Colouring, Constraint Satisfaction, and Complexity

For Aleš Pultr on the Occasion of his 70th Birthday

Pavol Hell [*]
School of Computing Science
Simon Fraser University
Burnaby, B.C., Canada V5A 1S6
pavol@cs.sfu.ca
and
Jaroslav Nešetřil [†]
KAM and ITI, Charles University
Malostranské Náměsti 22
Prague, Czech Republic
nesetril@kam.mff.cuni.cz

Abstract

Constraint satisfaction problems have enjoyed much attention since the early seventies, and in the last decade have become also a focus of attention amongst theoreticians. Graph colourings are a special class of constraint satisfaction problems; they offer a microcosm of many of the considerations that occur in constraint satisfaction. From the point of view of theory, they are well known to exhibit a dichotomy of complexity - the k-colouring problem is polynomial time solvable when $k \leq 2$, and NP-complete when $k \geq 3$. Similar dichotomy has been proved for the class of graph homomorphism problems, which are intermediate problems between graph colouring and constraint satisfaction. However, for general constraint satisfaction problems, dichotomy has only been conjectured. Although the conjecture remains unproven to this day, it has been driving much of the theoretical research on constraint satisfaction problems, which combines methods of logic, universal algebra, analysis, and combinatorics. Currently, this is a very active area of research, and it is our goal here to present some of the recent developments, updating some of the information in existing books and surveys, while focusing on both the mathematical and the computational aspects of the theory. Given the level of activity, we are only able to survey a fraction of the new work, with emphasis on our own areas of interest.

Contents

1 Introduction

Consider the following scheduling application. Each semester at a typical university, the courses taught need to be scheduled for examination: several courses can be examined in one examination period, but two courses that have common students must be scheduled at different times. The university aims to schedule all examinations in as few examination periods as possible.

The situation as simplified above leads to a classical formulation via graph colouring [32, 138, 178]. We form the graph G in which the vertices are the courses taught, and in which two courses are adjacent just if the courses *conflict*, i.e., have students in common. A schedule with k exam periods corresponds exactly to a k-colouring of G. In another view of this model, we may say that the courses represent *variables* with *values* from a certain *domain* of possibilities, which are *constrained* by requiring certain pairs of courses (those that conflict) to be assigned different values.

1.1 Constraint Satisfaction

A general *constraint satisfaction problem* is given by a set $\mathcal{V}$ of *variables* which are to be assigned *values* from a domain $\mathcal{D}$, and a set $\mathcal{C}$ of *constraints*, each of which restricts certain combinations of variables to certain sets of allowed values. Formally, a constraint satisfaction problem is defined a triple $\mathcal{V}, \mathcal{D}, \mathcal{C}$, where $\mathcal{V}$ is a set of variables, $\mathcal{D}$ is a domain (set) of values, and $\mathcal{C}$ is a set of constraints. Each constraint is a triple (r, t, U), where r is a positive integer called *arity* of the constraint, t is an r-tuple of variables, and U is a set of r-tuples of values. An *evaluation* of the variables is a mapping $f : \mathcal{V} \to \mathcal{D}$. Such an evaluation satisfies a constraint $(r, t, U) \in \mathcal{C}$ if $(f(t_1), f(t_2), \ldots, f(t_r)) \in U$. A *solution* is an evaluation that satisfies all constraints.

Obviously, such a general context allows much more refined models of the exam scheduling situation - we may express a much wider variety of restrictions on the schedule. For instance, suppose we wish to express the natural space limitation requiring that no more than three experimental physics exams are to be scheduled at any given time. This can be seen as a restriction on each quadruple of experimental physics courses, which limits their assigned values to those quadruples of time periods which do not fall within the same time period.

As another example, consider the *channel assignment problem* in which frequencies (channels) are assigned to transmitters for wireless communication, cf., e.g., [74].) Since the number of frequencies is limited, the operator must find a way to reuse the channels, in a way that minimizes interference. Interference will surely happen on two transmitters on the same antenna; these must receive different channels. Interference may also happen, to a smaller degree, on two transmitters on different antennas but in the same base transceiver station, or located in a particular way with respect to a geographical feature. Furthermore, there may be some interference even if two transmitters operate on adjacent, or nearby, channels, or on a channel and its harmonic. There may also be restrictions on which channels a particular transmitter may use—for instance at the edge of the operator's territory there may be constraints imposed by a neighbouring country or operator. We can model this problem by viewing the individual transmitters as variables whose values are the frequencies. To express the interference constraints we simply restrict pairs of more or less interfering transmitters to frequencies that can accommodate them. In addition to these binary constraints we also have the unary constraints that restrict the transmitters at the edge of the territory to frequencies available in such location.

These examples should make it clear that a great variety of natural problems can be expressed in this model, including problems in scheduling, planning, data bases, machine vision, belief maintenance, temporal reasoning, type reconstruction, and many other areas of artificial intelligence [37, 44, 126, 137, 145, 146, 183, 185]. The model was pioneered by Montanari [149], and enjoys wide popularity: there are journals entirely devoted to solution of such problems, and books have been written about them [182]. The recent Handbook of Constraint Satisfaction [1] witnesses the activity in the field. Our focus here is on the theoretical aspects of constraint satisfaction, which forms only a small but nevertheless very active part of this area. (In the Handbook [1], it is represented by one chapter.) Because of this rapid development, we feel another update of recent surveys such as [71, 98, 106, 99, 150, 152] is justified.

1.2 Homomorphisms

An alternate view of constraint satisfaction was proposed by Feder and Vardi [73]. Basically, it returns to the original definition of the graph colouring problem described earlier. In that case we were able to describe all constraints as a non-equality relation imposed

on certain pairs of vertices. With a suitably generalized context, we will be able to do something similar for all constraint satisfaction problems.

A *relational structure* G consists of a finite set $V(G)$, whose elements we shall call *vertices* in order to underscore our graph theoretic inspiration, and a finite number of relations $R_1, R_2, \ldots, R_p$, of arities $r_1, r_2, \ldots, r_p$ respectively. The vector $(r_1, r_2, \ldots, r_p)$ is called *the type* of G. Given two relational structures G (with vertex set $V(G)$ and relations $R_1, R_2, \ldots, R_p$) and H (with vertex set $V(H)$ and relations $S_1, S_2, \ldots, S_p$), of the same type (the arity of each R_i is the same as that of the corresponding S_i), a *homomorphism* of G to H is a mapping $f : V(G) \to V(H)$ which preserves all pairs of corresponding relations, i.e., such that $(v_1, v_2, \ldots, v_{r_i}) \in R_i$ implies $(f(v_1), f(v_2), \ldots, f(v_{r_i})) \in S_i$, for all $i = 1, 2, \ldots, p$.

In order to minimize the notation, we have adopted a natural extension of the graph theoretic notation. However, relational structures are complex objects, and we remark that it is more common to use a special notation, where a structure is written in bold font, say $\mathbf{A}$, its ground set by the corresponding letter, A, and the relational symbols are usually written as $R_{\mathbf{A}}$.

Given a constraint satisfaction problem with variables $\mathcal{V}$, domain $\mathcal{D}$, and a set $\mathcal{C}$ of constraints, we define the structure G with $V(G) = \mathcal{V}$ and the structure H with $V(H) = \mathcal{D}$, where each U occurring in some constraint $C = (r, t, U)$ defines an r-ary relation S_i on $V(H)$, for which the corresponding relation R_i consists of all r-tuples t' of variables with a constraint $C' = (r, t', U)$. Then a solution to the constraint satisfaction problem is simply a homomorphism of G to H. Of course, the reverse of this transformation succeeds in expressing the existence of any homomorphism of some general structures G to H as a constraint satisfaction problem. Therefore from now on we think of constraint satisfaction problems as problems seeking the existence of a homomorphism between relational structures [73]. This is not to say that we advocate the use of such reformulations in individual constraint satisfaction problems, quite to the contrary: the relational formulation doesn't have the same "feel" for individual data and its semantic meaning which can (and must) be used in solving concrete problems. On the theoretical side, however, this formulation of constraint satisfaction problems has proved crucial, by relating it to the techniques of categorial and universal algebra [28, 29] and more generally to algebraic combinatorics, as well as to statistical physics [19, 78, 106, 157]. For example, the study of homomorphisms between relational structures has a long tradition in the 'Prague School' of category theory [106], and many theoretical insights, such as recorded for instance in the monograph [172], are applicable to the present context.

A typical aspect of our example scheduling problem is the fact that the structure G of courses to be examined changes every semester, while the structure H of exam periods tends to be much more stable. (For instance, which courses are taught, and which pairs of courses have students in common, varies, while the total number of exam periods, the number of periods per day, and so on, may remain constant for several consecutive

semesters.) A similar situation arises in the channel assignment problem - again the values (frequencies) H are the same over a number of problems, while the variables (transmitters) G vary according to concrete locations. This occurs for many other applications (the structure H tends to be fixed), and it motivates the following variant of the constraint satisfaction problem.

Let H be a fixed relational structure. The *constraint satisfaction problem* CSP(H) asks whether or not an input relational structure G, of the same type as H, admits a homomorphism to H.

Consider the problem CSP(H), where H has only one (binary) relation, S_1, consisting of all ordered pairs of distinct vertices of $V(H)$. Since the fixed structure H has one symmetric binary relation, the input structure G may be assumed to have also just one symmetric binary relation, i.e., G is a graph. Thus we have the problem of colouring the input structure (graph) G with $|V(H)|$ colours, so that adjacent vertices of G obtain different colours. This is the *graph colouring problem*. Suppose more generally that H has any one symmetric binary relation, in other words, H is an undirected graph. Then the problem CSP(H) may be assumed to be restricted to undirected input graphs G; it asks whether or not G admits a homomorphism to H. This problem is called the *graph H-colouring problem*. Similarly, if H is a structure with one (not necessarily symmetric) binary relation, i.e., a digraph, each input G is also a digraph, and the problem reduces to the existence of digraph homomorphisms [106]. We emphasize that this view regards undirected graphs as a special class of digraphs, namely ones with a symmetric adjacency relation. Many (but not all) phenomena that occur for general constraint satisfaction are typified by their restrictions in the context of digraph homomorphisms. Definitions and results are often stated in the restricted domain of digraph homomorphisms [106], but with the understanding that they apply equally well in the general context of constraint satisfaction. Consider, for instance, the following observations ([106]). When H and H' are two digraphs (more generally, relational structures of the same type) which admit a homomorphism of H to H' as well as a homomorphism of H' to H, then the two problems CSP(H) and CSP(H') are obviously equivalent. (A structure G admits a homomorphism to H if and only if it admits a homomorphism to H', since the composition of two homomorphisms is again a homomorphism.) A digraph (relational structure) H with no proper substructure H' to which H admits a homomorphism is called a *core* [102, 106]. It is easy to check that every digraph (relational structure) H has, up to isomorphism, a unique core subdigraph (substructure) H' to which it admits a homomorphism [102, 106]. Thus we typically restrict our attention to problems CSP(H) where H is a core.

2   Dichotomy

The k-colouring problem is polynomial time solvable when $k = 1, 2$ and is NP-complete otherwise. This basic fact (established at the very onset of the theory of NP-completeness

[81]) illustrates the *dichotomy* of possible complexities of the class of k-colouring problems, as k varies. There is, in principle, no reason for each colouring problem to be polynomial time solvable (one of the easiest problems in NP) or NP-complete (one of the hardest problems in NP). Indeed, Ladner [130] has shown that if P $\neq$ NP, there are in NP problems that are neither polynomial nor NP-complete - in fact there must be an infinite hierarchy of such (non-polynomially-equivalent) problems. Since these "intermediate difficulty" problems must exist in NP (unless P=NP), dichotomies are always somewhat surprizing, especially if they apply to a broad class of problems.

Dichotomy for graph H-colouring problems has been proved by the authors in [104]. In other words, we have shown that, for every graph H, the problem of deciding the existence of a homomorphism of an input graph G to H is polynomial time solvable, or is NP-complete. In fact, we have classified which problems are polynomial time solvable and which are NP-complete.

Theorem 2.1 (Graph Dichotomy) [104] *Suppose H is a graph, i.e., a relational structure with a single relation which is binary and symmetric. Then $CSP(H)$ is NP-complete, except in the following, polynomial time solvable, cases.*

1. *H is bipartite; or*

2. *H has a loop*

The proof of the Graph Dichotomy in [104] was surprizingly complex. In the intervening years, a number of new proofs have appeared [26, 128, 187], based on a great variety of ideas and approaches. We discuss these in several places of this manuscript, as they present a testing ground of these techniques. In this sense, the Graph Dichotomy is a leitmotif of this survey.

Another early dichotomy has been proved for all *Boolean satisfiability problems* by Schaeffer [179]. These are the problems $CSP(H)$ where H has two vertices, say, 0 and 1. To describe Schaeffer's classification, we shall recall four well known operations on tuples. The *OR* operation on two tuples $(a_1, a_2, \ldots, a_s)$ and $(b_1, b_2, \ldots, b_s)$ is the tuple $(z_1, z_2, \ldots, z_s)$ where each $z_i = a_i \vee b_i$ ($z_i = 1$ unless both $a_i = b_i = 0$, in which case $z_i = 0$). The *AND* operation on two tuples $(a_1, a_2, \ldots, a_s)$ and $(b_1, b_2, \ldots, b_s)$ is the tuple $(z_1, z_2, \ldots, z_s)$ where each $z_i = a_i \wedge b_i$ ($z_i = 0$ unless both $a_i = b_i = 1$, in which case $z_i = 1$). The *MAJORITY* operation on three tuples $(a_1, a_2, \ldots, a_s)$, $(b_1, b_2, \ldots, b_s)$, and $(c_1, c_2, \ldots, c_s)$ is the tuple $(z_1, z_2, \ldots, z_s)$ where each z_i is the majority value (0 or 1) of a_i, b_i, c_i. The *XOR* (exclusive OR, also known as MINORITY) operation on three tuples $(a_1, a_2, \ldots, a_s)$, $(b_1, b_2, \ldots, b_s)$, and $(c_1, c_2, \ldots, c_s)$ is the tuple $(z_1, z_2, \ldots, z_s)$ where each z_i is the exclusive-or value of a_i, b_i, c_i (equal to 1 if the number of 1's amongst a_i, b_i, c_i is odd, and 0 otherwise). Schaeffer proved the following classification [179].

Theorem 2.2 (Boolean Dichotomy) [179] *Suppose H is a relational structure with $V(H) = \{0, 1\}$ and relations $S_1, S_2, \ldots, S_p$. Then $CSP(H)$ is NP-complete, except in the following, polynomial time solvable, cases:*

1. *each S_i contains the s_i-tuple $(0, 0, \ldots, 0)$; or*

2. *each S_i contains the s_i-tuple $(1, 1, \ldots, 1)$; or*

3. *each S_i is closed under the OR operation; or*

4. *each S_i is closed under the AND operation; or*

5. *each S_i is closed under the MAJORITY operation; or*

6. *each S_i is closed under the XOR operation.*

The polynomial time algorithms for these cases are well known. In cases 1 and 2, the core of H has one vertex, and any structure G admits a homomorphism to this core (and hence to H). Case 3 (respectively 4) corresponds to problems equivalent to the case where each S_i consists of all s_i-tuples with 1 in the first coordinate, plus possibly the tuple $(0, 0, \ldots, 0)$ (respectively all s_i-tuples with 0 in the first coordinate, plus possibly the tuple $(1, 1, \ldots, 1)$). Thus they can be expressed by *Horn clauses* (respectively *dual-Horn clauses*), i.e., disjunctions with at most one negated (respectively unnegated) variable, and solved as in [47, 111]. Case 5 corresponds to problems equivalent to H having just four binary relations, S_1 consisting of all pairs other than $(0, 0)$, S_2 consisting of all pairs other than $(0, 1)$, S_3, consisting of all pairs other than $(1, 0)$, and S_4, consisting of all pairs other than $(1, 1)$. Thus these are the problems expressible by disjunctions with two variables each, i.e., by 2-satisfiability [6]. The last case corresponds to systems of linear equations modulo two, solvable by Gaussian elimination.

