proof for $\mathrm{Expr}_{ep}(\Gamma)$ is a subset. So the input are formulas $\phi_0, \dots, \phi_n$ defining relations $R_0, \dots, R_n$, and we have to decide whether R_0 has a primitive positive definition from $R_1, \dots, R_n$. Let Θ be the structure which has $R_1, \dots, R_n$ as its relations. By Theorem 9, R_0 is not primitive positive definable from $R_1, \dots, R_n$ if and only if there is a finitary function $f \in \mathrm{Pol}(\Theta)$ which violates R_0. By the ideas of the last section, such a polymorphism can be chosen to be canonical as a function from $(\Delta, c_1, \dots, c_k)^m$ to Δ, where $c_i \in \Delta$. Such canonical functions are essentially finite objects since they can be represented as functions on types. Therefore, the algorithm can then check for a given canonical function whether it is a polymorphism of Θ and whether it violates R_0. Also, k and m can be calculated from the input, and so there are only finitely many complete behaviors to be checked. Finally, the additional assumption that Δ be finitely bounded allows the algorithm to check whether a function on types really comes from a function on Δ. We refer to [**16**] for details.

7. Interpretability

Many ω-categorical structures can be derived from other ω-categorical structures via first-order interpretations. In this section we will discuss the fact already mentioned in the introduction that bi-interpretations can be used to transfer the Ramsey property from one structure to another. A special type of interpretations, called *primitive positive interpretations*, will become important in Section 8. The definition of interpretability we use is standard, and follows [**28**].

When Δ is a structure with signature τ, and $\delta(x_1, \dots, x_k)$ is a first-order τ-formula with the k free variables $x_1, \dots, x_k$, we write $\delta(\Delta^k)$ for the k-ary relation that is defined by δ over Δ.

Definition 40. A relational σ-structure Γ has a *(first-order) interpretation* in a τ-structure Δ if there exists a natural number d, called the *dimension* of the interpretation, and

- a τ-formula $\delta(x_1, \dots, x_d)$ – called *domain formula*,
- for each k-ary relation symbol R in σ a τ-formula $\phi_R(\overline{x}_1, \dots, \overline{x}_k)$ where the $\overline{x}_i$ denote disjoint d-tuples of distinct variables – called the *defining formulas*,
- a τ-formula $\phi_=(x_1, \dots, x_d, y_1, \dots, y_d)$, and
- a surjective map $h : \delta(\Delta^d) \to \Gamma$ – called *coordinate map*,

such that for all relations R in Γ and all tuples $\overline{a}_i \in \delta(\Delta^d)$

$$(h(\overline{a}_1), \dots, h(\overline{a}_k)) \in R \Leftrightarrow \Delta \models \phi_R(\overline{a}_1, \dots, \overline{a}_k) \text{ , and}$$
$$h(\overline{a}_1) = h(\overline{a}_2) \Leftrightarrow \Delta \models \phi_=(\overline{a}_1, \overline{a}_2) \text{ .}$$

If the formulas δ, ϕ_R, and $\phi_=$ are all primitive positive, we say that Γ has a *primitive positive interpretation* in Δ; many primitive positive interpretations can be found in Section 8. We say that Γ is *interpretable in* Δ *with finitely many parameters* if there are $c_1, \dots, c_n \in \Delta$ such that Γ is interpretable in the expansion of Δ by the singleton relations $\{c_i\}$ for all $1 \le i \le n$. First-order *definitions* are a special case of interpretations: a structure Γ is *(first-order) definable* in Δ if Γ has an interpretation in Δ of dimension one where the domain formula is logically equivalent to true.

Lemma 41 (see e.g. Theorem 7.3.8 in [**28**]). *If Δ is an ω-categorical structure, then every structure Γ that is first-order interpretable in Δ with finitely many parameters is ω-categorical as well.*

The following nicely describes interpretability between structures in terms of the (topological) automorphism groups of the structures.

Theorem 42 (Ahlbrandt and Ziegler [**3**]; also see Theorem 5.3.5 and 7.3.7 in [**28**]). *Let Δ be an ω-categorical structure with at least two elements. Then a structure Γ has a first-order interpretation in Δ if and only if there is a continuous group homomorphism $f : \mathrm{Aut}(\Delta) \to \mathrm{Aut}(\Gamma)$ such that the image of f has finitely many orbits in its action on Γ.*

Note that if Γ_2 has a d-dimensional interpretation I in Γ_1, and Γ_3 has an e-dimensional interpretation J in Γ_2, then Γ_3 has a natural ed-dimensional interpretation in Γ_1, which we denote by $J \circ I$. To formally describe $J \circ I$, suppose that the signature of Γ_i is τ_i for $i = 1, 2, 3$, and that $I = (d, \delta, (\phi_R)_{R \in \tau_2}, \phi_=, h)$ where d is the dimension, δ the domain formula, $\phi_=$ and $(\phi_R)_{R \in \tau_2}$ the interpreting relations, and h the coordinate map. Similarly, let $J = (e, \gamma, (\psi_R)_{R \in \tau_3}, \psi_=, g)$. We use the following.

Lemma 43 (Theorem 5.3.2 in [**28**]). *Let Γ_1, Γ_2, I as in the preceding paragraph. Then for every first-order τ_2-formula $\phi(x_1, \dots, x_k)$ there is τ_1-formula*

$$\phi^I(x_1^1, \dots, x_1^d, \dots, x_k^1, \dots, x_k^d)$$

such that for all $a_1, \dots, a_k \in \delta((\Gamma_1)^d)$

$$\Gamma_2 \models \phi(h(a_1), \dots, h(a_k)) \Leftrightarrow \Gamma_1 \models \phi^I(a_1, \dots, a_k) \text{ .}$$

We can now define the interpretation $J \circ I$ as follows: the domain formula η is γ^I, and the defining formula for $R \in \tau_3$ is $(\psi_R)^I$. The coordinate map is from $\eta((\Gamma_1)^{ed}) \to \Gamma_3$, and defined by

$$(a_1^1, \ldots, a_1^d, \ldots, a_e^1, \ldots, a_e^d) \;\mapsto\; g(h(a_1^1, \ldots, a_1^d), \ldots, h(a_e^1, \ldots, a_e^d))\ .$$

Two interpretations of Γ in Δ with coordinate maps h_1 and h_2 are called *homotopic* if the relation $\{(\bar{x}, \bar{y}) \mid h_1(\bar{x}) = h_2(\bar{y})\}$ is definable in Δ. The *identity interpretation* of a structure Γ is the 1-dimensional interpretation of Γ in Γ whose coordinate map is the identity. Two structures Γ and Δ are called *bi-interpretable* if there is an interpretation I of Γ in Δ and an interpretation J of Δ in Γ such that both $I \circ J$ and $J \circ I$ are homotopic to the identity interpretation (of Γ and of Δ, respectively).

THEOREM 44 (Ahlbrandt and Ziegler [**3**]). *Two ω-categorical structures Γ and Δ are bi-interpretable if and only if* $\mathrm{Aut}(\Gamma)$ *and* $\mathrm{Aut}(\Delta)$ *are isomorphic as topological groups.*

As a consequence of this result and Theorem 23 we obtain the following.

