Promise Constraint Satisfaction and Width*

Albert Atserias
Universitat Politècnica de Catalunya

Víctor Dalmau
Universitat Pompeu Fabra

July 13, 2021

Abstract

We study the power of the bounded-width consistency algorithm in the context of the fixed-template Promise Constraint Satisfaction Problem (PCSP). Our main technical finding is that the template of every PCSP that is solvable in bounded width satisfies a certain structural condition implying that its algebraic closure-properties include weak near unanimity polymorphisms of all large arities. While this parallels the standard (non-promise) CSP theory, the method of proof is quite different and applies even to the regime of sublinear width. We also show that, in contrast with the CSP world, the presence of weak near unanimity polymorphisms of all large arities does not guarantee solvability in bounded width. The separating example is even solvable in the second level of the Sherali-Adams (SA) hierarchy of linear programming relaxations. This shows that, unlike for CSPs, linear programming can be stronger than bounded width. A direct application of these methods also show that the problem of q-coloring p-colorable graphs is not solvable in bounded or even sublinear width, for any two constants p and q such that $3 \leq p \leq q$. Turning to algorithms, we note that Wigderson's algorithm for $O(\sqrt{n})$-coloring 3-colorable graphs with n vertices is implementable in width 4. Indeed, by generalizing the method we see that, for any $\epsilon > 0$ smaller than $1/2$, the optimal width for solving the problem of $O(n^\epsilon)$-coloring 3-colorable graphs with n vertices lies between $n^{1-3\epsilon}$ and $n^{1-2\epsilon}$. The upper bound gives a simple $2^{\Theta(n^{1-2\epsilon}\log(n))}$-time algorithm that, asymptotically, beats the straightforward $2^{\Theta(n^{1-\epsilon})}$ bound that follows from partitioning the graph into $O(n^\epsilon)$ many independent parts each of size $O(n^{1-\epsilon})$.

*Work partially supported by Ministerio de Ciencia e Innovación (MICIN) through project PID2019-109137GB-C22 (PROOFS).

1 Introduction

The input to the constraint satisfaction problem (CSP) is a set of variables, each ranging over a specified domain of values, as well as a set of constraints, each binding a finite set of variables to take values in a relation from a specified set of relations. The problem asks to find an assignment of values to the variables in such a way that all the constraints are satisfied. It was first pointed out by Feder and Vardi [28] that the CSP can be modelled as the homomorphism problem for relational structures. In this view, the input to the CSP is a pair of relational structures, the *instance* $\mathbf{I}$ and the *constraint language* $\mathbf{S}$, and we are asked to find a homomorphism from $\mathbf{I}$ to $\mathbf{S}$; i.e., a map from the domain of $\mathbf{I}$ to the domain of $\mathbf{S}$ in such a way that all the relations are preserved. In the *fixed-template* variant of the problem, the constraint language is fixed and part of the definition of the problem, and the instance is the only input. The fixed-template CSP with template $\mathbf{S}$ is denoted by $\mathrm{CSP}(\mathbf{S})$.

As confirmed in retrospect, one of the most important problems raised by the seminal work of Feder and Vardi was that of characterizing the class of fixed-template CSPs of *bounded width*. In short, these are the fixed-template CSPs whose instances are always correctly solved by the so-called consistency algorithm with a fixed bound on its *width*. This class of CSPs was later studied in depth in the work of Kolaitis and Vardi [37, 36], where it was shown to constitute a robust fragment of the class of polynomial-time solvable CSPs that admits several equivalent reformulations.

After two decades of work in the area, the question of fully characterizing the CSPs of bounded-width was eventually answered in the work of Barto and Kozik [9]. Besides resolving one of the key problems in [28], the Barto-Kozik Theorem produced many important new insights for the theory of CSPs itself. This line of work eventually led to the proof of the celebrated Feder-Vardi Dichotomy Conjecture, due to Zhuk [51] and, independently, Bulatov [18]. This completed the program started by Feder and Vardi [28], Jeavons, Cohen, and Gyssens [33], and Bulatov, Jeavons and Krokhin [19], that aimed to characterize the class of all polynomial-time solvable fixed-template CSPs by the algebraic closure properties of their templates. By today, it is fair to say that almost all important questions raised by the early work of Feder and Vardi for the fixed-template CSP seem to have been resolved.