The Boolean and Graph Dichotomy theorems, Theorem 2.1, and Theorem 2.2 motivated Feder and Vardi [73] to formulate the following conjecture, which remains open to this day, and motivates much research in the area.

Conjecture 2.3 (The Dichotomy Conjecture) [73] *For any relational structure H, the problem $CSP(H)$ is NP-complete or polynomial time solvable.*

By now there is strong supporting evidence for the conjecture - in the intervening years it has been verified in many cases [11, 12, 13, 23, 24, 14, 38, 40, 51, 53, 54, 55, 60, 61, 62, 63, 67, 72, 136, 140, 171], cf. [52, 106], prominently including $CSP(H)$ for structures H with up to three vertices [25], extending the Boolean Dichotomy of [179], and for conservative structures [23], discussed in Section 5.

It is important to note that dichotomy is *not* known for the case of *digraphs*, i.e., for problems $CSP(H)$ where H has only one relation, S_1, which is binary (but not necessarily

symmetric). Many results have been proved, classifying the complexity of these *digraph H-colouring problems* for special families of digraphs H [11, 14, 12, 13, 51, 67, 92]. In fact, [11, 14] conjectured in 1989, a specific dichotomy classification for digraphs without sources and sinks (i.e., with all indegrees and outdegrees positive).

Conjecture 2.4 [11, 14] *Suppose H is a core digraph with all indegrees and outdegrees positive. If each component of H is a directed cycle, the digraph H-colouring problem is polynomial time solvable; otherwise it is NP-complete.*

After nearly two decades, this conjecture has been proved [15]. (See Section 3.3.) The fact that this longstanding open problem could now be solved, testifies to the success of the algebraic method outlined in the next sections.

With vertices of indegree or outdegree zero, very little is known. For instance, dichotomy is not known if the underlying undirected graph of H is a tree, not even if it is a union of three paths meeting at one vertex - for the so called *triads H* [107].

Finally, we note that it was shown by Feder and Vardi [73] that dichotomy for digraph H-colouring problems would imply the entire Dichotomy Conjecture 2.3. Thus there is a striking difference between the H-colouring problems for graphs and for digraphs.

2.1 A Combinatorial View of NP

The Dichotomy Conjecture 2.3 appears to be an important and difficult question. In a sense, the class of problems $\mathrm{CSP}(H)$ is a largest class in which dichotomy can be expected. This was concretely formulated and proved by Feder and Vardi [73]. Recently, the formulation has been combinatorially refined by Kun and Nešetřil [129]; we now review this research. The logic class SNP ("syntactic NP") consists of all problems expressible by an existential second-order formula with a universal first-order part [49, 73, 123]. For our purposes, we view the input of the problem as a relational structure S with the vertex set $V(S) = X$ and relational symbols $R(S) = R$, and view the existentially quantified relations as "proof relations" Π. It is shown in [73] that every problem (language) L in SNP is equivalent to a formula of the form

$$\exists P \forall \overline{x} \in S \bigwedge_i \neg(\alpha_i \wedge \beta_i \wedge \varepsilon_i),$$

where

- α_i is a conjunction of atoms or negated atoms involving variables and input relations (i.e., is of the form $R(\overline{x})$ and $\neg R(\overline{x})$ for a relational symbol R and $\overline{x}$ a tuple of elements of X);

8

- β_i is a conjunction of atoms and negated atoms involving variables and existentially quantified proof relations (i.e., is of the form $P(\bar{x})$ and $\neg P(\bar{x})$ for $P \in \Pi$ and $\bar{x}$ a tuple of elements of X); and

- ε_i is the conjunction of atoms involving variables and inequalities (i.e. of form $x \neq y$).

A formula of this type is called *canonical formula* of the language L. We say that the language L is *monotone* if there are no negations in the α_i's. (In such a language, more relations lead to fewer satisfiable formulas.) The language L is *monadic* if the relations in P are all monadic (all proof relations are unary). The language L is *without inequality* if no ε_i appears in the formula.

Example: Consider the language containing one binary symbol R and two unary proof relations P_1, P_2 and the following formula

$$\exists P_1 \exists P_2 \forall x_1, x_2, x_3, y \in X$$

$$(P_1(x_1) \wedge P_1(x_2) \wedge P_1(x_3)) \vee (P_2(x_1) \wedge P_2(x_2) \wedge P_2(x_3)) \wedge$$

$$(R(x_1, x_2) \wedge R(x_1, x_3) \wedge R(x_2, x_3)) \wedge ((x_1 \neq x_2) \wedge (x_1 \neq x_3) \wedge (x_2 \neq x_3))$$

$$\wedge [\neg(\neg P_1(y) \wedge \neg P_2(y))].$$

This formula correspond to the language of all relations whose vertices can be covered by two sets is such a way that neither of these sets contains a triple linearly ordered by R. If we in addition postulate that the relation R is symmetric then these are just graphs which can be vertex partitioned into two triangle-free graphs.

Feder and Vardi proved that the subclasses of SNP with any two of these syntactical restrictions have still full computational power of the class NP. These results can be viewed as a refinement of the classical result of Fagin [49].

Theorem 2.5 [73]

1. *Every problem in NP has a polynomially equivalent problem in monotone SNP without inequality. Moreover, we may assume that the existential relations are at most binary.*

2. *Every problem in NP has a polynomially equivalent problem in monotone, monadic SNP.*

3. *Every problem in NP has a polynomially equivalent problem in monadic SNP without inequality.*

The class with all the three restrictions is called *MMSNP* (for monotone monadic SNP without inequality). Feder and Vardi [73] have shown that each problem in the class MMSNP is polynomially equivalent to a finite union of classses $CSP(D_i), i = 1, \ldots, t$, for some templates D_i. Towards this end, for $\mathcal{D} = \{D_1, \ldots, D_t\}$, we denote by $CSP(\mathcal{D})$ the union of classes $CSP(D_i), i = 1, \ldots, t$. (Thus $CSP(\mathcal{D})$ can be viewed as a generalized constraint satisfaction problem.)

Theorem 2.6 [73] *For every language L in MMSNP there is a finite set of relational structures $\mathcal{D}$ (possibly of different types) and a positive integer k such that the following hold.*

1. *L can be polynomially reduced to $CSP(\mathcal{D})$.*

2. *$CSP_{Girth_k}(\mathcal{D})$ can be polynomially reduced to L.*

The language $CSP_C(\mathcal{D})$ is the intersection of $CSP(\mathcal{D})$ with the class C; the class $Girth_k$ consists of all structures without cycles of length k or less.

In fact, the class MMSNP actually coincides with the class of all generalized constraint satisfaction problems.

Theorem 2.7 [127] *For every language L in MMSNP there is a finite set of relational structures $\mathcal{D}$ such that L and $CSP(\mathcal{D})$ are polynomially equivalent.*

This was first proved, by Feder and Vardi [73], for randomized polynomial reductions, and later on improved to polynomial reductions by Kun [127].

In [129] these results have been further refined. These refinements yields a surprising combinatorial setting for the entire class NP. The key notions in this formulation are the notions of *lift* and *shadow* of a relational structure.

Let Γ denote a finite set whose elements we refer to as colours. A Γ-*coloured graph (relational structure)* is a graph (or structure) together with either a colouring of its vertices, or a colouring of its pairs of vertices, by colours from Γ. (Only in Theorem 2.8 we shall consider colouring of all pairs - but there this will play an important role). Thus we shall understand by a coloured graph a graph with coloured vertices. We denote coloured digraphs (relational structures) by A', B' etc. Following the more general notions in category theory we call A' a *lift* of A and A is called the *shadow* of A'. (In model theory these notions are often called *expansion* and *reduct*.) Thus (vertex-) coloured digraphs (structures) can be also described as *monadic* lifts. A homomorphism of coloured digraphs (relational structures) is a homomorphism of the digraphs (structures) which also preserves the colour of vertices (pairs of vertices). A homomorphism between two (coloured) digraphs is a *full homomorphism* if the preimage of each edge is an edge, and

it is an *injective homomorphism* if distinct vertices have distinct images. Let $\mathcal{F}'$ be a finite set of coloured relational structures (digraphs). By $Forb(\mathcal{F}')$ we denote the set of all coloured relational structures (digraphs) A' satisfying $F' \not\longrightarrow A'$ for every $F' \in \mathcal{F}'$. (If we restrict to injective or full homomorphisms this will be denoted by $Forb_{inj}(\mathcal{F}')$, or $Forb_{full}(\mathcal{F}')$, respectively.)

Let A' be a colored structure. We denote by $\Phi(A')$ the natural *forgetful functor* that "forgets" colors in A'. If $\mathcal{K}$ is a class of colored structures then $\Phi(\mathcal{K}')$ denotes the class of all $\Phi(A)$, $A' \in \mathcal{K}'$.

Theorem 2.8 [129] *For every language $L \in NP$ there exist a finite set of colours Γ and a finite set of Γ-coloured digraphs $\mathcal{F}'$, where we colour all pairs of vertices, such that L is computationally equivalent to the membership problem for $\Phi(Forb(\mathcal{F}'))$.*

Theorem 2.9 [129] *For every language $L \in NP$ there exist a finite set of colours Γ and a finite set of Γ-coloured digraphs $\mathcal{F}'$ (where we colour the vertices), such that L is computationally equivalent to the membership problem for $\Phi(Forb_{inj}(\mathcal{F}'))$.*

Theorem 2.10 [129] *For every language $L \in NP$ there exist a finite set of colours Γ and a finite set of Γ-coloured digraphs $\mathcal{F}'$ (where we colour the vertices), such that L is computationally equivalent to the membership problem for $\Phi(Forb_{full}(\mathcal{F}'))$.*

Those monotone properties P of structures which can be described by finitely many forbidden substructures are of particular interest. This will be treated in this paper in a section devoted to dualities. Let us mention here that it has been proved recently by B. Rossman [175] that a homomorphism monotone problem is *first order definable* (FO definable) if and only if it is *positively* FO definable (FO+ definable), i.e., if the formula does not contain any negations (and so implications and inequalitites), and thus alternatively defined as Forb($\mathcal{F}$) for a finite set $\mathcal{F}$ of structures. Although FO-definability is not a rare fact (and extremely useful in database theory), still FO-definability cannot express most combinatorial problems (compare [168],[5] which characterize all CSP which are FO-definable; see also Theorem 4.4). Thus it may seem to be surprising that the classes of relational structures defined by existential second order formulas (i.e. the entire class NP) corresponds exactly to those canonical *lifts* of structures which are defined by a finite set of forbidden substructures. In other words, the finite sets of forbidden lifts determine all languages in NP.

It is interesting to note how nicely these results fit the combinatorial common sense about the difficulty of problems. On the one side, the problems in CSP correspond to and generalize ordinary (vertex) coloring problems. One expects a dichotomy here: every CSP problem should be either polynomial-time solvable or NP-complete. On the other side, the above formulations (Theorems 2.8,2.9,2.10) model the entire class NP, where we cannot

expect dichotomy. These facts are consistent with the combinatorial meaning of these classes. The formulation in Theorem 2.8 expresses coloring of edges, triples etc., and thus it involves Ramsey theory [93, 151]. We have already mentioned above, as an example, how to interpret the problem whether a graph can be partitioned into two triangle free graphs.

The formulation in Theorem 2.9, by means of injective forbidden substructures, allows one to express classes of graphs with bounded degrees, or with given girth. The coloring problems for such graph classes represent well known hard (and classical) combinatorial problems, see for example [104, 110, 169]. The third formulation, in Theorem 2.10, by means of full homomorphisms, relates to vertex partitions with a given pattern among classes. We can express not only the existence of edges but also the non-existence of non-edges between classes in the partition. Such partitions have been investigated for instance in [56, 57, 65], and play an important role in many graph decomposition techniques (such as those involved in the solution of the Perfect Graph Conjecture [33]). This point of view is further explored in Section 7.1. This clear distinction between combinatorial interpretations of syntactic restrictions on formulas expressing the computational power of NP is one of the pleasing consequences of this approach.

2.2 Polymorphisms

The greatest progress on the dichotomy conjecture resulted from an algebraic approach pioneered by Jeavons [117], based on the Galois correspondence in [18, 82]. It turns out that what determines whether a structure H has a hard or easy constraint satisfaction problem $CSP(H)$, is its set of polymorphisms. A *polymorphism* (of order, or arity, k) of H is a mapping $f : V(H^k) \to V(H)$, such that $(v_1^j, v_2^j, \ldots, v_{r_i}^j) \in S_i$ for $j = 1, 2, \ldots, k$ implies that

$$(f(v_1^1, v_1^2, \ldots, v_1^k), f(v_2^1, v_2^2, \ldots, v_2^k), \ldots, f(v_{r_i}^1, v_{r_i}^2, \ldots, v_{r_i}^k)) \in S_i,$$

for all relations S_i of H. (Thus a polymorphism of order k of H is just a homomorphism of a suitably defined direct product power H^k to H [172].) Denoting by $\mathcal{P}ol(H)$ the set of all polymorphisms of H, Jeavons [117] observed the following fact.

Theorem 2.11 [117] *Assume $V(H) = V(H')$. If $\mathcal{P}ol(H') \subseteq \mathcal{P}ol(H)$, then $CSP(H)$ is polynomially reducible to $CSP(H')$.*

In particular, if H and H' have exactly the same set of polymorphisms (of all orders), then the constraint satisfaction problems $CSP(H)$ and $CSP(H')$ are polynomially equivalent, regardless of the relations H and H' may have. The structures H and H' need not even have the same type.

These facts have many consequences in the complexity analysis of CSP and establish the basis of what is known as the *algebraic method*. Consider, for instance, the Boolean

Dichotomy classification, Theorem 2.2, above. We have stated the result in a form which emphasizes that what makes the problems $\mathrm{CSP}(H)$ polynomial time solvable, is the existence of certain operations under which all the relations of H are closed. Each of these operations corresponds to a polymorphism of H, as we shall discuss below.

Thus structures H having many polymorphisms are likely to have polynomial time solvable problems $\mathrm{CSP}(H)$, and structures H which have few polymorphisms can be expected to have NP-complete problems $\mathrm{CSP}(H)$. Note that every structure H admits some polymorphisms - at least the projections π_i taking $(v_1, v_2, \ldots, v_k)$ to v_i. If the structure H has any automorphisms, i.e., bijective homomorphisms f of H to itself, then any composition $f \circ \pi_i$ is also a polymorphism of H. A structure H (with more than one vertex) which has no other polymorphisms is called *projective*. It is worth noting that our definition of projectivity implies that the structure is a core, because we apply the requirement also to polymorphisms of order one.

Let us return to the problem of graph k-colourability, which we denote by $\mathrm{CSP}(K_k)$. (The complete graph K_k is the structure H with $|V(H)| = k$ and with one symmetric binary relation, of non-equality. Note that any permutation of the vertices of K_k is an automorphism of K_k.)

Theorem 2.12 *If $k \geq 3$, then K_k is projective.*

This result has an interesting history. It was discovered independently by Greenwell and L. Lovász [84] and V. Müller [147, 148] in the context of uniquely colourable graphs; we have also been informed by A. Blass that G. Lawvere found a related result in the context of the theory of ultrafilters. The result has been reproved in many different contexts, most notably using Fourier analysis [3], economic theory [121], and analysis [46]. These results lead directly to one recent alternative proof of the Graph Dichotomy, Theorem 2.1 [128].