COROLLARY 45. *For ordered bi-interpretable ω-categorical homogeneous structures Γ and Δ, one has the Ramsey property if and only if the other one has the Ramsey property.*

We give an example. This corollary can be used to deduce that an important structure studied in temporal reasoning in artificial intelligence has the Ramsey property. For the relevance of this fact in constraint satisfaction, see Section 8.

We have already mentioned that the age of $(\mathbb{Q}; <)$ has the Ramsey property. Let Γ be the structure whose elements are pairs $(x, y) \in \mathbb{Q}^2$ with $x < y$, representing *intervals*, and which contains all binary relations R over those intervals such that the relation $\{(x, y, u, v) \mid ((x, y), (u, v)) \in R\}$ is first-order definable in $(\mathbb{Q}; <)$. Hence, Γ has a 2-dimensional interpretation I in $(\mathbb{Q}; <)$, whose coordinate map h_1 is the identity map on $D := \{(x, y) \in \mathbb{Q}^2 \mid x < y\}$.

The structure Γ is known under the name *Allen's Interval Algebra* in artificial intelligence. We claim that its age has the Ramsey property. Using the homogeneity of $(\mathbb{Q}; <)$, it is easy to show that Γ is homogeneous as well. By Corollary 45, it suffices to show that Γ and $(\mathbb{Q}; <)$ are bi-interpretable. We first show that $(\mathbb{Q}; <)$ has an interpretation J in Γ. The coordinate map h_2 of J maps $(x, y) \in D$ to x. The formula $\phi_=(a, b)$ is $R_0(a, b)$ where R_0 is the binary relation $\{((x, y), (u, v)) \mid x = u\}$ from Γ. The formula $\phi_<(a, b)$ is $R_1(a, b)$ where R_1 is the binary relation $\{((x, y), (u, v)) \mid x < u\}$.

We prove that $J \circ I$ is homotopic to the identity interpretation of $(\mathbb{Q}; <)$ in $(\mathbb{Q}; <)$. This holds since the relation $\{(x, y, u) \in \mathbb{Q}^3 \mid h_2(h_1(x, y)) = u\}$ has the first-order definition $x = u$ in $(\mathbb{Q}; <)$. To show that $I \circ J$ is homotopic to the identity interpretation, observe that the relation $\{(a, b, c) \in D^3 \mid h_1(h_2(a), h_2(b)) = c\}$ has the first-order definition $R_0(a, c) \wedge R_3(b, c)$ in Γ, where R_0 is the binary relation from Γ as defined above, and R_3 is the binary relation $\{((x, y), (u, v)) \in \Gamma^2 \mid x = v\}$ from Γ. This shows that Γ and $(\mathbb{Q}; <)$ are bi-interpretable.

8. Complexity of Constraint Satisfaction

In recent years, a considerable amount of research concentrated on the computational complexity of $\mathrm{CSP}(\Gamma)$ for *finite* structures Γ. Feder and Vardi [**22**] conjectured that for such Γ, the problem $\mathrm{CSP}(\Gamma)$ is either in P, or NP-complete[1]. This conjecture has been fascinating researchers from various areas, for instance from graph theory [**40**] and from finite model theory [**4**, **22**, **33**]. It has been discovered that complexity classification questions translate to fundamental questions in universal algebra [**19**, **29**], so that lately also many researchers in universal algebra started to work on questions that directly correspond to questions about the complexity of CSPs.

For arbitrary infinite structures Γ it can be shown that there are problems $\mathrm{CSP}(\Gamma)$ that are in NP, but neither in P nor NP-complete, unless P=NP. In fact, it can be shown that for every computational problem $\mathcal{P}$ there is an infinite structure Γ such that $\mathcal{P}$ and $\mathrm{CSP}(\Gamma)$ are equivalent under polynomial-time Turing reductions [**9**]. However, there are several classes of infinite structures Γ for which the complexity of $\mathrm{CSP}(\Gamma)$ can be classified completely.

In this section we will see three such classes of computational problems; they all have the property that

- every problem in this class can be formulated as $\mathrm{CSP}(\Gamma)$ where Γ has a first-order definition in a *base structure* Δ;
- Δ is ordered homogeneous Ramsey with finite signature.

For all three classes, the classification result can be obtained by the same method, which we describe in the following two subsections.

8.1. Climbing up the lattice. Clearly, if we add relations to a structure Γ with a finite relational signature, then the CSP of the structure thus obtained is computationally at least as complex as the CSP of Γ. On the other hand, when we add a primitive positive definable relation to Γ, then the CSP of the resulting structure has a polynomial-time reduction to $\mathrm{CSP}(\Gamma)$. This is not hard to show, and has been observed for finite domain structures in [**30**]; the same proof also works for structures over an infinite domain.

LEMMA 46. *Let $\Gamma = (D; R_1, \ldots, R_l)$ be a relational structure, and let R be a relation that has a primitive positive definition in Γ. Then the problems* $\mathrm{CSP}(\Gamma)$ *and* $\mathrm{CSP}(D; R, R_1, \ldots, R_l)$ *are equivalent under polynomial-time reductions.*

When we study the CSPs of the reducts Γ of a structure Δ, we therefore consider the lattice of reducts of Δ which are closed under primitive positive definitions (i.e., which contain all relations that are primitive positive definable from the reduct), and describe the border between tractability and NP-completeness in this lattice. We remark that the reducts of Δ have, since we expand them by all primitive positive definable relations, infinitely many relations, and hence do not define a CSP; however, we consider Γ tractable if and only if all structures obtained from Γ by dropping all but finitely many relations have a tractable CSP. Similarly, we consider Γ hard if there exists a structure obtained from Γ by dropping all but finitely many relations that has a hard CSP. With this convention, it is interesting to determine the *maximal tractable* reducts, i.e., those reducts closed under primitive positive definitions which do not contain any hard relation and which cannot be further extended without losing this property.

Recall the notion of a *clone* from Section 2. By Theorem 9, the lattice of primitive positive closed reducts of Δ and the lattice of closed clones containing

[1] By Ladner's theorem [**34**], there are infinitely many complexity classes between P and NP, unless P=NP.

$\mathrm{Aut}(\Delta)$ are antiisomorphic via the mappings $\Gamma \mapsto \mathrm{Pol}(\Gamma)$ (for reducts Γ) and $\mathcal{C} \mapsto \mathrm{Inv}(\mathcal{C})$ (for clones $\mathcal{C}$). We refer to the introduction of [**8**] for a detailed exposition of this well-known connection. Therefore, the maximal tractable reducts correspond to *minimal tractable* clones, which are precisely the clones of the form $\mathrm{Pol}(\Gamma)$ for a maximal tractable reduct.

