Discrete Mathematics 311 (2011) 1543–1584

Contents lists available at ScienceDirect

Discrete Mathematics

journal homepage: www.elsevier.com/locate/disc

# Forbidden substructures and combinatorial dichotomies: WQO and universality

Gregory Cherlin

*Department of Mathematics, Rutgers University, 110 Frelinghuysen Road, Piscataway, N.J. 08854, United States*

ARTICLE INFO

*Article history:*
Available online 11 April 2011

*Keywords:*
Universal graph
Well-quasi-order
Antichain
Model theory
Decision problems

ABSTRACT

We discuss two combinatorial problems concerning classes of finite or countable structures of combinatorial type. We consider classes determined by a finite set of finite constraints (forbidden substructures). Questions about such classes of structures are naturally viewed as algorithmic decision problems, taking the finite set of constraints as the input. While the two problems we consider have been studied in a number of natural contexts, it remains far from clear whether they are decidable in their general form. This broad question leads to a number of more concrete problems. We discuss twelve open problems of varying levels of concreteness, and we point to the "Hairy Ball Problem" as a particularly concrete problem, which we give first in direct model theoretic terms, and then decoded as an explicit graph theoretic problem.



## 1. Introduction

### *1.1. Dichotomies for combinatorial structures*

We will discuss two problems which concern classes of combinatorial structures—in the first case finite structures, and in the second case countably infinite ones. The classes we consider are defined by finitely many constraints provided by "forbidden substructures".[1] Influenced by logic—complexity theory on the one hand, model theory on the other—we tend to put these problems in a very broad context, but open questions abound at all levels. A considerable body of concrete work has been undertaken on both problems in a number of contexts, but there is a great deal of similar territory remaining largely unexplored. Our survey includes some new results that we find clarifying. We have put most of the detailed discussion of the new results in three Appendices, referring to them as needed in the text, with an indication of the line of argument. This includes some results to the effect that "Here there be tygers", which are intended to justify some of the restrictions we impose.

One of the aims of general model theory has been to prove a dichotomy for the behavior of the most general classes of structures: the so-called "structure/nonstructure" alternative, in Shelah's parlance. According to this dichotomy, when one looks at large infinite models of first order theories, one either has a coherent structure theory which in the first instance allows one to estimate the number of models, and to proceed from there to more delicate results, or on the other hand one finds a degree of chaos which can be expressed in a number of ways, the essential point being that the behavior of the models in the nonstructured case is more a matter of set theory than of algebraic structure.

Are there any similar phenomena in the world of finite (or nearly finite) combinatorics? We will confine ourselves to classes of structures with very simple definitions, namely with classes defined by finitely many constraints of the simplest

*E-mail address:* cherlin@math.rutgers.edu.

1 Graph theorists and model theorists use the term "substructure" in distinct ways; see Note 2 in Section 4. We follow graph theoretic usage here.

0012-365X/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.disc.2011.03.014

kind: forbidden substructures. We consider notions of "tameness" and "wildness" appropriate to this context, and we undertake to analyze the gap between the tame and the wild.

The two notions of tameness with which we will work are the following: first, well-quasi-order; second, the existence of a countable universal object. If we followed the pattern of model theory exactly, we would be looking to show that the wild case is extremely wild in some sense; in the first of our two cases we doubt this, and in the second case, while it seems to be true, it is not really the point. For us, the natural question at this level is whether the separation between the tame and wild cases is *effective* (algorithmically decidable). Indeed, that is simply a precise way of stating that the two cases can be clearly separated. For our two interpretations of tameness—and no doubt, many others—it is completely unclear at this stage whether such a separation occurs. All one can really say to date is that when one works on instances of these problems, they seem difficult, and not entirely unlike some known undecidable problems.

Let us take up these two problems one at a time.