Recall that $\mathrm{CSP}(K_k)$ is NP-complete when $k \geq 3$. It follows from Theorem 2.11 and Theorem 2.12 that if a structure H with at least three vertices is projective, then the problem $\mathrm{CSP}(H)$ is NP-complete. In fact, this holds for all projective structures.

Theorem 2.13 [117, 118] *If a structure H is projective, then the problem $CSP(H)$ is NP-complete.*

Consider next the structure N which has two vertices $0, 1$ and one ternary relation $E(N) = \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 0), (1, 0, 1), (0, 1, 1)\}$. The following folklore result, cf. [179], makes a nice exercise [106].

Theorem 2.14 *The structure N is projective.*

Corollary 2.15 *The problem $CSP(N)$ is NP-complete.*

The problem $CSP(N)$ takes as input a structure G with triples (u, v, w); its vertices must be assigned values $0, 1$ so that each triple receives at least one 0 and at least one 1. This is the problem of Not-All-Equal Three-Satisfiability Without Negated Variables, another standard NP-complete problem [81].

From all available evidence one may guess that if a problem $CSP(H)$ cannot be shown NP-complete by an (iterated) reduction from $CSP(K_k)$ or $CSP(N)$, then $CSP(H)$ is polynomial time solvable. The concrete classifications are related to this intuition. On the other hand, it was shown by Łuczak and Nešetřil [140] that almost all structures are projective, and hence have NP-complete problems $CSP(H)$. So, in this sense, the Dichotomy Conjecture is asymptotically true (and even asymptotically trivial).

3 Classification

The original dichotomy conjecture, Conjecture 2.3, dates back to a time when the motivation and evidence for it was based mostly on examples, in particular, the Graph and Boolean Dichotomy theorems, i.e., Theorem 2.1 and Theorem 2.2. There seemed to be no candidate classification of the problems into NP-complete and polynomial time solvable; in fact, evidence from graph theory [11, 13, 14] was quite discouraging as far as a possible classification was concerned. However, recent intensive efforts have lead to some concrete propsals and explicit classification conjectures, refining the original dichotomy conjecture 2.3. Remarkably, these efforts have occured in very different fields - in universal algebra, in analysis (Fourier analysis in particular), in complexity theory, and in combinatorics. We present these approaches in this section. As we shall see (Theorem 3.19 all these approaches turn out to be equivalent, in the sense that they propose the same classification. This remarkable fact seems to lend credence to the these conjectures, especially in view of the varied backgrounds they come from.

Let us return to the idea that polymorphisms of H tend to make the problem $CSP(H)$ polynomial time solvable. For the purposes of this discussion, we shall focus on polymorphisms f that are *idempotent*, i.e., satisfy $f(x, x, \ldots, x) = x$ for all vertices $x \in V(H)$. In fact, we may restrict our attention to *idempotent structures*, i.e., structures that only admit idempotent polymorphisms. Indeed, there is a simple construction to make an arbitrary core relational structure H into an equivalent idempotent structure H', namely adding, for each vertex v of H, the unary relation R_v containing the single 1-tuple (v). It is clear that H' is idempotent; in particular, H' is a core, and admits no automorphisms other than identity. Clearly, an idempotent structure is projective if and only if it admits no polymorphism other than a projection. Dealing with idempotent structures is more convenient; we are able to restrict our attention to them because of the following result. (See [106] exercise 2 in chapter 5; see also [29].)

Proposition 3.1 *Let H be a core relational structure. Then the problem CSP(H) is polynomially equivalent to the problem CSP(H').*

Suppose H is an idempotent structure. Theorem 2.13 says that if H is projective (admits only projections as polymorphisms), then $\mathrm{CSP}(H)$ is NP-complete. Admitting any polymorphism which is not a projection is, in some cases, sufficient to ensure that $\mathrm{CSP}(H)$ is polynomial time solvable. This is the case, for instance, when H has only two vertices, as can be checked from the Boolean Dichotomy classification in Theorem 2.2. However, it is not sufficient for all relational structures H: there are idempotent structures H which admit a polymorphism other than projection, yet yield an NP-complete problem $\mathrm{CSP}(H)$ [26]. Nevertheless, there are several natural extension of Theorem 2.13 to weakened version of projectivity.

3.1   The Algebra of H

Recall that Theorem 2.11 implies that two structures H, H' on the same vertex set with exactly the same polymorphisms, i.e., satisfying $\mathcal{P}ol(H) = \mathcal{P}ol(H')$, have constraint satisfaction problems $\mathrm{CSP}(H)$, $\mathrm{CSP}(H')$ of the same complexity. Hence it makes sense to investigate the algebra consisting of the vertex set $V(H)$ and the set of polymorphisms $\mathcal{P}ol(H)$, abstracting away the structure H itself. A finite *algebra* is a pair $(V, \mathcal{F})$, where V is a finite set, and $\mathcal{F}$ a set of finitary operations on V. A *subalgebra* of $(V, \mathcal{F}$ is, as would be expected, defined as an algebra $(V', \mathcal{F}')$ where $V' \subseteq V$, such that the results of all operations in $\mathcal{F}$ restricted to V' are in V', and $\mathcal{F}'$ consists of the restrictions of the operations in $\mathcal{F}$ to V'. A *homomorphic image* of $(V, \mathcal{F}$ is an algebra obtained from $(V, \mathcal{F}$ by identifying certain pairs of elements and defining the operations in the natural way [29]. A *divisor* of $(V, \mathcal{F}$ is a homomorphic image of a subalgebra of $(V, \mathcal{F}$. An algebra $(V, \mathcal{F})$ is non-trivial if $|V| > 1$.

The *algebra of* a structure H, denoted $\mathcal{A}(H)$, is $(V(H), \mathcal{P}ol(H))$. As noted above, Theorem 2.11 implies that the complexity of $\mathrm{CSP}(H)$ is the same for all structures H with the same algebra.

Theorem 3.2 [29] *Suppose H is an idempotent relational structure such that for some non-trivial divisor $(V', \mathcal{F}')$ of $\mathcal{A}(H)$, all operations in $\mathcal{F}'$ are projections. Then the problem CSP(H) is NP-complete.*

This result covers all known NP-complete cases of $\mathrm{CSP}(H)$. Accordingly, Bulatov, Jeavons and Krokhin have refined the dichotomy conjecture as follows:

Conjecture 3.3 [29] *If H is an idempotent relational structure such that each non-trivial divisor $(V', \mathcal{F}')$ of $\mathcal{A}(H)$ has an operation in $\mathcal{F}'$ which is not a projection, then the problem CSP(H) is polynomial-time solvable.*

Bulatov [26] has used this approach to re-derive the Graph Dichotomy theorem, Theorem 2.1. In his proof, he avoids much of the details in [104] by finding a suitable divisor of $\mathcal{A}(H)$ (for a non-bipartite H without loops) to which Theorem 3.2 applies.

3.2 Taylor Polymorphisms

Another way to formulate the weakened projectivity property was presented by Larose and Zádori [135]. We say that a polymorphism f is *inclusive in position i*, if it satisfies an identity involving two variables, with different entries in position i. More precisely, there exist choices $u_j, v_j \in \{u, v\}, j = 1, 2, \ldots, k$, with $u_i \neq v_i$, such that the identity

$$f(u_1, u_2, \ldots, u_k) = f(v_1, v_2, \ldots, v_k)$$

holds for all $u, v \in V(H)$. Polymorphisms inclusive in each position are called *Taylor polymorphisms*. Clearly, the i-th projection is not inclusive in position i, so projections are not Taylor polymorphisms, and, in fact, a Taylor polymorphism is just what was sought - a polymorphism far enough from a projection.

Theorem 3.4 [135] *Suppose H is an idempotent relational structure. If H does not admit a Taylor polymorphism, then the problem $CSP(H)$ is NP-complete.*

As with the previous conjecture, this seems to cover all known NP-complete cases, and so a refined dichotomy conjecture would look as follows.

Conjecture 3.5 [29] *Suppose H is an idempotent relational structure. If H admits a Taylor polymorphism, then the problem $CSP(H)$ is polynomial time solvable.*

For instance, we consider the following special Taylor polymorphisms that are often helpful. A *near-unanimity polymorphism* is a polymorphism f of order at least three which has the *near-unanimity property*, namely that

$$f(u, u, \ldots, u, v) = f(u, u, \ldots, v, u) = \cdots = f(v, u \ldots, u, u) = u,$$

for all vertices u, v. In particular, a *majority polymorphism* is a polymorphism f of order three satisfying $f(u, u, v) = f(u, v, u) = f(v, u, u) = u$ for all u and v. Clearly, any near-unanimity polymorphism is inclusive in each position. Near-unanimity polymorphisms are fairly common [134, 22]. For instance, if H is a digraph (structure with one binary relation) whose underlying undirected graph is a path, then H admits a simple majority polymorphism - namely $f(u, v, w)$ being the middle of the vertices u, v, w on the path. It is a simple exercise to verify that this definition (which clearly satisfies the majority property) yields a polymorphism of H; we simply check that in all cases when uu', vv', ww'

16

are arcs of H, the middle vertices $f(u,v,w)f(u',v',w')$ also form an arc of H. For another example, consider the MAJORITY operation on Boolean tuples, defined just above Theorem 2.2. It is easy to check that MAJORITY defines a polymorphism (and hence a majority polymorphism) on H if and only if each S_i is closed under MAJORITY. The following result goes back to at least 1993 [73]; cf. also [118].

Theorem 3.6 [73] *If H admits a near-unanimity polymorphism, then $CSP(H)$ is polynomial time solvable.*

According to the above example, we have the following non-trivial consequence.

Corollary 3.7 [92] *If H is a digraph such that the underlying undirected graph of H is a path, then the H-colouring problem is polynomial time solvable.*

Before leaving this topic, we shall mention two other kinds of Taylor polymorphisms that are known to imply polynomial time algorithms. A *Maltsev polymorphism* is a polymorphism f of order three which satisfies $f(u,u,v) = f(v,u,u) = v$ for all vertices u,v. A *semilattice polymorphism* is a polymorphism f of order two which satisfies $f(u,u) = u, f(u,v) = f(v,u)$, and $f(a,f(b,c)) = f(f(a,b),c)$, for all vertices u,v. A semilattice polymorphism is clearly inclusive in each position; to see that a Maltsev polymorphism is inclusive in the second position consider the identity $f(u,u,v) = f(v,v,v)$. Recall our Boolean polymorphisms OR, AND, XOR, from Theorem 2.2. If each relation S_i in H is closed under OR (respectively AND, respectively XOR), these operations define a polymorphism of H, which is semilattice for OR, AND and a Maltsev operation for XOR. It is known that the existence of either a Maltsev or a semilattice polymorphism on a structure H ensures that $CSP(H)$ is polynomial time solvable [24, 27]. There is a recent common generalization of structures with a Maltsev polymorphism and structures with a near-unanimity polymorphism [39, 17, 115] - these are structures whose algebra is said to have 'few subpowers' [17, 115].

3.3   Weak Near-Unanimity Polymorphisms

The following generalization of near-unanimity polymorphisms has recently been proved to be particularly fruitful.

A *weak near-unanimity polymorphism* is an idempotent polymorphism which satisfies

$$f(u,u,\ldots,u,v) = f(u,u,\ldots,v,u) = \cdots = f(v,u\ldots,u,u),$$

for all vertices u,v.

Even though weak near-unanimity polymorphisms seem only slightly weaker than near-unanimity polymorphisms, it was recently proved by Mároti and McKenzie that not having one implies intractability.

Theorem 3.8 [142] *If H admits no weak near-unanimity polymorphism, then $CSP(H)$ is NP-complete.*

This suggests another very attractive form of a possible dichotomy classification (to be compared with Theorem 3.6.

Conjecture 3.9 [142] *If H admits a weak near-unanimity polymorphism, then $CSP(H)$ is polynomial time solvable.*

Theorem 3.8 has important applications, for instance, to the proof of Conjecture 2.4.

Theorem 3.10 [15] *Suppose H is a core digraph with all indegrees and outdegrees positive. If each component of H is a directed cycle, then the digraph H-colouring problem is polynomial time solvable; otherwise it is NP-complete.*

The authors have shown that if H has a component which is not a directed cycle, then it cannot have a weak near-unanimity polymorphism.

3.4   Fibre Gadgets

We now turn to a discussion of combinatorial approaches, which are simple enough to describe in full.

This is the first combinatorial approach, and perhaps the simplest one. It proves the NP-completeness of a concrete problem $CSP(H)$ directly, by constructing a single gadget for a reduction from 3-colourability. The gadget itself of course depends on H, but otherwise the construction is simple and uniform.

We will often define sets of indexed vertices such as $W^* = \{w_1^*, \ldots, w_d^*\}$. A copy W^a of the set W^* will mean the set $W^a = \{w_1^a, \ldots, w_d^a\}$. Given two copies W^a and W^b of the same set W^* we say that we *identify W^a and W^b index-wise* to mean we identify the vertices w_i^a and w_i^b for $i = 1, \ldots, d$. When we define a function f on W^*, we will assume it to be defined on any copy W^a of W^* by $f(w_\alpha^a) = f(w_\alpha^*)$ for all $\alpha = 1, \ldots, d$. We refer to a function f on an set W^* as a *pattern* of W^*. In the case that the image of f is contained in the vertex set of some structure H we speak about *H-pattern* of W^*. We will often describe H-patterns of W^* explicitly as vectors of elements of H. For example:

$$f(\{w_1^*, w_2^*, w_3^*, w_4^*\}) = (h, h, h', h),$$

for $h, h' \in V(H)$. Moreover, patterns will often be denoted by capital letters such as P, so one will see P as a function.

Let H be a core relational structure. An instance M of $CSP(H)$ is called a K_3-*partition* if $V(M)$ contains two disjoint copies W^1 and W^2 of some set W^* of indexed vertices, and there are three disjoint sets $\P_1, \P_2, \P_3$ of H-patterns of W^* such that the following properties are met.

1. Under every H-colouring ϕ of M, $\phi|_{W^1}$ and $\phi|_{W^2}$ are in different sets in $\{\P_1, \P_2, \P_3\}$.

2. There are representative H-patterns P_1, P_2, and P_3 of $\P_1, \P_2$, and $\P_3$ respectively such that for every choice of $i \neq j \in \{1, 2, 3\}$ there is an H-colouring ϕ_{ij} of M for which $\phi_{ij}|_{W^1} = P_i$ and $\phi_{ij}|_{W^2} = P_j$.

If H has a K_3-partition, it is called K_3-*partitionable*.

The following theorem and corollary are the main results of [165].

Theorem 3.11 (Fibre construction) [165] *Suppose H is a core relational structure. If H has a K_3-partition, then $CSP(H)$ is NP-complete.*

Proof. Let M be a K_3-partition of H, and let G be any instance of $CSP(K_3)$. We construct an instance $M(G)$ of $CSP(H)$ as follows.

1. For each vertex v of G let W^v be a copy of W^*.

2. For each edge $e = \{u, v\}$ of G let M^e be a copy of M. Identify, index-wise, W^u and W^v with W^1 and W^2 of M^e respectively.

Thus $M(G)$ consists of $|V(G)|$ copies of W^* and $|E(G)|$ copies of M. All vertices are distinct unless identified above.

Observe that this is not the amalgamation of M and G which is often used in indicator constructions. In fact, for a given copy W^v of W^* in $M(G)$ that has been identified with copies of W^1 and W^2 from different copies of M, W^v induces the union of the edges induced by W^1 and W^2.

We now show that $G \to K_3 \iff M(G) \to H$. Since the construction of $M(G)$ is polynomial in $|V(G)|$, this will prove the theorem. $\square$

Accordingly, in this case, the complementary conjecture would take the following form.