The proof strategy of the classification results presented in Sections 8.3, 8.4, and 8.5 is as follows. We start by proving that certain reducts Γ have an NP-hard CSP. How to show this, and how to find those 'basic hard reducts' will be the topic of the next subsection. Let R be one of the relations from those hard reducts. If R does not have a primitive positive definition in Γ, then Theorem 9 implies that Γ has a polymorphism f that does not preserve R. We are now in a similar situation as in Section 5. Introducing constants, we can show that f generates an operation g that still does not preserve R but is canonical with respect to the expansion of Γ by constants. There are only finitely many canonical behaviours that g might have, and therefore we can start a combinatorial analysis. In the three classifications that follow, this strategy always leads to polymorphisms that imply that $\mathrm{CSP}(\Gamma)$ can be solved in polynomial time.

8.2. Primitive positive interpretations, and adding constants. Surprisingly, in all the classification results that we present in Sections 8.3, 8.4, and 8.5, there is a single condition that implies that a CSP is NP-hard. Recall that an interpretation is called *primitive positive* if all formulas involved in the interpretation (the domain formula, the formulas ϕ_R and $\phi_=$) are primitive positive. The relevance of primitive positive interpretations in constraint satisfaction comes from the following fact, which is known for finite domain constraint satisfaction, albeit not using the terminology of primitive positive interpretations [**19**]. In the present form, it appears first in [**6**].

THEOREM 47. *Let Γ and Δ be structures with finite relational signatures. If there is a primitive positive interpretation of Γ in Δ, then there is a polynomial-time reduction from* $\mathrm{CSP}(\Gamma)$ *to* $\mathrm{CSP}(\Delta)$*.*

All hardness proofs presented later can be shown via primitive positive interpretations of Boolean structures (i.e., structures with the domain $\{0,1\}$) with a hard CSP. In fact, in all such Boolean structures the relation NAE defined as

$$\mathrm{NAE} = \{0,1\}^3 \setminus \{(0,0,0),(1,1,1)\}$$

is primitive positive definable. This fact has not been stated in the original publications; however, it deserves to be mentioned as a unifying feature of all the classification results presented here. It is often more convenient to interpret other Boolean structures than $(\{0,1\}; \mathrm{NAE})$, and to then apply the following Lemma. An operation $f : D^k \rightarrow D$ is called *essentially a permutation* if there exists an i and a bijection $g : D\hat{E} \rightarrow D$ so that $f(x_1,\ldots,x_k) = g(x_i)$ for all $(x_1,\ldots,x_k) \in D^k$.

LEMMA 48. *Let Δ be a structure that interprets a Boolean structure Γ such that all polymorphisms of Γ are essentially a permutation. Then the structure* $(\{0,1\}; \mathrm{NAE})$ *has a primitive positive interpretation in Δ, and* $\mathrm{CSP}(\Delta)$ *is NP-hard.*

PROOF. Since the polymorphisms of Γ preserve the relation NAE, and by the well-known finite analog of Theorem 9 (due to [**25**] and independently, [**17**]), NAE is primitive positive definable in Γ. When ϕ is such a primitive positive definition, by substituting all relations in ϕ by their defining relations in Δ we obtain an interpretation of $(\{0,1\};\mathrm{NAE})$ in Δ. Hardness of $\mathrm{CSP}(\Delta)$ follows from the NP-hardness of $\mathrm{CSP}((\{0,1\};\mathrm{NAE}))$ (this problem is called NOT-ALL-3-EQUAL-3SAT in [**24**]) and Theorem 47. □

Typical Boolean structures Γ such that all polymorphisms of Γ are essentially a permutation are the structure $(\{0,1\};\{(t_1,t_2,t_3,t_4)\in\{0,1\}\mid t_1+t_2+t_3+t_4=2\}$, the structure $(\{0,1\};1\mathrm{IN}3)$, or the structure $(\{0,1\};\mathrm{NAE})$ itself.

Sometimes it is not possible to give a primitive positive interpretation of the structure $(\{0,1\};\mathrm{NAE})$ in Γ, but it is possible after expanding Γ with constants. Under an assumption about the endomorphism monoid of Γ, however, introducing constants does not change the computational complexity of Γ. More precisely, we have the following.

THEOREM 49 (Theorem 19 in [**5**]). *Let Γ be an ω-categorical structure with a finite relational signature such that* $\mathrm{Aut}(\Gamma)$ *is dense in* $\mathrm{End}(\Gamma)$*. Then for any finite number of elements $c_1,\dots,c_k$ of Γ there is a polynomial-time reduction from* $\mathrm{CSP}((\Gamma,\{c_1,\},\dots,\{c_k\}))$ *to* $\mathrm{CSP}(\Gamma)$.

8.3. Reducts of equality. One of the most fundamental classes of ω-categorical structures is the class of all reducts of $(X;=)$, where X is an arbitrary countably infinite set. Up to isomorphism, this is exactly the class of countable structures that are preserved by all permutations of their domain. The other two classes of ω-categorical structures that we will study here both contain this class.

We go straight to the statement of the complexity classification in terms of primitive positive interpretations. This is essentially a reformulation of a result from [**10**] which has been formulated without primitive positive interpretations. It turns out that when Γ is preserved by the operations from one of the minimal clones above the clone generated by all the permutations of X, then $\mathrm{CSP}(\Gamma)$ can be solved in polynomial time, and otherwise $\mathrm{CSP}(\Gamma)$ is NP-hard.

THEOREM 50 (essentially from [**10**]). *Let Γ be a reduct of $(X;=)$. Then exactly one of the following holds.*

- *Γ has a constant endomorphism. In this case,* $\mathrm{CSP}(\Gamma)$ *is trivially in P.*
- *Γ has a binary injective polymorphism. In this case,* $\mathrm{CSP}(\Gamma)$ *is in P.*
- *All relations with a first-order definition in $(X;=)$ have a primitive positive definition in Γ. Furthermore, the structure $(\{0,1\};\mathrm{NAE})$ has a primitive positive interpretation in Γ, and* $\mathrm{CSP}(\Gamma)$ *is NP-complete.*

PROOF. It has been shown in [**10**] that $\mathrm{CSP}(\Gamma)$ is in P when Γ has a constant or a binary injective polymorphism. Otherwise, by Theorem 36, every polymorphism of Γ is generated by the permutations of X. Hence, every relation R with a first-order definition in $(X;=)$ is preserved by all polymorphisms of Γ, and it follows from Theorem 9 that every relation is primitive positive definable in Γ.

This holds in particular for the relation E_6 defined as follows.

$$\begin{aligned} E_6=\{(x_1,x_2,y_1,y_2,z_1,z_2)\in X^6 \mid {} & (x_1=x_2\wedge y_1\neq y_2\wedge z_1\neq z_2)\\ &\vee\ (x_1\neq x_2\wedge y_1=y_2\wedge z_1\neq z_2)\\ &\vee\ (x_1\neq x_2\wedge y_1\neq y_2\wedge z_1=z_2)\}\end{aligned}$$

We now show that the structure $(\{0,1\}; \mathrm{1IN3})$ has a primitive positive interpretation in $(X; E_6)$, which by Lemma 48 also shows that $(\{0,1\}; \mathrm{NAE})$ has a primitive positive interpretation in $(X; E_6)$ and that $\mathrm{CSP}(\Gamma)$ is NP-hard.