Recently, Brakensiek and Guruswami [12] have put forward a generalization of the fixed-template CSP as a natural next step in the development of the theory. In the *Fixed Template Promise Constraint Satisfaction Problem* (PCSP) the template is a fixed pair $(\mathbf{S}, \mathbf{T})$ of relational structures such that there is a homomorphism from $\mathbf{S}$ to $\mathbf{T}$. The problem has two variants:

> *Search variant of* $\mathrm{PCSP}(\mathbf{S}, \mathbf{T})$: The instance is a relational structure $\mathbf{I}$ with the promise that there is a homomorphism from $\mathbf{I}$ to $\mathbf{S}$ and we are asked to find a homomorphism from $\mathbf{I}$ to $\mathbf{T}$.
>
> *Decision variant of* $\mathrm{PCSP}(\mathbf{S}, \mathbf{T})$: The instance is a relational structure $\mathbf{I}$ and we are asked to distinguish the case in which there is a homomorphism from $\mathbf{I}$ to $\mathbf{S}$ from the case in which there is not even a homomorphism from $\mathbf{I}$ to $\mathbf{T}$. For this variant the promise is that $\mathbf{I}$ fullfils one of these two conditions.

To motivate the generalization, it is useful to look at some special cases. First, it is obvious that PCSP($\mathbf{S}, \mathbf{S}$) is just the same problem as CSP($\mathbf{S}$), hence any fixed-template CSP is a fixed-template PCSP. Second, consider the following graph-coloring problem. We are given a graph $\mathbf{G}$ with the promise that it has an unknown proper p-coloring for a certain number of colors p, and we are asked to find a proper q-coloring for a fixed put potentially larger target number of colors q. By viewing the proper q-colorings of $\mathbf{G}$ as the homomorphisms into the complete graph $\mathbf{K}_q$ with q vertices, PCSP($\mathbf{K}_p, \mathbf{K}_q$) models the problem of approximating the chromatic number of a graph. We refer to this as the Approximate Graph Coloring Problem with parameters p and q. The special case with $p = q$ is exactly the same as CSP($\mathbf{K}_q$), the standard Graph q-Coloring Problem for undirected graphs, which, for $q \geq 3$, is one of the twenty-one NP-complete problems of Karp [34]. The problem of 4-coloring 3-colorable graphs, PCSP($\mathbf{K}_3, \mathbf{K}_4$), was shown NP-hard in the early 1990's, using the theory of probabilistically checkable proofs (PCP) [35]. Not until two decades later has it been shown that the problem of 5-coloring 3-colorable graphs is NP-hard, and, more generally, that PCSP($\mathbf{K}_p, \mathbf{K}_{2p-1}$) is NP-hard for any $p \geq 3$ [8]. The problem of exactly identifying the pairs (p, q) with $p \leq q$ for which PCSP($\mathbf{K}_p, \mathbf{K}_q$) is NP-hard is one of the leading open problems in the area, with deep connections with the theories of PCPs and hardness of approximation [43], and the notorious Unique Games Conjecture [27].

The theory of Promise PCSP has been further developped (see [6, 7, 8, 13, 14, 15, 16, 17, 29, 39, 48] for example) and applied to other problems beyond graph coloring. The goal of this paper is to initiate a study of the power of the consistency algorithm for the fixed-template PCSP. Following [8], in this new context, the consistency algorithm can be thought of in the following terms. Fix a PCSP template ($\mathbf{S}, \mathbf{T}$) and an integer $k \geq 1$. We say that PCSP($\mathbf{S}, \mathbf{T}$) is *solvable in width* k if, for every instance $\mathbf{I}$, it holds that if $\mathbf{I}$ is k-consistent with respect to the left-template $\mathbf{S}$, then there is a homomorphism from $\mathbf{I}$ to the right-template $\mathbf{T}$. Here, as in the standard consistency algorithm for the standard CSP, an instance $\mathbf{I}$ is said to be k*-consistent with respect to* $\mathbf{S}$ if the algorithm that checks if the set of all subinstances of $\mathbf{I}$ with at most k elements admits a system of compatible homomorphisms into $\mathbf{S}$ does not find a blatant contradiction. For each fixed integer k, this algorithm runs in polynomial time. When PCSP($\mathbf{S}, \mathbf{T}$) is solvable in width k we also say that the template ($\mathbf{S}, \mathbf{T}$) *has* width k.