Conjecture 3.12 [165] *Suppose H is a core relational structure. If H has no K_3-partition, then CSP(H) is polynomial time solvable.*

In support of such a daring conjecture, let us at least indicate why a projective relational structure H admits a K_3-partition [165]. We may assume, without loss of generality (see Theorem 3.1) that H has only the identity automorphism, and let a, b be distinct vertices of H. In the structure H^6, we consider the vertex sets

$$W^1 = \{(a, a, b, b, b, b), (b, b, a, a, b, b), (b, b, b, b, a, a)\}$$

$$W^2 = \{(b, b, a, b, a, b), (a, b, b, b, b, a), (b, a, b, a, b, b)\},$$

and the H-patterns

$$P_1 = (0, 1, 1), P_2 = (0, 1, 0), P_3 = (0, 0, 1).$$

One can check that any homomorphism $f : H^6 \to H$ when restricted to any of the sets W^i coincides with a pattern $P_{j(i)}$. Furthermore, projectivity implies that $j(i) \neq j(i')$ for $i \neq i'$. Indeed, we know that f is a projection; if, say, f is the projection on the third coordinate, then $f|W^1 = (b, a, b)$, while $f|W^2 = (a, b, b)$. On the other hand, the sets W^1 and W^2 have been constructed so that for any patterns $P_j, P_{j'}$ with $j \neq j'$, there is a homomorphism (in fact, a projection) $f : H^6 \to H$ which has pattern P_j on W^1 and $P_{j'}$ on W^2. Thus any projective structure admits a K_3-partition.

In a remarkable twist, M. Siggers [187] has very recently succeeded in modifying this proof to imply Theorem 2.1. This is at present the simplest proof of the undirected graph dichotomy [104].

We also note here that this K_3-partition construction is motivated by an elegant result of V. Müller [147, 148], which we state here as given in [106].

Theorem 3.13 (Müller's Extension Theorem) [147, 148]

Let k, ℓ, t be positive integers, $k > 2$.

Let $\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_t$ be distinct partitions of a finite set A, each with k (possibly empty) parts. Then there exists a graph G of girth at least ℓ and chromatic number k, such that

- *A is a subset of $V(G)$,*

- *G has precisely t k-colourings $c_1, c_2, \cdots, c_t$, and*

- *the partition associated with the k-colouring c_i, restricted to the set A, is precisely $\mathcal{A}_i$.*

We observe that for $t = 1$, Müller's Extension Theorem yields uniquely colourable graphs without short cycles, answering a question posed by Erdos. The theorem has several important applications, and has since been generalized to H-colourings, being the basis of the so-called Sparse Incomparability Lemma [163, 106]. The validity of the Sparse Incomparability Lemma was characterized on a general level by Nešetřil and Zhu [169]. The characterization is related to projectivity.

This combinatorial approach to the dichotomy has several pleasing consequences. The reduction implicit in the concept of K_3-partitionability is easily seen to preserve bounded degrees, and can also be shown to preserve large girth. Thus we were able to derive from the theorem a proof of a metaconjecture of Feder, Hell, and Huang [62], which states that all homomorphism problems which are NP-complete for graphs in general, remain NP-complete for graphs with bounded degrees, as long as the bound is high enough. (This also follows from the results of [127].) One can also derive results related to the problem of Kostochka, Nešetřil, and Smolíková, on problems for graphs with large girth [124].

3.5 Block Projectivity

This constitutes the second combinatorial approach to dichotomy classification. As projectivity alone does not fully capture all NP-complete instances, the authors of [165] first introduced a new notion of *subprojectivity* as a candidate classifier. Since this turned out still not to be sufficient [141], the authors extended it [164] to the following notion of block projectivity.

A pair a, b of vertices of H is *block projective* if there exist disjoint sets $H_a, H_b \subset V(H)$ (called *blocks*), such that the following is true. For any polymorphism $\phi : H^d \to H$ of H, there is an integer $i, 1 \leq i \leq d$, such that

$$\phi\left((s_1, \ldots, s_d)\right) \in H_{s_i},$$

for any $(s_1, \ldots s_d) \in \{a, b\}^d$.

A relational system is *block projective* if it is a core and contains a block projective pair.

One can prove the following result.

Theorem 3.14 [164] *If a structure H is block projective, then CSP(H) is NP-complete.*

This in turn leads to the following conjecture.

Conjecture 3.15 [164] *If a structure H is not block projective, then $CSP(H)$ is polynomial time solvable.*

We note that it is known [139] that almost all large structures are projective and hence block projective. However, to prove the block projectivity of a particular structure H is another matter. For non-bipartite graphs, this was recently achieved in [187], as a byproduct of a recent new proof of Graph Dichotomy, Theorem 2.1.

3.6 An Analytic Approach

Recently, a new approach to the CSP dichotomy was proposed by G. Kun and M. Szegedy. In these notes, we only offer a hint of the main ingredients of this new approach, based on dynamic systems and probability measures, in a way resembling the Fourier analytic techniques used in the PCP theorem [4, 45, 174].

Let X be a finite set, k a positive integer, and f a function $X^k \to X$. Iterated functions $f^{(i)}$ are defined recursively by $f^{(1)} = f$ and

$$f^{(i+1)} = f^{(i)}(f, f, \ldots, f).$$

Given two probability measures μ and ν on X, the *variational distance* between μ and ν is defined as

$$\delta(\mu, \nu) = \frac{1}{2} \sum_{x \in X} |\mu(x) - \nu(x)|.$$

If $\mu_1, \mu_2, \ldots, \mu_k$ are probability measures on X, we denote by $f(\mu_1, \mu_2, \ldots, \mu_k)$ the measure on X^k defined by the composition of f and the μ_i's. The key notion of the Kun-Szegedy approach is the notion of asymptotic resilience [128].

A function $f : X^k \to X$ is *asymptotically resilient* if for every measure μ on X, positive integer ℓ, and positive ϵ, there exists a constant i_0, such that for all $i \geq i_0$ we have

$$\delta(f^{(i)}(\mu, \mu, \ldots, \mu), f^{(i)}(\mu_1, \mu_2, \ldots, \mu_k)) < \epsilon,$$

for any measures $\mu_1, \mu_2, \ldots, \mu_k$ with support contained in the support of μ, such that at most ℓ of them are different from μ.

Theorem 3.16 [128] *A polymorphism is a weak near unanimity polymorphism if and only if it is asymptotically resilient.*

Corollary 3.17 *If H does not admit an asymptotically resilient polymorphism, then $CSP(H)$ is NP-complete.*

Accordingly, we obtain the following conjecture.

Conjecture 3.18 [128] *If H admits an asymptotically resilient polymorphism, then CSP(H) is polynomial time solvable.*

G. Kun and M. Szegedy also re-derive Theorem 2.1, illustrating their techniques; specifically, they prove that a non-bipartite graph without loops does not admit an asymptotically resilient polymorphism. This is a new and very different approach, in which they apply the characterizations of large independent sets in graph products, due to Dinur, Friedgut, and Regev [46]. In this sense, this approach is also related to a weaker form of projectivity; of course the details of [128, 46] are much more involved. As a whole this represents a very elegant and original approach to the dichotomy problem, and one building on the experience of other developments in theoretical computer science, long code testing, PCP theorem, and dynamical systems.

3.7 Equivalence of Conjectures

Surprizingly, all the various dichotomy conjectures turn out to be mutually equivalent.

Theorem 3.19 *For an idempotent relational structure H, the following statements are equivalent:*

- *the algebra $(V(H), \mathcal{P}ol(H))$ has a non-trivial divisor in which all operations are projections;*

- *H has no Taylor polymorphism;*

- *H has no weak near unanimity polymorphism;*

- *H is a block projective relational structure;*

- *H is K_3-partitionable;*

- *H has no asymptotically resilient polymorphism.*

In these cases $CSP(H)$ is NP-complete.

This easily stated theorem is a culmination of several results.

In [28], Bulatov and Jeavons show that an idempotent algebra **B** has a non-trivial divisor all of whose term operations are projective if and only if, in the language of tame congruence theory, the variety generated by **B** admits type one. The relation to Taylor operations is developed in [135]. In [142], it is shown that this is true if and only if **B** admits no weak unanimity term operation. The equivalence of the asymptotic resilience is proved in [128].

23

Perhaps most surprisingly, the equivalence of these algebraic notions with the existence of K_3-partitions and block projectivity is proved recently in [164].

Corollary 3.20 *All the conjectures 3.3, 3.5, 3.9, 3.12, 3.15, and 3.18 are equivalent.*

The fact that such a variety of different formulations all turn out to be equivalent, lends clear support to the current version of the dichotomy conjecture, which is simply that all other problems $\mathrm{CSP}(H)$ are polynomial time solvable.

4   Duality

Consider the problem $\mathrm{CSP}(H)$. What can be better than a characterization by a finite set of obstructions? Surely having such a characterization is a desirable goal, but is it realistic, and with interesting instances? It is reasonable to be skeptical, as for undirected graphs, the answer is negative [162], i.e., there are only trivial instances of finite obstruction theorems.

However the properties characterized by a finite set $\mathcal{F}$ of obstructions are very interesting if we consider them for more complicated structures than undirected graphs only. The evidence from Section 2.1 shows that by means of lifts and shadows, forbidden homomorphisms from finitely many structures suffice to capture the entire class NP. In this section we complement this by considering CSP problems defined by finitely many obstructions. Towards this end we define the notion of *finite (homomorphism) duality*.

Let $\mathcal{F}$, $\mathcal{D}$ be finite sets of structures (of type t). We say that sets $\mathcal{F}$ and $\mathcal{D}$ establish *finite duality* if the following holds for every structure A of type t.

$$F \not\longrightarrow A \text{ for every } F \in \mathcal{F} \Longleftrightarrow A \longrightarrow D \text{ for some } D \in \mathcal{D}.$$

The term duality refers to the fact that we are describing the structures which admit homomorphisms to H dually, by means of forbidden subobjects. This is particularly fitting in the language of category theory [106, 150].

In this case we say that $(\mathcal{F}, \mathcal{D})$ is *dual pair*, and that $\mathcal{D}$ is *dual set* of $\mathcal{F}$.

The simplest non-trivial instance of dualities is for oriented graphs and it is usually expressed in terms of orientations of graphs. The connection between chromatic number and orientations goes back to Gallai and Roy [79, 176]. These pioneering works provided a name for the result although both of these papers were anticipated by M. Hasse [94] and L. M. Vitaver [177], where the same result is proved (in the more algebraic language). For our purposes the Gallai-Hasse-Roy-Vitaver result takes the following compact form.

24

Theorem 4.1 *For any directed graph G the following holds:*

$$P_k \nrightarrow G \Longleftrightarrow G \to T_k$$

Here P_k denotes the directed path of length k (i.e. with $k+1$ vertices), and T_k denotes the transitive tournament with k vertices.

It may be seen easily that for undirected graph this has the following consequence.

Corollary 4.2 [103, 106] *For an undirected graph G the following statements are equivalent:*

1. *$\chi(G) \leq k$ (which is equivalent to $G \to K_k$);*

2. *There exists an orientation $\vec{G}$ of G such that $\vec{G} \to T_k$;*

3. *There exists an orientation $\vec{G}$ of G such that $P_k \nrightarrow \vec{G}$.*

This particular result was one of the starting points ([162]) for the following result which characterizes homomorphisms duality in classes [168]. First we state the theorem for singleton sets only:

Theorem 4.3 (Singleton Homomorphism Dualities) [168]

1. *For every relational tree T there exists a structure D_T (called the* dual *of T) such that the following holds (for every structure A):*

$$T \nrightarrow A \Longleftrightarrow A \to D_T.$$

2. *Up to a homomorphism equivalence there are no other dual pairs (of singleton structures).*

The characterization of singleton dualities is the basis of the characterization of dual pairs of sets and of finite dualities:

Theorem 4.4 (Finite Homomorphism Dualities) [168, 77]

1. *For every finite set of relational trees $\mathcal{T}$ there exists a dual set of structures $\mathcal{D}_\mathcal{T}$ such that the following holds (for every structure A):*

$$\mathcal{T} \nrightarrow A \Longleftrightarrow A \to \mathcal{D}_\mathcal{T}.$$

2. *Up to a homomorphism equivalence there are no other dual pairs (of singleton structures).*

(Here we write $\mathcal{T} \not\to A$ if $T \not\to A$ for every $T \in \mathcal{T}$. Similarly, we write $A \to \mathcal{D}_\mathcal{T}$ if $A \to D$ for every $D \in \mathcal{D}_\mathcal{T}$.)

These theorems are nontrivial in both directions and the existence of duals is nontrivial. In a certain sense the duals are well understood. This can be illustrated by the fact that there are several constructions of dual objects by means of

- explicit construction, [166, 167];

- construction using the connection between gaps (in the homomoprhism order) and dualities, [168];

- deletion method using lifts and shadows [152];

- generic method using shadows of universal structures, [112].

Let us remark that for a tree T the dual structure D_T may have a size exponential in the size of T [166, 167], yet to decide whether a given core graph G is a dual of a tree T can be decided polynomially, [133].

It has been also proved that finite dualities correspond exactly to those Constraint Satisfaction Problems which are first order (FO) definable. This follows by a combination of results [168, 77] and [5] (and also [175]). For further reading on this subject we refer to [152, 30].

4.1   Restricted Duality

Finite dualities correspond to first order CSP's. Finite dualities become much more abundant if we demand the validity of the duality formula just for all graphs from a given class $\mathcal{K}$. In such a case we speak about $\mathcal{K}$-*restricted duality*, defined as follows. We say a class of structures $\mathcal{K}$ admits *all restricted dualities* if, for any finite set of connected graphs $\mathcal{F} = \{F_1, F_2, \ldots, F_t\}$, there exists a finite structure $D_\mathcal{F}^\mathcal{K}$ such that $F_i \not\longrightarrow D_\mathcal{F}^\mathcal{K}$ for $i = 1, \ldots, t$ and for all $G \in \mathcal{K}$,

$$(F_i \not\longrightarrow G), i = 1, 2, \ldots, t, \iff (G \longrightarrow D_\mathcal{F}^\mathcal{K}). \tag{1}$$

Any instance of (1) is called a *restricted duality* (for the class $\mathcal{K}$).

To motivate this definition let us consider the following example.

The Grötzsch's celebrated theorem (see e.g. [189]) says that every triangle-free planar graph is 3-colorable. In the language of homomorphisms this says that for every triangle-free planar graph G there is a homomorphism of G into K_3. Using the partial order terminology, Grötzsch's theorem says that K_3 is an upper bound (in the homomorphism

order) for the class $\mathcal{P}_3$ of all planar triangle-free graphs. The fact that $K_3 \notin \mathcal{P}_3$ suggests a natural question (first formulated in [150]): is there yet a smaller bound?

The answer, which may be viewed as a strengthening of Grötzsch's theorem, is positive [154],[155]: there exists a triangle free 3-colorable graph H such that $G \longrightarrow H$ for every graph $G \in \mathcal{P}_3$. One can view these results as restricted dualities (which hold in the class of planar graphs). Restricted duality results have since been generalized not only to proper minor closed classes of graphs and but also to other forbidden subgraphs, in fact to any finite set of connected graphs thus yielding all restricted dualities for the class of planar graphs[156]. This then implies that Grötzsch's theorem can be strengthened by a sequence of even stronger bounds and that the supremum (in the homomorphism order) of the class of all triangle free planar graphs does not exist, [153].