The dimension of the interpretation is 2, and the domain formula is *'true'*. The formula $\phi_{1\mathrm{IN}3}(x_1, x_2, y_1, y_2, z_1, z_2)$ is $E_6(x_1, x_2, y_1, y_2, z_1, z_2)$, and The formula $\phi_=(x_1, x_2, y_1, y_2)$ is

$$\exists a_1, a_2, u_1, u_2, u_3, u_4, z_1, z_2.\ a_1 = a_2 \wedge E_6(a_1, a_2, u_1, u_2, u_3, u_4) \\ \wedge E_6(u_1, u_2, x_1, x_2, z_1, z_2) \wedge E_6(u_3, u_4, z_1, z_2, y_1, y_2).$$

Note that the primitive positive formula $\phi_=(x_1, x_2, y_1, y_2)$ is equivalent to $x_1 = x_2 \Leftrightarrow y_1 = y_2$. The map h maps (a_1, a_2) to 1 if $a_1 = a_2$, and to 0 otherwise. $\square$

Note that both the constant and the binary injective operation are canonical as functions over $(X; =)$.

8.4. Reducts of the dense linear order. An extension of the result in the previous subsection has been obtained in [**11**]; there, the complexity of the CSP for all reducts of $(\mathbb{Q}; <)$ has been classified. By a theorem of Cameron, those reducts are (again up to isomorphism) exactly the structures that are *highly set-transitive* [**20**], i.e., structures Γ such that for any two finite subsets A, B with $|A| = |B|$ of the domain there is an automorphism of Γ that maps A to B.

The corresponding class of CSPs contains many computational problems that have been studied in Artificial Intelligence, in particular in temporal reasoning [**18**, **37**, **48**], but also in scheduling [**36**] or general theoretical computer science [**23**, **43**]. The following theorem is a consequence of results from [**11**]. Again, we show that the hardness proofs in this class are captured by interpreting Boolean structures with few polymorphisms via primitive positive interpretations with finitely many parameters; this has not appeared in [**11**], so we provide the proof. The central arguments in the classification follow the reduct classification technique based on Ramsey theory that we present in this survey; see Figure 1 for an illustration of the bottom of the lattice of reducts of $(\mathbb{Q}; <)$, and the border of tractability for such reducts.

THEOREM 51 (essentially from [**11**]). *Let Γ be a reduct of $(\mathbb{Q}; <)$. Then exactly one of the following holds.*

- *Γ has one out of 9 binary polymorphisms (for a detailed description of those see [**11**]), and $\mathrm{CSP}(\Gamma)$ is in P.*
- *$\mathrm{Aut}(\Gamma)$ is dense in $\mathrm{End}(\Gamma)$, and the structure $(\{0,1\}; \mathrm{NAE})$ has a primitive positive interpretation with finitely many parameters in Γ. In this case, $\mathrm{CSP}(\Gamma)$ is NP-complete.*

Before we derive Theorem 51 from what has been shown in [**11**], we would like to point to Figure 1 for an illustration of the clones that correspond to maximal tractable reducts. The diagram also shows the constraint languages that just contain one of the important relations *Betw* (introduced in the introduction), *Cycl*, *Sep* (*Cycl* and *Sep* already appeared in Section 2), E_6 (which appeared earlier in this section), T_3, and $-T_3$. Here, T_3 stands for the relation

$$\{(x, y, z) \in \mathbb{Q}^3 \mid (x = y < z) \vee (x = z < y)\},$$

and when $R \subseteq \mathbb{Q}^k$, then $-R$ denotes $\{(-t_1, \ldots, -t_k) \mid (t_1, \ldots, t_k) \in R\}$.

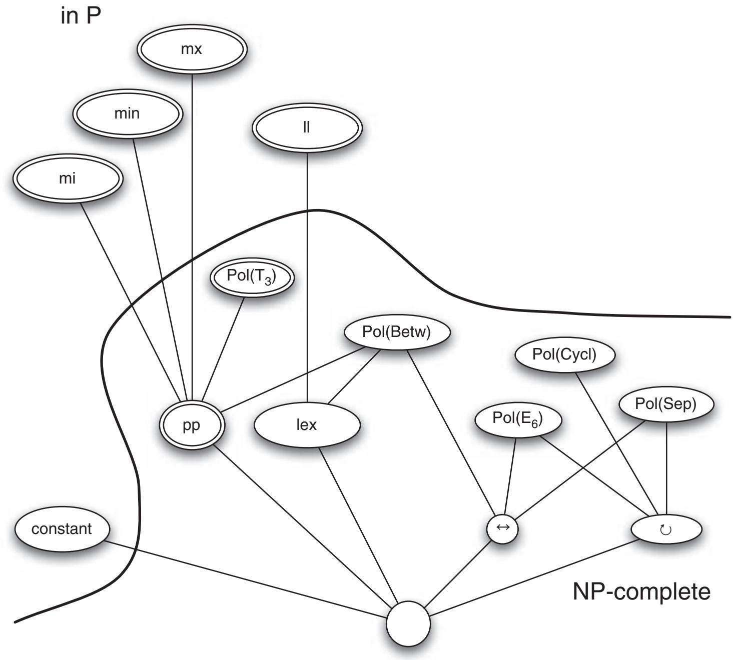


FIGURE 1. An illustration of the classification result for temporal constraint languages. Double-circles mean that the corresponding operation has a dual generating a distinct clone which is not drawn in the figure. For the definition of mi, min, mx, and ll, see [**11**].

The importance of those relations comes from the fact (shown in [**11**]) that unless Γ has one out of the 9 binary polymorphisms mentioned in Theorem 51 then there is a primitive positive definition of at least one of the relations *Betw*, *Cycl*, *Sep*, E_6, T_3, or $-T_3$.

PROOF OF THEOREM 51. It has been shown in [**11**] that unless Γ has a constant endomorphism, $\mathrm{Aut}(\Gamma)$ is dense in $\mathrm{End}(\Gamma)$. We have already seen that there is a primitive positive interpretation of $(\{0,1\};\mathrm{NAE})$ in structures isomorphic to $(\mathbb{Q};E_6)$.

Now suppose that T_3 is primitive positive definable in Γ. We give below a primitive positive interpretation of the structure $(\{0,1\};1\mathrm{IN}3)$ in $\Delta = (\mathbb{Q};T_3,0)$. Hence, there is also a primitive positive definition of $(\{0,1\};1\mathrm{IN}3)$ in the expansion of Γ by the constant 0. Expansions by constants do not change the computational complexity of $\mathrm{CSP}(\Gamma)$ since $\mathrm{Aut}(\Gamma)$ is dense in $\mathrm{End}(\Gamma)$. Thus, Lemma 48 shows NP-hardness of $\mathrm{CSP}(\Gamma)$, and that $(\{0,1\};\mathrm{NAE})$ has a primitive positive interpretation in $(\Gamma,0)$.