### *1.2. The WQO problem*

Here we deal with the class $\mathcal{Q}$ of all finite structures of a particular combinatorial type. This may be the class of (finite) graphs, tournaments, digraphs, permutation patterns, matroids, and such. We take a finite subset $\mathcal{C}$ of $\mathcal{Q}$, the *forbidden substructures*, and consider the subclass $\mathcal{Q}_{\mathcal{C}}$ of structures in $\mathcal{Q}$ containing no substructure isomorphic to any $C$ in $\mathcal{C}$. A note on terminology: we use the term "substructure" here in much the same way that graph theorists use the term "subgraph;" and this is not consistent with standard model theoretic terminology. See Note 2, Section 4.3 for more on this point, and also Section 1.6 and Appendix C.

As $\mathcal{Q}$ is not actually a set, one may prefer to cut it down by taking all structures under consideration to have their elements in a fixed countable set; or indeed by working with isomorphism types rather than structures. We will not concern ourselves with the choice of formalism.

The relation that interests us here is the embeddability relation on $\mathcal{Q} : a \leq b$ if $a$ is isomorphic with a substructure of $b$. Then $\mathcal{Q}$ is a quasi-order, and the equivalence relation given by $a \leq b \leq a$ is the relation of isomorphism. All of these quasi-orders are *well-founded*, that is there is no infinite strictly descending sequence $a_1 > a_2 > \cdots$.

In general, a quasi-order is said to be well-quasi-ordered (wqo) if it is both well-founded and contains no infinite antichain (i.e., set of pairwise incomparable elements). The problem we wish to consider—in its first formulation—is the following.

**Problem** $(A)$**.** With $\mathcal{Q}$ and $\mathcal{C}$ specified, is $\mathcal{Q}_{\mathcal{C}}$ wqo? In other words, does $\mathcal{Q}_{\mathcal{C}}$ contain an infinite antichain?

We consider some illustrative examples.

**Fact 1.1.**

1. *Let L be a finite linear tournament. Then the L-free tournaments are wqo (in fact of bounded size, by Ramsey's theorem) ....*
2. *But if T is a nonlinear tournament, with at least 7 vertices, then the T-free tournaments are not wqo (by* [29]*, because of two very special antichains serving to witness this in all cases).*

This gives the following corollary.

**Corollary 1.2.** *The finite tournaments T for which the class of T-free tournaments is wqo can be recognized in polynomial time.*

Results of this kind often have a paradoxical quality: Fact 1.1 does not actually tell us how to determine which side of the fence a particular constraint $T$ will actually fall, if $T$ is nonlinear and very small, nor does it give us any hint as to how one should find out in such cases. But once the number of cases left unsettled is finite, and the others are cleanly handled, the problem becomes polynomial time decidable. At the same time, it is precisely the finitely many cases left over that tend to be the real challenges in practice, and in the present instance it took extensive structural analyses of the classes $\mathcal{Q}_T$ associated with two of these "left over" tournaments $T$, and then an application of Kruskal's tree theorem [27], to convert this abstract statement into a definite answer.

Thus a proof that a problem is solvable is not at all the same thing as a solution, and the distinction is worth bearing in mind. But we find the question, whether such combinatorial problems are solvable in principle at a systematic level, to be one with its own interest.

At the level of generality of the problems we consider, algorithmic decidability per se is the natural question. But one curious feature of the wqo problem is that decidability results are obtained by noneffective methods, and that the resulting algorithms whose existence is proved are "good" in the conventional sense of polynomial time computability, even though no single correct algorithm is produced, and for that matter in certain cases no explicit bound on the degree of the associated polynomial can be extracted from the decidability proof. This is not a new phenomenon; it comes with the general territory of wqo theory [16, Section 8].

We restate our problem in the form that actually concerns us.

**Problem** $(A_{\mathcal{Q}})$**.** With $\mathcal{Q}$ fixed, for example the class of finite tournaments, and with $\mathcal{C}$ varying, is Problem $A$ effectively solvable (and if so, in polynomial time)? That is the function taking us from the specification of $\mathcal{C}$ to the answer, a computable function?