What is the proper setting for the restricted dualities? This is presently an open problem but the strongest result in this direction is the notion of a class with bounded expansion. Such a class may be defined in several (very) different ways, see [158, 159, 160]. Here we selected the perhaps the most intuitive definition (and chronologically the first definition) by means of densities of shallow minors of graphs ([158]):

The *maximum average degree* $\mathrm{mad}(G)$ of a graph G is the maximum over all subgraphs H of G of the average degree of H, that is $\mathrm{mad}(G) = \max_{H \subseteq G} \frac{2|E(H)|}{|V(H)|}$. The *distance* $d(x, y)$ between two vertices x and y of a graph is the minimum length of a path linking x and y, or ∞ if x and y do not belong to same connected component. Also we denote by $G[A]$ the subgraph of G induced by a subset A of its vertices.

We introduce several notations:

- The *radius* $\rho(G)$ of a connected graph G is:

$$\rho(G) = \min_{r \in V(G)} \max_{x \in V(G)} \mathrm{d}(r, x)$$

- A *center* of G is a vertex r such that $\max_{x \in V(G)} \mathrm{d}(r, x) = \rho(G)$.

Let G be a graph. A *ball* of G is a subset of vertices inducing a connected subgraph. The set of all the families of pairwise disjoint balls of G is noted $\mathcal{B}(G)$.

Let $\mathcal{P} = \{V_1, \ldots, V_p\}$ be a family of pairwise disjoint balls of G.

- The *radius* $\rho(\mathcal{P})$ of $\mathcal{P}$ is $\rho(\mathcal{P}) = \max_{X \in \mathcal{P}} \rho(G[X])$.

- The *quotient* $G/\mathcal{P}$ of G by $\mathcal{P}$ is a graph with vertex set $\{1, \ldots, p\}$ and edge set $E(G/\mathcal{P}) = \{\{i, j\} : (V_i \times V_j) \cap E(G) \neq \emptyset \text{ or } V_i \cap V_j = \emptyset\}$.

The following invariants generalize maximum average degree. The *greatest reduced average density (grad)* of a graph G with rank r is

$$\nabla_r(G) = \max \frac{|E(G/\mathcal{P})|}{|\mathcal{P}|}$$

.

where maximum is taken over all $\mathcal{P} \in \mathcal{B}(G)$ satisfying $\rho(\mathcal{P}) \leq r$.

Finally, here is our key definition.

A class of graphs $\mathcal{C}$ has *bounded expansion* if there exists a function $f : N \rightarrow N$ such that for every graph $G \in \mathcal{C}$ and every r,

$$\nabla_r(G) \leq f(r). \tag{2}$$

f is called the *expansion function.*

The definition of bounded expansion can be carried over to general structures by means of incidence graphs. (In most cases this is equivalent to considering 2-section which is in the logical context usually called Gaifman graph.) Thus we may speak about classes of structures with bounded expansion.

After this the main result can be briefly stated.

Theorem 4.5 [156, 160] *Any class of structures with bounded expansion has all restricted dualities.*

By a combination of various results in model theory (so called *homomorphism preservation theorem*) it has been recently showed that every restriction of a homomorphism closed class defined by an FO formula to a bounded expansion class is induced by a duality, [161]. It follows that every restriction of a homomorphism closed class defined by an FO formula to a bounded expansion class is a restriction of a CSP problem.

5 Conservative Structures and List CSP's

A relational structure H is *conservative* if each subset $U \subseteq V(H)$ is a unary relation in H. What this means, is that the input structures G have all possible unary relations U', each corresponding to a subset U of $V(H)$. Imposing U' on a vertex $v \in V(G)$ amounts to restricting its image to be in the corresponding set $U \subseteq V(H)$. Thus we may equivalently view the structures G, H without these unary relations but with each vertex v of G being equipped with a *list* $L(v) \subseteq V(H)$ of allowed images in H. In other words, a *list homomorphism* of G to H is a homomorphism f of G to H such that $f(v) \in L(v)$

28

for each vertex v of G. A *List CSP(H)* is the problem CSP(H) in which inputs G are structures of the same type as H whose vertices are moreover equipped with list; and we seek a list homomorphism to H.

The first result about dichotomy of list CSP's [55] dealt with the case when H has one relation, $E(H)$, which is binary, symmetric, and reflexive. In other words, H is a reflexive undirected graph (each vertex has a loop). In this case, the dichotomy takes on the following attractive form.

Theorem 5.1 (Reflexive Graph List Dichotomy) [55] *Suppose H is a reflexive undirected graph. Then List CSP(H) is NP-complete, except when H is an interval graph, in which case List CSP(H) is polynomial time solvable.*

An *interval graph* is a graph H whose vertices can be associated with real intervals so that two vertices are adjacent in H if and only if the corresponding intervals intersect. Similarly, a *circular arc graph* is a graph H whose vertices can be associated with arcs on a circle so that two vertices are adjacent in H if and only if the corresponding arcs intersect. The next result [60] classifies the complexity of List CSP(H) when H is an irreflexive undirected graph (no vertex has a loop).

Theorem 5.2 [60] *Suppose H is an irreflexive undirected graph. Then List CSP(H) is NP-complete, except when H is a bipartite graph and $\overline{H}$ is a circular arc graph, in which case List CSP(H) is polynomial time solvable.*

For general undirected graphs, where vertices may have loops or not, we have the following unified statement [61]. *Bi-arc graphs* are defined in [61]; they include the reflexive interval graphs as well as the bipartite complements of circular arc graphs from the previous two theorems.

Theorem 5.3 [61] *Suppose H is an undirected graph. Then List CSP(H) is NP-complete, except when H is a bi-arc graph, in which case List CSP(H) is polynomial time solvable.*

Bi-arc graphs are shown to be precisely the graphs with a near-unanimity polymorphism in [22], and have recently been shown also precisely the graphs with a Taylor operation [48]. (The latter fact confirms Conjecture 3.5 for conservative structures that are graphs.)

These results have inspired the following dichotomy result for conservative structures [23]. We state the result in the context of list CSP's as follows.

Theorem 5.4 (Conservative Dichotomy) [23] *For any relational structure H, the problem List CSP(H) is NP-complete or polynomial time solvable.*

For a given structure H, the proof of [23] gives an algorithm to decide whether $\text{CSP}(H)$ is NP-complete or polynomial time solvable. It is polynomial time solvable when H admits, for each pair of vertices u, v, a polymorphism f such that the restriction of f to tuples of u's and v's is a majority operation, a semilattice operation, or a Maltsev operation.

By comparison, the algorithms inherent in the previous three theorems, classifying the complexity of list CSP's for undirected graphs, are efficient low degree polynomial time algorithms [83]. Moreover, they are elegantly characterized by forbidden substructures [83]. It would be desirable to have similar concrete dichotomy classifications for other list CSP problems, at least in the case when H is a digraph. In [63] we proposed the following conjecture (earlier versions appeared in [69, 98]).

Conjecture 5.5 [63] *Suppose H is a reflexive digraph. Then List $\text{CSP}(H)$ is NP-complete, except when H is an adjusted interval digraph, in which case list $\text{CSP}(H)$ is polynomial time solvable.*

An *adjusted interval digraph* is a digraph H whose vertices v can be associated with pairs of real intervals I_v, J_v with the same left endpoint, so that $v \to w$ if and only if I_v intersects J_w. (Ordinary *interval digraphs* have been investigated by West and others [186]; they are defined analogously but do not require each I_v, J_v to have the same left endpoint.) The authors of [63] have recently succeeded in obtaining a forbidden structure characterization of adjusted interval digraphs, and a polynomial time algorithm for their recognition [63]. It is hoped that it will allow us to prove Conjecture 5.5; we have already done this for reflexive digraphs whose underlying graphs are trees and complete graphs. Furthermore, the conjecture also holds for triangle-free graphs [131]. (We also refer the reader to [69].)

The Conservative Dichotomy classification of list CSP's suggests the importance of majority, Maltsev, and semilattice polymorphisms. For the examples discussed above, the role of semilattice polymorphisms is most important, and it can be replaced in this context by the *min polymorphism*, which is a polymorphism f of order two such that there exists a linear ordering of $V(H)$ in which $f(u, v) = \min(u, v)$. (An *X-underbar enumeration* of a digraph H is an ordering of its vertices as $v_1, v_2, \ldots, v_n$ such that whenever the arcs $v_i v_j$ and $v_{i'} v_{j'}$ are in H, then the arc $v_{i''} v_{j''}$, with $i'' = \min(i, i'), j'' = \min(j, j')$, is also in H. It is clear from these definitions that a digraph admits a min polymorphism if and only if it admits an X-underbar enumeration.)

In particular, for reflexive graphs and digraphs we have the following easy observations.

Proposition 5.6 *A reflexive graph H is an interval graph if and only if it admits a min polymorphism.*

Proposition 5.7 *A reflexive digraph H is an adjusted interval digraph if and only if it admits a min polymorphism.*

Thus for reflexive structures H the problem List $\mathrm{CSP}(H)$ seems polynomial if and only if H admits a min polymorphism. For irreflexive structures, the situation is much more complicated, and the full range of the possibilities identified by Bulatov in [23] is possible [68]. (In particular, [68] shows that this is contrary to what was proposed as Conjecture 6.2 in [69] and the Conjecture 3.5 in [98]; this was independently observed by C. Carvalho.)

5.1 Retraction and Extension Problems

The *retraction problem* (or the *precolouring extension problem*) is a classical constraint satisfaction (or homomorphism) problem, which can be seen as a list constraint satisfaction problem in which the lists are restricted to be either singletons (the vertices with singleton lists are essentially "precoloured", or equipped with unary relations $R_v = \{(v)\}$, as in the construction of idempotent H' in Proposition 3.1), or the entire set $V(H)$ (the vertices with lists $V(H)$ are essentially unrestricted). Alternately, if H is a substructure of G, and if the lists of each vertex $v \in V(H)$ is $\{v\}$ while all other lists are $V(H)$, then we seek a homomorphism f of G to its substructure H such that $f(h) = h$ for all $h \in V(H)$. Such homomorphisms are called *retractions* and they have been of interest since Borsuk studied them in the context of topological spaces [20, 95, 116, 97, 135, 170, 173]. There has been a lot of interest in retraction problems in graphs and other relational structures [8, 9, 10, 42, 16, 52, 59, 96, 109, 113, 114, 131, 143, 170, 184].

Of particular interest has been the class of reflexive graphs H for which the retraction of any supergraph G only depends on H being an *isometric subgraph* of G. (In other words, if the distances in H are not shortcut in G, then H is a retract of G.) Of course, is this condition is satisfied, then retraction to H can be tested in polynomial time. Many results about these 'absolute retracts' can be found in [95, 116, 97, 170, 173, 109, 10], and are discussed in detail in [106]. A larger class of reflexive graphs with polynomial time solvable retraction problems would be the class in which the non-existence of a retraction is first order (FO) definable. It has recently been proved by Dalmau, Krokhin, and Larose, that a reflexive graph has an FO definable retraction problem if and only if it is connected and admits a near-unanimity polymorphism [41]. (A similar result holds also for partial orders.)

Many other partial results on dichotomy of retraction problems, especially in the context of graphs, digraphs, and partial orders, can be found in [8, 9, 10, 59, 55, 60, 131, 170]; nevertheless, dichotomy is *not* known for all retraction problems, and it was shown by Feder and Vardi [73] that dichotomy for retraction problems when H^- is a graph (or even a bipartite graph, or a graph with some other properties) would imply the entire Dichotomy Conjecture. On the other hand, there seems to be hope of characterizing graphs (or at least reflexive graphs) H for which the following *subretraction problem* has a polynomial solution: given a graph G with lists restricted to be either singletons, or a fixed set $S \subseteq V(H)$, decide if there exists a list homomorphism of G to H [59].

6 Minimum Cost and Soft CSP's

We have observed earlier that each List $\text{CSP}(H)$ is equivalent to some $\text{CSP}(H')$, where H' is obtained from H by the addition of all unary constraints. On the other hand, every problem $\text{CSP}(H)$ can be viewed as a restriction of the corresponding problem List $\text{CSP}(H)$, where the inputs G have all lists $L(v)$ equal to the entire vertex set of H. In turn, List $\text{CSP}(H)$ is a restriction of the following optimization problem MinCost $\text{CSP}(H)$. The input to the problem MinCost $\text{CSP}(H)$ is a structure G of the same type as H, together with "costs" $c_v(x)$ for all $x \in V(G)$ and $v \in V(H)$, and a target cost K. The *cost* of a homomorphism $f : G \to H$ is the sum $\sum_{x \in V(G)} c_{f(x)}(x)$. The problem is to decide if there exists a homomorphism of G to H of cost at most K.

The problem MinCost $\text{CSP}(H)$ was introduced in [90] (for undirected graphs), where it was motivated by a real-world problem in defence logistics. We believe it offers a practical and natural model for optimization of weighted homomorphisms. Each problem List $\text{CSP}(H)$ is a restriction of the corresponding problem MinCost $\text{CSP}(H)$, where the inputs have $c_i(u) = 0$ if $i \in L(u)$, and $c_i(u) = 1$ otherwise (and $K = 0$).

Thus the unary constraints (which we viewed as lists) are no longer hard in this model: a vertex v can map to any vertex but at different costs. In this sense, the unary constraints have become 'soft'. One can define a more general model where all constraints are 'soft' [36].

Specifically, suppose H is a relational structure which is *complete*, in the sense that each of its relations $S_i, i = 1, \ldots, p$, is equal to $V(H)^{s_i}$. (Recall that s_i is the arity of the relation S_i.) Suppose we additionally have, for each $i = 1, \ldots, p$, a set F_i of real-valued functions on S_i. The structure H, together with the sets F_i is fixed. The problem Soft $\text{CSP}(H)$ has inputs consisting of a structure G of the same type as H, with relations $R_i, i = 1, \ldots, p$, and functions $f_i : R_i \to F_i, i = 1, \ldots, p$, together with a target cost K. The question is whether or not there exists a mapping $f : V(G) \to V(H)$ for which the sum over all $i = 1, \ldots, p$ and over all s_i-tuples $t \in R_i$, of $f_i(t)(f(t))$ does not exceed K. (Here $f(t)$ denotes the coordinatewise evaluation of t, i.e., if $t = (t_1, t_2, \ldots, t_{s_i})$, then $f(t) = (f(t_1), f(t_2), \ldots, f(t_{s_i}))$.)

It is again not difficult to see that each MinCost $\text{CSP}(H)$ is a restriction of a corresponding Soft $\text{CSP}(H)$ (where only 0-1 functions of S_i are allowed in F_i and $K = 0$). Thus the hierarchy of consecutive generalizations goes from $\text{CSP}(H)$ to List $\text{CSP}(H)$, to MinCost $\text{CSP}(H)$, and to Soft $\text{CSP}(H)$.

The following dichotomy for Soft $\text{CSP}(H)$ where H are binary structures was proved in [36]. Note that it applies, in particular, to digraphs.

Theorem 6.1 [36] *Suppose H is a relational structure with only unary and binary relations. If all binary functions in all sets $F_i, i = 1, \ldots, p$ are submodular, then Soft CSP(H)*

is polynomial time solvable.

Otherwise, the problem Soft CSP(H) is NP-complete.

We now discuss MinCost CSP(H) where H is a graph or a digraph. Since MinCost CSP's are restrictions of 'soft' CSP's, some NP-complete Soft CSP's may become polynomial time solvable in the MinCost version. (Thus Theorem 6.1 does not imply dichotomy for MinCost CSP's for digraphs.)

For undirected graphs (with possible loops) we have the following dichotomy classification for MinCost CSP(H).