The interpretation of $(\{0,1\};1\mathrm{IN}3)$ in Δ

- has dimension 2;
- the domain formula $\delta(x_1,x_2)$ is $T_3(0,x_1,x_2)$;

- the formula $\phi_{1\text{IN}3}(x_1, x_2, y_1, y_2, z_1, z_2)$ is

$$\exists u.\ T_3(u, x_1, y_1) \wedge T_3(0, u, z_1)\ ;$$

- the formula $\phi_=(x_1, x_2, y_1, y_2)$ is $T_3(0, x_1, y_2)$;
- the coordinate map $h : \delta(\Delta^2) \to \{0, 1\}$ is defined as follows. Let (b_1, b_2) be a pair of elements of Δ that satisfies δ. Then exactly one of b_1, b_2 must have value 0, and the other element is strictly greater than 0. We define $h(b_1, b_2)$ to be 1 if $b_1 = 0$, and to be 0 otherwise.

To see that this is the intended interpretation, let $(x_1, x_2), (y_1, y_2), (z_1, z_2) \in \delta(\Delta^2)$, and suppose that $t := (h(x_1, x_2), h(y_1, y_2), h(z_1, z_2)) = (1, 0, 0) \in 1\text{IN}3$. We have to verify that $(x_1, x_2, y_1, y_2, z_1, z_2)$ satisfies $\phi_{1\text{IN}3}$ in Δ. Since $h(x_1, x_2) = 1$, we have $x_1 = 0$, and similarly we get that $y_1, z_1 > 0$. We can then set u to 0 and have $T_3(u, x_1, y_1)$ since $0 = u = x_1 < y_1$, and we also have $T_3(0, u, z_1)$ since $0 = u < z_1$. The case that $t = (0, 1, 0)$ is analogous. Suppose now that $t = (0, 0, 1) \in 1\text{IN}3$. Then $x_1, y_1 > 0$, and $z_1 = 0$. We can then set u to $\min(x_1, y_1)$, and therefore have $T_3(u, x_1, y_1)$, and $T_3(0, u, z_1)$ since $0 = z_1 < u$. Conversely, suppose that $(x_1, x_2, y_1, y_2, z_1, z_2)$ satisfies $\phi_{1\text{IN}3}$ in Δ. Since $T_3(0, u, z_1)$, exactly one out of u, z_1 equals 0. When $u = 0$, then because of $T_3(u, x_1, y_1)$ exactly one out of x_1, y_1 equals 0, and we get that $(h(x_1, x_2), h(y_1, y_2), h(z_1, z_2)) \in \{(0, 1, 0), (1, 0, 0)\} \subseteq 1\text{IN}3$. When $u > 0$, then $x_1 > 0$ and $y_1 > 0$, and so $(h(x_1, x_2), h(y_1, y_2), h(z_1, z_2)) = (0, 0, 1) \in 1\text{IN}3$.

An interpretation of $(\{0, 1\}; 1\text{IN}3)$ in $(\mathbb{Q}; -T_3, 0)$ can be obtained in a dual way.

Next, suppose that *Betw* is primitive positive definable in Γ. We will give a primitive positive interpretation of $(\{0, 1\}; \text{NAE})$ in $(\mathbb{Q}; \mathit{Betw}, 0)$. Hence, when *Betw* has a primitive positive definition in Γ, then by Theorem 49 (since $\text{Aut}(\Gamma)$ is dense in $\text{End}(\Gamma)$) and Lemma 48 we obtain NP-hardness of $\text{CSP}(\Gamma)$.

The dimension of the interpretation is one, and the domain formula is $x \neq 0$, which is clearly equivalent to a primitive positive formula over $(\mathbb{Q}; \mathit{Betw}, 0)$. The map h maps positive points to 1, and all other points from $\mathbb{Q}$ to 0. The formula $\phi_=(x_1, y_1)$ is

$$\exists z.\ \mathit{Betw}(x_1, 0, z) \wedge \mathit{Betw}(z, 0, y_1)$$

Note that the primitive positive formula $\phi_=$ is over $(\mathbb{Q}; \mathit{Betw}, 0)$ equivalent to $(x_1 > 0 \Leftrightarrow y_1 > 0)$. Finally, $\phi_{\text{NAE}}(x_1, y_1, z_1)$ is

$$\exists u.\ \mathit{Betw}(x_1, u, y_1) \wedge \mathit{Betw}(u, 0, z_1)\ .$$

If *Sep* has a primitive positive definition in Γ, then the statement follows easily from the previous argument since $\mathit{Betw}(x, y, z)$ has a 1-dimensional primitive positive interpretation in $(\mathbb{Q}; \mathit{Sep})$ (the formula $\phi_{\mathit{Betw}}(x, y, z)$ is $\exists u.\mathit{Sep}(u, x, y, z)$).

Finally, if *Cycl* is primitive positive definable in Γ, we give a 3-dimensional primitive positive interpretation of the structure $(\{0, 1\}; R, \neg)$ where $R = \{0, 1\}^3 \setminus \{(0, 0, 0)\}$ and $\neg = \{(0, 1), (1, 0)\}$. The idea of the interpretation is inspired by the NP-hardness proof of [**23**] for the 'Cyclic ordering problem' (see [**24**]).

The dimension of our interpretation is three, and the domain formula $\delta(x_1, x_2, x_3)$ is $x_1 \neq x_2 \wedge x_2 \neq x_3 \wedge x_3 \neq x_1$, which clearly has a primitive positive definition in $(\mathbb{Q}; \mathit{Cycl})$. The coordinate map h sends (x_1, x_2, x_3) to 0 if $\mathit{Cycl}(x_1, x_2, x_3)$, and to 1 otherwise.

Let $\phi(x_1, x_2, x_3, y_1, y_2, y_3)$ be the formula

$$\begin{aligned}&\mathit{Cycl}(x_1, y_1, x_2) \wedge \mathit{Cycl}(y_1, x_2, y_2) \wedge \mathit{Cycl}(x_2, y_2, x_3)\\ &\mathit{Cycl}(y_2, x_3, y_3) \wedge \mathit{Cycl}(x_3, y_3, x_1) \wedge \mathit{Cycl}(y_3, x_1, y_1)\ .\end{aligned}$$

When $(a_1, \ldots, a_6)$ satisfies ϕ, we can imagine $a_1, \ldots, a_6$ as points that appear clockwise in this order on the unit circle. In particular, we then have that $\mathit{Cycl}(a_1, a_3, a_5)$ holds if and only if $\mathit{Cycl}(a_2, a_4, a_6)$ holds. The formula $\phi_=(x_1, x_2, x_3, y_1, y_2, y_3)$ is

$$\begin{aligned}&\exists u_1^1, \ldots, u_3^4.\ \phi(x_1, x_2, x_3, u_1^1, u_2^1, u_3^1) \wedge \\ &\qquad \bigwedge_{i=1}^{3} \phi(u_1^i, u_2^i, u_3^i, u_1^{i+1}, u_2^{i+1}, u_3^{i+1}) \wedge \phi(u_1^4, u_2^4, u_3^4, y_1, y_2, y_3)\,,\end{aligned}$$

which is equivalent to

$$\delta(x_1, x_2, x_3) \wedge \delta(y_1, y_2, y_3) \wedge (\mathit{Cycl}(x_1, x_2, x_3) \Leftrightarrow \mathit{Cycl}(y_1, y_2, y_3))\,;$$

this is tedious, but straightforward to verify, and we omit the proof.