A key insight from the theory of CSPs is that the computational complexity of a fixed-template CSP is governed by the algebraic structure of the polymorphisms of its template [33]. This phenomenon led to the so-called *algebraic approach* to CSPs, of which the aforementioned Barto-Kozik, Zhuk, and Bulatov Theorems are prime examples. Indeed, the governing power of polymorphisms is so general that the phenomenon has been observed to hold for other ways of measuring the complexity of the problem. In particular, it applies to the analysis of bounded width [42], to the more general settings of descriptive complexity [1] and propositional proof complexity [3], and even to different variants of the CSP itself. For PCSPs, a suitable definition of polymorphisms has been put forward to show that the computational and width complexity of a PCSP is, again, governed by the polymorphisms of its template [8].

A natural question at this point is whether the analogue of the Barto-Kozik Theorem holds for PCSPs. In the language of polymorphisms, the Barto-Kozik Theorem for CSPs can be stated as the equivalence of the following two statements:

(a) CSP($\mathbf{S}$) is solvable in some bounded width,

(b) CSP($\mathbf{S}$) admits weak near unanimity (WNU) polymorphisms of all large arities.

In definition, this last condition means that, for any large enough integer m, there exists a homomorphism $f : \mathbf{S}^m \to \mathbf{S}$ from the m-power $\mathbf{S}^m$ into $\mathbf{S}$ that satisfies the *weak near unanimity* (WNU) identities:

$$f(y, x, x, \ldots, x) = f(x, y, x, \ldots, x) = \cdots = f(x, \ldots, x, x, y), \tag{1}$$

for all choices of x and y in the domain of $\mathbf{S}$.

Turning to PCSP, for a fixed template $(\mathbf{S}, \mathbf{T})$, consider the following statements:

(a') PCSP($\mathbf{S}, \mathbf{T}$) is solvable in some bounded width,

(b') PCSP($\mathbf{S}, \mathbf{T}$) admits WNU polymorphisms of all large arities.

Following [8], a WNU polymorphism of PCSP($\mathbf{S}, \mathbf{T}$) is a homomorphism $f : \mathbf{S}^m \to \mathbf{T}$ from the m-power $\mathbf{S}^m$ of the left-template $\mathbf{S}$ into the right-template $\mathbf{T}$ that satisfies Equation (1) for all x and y in the domain of $\mathbf{S}$.

The first technically novel result of this paper is that (b') is a *necessary* but *not sufficient* condition for (a'):

Main Result: (a')$\Longrightarrow$ (b') $\not\Longrightarrow$ (a').

To prove that (a') $\Longrightarrow$ (b') we first establish a structural result about the template that may be of independent interest. We show that any template $(\mathbf{S}, \mathbf{T})$ whose PCSP is solvable in some bounded width satisfies the following property: either a proper subset of $S \times S$ can be obtained by composing two binary projections of the relations in $\mathbf{S}$, where S is the domain of $\mathbf{S}$, or else the problem is trivial because $\mathbf{T}$ contains a reflexive tuple and then *any* instance admits a homomorphism to $\mathbf{T}$. In its proof we need to resort to a probabilistic construction of a large instance $\mathbf{I}$ that is *sparse enough* to be k-consistent with respect to $\mathbf{S}$, for any fixed k, but *dense enough* to not admit a homomomorphisms into $\mathbf{T}$.

The reasoning that goes into the analysis of the probabilistic construction of this instance is reminiscent of the methods for proving lower bounds for Resolution in propositional proof complexity, going back to the influential work of Chvátal and Szeméredi [23] and posterior follow-ups (particularly, [10] and [44]). Indeed, as in these related works, our analysis shows that the random instance $\mathbf{I}$ is k-consistent with respect to $\mathbf{S}$ for k as large as ϵn, where n is the number of elements in $\mathbf{I}$, and ϵ is a fixed positive constant that depends only on $\mathbf{S}$ and $\mathbf{T}$. This means that the necessary condition (b') applies also to all PCSPs that are solvable in *sublinear width*; i.e., in width $k = k(n) = o(n)$, where n is the number of elements in the instance. It should be pointed out that, for standard CSPs, it was already known that the

Barto-Kozik Theorem can also be strengthened to show that a fixed-template CSP is solvable in bounded width if and only if it is solvable in sublinear width (see, e.g., [3]).