Theorem 6.2 [88] *Let H be any graph. If each component of H is a reflexive proper interval graph or an irreflexive proper interval bigraph, then the problem MinCost CSP(H) is polynomial time solvable. In all other cases, the problem MinCost CSP(H) is NP-complete.*

A *proper interval graph* is a graph H whose vertices can be associated with an inclusion-free family of real intervals so that two vertices are adjacent in H if and only if the corresponding intervals intersect. An *proper interval bigraph* is a bipartite graph H (with a fixed two-colouring) whose vertices can be associated with two inclusion-free families of real intervals (one for each colour class of H) so that two vertices of opposite colour in H are adjacent in H if and only if the corresponding intervals intersect.

What distinguishes the reflexive proper interval graphs and the irreflexive proper interval bigraphs is the fact that they admit so-called min-max orderings. A linear ordering $<$ of $V(H)$ is a *min-max ordering* if $i < j, s < r$ and $ir, js \in A(H)$ imply that $is \in A(H)$ and $jr \in A(H)$. (These are submodular flows [36].) These play for minimum cost homomorphisms the role played by min polymorphisms in list CSP's. Min-max orderings are directly related to supermodular functions, which have been identified in Theorem 6.1 as the reason for polynomiality of soft CSP's, cf. [36, 125, 43]. (They are also a particular kind of *multimorphism* [36].)

There has also been much activity for digraphs, where the situation is more complicated [89, 91, 85, 86, 87]. In particular, there are polynomial time solvable MinCost CSP's which do not admit a min-max ordering [89], and hence are not covered by a polynomial case of Soft CSP(H). Here we just describe one recent dichotomy - for reflexive digraphs [85]. It turns out that in this case the polynomial cases are exactly those covered by min-max orderings. The key to proving the dichotomy is a forbidden structure characterization of the class of reflexive digraphs H admitting a min-max ordering. The characterization also implies a polynomial time recognition algorithm for the class. One economical way to describe the class is to use two undirected graphs associated with each digraph H - the *symmetric graph $S(H)$* which has $V(S(H)) = V(H)$ and $uv \in E(S(H))$ just if $uv \in H$
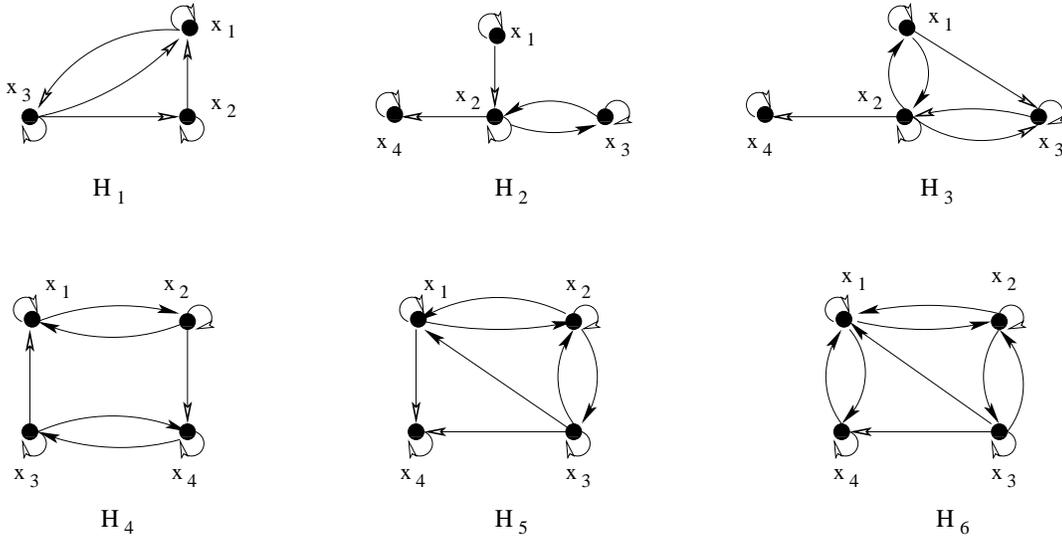
Figure 1: The obstructions H_i with $i = 1, 2, 3, 4, 5, 6$.

and $vu \in H$, and the *associated bipartite graph* $B(H)$ with $V(B(H)) = V(H) \times \{0, 1\}$ in which $(u, 0)(v, 1) \in E(B(H))$ just if $uv \in H$.

Theorem 6.3 [85] *Let H be a reflexive digraph H. If*

- $S(H)$ *is a proper interval graph, and*

- $B(H)$ *is a proper interval bigraph, and*

- H *does not contain an induced subgraph isomorphic to H_i with $i = 1, 2, 3, 4, 5, 6$,*

then H admits a min-max ordering and mincost $CSP(H)$ is polynomial time solvable.

Otherwise, mincost $CSP(H)$ is NP-complete.

7 Full CSP's

Consider the relational structure H with $V(H) = \{0, 1, 2\}$ with three symmetric binary relations S_0, S_1, S_2 where $S_i = [V(H) \times V(H)] - (i, i)$. The problem $CSP(H)$ takes as input structures G with three binary relations R_0, R_1, R_2 (which can be assumed without loss of generality to be symmetric as well). It is not difficult to prove that $CSP(H)$ is NP-complete, i.e., it is NP-complete to decide if input G admits a homomorphism to H.

34

The situation changes drastically if we restrict the inputs to structures G in which the union of the relations R_0, R_1, R_2 contains all pairs of distinct vertices of G. (Such a structure is called 2-*full* [56].) Without going into details [56], we shall informally call a *full CSP* any CSP(H) restricted to instances in which the union of all relations of certain specified arities is complete. In this section we focus on 2-full CSP's. The first kind of 2-full CSP's, illustrated in the first example, has H consist of $V(H) = \{0, 1, \ldots, k-1\}$ and k symmetric binary relations $S_0, S_1, \ldots, S_{k-1}$, where $S_i = [V(H) \times V(H)] - (i, i)$. The inputs G are restricted to have relations $R_0, R_1, \ldots, R_{k-1}$ whose union contains all distinct pairs of vertices. For simplicity, let us further assume that the relations $R_0, R_1, \ldots R_{k-1}$ are also disjoint. (This turns out not to be a real restriction [56].) We call this the *compatible k-colouring problem* [56]. It can be restated as follows.

Given a complete graph G whose edges are coloured by $0, 1, \ldots, k-1$, decide if G admits a *compatible vertex colouring* of G by colours $0, 1, \ldots, k-1$ so that no edge $e = uv$ of G has the same colour on $e, u,,$ and v.

The compatible 2-colouring problem has a nice solution. Indeed, a complete graph G with edges coloured $0, 1$ may be viewed as just an ordinary graph (by taking the edges of colour 0 to be absent); thus the problem has become one of partitioning the vertices of G into two sets, those coloured 0 - which must form a clique - and those coloured 1 - which must form an independent set. Graphs G which admit such a partition are called *split* graphs [83] and can be recognized in linear time. (In fact G is a split graph if and only if it is chordal and contains no induced $2K_2$ [83].) This also works in the presence of lists. as the addition of lists makes only a very small difference and both linear time algorithms and forbidden induced subgraph characterizations still apply [100].

The compatible 4-colouring problem (and similarly for k-colourings with $k > 4$) is NP-complete. Indeed, given a graph G, it is easy to construct a graph G' with edges coloured $0, 1, 2, 3$, which can be compatibly coloured if and only if G is 3-colourable (in the usual sense). It is enough to let G' consist of two disjoint copies of G with each edge of G replaced by three parallel edges coloured $0, 1, 2$, and colouring all remaining pairs of vertices by 3. Indeed, any 3 colouring of G (repeated on each copy of G in G') is a compatible colouring of G'; and conversely, any compatible colouring of G' must not use colour 3 on the vertices of one copy of G (since they are linked by all edges coloured 3), and hence induce a valid 3-colouring on G. (Note that the graph G' has parallel edges, or equivalently, some pairs of vertices of G are related in several relations. It is not hard to replace these multiple edges by substituting p independent vertices for each vertex of G, and joining two independent sets corresponding to a pair of vertices joined by edges of colours $0, 1, 2$, by a complete bipartite graph with edges coloured $0, 1, 2$, so that between any two subsets of at least $p/2$ vertices there is an edge of each colour $0, 1, 2$. If p is large enough, random colourings have this property.)

The complexity of the compatible 3-colouring problem is not known. It is not likely to be NP-complete, because of the following result [56].

Theorem 7.1 [56] *There is an $n^{O(\log n)}$ algorithm to solve the compatible 3-colouring problem, with lists.*

Proof. Initial lists are given, $L(v) \subseteq \{0, 1, 2\}$, for all vertices v of G. (In the problem without lists, we can take all initial $L(v) = \{0, 1, 2\}$.) At each stage, the lists will be reduced, by replacing a problem by a set of subproblems with smaller lists, until we reach a stage when all lists have at most two vertices. Such problems can be solved by a standard application of a two-satisfiability algorithm [6]; indeed choosing an image from a list of size two represents a Boolean choice, and all constraints are defined over pairs of vertices, yielding clauses of size two [64].

Let S denote the (changing) set of vertices v with $L(v) = \{0, 1, 2\}$. We say that i is the *majority colour* at $v \in S$ if it occurs on at least a third of the edges from v to the other vertices of S.

We reduce the current problem to $|S| + 1$ subproblems as follows. In the first subproblem, we avoid giving any vertex $v \in S$ its majority colour. This results in all lists of size at most two, and can be tested in linear time. In the remaining $|S|$ subproblems, we assume for each $v \in S$ in turn, that (at least) v receives its majority colour i; this allows us to remove i from the list of at least $|S|/3$ other vertices.

We obtain the recurrence $T(s) \leq (1 + sT(2s/3))T_2(n)$, where $s = |S|$ and $T_2(n)$ is the time for solving an instance of two-satisfiability on n variables and at most n^2 clauses. It is easy to see that the solution is $T(n) = n^{O(\log n)}$. $\qquad\square$

No NP-complete problem is known to have an $n^{O(\log n)}$ algorithm, and the nature of polynomial time reductions ensures that if one NP-complete problem did have such an algorithm, then so would all the others. Thus we take our result as evidence that the compatible 3-colouring problem is not likely to be NP-complete. Yet no polynomial time algorithm is known for the problem. There are better algorithms than the simple example above - the currently best algorithm has complexity $n^{O(\log n / \log \log n)}$ [66].

7.1 Matrix Partitions

Here we discuss another class of 2-full CSP's, in which H has just two binary relations [56]. These problems have been studied as so-called *matrix partition problems* [64], and can be best formulated as follows.

Let M be a symmetric m by m matrix over $0, 1, *$. An M-partition of a graph G is a partition of $V(G)$ into parts $V_1, V_2, \ldots, V_m$ such that for distinct vertices $u \in V_i$, $v \in V_j$, we have $uv \in E(G)$ if $M(i, j) = 1$, and $uv \notin E(G)$ if $M(i, j) = 0$. Note that we admit $i = j$; in particular, if $M(i, i) = 0$, the set V_i is independent in G, and if $M(i, i) = 1$, it is a clique. Also note that $*$ means no restriction. For each fixed matrix M we obtain the M-*partition problem* - to decide whether or not a given graph G admits an M-partition. For

36

instance, if C_m is the matrix in which the diagonal entries are 0 and all other entries are *, then the C_m-partition problem asks whether or not the graph G is m-colourable, in the usual sense. Many other graph partition problems, especially those arising in the study of perfect graphs [33, 34, 64]), can be formulated as matrix partition problems; these include problems such as deciding the existence of a clique cutset, or a skew cutset [34, 188], and so on [33]. Take, for instance, the problem of recognizing split graphs [76]. These are precisely the M-partitionable graphs where M is the two by two matrix with rows 0* and *1. For this problem there is a well-known polynomial time recognition algorithm and a characterization by forbidden induced subgraphs - G is partitionable (is a split graph) if and only if it does not have an induced $2K_2, C_4, C_5$ [76]. Similarly, complete bipartite graphs are precisely the M-partitionable graphs where M is the two by two matrix with rows 01 and 10. In this case, it is easy to check that G is complete bipartite, i.e., M-partitionable, if and only if it does not have an induced copy of $K_1 \cup K_2$.

Thus we may ask, for any given matrix M, whether M-partitionable graphs be recognized in polynomial time, and whether M-partitionable graphs be characterized by a finite set of forbidden induced subgraphs. (The latter case is just finite duality in this context [7].) An intermediate question might be whether M-partitionable graphs are FO definable [5, 175, 190]. However, unlike the case of CSP's it is not known whether FO definability is equivalent to finite duality.

There are several variants of the basic M-partition problem: these include partitioning digraphs (the matrix M is not necessarily symmetric) [69, 180], equipping the vertices of G with lists, or insisting that each part be nonempty [31, 64, 75, 106], generalizing to certain constraint satisfaction problems [56], or restricting the input graphs to have special structure [57, 65, 70, 100, 106].

For the basic problem discussed here, it makes sense to assume that all $M(i,i) \neq *$. Indeed, if some $M(i,i) = *$, then every graph G is M-partitionable, and all our questions are trivial. If M has no diagonal *, then we may assume that $M(1,1) = \ldots = M(k,k) = 0, M(k+1,k+1) = \ldots = M(m,m) = 1$. Let A denote the submatrix of M with rows and columns $1, \ldots, k$; let B denote the submatrix with rows and columns $k+1, \ldots, m$; and let C denote the submatrix with rows $1, \ldots, k$ and columns $k+1, \ldots, m$. A matrix which has no * in the submatrices A and B is called *friendly*.

Theorem 7.2 [190] *If M is not a friendly matrix, then M-partitionble graphs cannot be characterized by a finite set of forbidden induced subgraphs.*

We have several classes of friendly matrices M for which M-partitionable graphs are known to have a characterization by finitely many forbidden subgraphs. The simplest case occurs when M has no * entries at all.

In [56] we have shown the following fact.

Theorem 7.3 [56] *If M has no $*$ entries, then all minimal obstructions to M-partition have at most $(k+1)(m-k+1)$ vertices (and there are at most two minimal obstructions with precisely $(k+1)(m-k+1)$ vertices).*

Different proofs of the fact that there are only finitely many minimal obstructions in this case can also be found in [7] and [190].

We can extend the validity of this result by applying the so-called sparse-dense technique from [57], cf. [71, 190].

Theorem 7.4 [190] *If M is a friendly matrix in which C has only $*$ entries, then M-partitionable graphs can be characterized by a finite set of forbidden induced subgraphs.*

In certain other cases, while we cannot prove the existence finitely many forbidden induced subgraphs, we can at least guarantee a polynomial time algorithm for M-partitionability. We say that M is *triangle-free* if there do not exist subscripts $i \leq i' \leq i'' \leq k$ with $M(i, i') = M(i, i'') = M(i', i'') = 0$, and there do not exist subscripts $k < j \leq j' \leq j''$ with $M(j, j') = M(j, j'') = M(j', j'') = 1$.

Theorem 7.5 [190] *If M is a triangle-free friendly matrix, then the M-partition problem is polynomial time solvable.*

Full CSP's appear less likely to enjoy dichotomy - even for the simple compatible 3-colouring problem we do not know a polynomial time algorithm. A relative of the compatible 3-colouring problem from [31], called the *stubborn problem*, is a list version of an M-partition problem where the matrix has size only four. For this problem there is an $n^{O(\log n)}$ algorithm similar to the one described above for the compatible 3-colouring problem, but no polynomial algorithm is known. This suggests that proving dichotomy for all matrix partition problems must be hard - even such a concrete small problem is causing difficulty. However, it turns out that the situation does not get any worse.

Let us say that a problem is *quasi-polynomial*, if it admits an $n^{O(\log n)}$ algorithm [64].