The formula $\phi_\neg(x_1, x_2, x_3, y_1, y_2, y_3)$ is $\phi_=(x_1, x_2, x_3, z_1, z_3, z_2)$.

The formula $\phi_R(x_1, x_2, x_3, y_1, y_2, y_3, z_1, z_2, z_3)$ is

$$\begin{aligned}\exists a, b, c, d, e, f, g, h, i, j, k, l, m, n.\ &\mathit{Cycl}(a, c, j) \wedge \mathit{Cycl}(b, j, k) \wedge \mathit{Cycl}(c, k, l)\\ &\wedge \mathit{Cycl}(d, f, j) \wedge \mathit{Cycl}(e, j, l) \wedge \mathit{Cycl}(f, l, m)\\ &\wedge \mathit{Cycl}(g, i, k) \wedge \mathit{Cycl}(h, k, m) \wedge \mathit{Cycl}(i, m, n)\\ &\wedge \mathit{Cycl}(n, m, l) \wedge \phi_=(x_1, x_2, x_3, a, b, c)\\ &\wedge \phi_=(y_1, y_2, y_3, d, e, f) \wedge \phi_=(z_1, z_2, z_3, g, h, i)\end{aligned}$$

The proof that for all tuples $\bar{a}_1, \bar{a}_2, \bar{a}_3 \in \mathbb{Q}^3$

$$(h(\bar{a}_1), h(\bar{a}_3), h(\bar{a}_3)) \in R \ \Leftrightarrow\ (\mathbb{Q}; \mathit{Cycl}) \models \phi_R(\bar{a}_1, \bar{a}_2, \bar{a}_3)$$

follows directly the correctness proof of the reduction presented in [**23**]. $\square$

8.5. Reducts of the random graph. The full power of the technique that is developed in this paper can be used to obtain a full complexity classification for all reducts of the random graph $G = (V; E)$ [**15**]. Again, the result can be stated in terms of primitive positive interpretations – this is not obvious from the statement of the result in [**15**], therefore we provide the proofs.

THEOREM 52 (essentially from [**15**]). *Let Γ be a reduct of the countably infinite random graph G. Then exactly one of the following holds.*

- *Γ has one out of 17 at most ternary canonical polymorphisms (for a detailed description of those see [**15**]), and* $\mathrm{CSP}(\Gamma)$ *is in P.*
- *Γ admits a primitive positive interpretation of* $(\{0,1\}; \mathrm{1IN3})$*. In this case,* $\mathrm{CSP}(\Gamma)$ *is NP-complete.*

PROOF. It has been shown in [**15**] that Γ has one out of 17 at most ternary canonical polymorphisms, and $\mathrm{CSP}(\Gamma)$ is in P, or one of the following relations has a primitive positive definition in Γ: the relation E_6, or the relation T, H, or $P^{(3)}$, which are defined as follows. The 4-ary relation T holds on $x_1, x_2, x_3, x_4 \in V$ if x_1, x_2, x_3, x_4 are pairwise distinct, and induce in G either

- a single edge and two isolated vertices,
- a path with two edges and an isolated vertex,

- a path with three edges, or
- a complement of one of the structures stated above.

To define the relation H, we write $N(u,v)$ as a shortcut for $E(u,v) \wedge u \neq v$. Then $H(x_1,y_1,x_2,y_2,x_3,y_3)$ holds on V if

$$\begin{aligned}
&\bigwedge_{i,j\in\{1,2,3\},i\neq j,u\in\{x_i,y_i\},v\in\{x_j,y_j\}} N(u,v)\\
&\wedge \big(((E(x_1,y_1)\wedge N(x_2,y_2)\wedge N(x_3,y_3))\\
&\quad\vee\ (N(x_1,y_1)\wedge E(x_2,y_2)\wedge N(x_3,y_3))\\
&\quad\vee\ (N(x_1,y_1)\wedge N(x_2,y_2)\wedge E(x_3,y_3))\big)\,.
\end{aligned}$$

The ternary relation $P^{(3)}$ holds on x_1,x_2,x_3 if those three vertices are pairwise distinct and do not induce a clique or an independent set in G.

Suppose first that T is primitive positive definable in Γ. Let R be the relation $\{(t_1,t_2,t_3,t_4)\in\{0,1\} \mid t_1+t_2+t_3+t_4=2\}$. We have already mentioned that all polymorphisms of $(\{0,1\};R)$ are essentially permutations. To show that $(\{0,1\};\mathrm{NAE})$ has a primitive positive interpretation in Γ, we can therefore use Lemma 48 and it suffices to show that there is a primitive positive interpretation of the structure $(\{0,1\};R)$ in $(V;T)$. For a finite subset S of V, write $\#S$ for the parity of edges between members of S. Now we define the relation $L\subseteq V^6$ as follows.

$$\begin{aligned}L := \big\{x\in V^6 \mid{}& \text{the entries of } x \text{ are pairwise distinct, and}\\ &\#\{x_1,x_2,x_3\} = \#\{x_4,x_5,x_6\}\big\}\end{aligned}$$

It has been shown in [**15**] that the relation L is pp-definable in $(V;T)$. We therefore freely use the relation L (and similarly $\neq$, the disequality relation) in primitive positive formulas over $(V;T)$.

Our primitive positive interpretation of $(\{0,1\};R)$ has dimension three. The domain formula $\delta(x_1,x_2,x_3)$ is $x_1\neq x_2 \ \wedge\ x_1\neq x_3\ \wedge\ x_2\neq x_3$. The formula $\phi_R(x_1^1,x_2^1,x_3^1,\dots,x_1^4,x_2^4,x_3^4)$ of the interpretation is

$$\begin{aligned}\exists y_1,y_2,y_3,y_4.\ T(y_1,\dots,y_4)&\\
\wedge L(x_1^1,x_2^1,x_3^1,y_2,y_3,y_4)&\\
\wedge L(x_1^2,x_2^2,x_3^2,y_1,y_3,y_4)&\\
\wedge L(x_1^3,x_2^3,x_3^3,y_1,y_2,y_4)&\\
\wedge L(x_1^4,x_2^4,x_3^4,y_1,y_2,y_3)&\end{aligned}$$

The formula $\phi_=$ is $L(x_1,x_2,x_3,y_1,y_2,y_3)$. Finally, the coordinate map sends a tuple (a_1,a_2,a_3) for pairwise distinct a_1,a_2,a_3 to 1 if $P^{(3)}(a_1,a_2,a_3)$, and to 0 otherwise.