To prove that (b') $\not\Longrightarrow$ (a'), we analyze the polymorphisms of a specific *Boolean* PCSP template, i.e., one with a two-element domain $\{0, 1\}$, namely $(2\text{-IN-4-SAT}, 4\text{-NAE-SAT})$. We first show that this template admits WNU polymorphisms of all large arities, and then apply the structural result of the previous paragraph to conclude that it is not solvable in bounded width. This analysis also led us to conclude that, for any two integers s and r such that $0 < s < r$ and $r > 2$, the Boolean PCSP(s-IN-r-SAT, r-NAE-SAT) is not solvable in bounded width, but is solvable in the second level of the Sherali-Adams hierarchy applied to its basic linear programming relaxation. This is in sharp constrast with the status for standard CSPs for which, as is known, the fixed-template CSPs that are solvable in bounded-width and those that are solvable in some fixed-level of the Sherali-Adams hierarchy coincide (this follows, e.g., from the Barto-Kozik Theorem combined with the results in [1] and [2]). To show that PCSP(s-IN-r-SAT, r-NAE-SAT) can be solved in the second level of the Sherali-Adams hierarchy we build on the recent results on Boolean PCSPs from [12].

As a corollary to the aforementioned structural result we obtain a complete classification of the Approximate Graph Coloring Problems that are solvable in bounded width. We show that $\mathrm{PCSP}(\mathbf{K}_p, \mathbf{K}_q)$ is not solvable in bounded width for *any* two constants p and q, such that $p \leq q$, unless $p = 1$ or $p = 2$.

Corollary 1. *For any two integers p and q such that $1 \leq p \leq q$, the following statements are equivalent:*

(a) $\mathrm{PCSP}(\mathbf{K}_p, \mathbf{K}_q)$ *is solvable by the consistency algorithm in width* 3*,*

(b) $\mathrm{PCSP}(\mathbf{K}_p, \mathbf{K}_q)$ *is solvable by the consistency algorithm in bounded width,*

(c) $\mathrm{PCSP}(\mathbf{K}_p, \mathbf{K}_q)$ *is solvable by the consistency algorithm in sublinear width,*

(d) $p = 1$ *or* $p = 2$*.*

We note that this classification agrees with the one predicted by the NP-hardness results that would follow from certain variants of the Unique Games Conjecture [27], but ours is unconditional.

Turning to upper bounds for the Approximate Graph Coloring Problem, we observe that the well-known algorithm due to Wigderson [47] that properly colors any 3-colorable graph with n vertices with $O(\sqrt{n})$ colors is implementable in width 4. We generalize Wigderson's algorithm to show that, for any fixed real ϵ in the interval $(0, 1/2)$, the problem of distinguishing 3-colorable from non-$O(n^\epsilon)$-colorable graphs with n vertices can be solved in width $n^{1-2\epsilon}$. This leads to a simple algorithm that properly colors any 3-colorable graph with $O(n^\epsilon)$ colors in time $2^{\Theta(n^{1-2\epsilon}\log(n))}$. Asymptotically, this beats the straightforward $2^{\Theta(n^{1-\epsilon})}$ time-bound that follows from partitioning the graph into $O(n^\epsilon)$ many independent parts each of size $O(n^{1-\epsilon})$. As a nearly matching lower bound, we show that the same problem cannot be solved in width less than $n^{1-3\epsilon}$. The problem of closing the gap between the $1 - 3\epsilon$ in the lower bound and the $1 - 2\epsilon$ in the upper bound is left as an intriguing open problem.

2 Preliminaries

For an integer n, we shall use $[n]$ to denote the set $\{1, 2, \ldots, n\}$.

Tuples and relations Let A be a set and let k be a positive integer. A *k-tuple over A* is as a sequence $\mathbf{t} = (\mathbf{t}(1), \ldots, \mathbf{t}(k))$, where $\mathbf{t}(1), \ldots, \mathbf{t}(k)$ are elements of A; equivalently, a k-tuple over A can be seen as a map $\mathbf{t} : [k] \to A$ with domain $[k]$ and range in A. The set of k-tuples over A is denoted by A^k. If $\mathbf{j} = (\mathbf{j}(1), \ldots, \mathbf{j}(m))$ is an m-tuple over $[k]$ and $\mathbf{a}$ is a k-tuple, then the *projection* $\mathrm{pr}_{\mathbf{j}}\,\mathbf{a}$ of $\mathbf{a}$ on $\mathbf{j}$ is the m-tuple $(\mathbf{a}(\mathbf{j}(1)), \ldots, \mathbf{a}(\mathbf{j}(m)))$. If $J = \{j_1, \ldots, j_m\}$ is a subset of $[k]$ with $j_1 < \cdots < j_m$, then the projection $\mathrm{pr}_J\,\mathbf{a}$ of $\mathbf{a}$ on J is the m-tuple $(\mathbf{a}(j_1), \ldots, \mathbf{a}(j_m))$.