Theorem 7.6 [56] *For each matrix M, the list M-partition problem is NP-complete or quasi-polynomial.*

8 Conclusions

The most important open theoretical problem is certainly the Dichotomy Conjecture. We see this conjecture taking now a much more concrete and perhaps definitive form,

and see hopeful signs that the new tools [29, 128, 142, 165] might be bringing us nearer to its solution. It is worth noting that while the NP-completeness of CSP's appears well understood (and seems related to various forms of projectivity), we are lacking a good understanding of the polynomial algorithms - which is of course also of practical importance. Indeed, much of the current research on the Dichotomy Conjecture amounts to a hunt for new polynomial algorithms.

We are grateful to G. Kun and M. Szegedy, as well as to M. Siggers, for providing us with early drafts of their papers [128, 187].

References

[1] Handbook of Constraint Programming (F. Rossi, P. van Beek, T. Walsh, eds.), Elsevier, 2006, xix+995p.

[2] N. Alon, A. Shapira, "Homomorphisms in graph property testing", in: Topics in Discrete Mathematics (M. Klazar, J. Kratochvíl, J. Matoušek, R. Thomas, P. Valtr, eds), Springer 2006, pp. 281–314.

[3] N. Alon, I. Dinur, E. Friedgut, B. Sudakov, "Graph products, Fourier analysis and spectral techniques", Geom. Funct. Anal. 14 (2004) 913–940.

[4] S. Arora, C. Lund, R. Motwani, M. Sudan, M. Szegedy, "Proof verification and the hardness of approximation problems", Journal of ACM, 45 (1998) 501–555.

[5] A. Atserias, "On digraph coloring problems and treewidth duality", European J. Comb. 29 (2008) 796–820, also in: 20th IEEE Symposium on Logic in Computer Science (LICS), 2005, pp. 106–115.

[6] B. Aspvall, F. Plass and R. E. Tarjan, "A linear time algorithm for testing the truth of certain quantified Boolean formulas", Information Processing Letters, Vol. 8, 1979 121–123.

[7] R. N. Ball, J. Nešetřil, and A. Pultr, "Dualities in full homomorphisms", mansucript 2006.

[8] H.-J. Bandelt, A. Dählmann, H. Schütte, "Absolute retracts of bipartite graphs", Discrete Appl. Math. 16 (1987) 191 – 215.

[9] H.-J. Bandelt, M. Farber, and P. Hell, "Absolute reflexive retracts and absolute bipartite retracts", Discrete Applied Math. 44 (1993) 9 - 20.

[10] H.-J. Bandelt, E. Pesch, "Efficient characterizations of n-chromatic absolute retracts", J. Combinatorial Theory B 53 (1991) 5 – 31.

[11] J. Bang-Jensen and P. Hell, "The effect of two cyles on the complexity of colourings by directed graphs", Discrete Applied Math. 26 (1990) 1–23.

[12] J. Bang-Jensen, P. Hell, and G. MacGillivray, "The complexity of colouring by semicomplete digraphs", SIAM J. on Discrete Math. 1 (1988) 281-298.

[13] J. Bang-Jensen, P. Hell, and G. MacGillivray, "On the complexity of colouring by superdigraphs of bipartite graphs," Discrete Math. 109 (1992) 27–44.

[14] J. Bang-Jensen, P. Hell, and G. MacGillivray, "Hereditarily hard H-colouring problems," Discrete Math. 138 (1995) 75 – 92.

[15] L. Barto, M. Kozik, and T. Niven, "The CSP dichotomy holds for digraphs with no sources and sinks (a positive answer to a conjecture of Bang-Jensen and Hell)", SIAM J. Computing, to appear. Also, Proc. of the 40th ACM Symposium on Theory of Computing, STOC'08 (2008), 789–796.

[16] M. Birc, M. Hujter, and Zs. Tuza, "Precoloring extension I; interval graphs", Discrete Math. 100 (1992) 267 – 279.

[17] J. Berman, P. Idziak, P. Markovic, R. McKenzie, M. Valeriote and R. Willard, "Varieties with few subalgebras of powers", to appear in the Transactions of the American Mathematics Society, 2008.

[18] V.G. Bodnarčuk, L.A. Kaluzhnin, V.N. Kotov, B.A. Romov, "Galois theory for Post algebras I-II (russian)", Kibernetika 3 (1969), 1–10, and 5 (1969), 1–9; english version Cybernetics, (1969) 243–252 and 531–539.

[19] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, K. Vesztergombi, "Counting graph homomorphisms", in: Topics in Discrete Mathematics (M. Klazar, J. Kratochvíl, J. Matoušek, R. Thomas, P. Valtr, eds), Springer 2006, pp. 315 – 372.

[20] K. Borsuk, "Sur les rétractes", Fund. Math. 17 (1931) 152 – 170.

[21] A. Branstädt, P. Hammer, V. Bang Le, and V. Lozin, "Bisplit graphs", DIMACS Technical Report 2002–44 (2002).

[22] R. Brewster, T. Feder, P. Hell, J. Huang, and G. MacGillivray, "Near-unanimity functions and varieties of graphs", SIAM Journal on Discrete Math. 22 (2008) 938 – 960..

[23] A. A. Bulatov, "Tractable conservative constraint satisfaction problems," In Proceedings of the 18th IEEE Annual Symposium on Logic in Computer Science (LICS 2003), 321–330.

[24] A. A. Bulatov, "Maltsev constraints are tractable", Technical report PRG-RR-02-05 Oxford University (2002).

[25] A. A. Bulatov, "A dichotomy constraint on a three-element set," In Proceedings of the 43rd IEEE Symposium on Theory of Computing (2002), 649–658.

[26] A. A. Bulatov, "H-colouring dichotomy revisited", Theoretical Computer Science 349 (2005) 31–39.

[27] A. A. Bulatov and V. Dalmau, "Maltsev constraints are tractable", SIAM J. on Computing, 36 (2006) 16–27.

[28] A. A. Bulatov and P. Jeavons, "Algebraic structures in combinatorial problems," Technical Report MATH-AL-4-2001m Technische universität Dresden, Germany, 2001.

[29] A. A. Bulatov, P. Jeavons, and A. A. Krokhin, "Classifying complexity of constraints using finite algebras," SIAM J. on Computing 34 (2005) 720–742.

[30] A. Bulatov, A. Krokhin, B. Larose, "Dualities for constraint satisfaction problems", Proceedings of the Dagstuhl Seminar on the Complexity of Constraints, Springer Lecture Notes in Computer Science, 2008, to appear.

[31] K. Cameron, E. E. Eschen, C. T. Hoang and R. Sritharan, "The list partition problem for graphs," Proc. 15th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA) 2004, 391–399.

[32] E. Cheng, R. P. Kleinberg, S. G. Kruk, W. A. Lindsey, and D. E. Steffy, "A strictly combinatorial approach to a university exam scheduling problem", Congr. Numer. 167 (2004) 121 – 132.

[33] M. Chudnovsky, N. Robertson, P. Seymour, and R. Thomas, "The strong perfect graph theorem", Annals of Mathematics 164 (2006), 51–229.

[34] V. Chvátal, "Star-cutsets and perfect graphs", J. Combinatorial Theory B 39 (1985) 189–199.

[35] D. Cohen, M. Cooper, and P. Jeavons, "Characterizing tractable constraints", Artificial Intelligence 65 (1994) 347 – 361.

[36] D. Cohen, M. Cooper, P. Jeavons, and A. Krokhin, "A maximal tractable class of soft constraints", J. of Artificial Intelligence Research 22 (2004) 1 – 22.

[37] N. Creignou, S. Khanna, and M. Sudan, **Complexity Classifications of Boolean Constraint Satisfaction Problems**, SIAM Monographs on Discrete Math. and Applications, vol. 7 (2001).

[38] V. Dalmau, "A new tractable class of constraint satisfaction problems", in Proceedings 6th International Symposium on Artificial Intelligence and Mathematics, 2000.

[39] V. Dalmau, "Generalized majority-minority operations are tractable", in Proceedings of the 20th Annual IEEE Symposium on Logic in Computer Science (LICS 05), pages 438447.

[40] V. Dalmau and D. Ford, "Generalized satisfiability with k occurrences per variable: a study through delta-matroid parity," Mathematical Foundations of Computer Science (MFCS 2003), Lecture Notes in Computer Science 2747, Springer 2003, 358–367.

[41] V. Dalmau, A. Krokhin, and B. Larose, "First-order definable retraction problems for posets and reflexive graphs", J. Logic Comput. 17 (2007), 31-51; also in LICS 2004, pp. 232–241.

[42] V. Dalmau, A. Krokhin, B. Larose, "Retractions onto series-parallel posets", Discrete Math., to appear.

[43] M. Datar, T. Feder, A. Gionis, R. Motwani, and R. Panigrahy, "A combinatorial algorithm for MAX CSP", Information Processing Letters 85 (2003) 307 – 315.

[44] R. Dechter, "Constraint networks," In Encyclopedia of Artificial Intelligence, 1992, 276–285.

[45] I. Dinur, "The PCP theorem by gap amplification", J. of ACM 54 (2007) 12.

[46] I. Dinur, E. Friedgut, O. Regev "Independent sets in graph powers are almost contained in juntas", to appear in GAFA.

[47] W. F. Dowling and J. H. Gallier, "Linear-time algorithms for testing the satisfiability of propositional Horn formulae", J. Logic Programming 1 (1984) 267 – 284.

[48] L. Egri, B. Larose, P. Tesson "The complexity of list homomorphism problems for graphs", in preparation.

[49] R. Fagin: Generalized first-order spectra and polynomial-time recognizable sets. in: Complexity of Computation (ed. R. Karp), SIAM-AMS Proceedings 7, 1974, pp. 43–73.

[50] T. Feder, "Fanout limitations on constraint systems," Theoretical Computer Science 255 (2001) 281–293.

[51] T. Feder, "Homomorphisms to oriented cycles and k-partite satisfiability," SIAM J. Discrete Math. 14 (2001) 471–480.

[52] T. Feder, "Constraint satisfaction: a personal perspective", manuscript 2004.

[53] T. Feder, "A dichotomy theorem on fixed points of several nonexpansive mappings", to appear in SIAM J. Discrete Math.

[54] T. Feder and D. Ford, "Classification of bipartite Boolean constraint satisfaction through delta-matroid intersection," manuscript.

[55] T. Feder and P. Hell, "List homomorphisms to reflexive graphs," J. Comb. Theory Series B 72 (1998) 236–250.

[56] T. Feder and P. Hell, "Full constraint satisfaction problems", SIAM J. on Computing 36 (2006) 230–246.

[57] T. Feder and P. Hell, "Matrix partitions of perfect graphs", Discrete Math. 306 (2006) 2450–2460.

[58] T. Feder and P. Hell, "On realizations of point determining graphs, and obstructions to full homomorphisms", Discrete Math. 308 (2008) 1639 – 1652.

[59] T. Feder and P. Hell, "The complexity of retraction and subretraction problems for reflexive digraphs", manuscript 2007.

[60] T. Feder, P. Hell, and J. Huang, "List homomorphisms and circular arc graphs," Combinatorica 19 (1999) 487–505.

[61] T. Feder, P. Hell, and J. Huang, "Bi-arc graphs and the complexity of list homomorphisms," J. Graph Theory 42 (1999) 61–80.

[62] T. Feder, P. Hell, and J. Huang, "List homomorphisms of graphs with bounded degrees," to appear in Discrete Math.

[63] T. Feder, P. Hell, J. Huang, and A. Rafiey, "Adjusted interval digraphs", manuscript 2008.

[64] T. Feder, P. Hell, S. Klein, and R. Motwani, "Complexity of list partitions," SIAM J. Discrete Mathematics 16 (2003) 449-478.

[65] T. Feder, P. Hell, S. Klein, L. Nogueira, and F. Protti, "List matrix partitions of chordal graphs", Theoretical Computer Science 349 (2005) 52 – 66.

[66] T. Feder, P. Hell, D. Kral, and J. Sgall, "Two algorithms for list matrix partitions." Symposium on Discrete Algorithms 2005.

[67] T. Feder, P. Hell, and B. Mohar, "Acyclic homomorphisms and circular colorings of digraphs", SIAM J. Discrete Math. 17 (2003) 161–169.

[68] T. Feder, P. Hell and A. Rafiey, "List homomorphisms to irreflexive digraphs", manuscript 2008.

[69] T. Feder, P. Hell, and K. Tucker-Nally, "Digraph matrix partitions and trigraph homomorphisms", manuscript 2004.

[70] T. Feder, P. Hell, and W. Hochstättler, "Generalized colourings (matrix partitions) of cographs", manuscript 2005.

[71] T. Feder, P. Hell, and W. Xie, "Obstructions to generalized colourings", ICDM 2006, Bangalore, India, Lecture Notes of the Ramanujan Math. Society 2007.

[72] T. Feder, F. Madelaine, and I.A. Stewart, "Dichotomies for classes of homomorphism problems involving unary functions," Theoretical Computer Science 314 (2004), 1–43.

[73] T. Feder and M.Y. Vardi, "The computational structure of monotone monadic SNP and constraint satisfaction: a study through Datalog and group theory," SIAM J. Comput. 28 (1998) 236–250 (ALSO in STOC 25 (1993) 612 – 622).

[74] J. Fiala, J. Kratochvil, "Locally constrained graph homomorphisms", Comp. Sci. Review 2 (2008) 97–112.

[75] C. M. H. de Figueiredo, S. Klein, Y. Kohayakawa and B. Reed, "Finding skew partitions efficiently", J. Algorithms 37 (2000) 505 – 521.

[76] S. Foldes and P. Hammer, "Split graphs", Congr. Numer. 19 (1977) 311–315.

[77] J. Foniok, J. Nešetřil, C. Tardif, "Generalized dualities and maximal finite antichains in the homomorphism order of relational structures", European J. Comb. 29 (2008) 881–899.

[78] M. H. Freedman, L. Lovász, A. Schrijver, Reflection positivity, rank connectivity and homomorphisms of graphs, J. of Amer. math. Soc. 20 (2007) 37–51.

[79] T. Gallai, "On directed paths and circuits", Theory of Graphs (Proc. Colloq., Tihany, 1966) Academic Press, New York, 1968, 115–118.

[80] A. Galluccio, P. Hell, J. Nešetřil, "The complexity of H-colouring of bounded degree graphs", Discrete Math. 222 (2000) 101 - 109.

[81] M. Garey and D. S. Johnson, **Computers and Intractability: a Guide to the Theory of NP-completeness**, Freeman and Company, 1979.

[82] D. Geiger, "Closed systems of functions and predicates", Pacific Journal of Math. 27 (1968) 95–100.

[83] M. C. Golumbic, **Algorithmic Graph Theory and Perfect Graphs**, Academic Press, New York, 1980.

[84] D. Greenwell and L. Lovász, "Applications of product coloring", Acta Math. Acad. Sci. Hungar. 25 (1974) 335–340.

[85] A. Gupta, P. Hell, M. Karimi, and A. Rafiey, "Minimum cost homomorphisms to reflexive digraphs", LATIN 2008.

[86] A. Gupta, G. Gutin, M. Karimi, E. J. Kim and A. Rafiey, "Minimum cost homomorphisms to locally semicomplete and quasi-transitive digraphs", manuscript 2008.

[87] A. Gupta, M. Karimi, E. J. Kim and A. Rafiey, Minimum cost homomorphism dichotomy for locally in-semicomplete digraphs, COCOA 2008.

[88] G. Gutin, P. Hell, A. Rafiey and A. Yeo, Minimum Cost "Homomorphisms to proper interval graphs and bigraphs", Europ. J. Combin. 29 (2008) 900–911.

[89] G. Gutin, A. Rafiey and A. Yeo, Minimum Cost and List Homomorphisms to Semicomplete Digraphs. *Discrete Appl. Math.* 154 (2006), 890–897.