Next, suppose that H is primitive positive definable in Γ. We give a 2-dimensional interpretation of $(\{0,1\};\mathrm{1IN3})$ in Γ. The domain formula is *'true'*. The formula $\phi_=(x_1,x_2,y_1,y_2)$ is

$$\begin{aligned}\exists z_1,z_2,u_1,u_2,v_1,v_2.\ &H(x_1,x_2,u_1,u_2,z_1,z_2)\wedge N(u_1,u_2)\\ &\wedge H(z_1,z_2,v_1,v_2,y_1,y_2)\wedge N(v_1,v_2)\,.\end{aligned}$$

This formula is equivalent to a primitive positive formula over Γ since $N(x,y)$ is primitive positive definable by H. The formula $\phi_{1\mathrm{IN}3}(x_1,x_2,y_1,y_2,z_1,z_2)$ is

$$\exists x_1',x_2',y_1',y_2',z_1',z_2'.\, H(x_1',x_2',y_1',y_2',z_1',z_2') \\ \wedge \phi_=(x_1,x_2,x_1',x_2') \wedge \phi_=(x_1,x_2,x_1',x_2') \wedge \phi_=(x_1,x_2,x_1',x_2')\,.$$

The coordinate map sends a tuple (x_1,x_2) to 1 if $E(x_1,x_2)$ and to 0 otherwise.

Finally, suppose that $P^{(3)}$ has a primitive positive definition in Γ. We give a 2-dimensional primitive positive interpretation of $(\{0,1\};\mathrm{NAE})$. For $k \geq 3$, let $Q^{(k)}$ be the k-ary relation that holds for a tuple $(x_1,\ldots,x_k) \in V^k$ iff $x_1,\ldots,x_k$ are pairwise distinct, and $(x_1,\ldots,x_k) \notin P^{(k)}$. It has been shown in [**15**] that the relation $Q^{(4)}$ is primitive positive definable by the relation $P^{(3)}$. Now, the formula $\phi_=(x_1,x_2,y_1,y_2)$ is $\exists z_1,z_2.Q^{(4)}(x_1,x_2,z_1,z_2) \wedge Q^{(4)}(z_1,z_2,y_1,y_2)$. The formula $\phi_{\mathrm{NAE}}(x_1,x_2,y_1,y_2,z_1,z_2)$ is

$$\exists u,v,w.\ P^{(3)}(u,v,w) \wedge Q^{(4)}(x_1,x_2,u,v) \\ \wedge \hat{E}\ Q^{(4)}(y_1,y_2,v,w) \wedge Q^{(4)}(z_1,z_2,w,u)\,.$$

The coordinate map sends a tuple (x_1,x_2) to 1 if $E(x_1,x_2)$ and to 0 otherwise. □

9. Concluding Remarks and Further Directions

We have outlined an approach to use Ramsey theory for the classification of reducts of a structure, considered up to existential positive, or primitive positive interdefinability. The central idea in this approach is to study functions that preserve the reduct, and to apply structural Ramsey theory to show that those functions must act regularly on large parts of the domain. This insight makes those functions accessible to combinatoral arguments and classification.

Our approach has been illustrated for the reducts of $(\mathbb{Q};<)$, and the reducts of the random graph $(V;E)$. One application of the results is complexity classification of constraint satisfaction problems in theoretical computer science. Interestingly, the hardness proofs in those classifications all follow a common pattern: they are based on primitive positive interpretations. In particular, we proved complete complexity classifications without the typical computer science hardness proofs – rather, the hardness results follow from mathematical statements about primitive positive interpretability in ω-categorical structures.

There are many other natural and important ω-categorical structures besides $(\mathbb{Q};<)$ and $(V;E)$ where this approach seems promising. We have listed some of the simplest and most basic examples in Figure 2. In this table, the first column specifies the 'base structure' Δ, and we will be interested in the class of all structures definable in Δ. The second column lists what is known about this class, considered up to first-order interdefinability. The third column describes the corresponding Ramsey result, when Δ is equipped with an appropriate linear order. The fourth column gives the status with respect to complexity classification of the corresponding class of CSPs. The fifth class indicates in which areas in computer science those CSPs find applications.

References

[1] Samson Adepoju Adeleke and Dugald Macpherson. Classification of infinite primitive Jordan permutation groups. *Proceedings of the London Mathematical Society*, 72(1):63–123, 1996.

Reducts of	First-order Reducts	Ramsey Class	CSP Dichotomy	Application, Motivation
$(X;=)$	Trivial	Ramsey's theorem	Yes	Equality Constraints
$(\mathbb{Q};<)$	Cameron [**21**]	Ramsey's theorem	Yes	Temporal Reasoning
$(V;E)$	Thomas [**46**]	Nešetřil + Rödl [**42**]	Yes	Schaefer's theorem for graphs
Homogeneous universal poset	?	Nešetřil + Rödl [**41**]	?	Temporal Reasoning
Homogeneous C-relation	Adeleke, Macpherson, Neumann [**1**, **2**]	Deuber, Miliken	?	Phylogeny Reconstruction
Countable atomless Boolean algebra	?	Graham, Leeb, Rothschild (see [**32**])	?	Set Constraints
Allen's Interval Algebra	?	This paper, Section 7	?	Temporal Reasoning

FIGURE 2. A diagram suggesting future research.

[2] Samson Adepoju Adeleke and Peter M. Neumann. *Relations related to betweenness: their structure and automorphisms*, volume 623 of *Memoirs of the AMS 131*. American Mathematical Society, 1998.
[3] Gisela Ahlbrandt and Martin Ziegler. Quasi-finitely axiomatizable totally categorical theories. *Annals of Pure and Applied Logic*, 30(1):63–82, 1986.
[4] Albert Atserias, Andrei A. Bulatov, and Anuj Dawar. Affine systems of equations and counting infinitary logic. *Theoretical Computer Science*, 410(18):1666–1683, 2009.
[5] Manuel Bodirsky. Cores of countably categorical structures. *Logical Methods in Computer Science*, 3(1):1–16, 2007.
[6] Manuel Bodirsky. Constraint satisfaction problems with infinite templates. In Heribert Vollmer, editor, *Complexity of Constraints (a collection of survey articles)*, pages 196–228. Springer, LNCS 5250, 2008.
[7] Manuel Bodirsky and Hubert Chen. Oligomorphic clones. *Algebra Universalis*, 57(1):109–125, 2007.
[8] Manuel Bodirsky, Hubie Chen, and Michael Pinsker. The reducts of equality up to primitive positive interdefinability. *Journal of Symbolic Logic*, 75(4):1249–1292, 2010.
[9] Manuel Bodirsky and Martin Grohe. Non-dichotomies in constraint satisfaction complexity. In *Proceedings of ICALP'08*, pages 184–196, 2008.
[10] Manuel Bodirsky and Jan Kára. The complexity of equality constraint languages. *Theory of Computing Systems*, 3(2):136–158, 2008. A conference version appeared in the proceedings of CSR'06.
[11] Manuel Bodirsky and Jan Kára. The complexity of temporal constraint satisfaction problems. *Journal of the ACM*, 57(2), 2009. An extended abstract appeared in the proceedings of STOC'08.
[12] Manuel Bodirsky and Jaroslav Nešetřil. Constraint satisfaction with countable homogeneous templates. *Journal of Logic and Computation*, 16(3):359–373, 2006.
[13] Manuel Bodirsky and Diana Piguet. Finite trees are ramsey with respect to topological embeddings. Preprint, arXiv:1002:1557, 2008.

[14] Manuel Bodirsky and Michael Pinsker. Minimal functions on the random graph. Preprint, arXiv:1003.4030, 2010.
[15] Manuel Bodirsky and Michael Pinsker. Schaefer's theorem for graphs. In *Proceedings of STOC'11*, 2011. Preprint of the long version available from http://www.dmg.tuwien.ac.at/pinsker/.
[16] Manuel Bodirsky, Michael Pinsker, and Todor Tsankov. Decidability of definability. In *Proceedings of LICS'11*, 2011.
[17] V. G. Bodnarčuk, L. A. Kalužnin, V. N. Kotov, and B. A. Romov. Galois theory for post algebras, part I and II. *Cybernetics*, 5:243–539, 1969.
[18] Mathias Broxvall and Peter Jonsson. Point algebras for temporal reasoning: Algorithms and complexity. *Artificial Intelligence*, 149(2):179–220, 2003.
[19] Andrei Bulatov, Andrei Krokhin, and Peter G. Jeavons. Classifying the complexity of constraints using finite algebras. *SIAM Journal on Computing*, 34:720–742, 2005.
[20] Peter J. Cameron. Transitivity of permutation groups on unordered sets. *Mathematische Zeitschrift*, 148:127–139, 1976.
[21] Peter J. Cameron. The random graph. *R. L. Graham and J. Nešetřil, Editors, The Mathematics of Paul Erdös*, 1996.
[22] Tomás Feder and Moshe 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, 1999.
[23] Zvi Galil and Nimrod Megiddo. Cyclic ordering is NP-complete. *Theoretical Computer Science*, 5(2):179–182, 1977.
[24] Michael Garey and David Johnson. *A guide to NP-completeness.* CSLI Press, Stanford, 1978.
[25] David Geiger. Closed systems of functions and predicates. *Pacific Journal of Mathematics*, 27:95–100, 1968.
[26] Martin Goldstern and Michael Pinsker. A survey of clones on infinite sets. *Algebra Universalis*, 59:365–403, 2008.
[27] Edmond E. Granirer. Extremely amenable semigroups 2. *Mathematica Scandinavica*, 20:93–113, 1967.
[28] Wilfrid Hodges. *Model theory.* Cambridge University Press, 1993.
[29] Pawel M. Idziak, Petar Markovic, Ralph McKenzie, Matthew Valeriote, and Ross Willard. Tractability and learnability arising from algebras with few subpowers. In *Proceedings of LICS'07*, pages 213–224, 2007.
[30] Peter Jeavons, David Cohen, and Marc Gyssens. Closure properties of constraints. *Journal of the ACM*, 44(4):527–548, 1997.
[31] Markus Junker and Martin Ziegler. The 116 reducts of $(\mathbb{Q}, <, a)$. *Journal of Symbolic Logic*, 74(3):861–884, 2008.
[32] Alexander Kechris, Vladimir Pestov, and Stevo Todorcevic. Fraissé limits, Ramsey theory, and topological dynamics of automorphism groups. *Geometric and Functional Analysis*, 15(1):106–189, 2005.
[33] Phokion G. Kolaitis and Moshe Y. Vardi. Conjunctive-query containment and constraint satisfaction. In *Proceedings of PODS'98*, pages 205–213, 1998.
[34] Richard E. Ladner. On the structure of polynomial time reducibility. *Journal of the ACM*, 22(1):155–171, 1975.
[35] Dugald Macpherson. A survey of homogeneous structures. *Discrete Mathematics*, 311(15):1599–1634, 2011.
[36] Rolf H. Möhring, Martin Skutella, and Frederik Stork. Scheduling with and/or precedence constraints. *SIAM Journal on Computing*, 33(2):393–415, 2004.
[37] Bernhard Nebel and Hans-Jürgen Bürckert. Reasoning about temporal relations: A maximal tractable subclass of Allen's interval algebra. *Journal of the ACM*, 42(1):43–66, 1995.
[38] Jaroslav Nešetřil. Ramsey theory. *Handbook of Combinatorics*, pages 1331–1403, 1995.
[39] Jaroslav Nešetřil. Ramsey classes and homogeneous structures. *Combinatorics, Probability & Computing*, 14(1-2):171–189, 2005.
[40] Jaroslav Nešetřil and Pavol Hell. Colouring, constraint satisfaction, and complexity. *Computer Science Review 2*, pages 143–163, 2008.
[41] Jaroslav Nešetřil and Vojtech Rödl. Combinatorial partitions of finite posets and lattices. *Algebra Universalis*, 19:106–119, 1984.
[42] Jaroslav Nešetřil and Vojtech Rödl. *Mathematics of Ramsey Theory.* Springer, Berlin, 1998.

[43] Jaroslav Opatrny. Total ordering problem. *SIAM Journal on Computing*, 8(1):111–114, 1979.
[44] Michael Pinsker. More sublattices of the lattice of local clones. *Order*, 27(3):353–364, 2010.
[45] Ivo G. Rosenberg. Minimal clones I: the five types. *Lectures in Universal Algebra (Proc. Conf. Szeged, 1983), Colloq. Math. Soc. J. Bolyai*, 43:405–427, 1986.
[46] Simon Thomas. Reducts of the random graph. *Journal of Symbolic Logic*, 56(1):176–181, 1991.
[47] Simon Thomas. Reducts of random hypergraphs. *Annals of Pure and Applied Logic*, 80(2):165–193, 1996.
[48] Peter van Beek. Reasoning about qualitative temporal information. *Artificial Intelligence*, 58:297–326, 1992.
[49] Ross Willard. Testing expressibility is hard. In *Proceedings of CP'10*, 2010.

Laboratoire d'Informatique (LIX), CNRS UMR 7161, École Polytechnique, 91128 Palaiseau, France
E-mail address: bodirsky@lix.polytechnique.fr
URL: http://www.lix.polytechnique.fr/~bodirsky/

Équipe de Logique Mathématique, Université Denis Diderot - Paris 7, UFR de Mathématiques, 75205 Paris Cedex 13, France
E-mail address: marula@gmx.at
URL: http://dmg.tuwien.ac.at/pinsker/

This volume contains the proceedings of the AMS-ASL Special Session on Model Theoretic Methods in Finite Combinatorics, held January 5–8, 2009, in Washington, DC.

Over the last 20 years, various new connections between model theory and finite combinatorics emerged. The best known of these are in the area of 0-1 laws, but in recent years other very promising interactions between model theory and combinatorics have been developed in areas such as extremal combinatorics and graph limits, graph polynomials, homomorphism functions and related counting functions, and discrete algorithms, touching the boundaries of computer science and statistical physics.

This volume highlights some of the main results, techniques, and research directions of the area. Topics covered in this volume include recent developments on 0-1 laws and their variations, counting functions defined by homomorphisms and graph polynomials and their relation to logic, recurrences and spectra, the logical complexity of graphs, algorithmic meta theorems based on logic, universal and homogeneous structures, and logical aspects of Ramsey theory.

ISBN 978-0-8218-4943-9

9 780821 849439

CONM/558