[90] G. Gutin, A. Rafiey, A. Yeo and M. Tso, Level of repair analysis and minimum cost homomorphisms of graphs, *Discrete Appl. Math.* (2007)

[91] G. Gutin and E.J. Kim, Introduction to the minimum cost homomorphism problem for directed and undirected graphs, Lecture Notes of the Ramanujan Math. Society, 2007.

[92] W. Gutjahr, E. Welzl and G. Woeginger, "Polynomial graph-colorings", Discrete Applied Math, 35 (1992) 29 – 45.

[93] R. L. Graham, B. Rothschild, J. Spencer: **Ramsey Theory**, Wiley, 1980.

[94] M. Hasse, "Zur algebraischen Begründung der Graphentheorie I.", Math Nachr. 28 (1964/1965), 275–290.

[95] P. Hell, Rétractions des graphes, PhD thesis, Université de Montréal, 1972.

[96] P. Hell, "Absolute planar retracts and the four color conjecture", J. Combinatorial Theory B 17 (1974) 5-10.

[97] P. Hell, "Absolute retracts in graphs", in **Graphs and Combinatorics** (R.A. Bari, F. Harary, eds.), Springer-Verlag Lecture Notes in Mathematics 406 (1974) 291 – 301.

[98] P. Hell, "From graph colouring to constraint satisfaction: there and back again", **Topics in Discrete Mathematics**, Springer Verlag Algorithms and Combinatorics Series, volume 26, 2006, pp. 407 – 432.

[99] P. Hell "Algorithmic aspects of graph homomorphisms", in **Surveys in Combinatorics 2003** (C.D. Wensley ed.), London Math. Society Lecture Note Series 307, Cambridge University Press 2003, pp. 239-276.

[100] P. Hell, S. Klein, L. Tito Nogueira, and F. Protti, "Partitioning chordal graphs into independent sets and cliques", Discrete Applied Math. 141 (2004) 185-194.

[101] P. Hell, S. Klein, L. Tito Nogueira, and F. Protti, "Packing r-cliques in chordal graphs", to appear in ANOR.

[102] P. Hell, J. Nešetřil, "The core of a graph", Discrete Math. 109 (1992) 117 – 126.

[103] P. Hell, J. Nešetřil, and X. Zhu, "Duality of graph homomorphisms", in: *Combinatorics, Paul Erdos is Eighty* Bolyai Society Mathematical Studies, vol.2, 1996, 271–282.

[104] P. Hell and J. Nešetřil, "On the complexity of H-coloring," J. Comb. Theory, Series B 48 (1990), 92–110.

[105] P. Hell and J. Nešetřil, "Counting list homomorphisms and graphs with bounded degrees", in **Graphs, Morphisms and Statistical Physics** (J. Nešetřil and P. Winkler, eds.) DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Volume 63 (2004) 105 – 112.

[106] P. Hell and J. Nešetřil, **Graphs and Homomorphisms**, Oxford University Press 2004.

[107] P. Hell, J. Nešetřil, and X. Zhu, "Complexity of tree homomorphisms", Discrete Applied Math. 70 (1996) 23–36.

[108] P. Hell, J. Nešetřil, and X. Zhu, "Duality and polynomial testing of tree homomorphisms", Transactions of the A.M.S. 348 (1996) 1281–1297.

[109] P. Hell, I. Rival, "Absolute retracts and varieties of reflexive graphs", Canad. J. Math. 39 (1987) 544 – 567.

[110] S. Hoory, N. Linial, A. Wigderson, "Expander graphs and their applications", Bull. Amer. Math. Soc. 43 (2006) 439–561.

[111] A. Horn, "On sentences which are true of direct unions of algebras", J. Symbolic Logic 16 (1951) 14 – 21.

[112] J. Hubička, J. Nešetřil, "Universal structures as shadows of ultrahomogeneous structures" (submitted)

[113] M. Hujter and Zs. Tuza, "Precoloring extension II; graph classes related to bipartite graphs", Acta Math. Universitatis Comenianae 62 (1993) 1 – 11.

[114] M. Hujter and Zs. Tuza, "Precoloring extension III; classes of perfect graphs", Combin. Probab. Comput. 5 (1996) 35 – 56.

[115] P. Idziak, P. Markovic, R. McKenzie, M. Valeriote and R. Willard, Tractability and learnability arising from algebras with few subpowers (extended abstract), Proceedings of the 22nd Annual IEEE Symposium on Logic in Computer Science, LICS 2007, 213–224.

[116] E.M. Jawhari, D. Misane, M. Pouzet, "Retracts: graphs and ordered sets from the metric point of view", Contemp. Math. (Amer. Math. Soc.) 57 (1986) 175 – 226.

[117] P. Jeavons, "On the algebraic structure of combinatorial problems," Theoretical Computer Science 200 (1998) 185–204.

[118] P. Jeavons, D. Cohen, and M. Gyssens, "Closure properties of constraints," J. of the ACM 44 (1997) 527–548.

[119] P. Jeavons, D. Cohen, and M. Gyssens, "A unifying framework for tractable constraints," Proc. 1st International Conference on Constraint Programming, CP'95, Cassis 1995.

[120] D. S. Johnson, "The NP-completeness columns: an ongoing guide", J. Algorithms 3 (1982) 89 – 99.

[121] G. Kalai, "A Fourier-theoretic perspective for the Condorcet parados and Arrow's theorem", Adv. in Appl. Math 29 (2002) 412–426.

[122] J. H. Kim, "The Ramsey number $R(3,t)$ has order of magnitude $t^2/logt$", Random Structures and Algorithms, 7 (1995) 173–207.

[123] P. Kolaitis and M. Vardi, "Conjunctive-query containment and constraint satisfaction", In Proc. of the 17th ACM SIGACT SIGMOD SIGART Symp. on Principles of Database Systems 1998 (PODS'98), 205–213.

[124] A. Kostochka, J. Nešetřil, P. Smolíková, "Colorings and homomorphisms of degenerate and bounded degree graphs" Discrete Math. 233 (2001) 257–276.

[125] A. Krokhin and B. Larose, "Maximizing supermodular functions on product lattices, with application to maximum constraint satisfaction", SIAM J. on Discrete Math., to appear.

[126] V. Kumar, "Algorithms for constraint-satisfaction problems," AI Magazine 13 (1992), 32–44.

[127] G. Kun, "Constraints, MMSNP and expander structures", Combinatorica, submitted.

[128] G. Kun, M. Szegedy, "A new line of attack on the dichotomy conjecture", manuscript 2008.

[129] G. Kun, J. Nešetřil: Forbidden lifts (NP and CSP for combinatorists), European J. Comb. 29 (2008) 930–945.

[130] R. E. Ladner, "On the structure of polynomial time reducibility", J. Assoc. Comput. Mach. 22 (1975) 155–171.

[131] B. Larose, "Taylor operations on finite reflexive structures", International Journal of Mathematics and Computer Science, 1 (2006) 1–26.

[132] B. Larose and C. Tardif, "Projectivity and independent sets in powers of graphs", J. Graph Theory 40 (2002) 162 – 171.

[133] B. Larose, C. Loten, C. Tardif, "A characterization of first-order constraint satisfaction problems", Logical Methods in Computer Science 3 (2007), no. 4, 6, 22pp. (Electronic)

[134] B. Larose, C. Loten, L. Zádori, "A polynomial-time algorithm for near-unanimity graphs", Journal of Algorithms 55 (2005) 177–191.

[135] B. Larose and L. Zádori, "Finite posets and topological spaces in locally finite varieties", Algebra Universalis 52 (2004) 119 – 136.

[136] B. Larose and L. Zádori, "The complexity of the extendibility problem for finite posets", SIAM J. on Discrete Math. 17 (2003) 114 – 121.

[137] P. Lincoln and J. C. Mitchell, "Algorithmic aspects of type inference with subtypes," in Conf. Rec. 19th ACM Symp. on Principles of Programming Languages (1992), 293–304.

[138] V. Lotfi and S. Sarin, "A graph coloring algorithm for large scale scheduling problems", Comput. Oper. Res. 13 (1986) 27 – 32.

[139] T. Łuczak, J, Nešetřil, "A note on projective graphs", J. Graph Theory 47 (2004) 81–86.

[140] T. Łuczak, J. Nešetřil, "Towards probabilistic analysis of the dichotomy problem", KAM-DIMATIA Series 2003-640, Charles Univ. Prague.

[141] R. McKenzie, Personal communication, 2007.

[142] M. Maŕoti, R. McKenzie, "Existence theorems for weakly symmetric operations", manuscript 2006, to appear in Algebra Universalis.

[143] D. Marx, "Precoloring extension on chordal graphs", manuscript 2003.

[144] J. Matoušek, J. Nešetřil, "Constructions of sparse graphs with given homomorphisms", to appear.

[145] P. Meseguer, "Constraint satisfaction problem: an overview," AICOM 2 (1989), 3–16.

[146] J. C. Mitchell, "Coercion and type inference (summary)," in Conf. Rec. 11th ACM Symp. on Principles of Programming Languages (1984), 175–185.

[147] V. Müller, "On colorable critical and uniquely colorable critical graphs", in: Recent Advances in Graph Theory (ed. M. Fiedler), Academia, Prague 1975.

[148] V. Müller, "On coloring of graphs without short cycles", Discrete Math., 26 (1979) 165–179.

[149] U. Montanari, "Networks of constraints: Fundamental properties and applications to picture processing," Information Sciences 7 (1974) 95–132.

[150] J. Nešetřil, "Aspects of Structural Combinatorics", Taiwanese J. Math. 3, 4 (1999), 381–424.

[151] J. Nešetřil, "Ramsey Theory.", in: **Handbook of Combinatorics** (ed. R. L. Graham, M. Grötschel, L. Lovász), Elsevier, 1995, pp. 1331–1403.

[152] J. Nešetřil, "Many facets of dualities", in: Research Trends in Combinatorial Optimization (W. Cook, L. Lovász, J. Vygen, eds.), Springer, Berlin 2009.

[153] J. Nešetřil, P. Ossona de Mendez, "Cuts and Bounds", Discrete Math. 302 (2005) 211–224.

[154] J. Nešetřil, P. Ossona de Mendez, "Folding", J. Comb. Theory B 96 (2006) 730–739.

[155] J. Nešetřil, P. Ossona de Mendez, "Tree depth, subgraph coloring and homomorphism bounds", European J. Comb. 27 (2006) 1022–1041.

[156] J. Nešetřil, P. Ossona de Mendez, "Low tree-width decompositions and algorithmic consequences", in STOC'06, Proceedings of the 38th Annual ACM Symposium on Theory of Computing, ACM Press 2006, pp. 391–400.

[157] J. Nešetřil, P. Ossona de Mendez, "Structural properties of sparse graphs", in: Building bridges - between mathematics and computer science (ed. M. Grötschel, G. O. H. Katona), Springer 2008, pp. 369–426.

[158] J. Nešetřil, P. Ossona de Mendez, "Grad and classes with bounded expansion I. Decompositions", European Journal of Combinatorics 29(3) (2008), 760–776.

[159] J. Nešetřil, P. Ossona de Mendez, "Grad and classes with bounded expansion II. Algorithmic aspects", European Journal of Combinatorics 29 (2008), 777–791.

[160] J. Nešetřil, P. Ossona de Mendez, "Grad and Classes with bounded expansion III. Restricted Dualities", European J. Comb. 29 (2008) 1012–1024.

[161] J. Nešetřil, P. Ossona de Mendez, "First order properties of nowhere dense structures", J. Symb. Logic submitted (2008).

[162] J. Nešetřil, A. Pultr. "On classes of relations and graphs determined by subobjects and factorobjects", Discrete Math. 22 (1978), 287–300.

[163] J. Nešetřil, V. Rödl, "Chromatically optimal rigid graphs", J. Combinatorial Th. B 46 (1989) 122–141.

[164] J. Nešetřil, M. Siggers, L. Zádori, "A Combinatorial constraint satisfaction problem dichotomy classification conjecture, manuscript 2007.

[165] J. Nešetřil, M. Siggers, "A combinatorial proof that subprojective constraint satisfaction problems are NP-complete", in: 32nd International Symposium MFCS 2007, Lecture Notes in Computer Science 4708, Springer 2007, pp. 159–170.

[166] J. Nešetřil, C. Tardif, "A dualistic approach to bounding the chromatic number of a graph", European J. Comb. 29 (2008) 254–260.

[167] J. Nešetřil, C. Tardif, "Homomorphism duality: on short answers to exponentially long questions", SIAM J. Discrete Math. 19 (2005), 914–920.

[168] J. Nešetřil and C. Tardif, "Duality theorems for finite structures (characterising gaps and good characterizations)", J. Combin. Theory B 80 (2000), 80–97.

[169] J. Nešetřil, X. Zhu, "On sparse graphs with given colorings and homomorphisms", J. Comb. Th. B 90 (2004) 161–172.

[170] E. Pesch, **Retracts of Graphs**, Athenaeum Verlag, Frankfurt 1988.

[171] V. Pratt and J. Tiuryn, "Satisfiability of inequalities in a poset", Fundamenta Informaticae 28 (1996) 165–182.

[172] A. Pultr, V. Trnková, **Combinatorial, Algebraic and Topological Representations of Groups, Semigroups and Categories**, North-Holland, Amsterdam, 1980.

[173] A. Quilliot, "A retraction problem in graph theory", Discrete Math. 54 (1985) 61 – 71.

[174] J. Radhakrishnan, M. Sudan, "On Dinur's proof of the PCP theorem", Bull. Amer. Math. Soc. 44 (2007) 19–61.

[175] B. Rossman "Existential positive types and preservation under homomorphisms", in: 20th IEEE Symposium on Logic in Computer Science (LICS), 2005, pp. 467–476.

[176] B. Roy, "Nombre chromatique et plus longs chemins d'un graph", Rev. Francaise Informat. Recherche Opérationelle 1 (1967), 129–132.

[177] L. M. Vitaver, "Determination of minimal coloring of vertices of a graph by means of Boolean powers of the incidence matrix" (in Russian), Dokl. Akad. Nauk SSSR 147 (1962), 758–759.

[178] A. Salazar and R. V. Oakford, "A graph formulation of a school scheduling algorithm", Comm. ACM 17 (1974) 696–698.

[179] T. J. Schaefer, "The complexity of satisfiability problems," Proc. 10th ACM Symp. on Theory of Computing (1978), 216–226.

[180] G. Schell, Tridiagonal matrix partitions, M.Sc. thesis, SFU 2008.

[181] R.E. Tarjan, Decomposition by clique separators, *Discrete Mathematics* 55 (1985) 221–232.

[182] E. P. K. Tsang, **Foundations of Constraint Satisfaction**, Academic Press, London and San Diego (1993).

[183] M.Y. Vardi, "Constraint satisfaction and database theory: a tutorial," Proceedings of the 19th Symposium on Principles of Database Systems (PODS 2000), 76–85.

[184] N. Vikas, "Computational complexity of compaction to reflexive cycles" SIAM J. on Computing 32 (2003) 253 – 280.

[185] M. Wand and P. M. O'Keefe, "On the complexity of type inference with coercion," in Conf. on Functional Programming Languages and Computer Architecture (1989), 293–298.

[186] M. Sen, S. Das, A.B. Roy, and D.B. West, Interval digraphs: An analogue of interval graphs, J. Graph Theory 13 (1989) 581 - 592.

[187] M. H. Siggers "H-colouring dichotomy re-revisited", manuscript 2008.

[188] R. E. Tarjan, "Decomposition by clique separators", Discrete Math. 55 (1985) 221 – 232.

[189] C. Thomassen, "Grötzsch's 3-color theorem and its counterparts for torus and the projective plane", J. Comb. Th. B, 62 (1994), 268-279.

[190] W. Xie, "Forbidden subgraph characterizations of matrix partitions", M.Sc. thesis, Simon Fraser University 2006.