A *relation* R of arity k over A is a subset $R \subseteq A^k$ of k-tuples over A. For every non-negative integer r, every relation $R \subseteq A^r$ of arity r over A, and every k-tuple or k-subset J over $[r]$, the *projection* $\mathrm{pr}_J\,R$ of R on J is the k-ary relation $\{\mathrm{pr}_J\,\mathbf{a} \mid \mathbf{a} \in R\}$. For any two binary relations $R \subseteq A^2$ and $S \subseteq A^2$ over A, their *composition* $R \circ S$ is the binary relation $\{(a, b) \mid \text{ there exists } c \in A \text{ such that } (a, c) \in R \text{ and } (c, b) \in S\}$. For any two relations $R \subseteq A^r$ and $S \subseteq A^s$ over A of arities r and s, their *product* $R \times S$ is the relation $\{\mathbf{t} \in A^{r+s} \mid \mathrm{pr}_{[r]}\,\mathbf{t} \in R \text{ and } \mathrm{pr}_{[r+s]\setminus[r]}\,\mathbf{t} \in S\}$. The iterated product of more than two relations $R_1, \ldots, R_m$ is denoted by $\prod_{i=1}^m R_i$. If all R_i are the same relation R, then we call it the m-power of R and denote it by R^m.

Let $f : A \to B$ be a map. We shall use $\mathrm{dom}(f)$ to denote its domain A and $\mathrm{img}(f)$ to denote its image $f(A)$. It will be convenient to allow functions with empty domain (but note that there is a unique function with empty domain). If $X \subseteq \mathrm{dom}(f)$ then we shall use $f|_X$ to denote the *restriction* of f to X, i.e., the unique map g with $\mathrm{dom}(g) = X$ that agrees with f on X. If $g = f|_X$ for some X we shall say that f is an *extension* of g. We write $g \subseteq f$ to denote the fact that f is an extension of g. For any k-tuple $\mathbf{a} \in A^k$ we shall use $f(\mathbf{a})$ to denote the k-ary tuple obtained by applying f to $\mathbf{a}$ component-wise, i.e., $f(\mathbf{a}) = (f(\mathbf{a}(1)), \ldots, f(\mathbf{a}(k)))$. For two finite sets A and B and a non-negative integer k, we write $M_k(A, B)$ to denote the set of maps from a subset of A into a subset of B with a domain of cardinality at most k. For each $f \in M_k(A, B)$, let $v_f = (v_{f,a,b} \mid (a, b) \in A \times B)$ denote the *characteristic vector* of f, i.e., $v_{f,a,b} = 1$ if $a \in \mathrm{dom}(f)$ and $f(a) = b$, and $v_{f,a,b} = 0$ otherwise, for all $a \in A$ and $b \in B$.

Relational structures A *signature* is a finite collection of relation symbols R, each of them with an associated non-negative integer called the *arity* of R. If σ is a signature, then a *σ-structure* $\mathbf{A}$ consists of a set A, called the *domain* of $\mathbf{A}$, and a relation $R^{\mathbf{A}} \subseteq A^r$ for each R in σ, where r is the arity of R, called the *interpretation* of R in $\mathbf{A}$. We shall use the same capital letter to denote the universe of a σ-structure, e.g., A is the universe of $\mathbf{A}$. All signatures and structures in this paper are assumed to be finite, i.e., they have a finite number of relations over a finite domain.

Let σ be a signature. For any two σ-structures $\mathbf{A}$ and $\mathbf{B}$ with domains A and B, their *union* $\mathbf{A} \cup \mathbf{B}$ is the σ-structure with domain $A \cup B$ and interpretations $R^{\mathbf{A}\cup\mathbf{B}} = R^{\mathbf{A}} \cup R^{\mathbf{B}}$ for all $R \in \sigma$. We say that $\mathbf{A}$ is a *substructure* of $\mathbf{B}$ if $A \subseteq B$ and $R^{\mathbf{A}} \subseteq R^{\mathbf{B}}$ hold for all $R \in \sigma$. If, furthermore, $R^{\mathbf{A}} = R^{\mathbf{B}} \cap A^r$ for all $R \in \sigma$, where r is the arity of R, then we say that $\mathbf{A}$ is an *induced substructure* of $\mathbf{B}$. We say that the set A *induces* $\mathbf{A}$ in $\mathbf{B}$ and write $\mathbf{B}|_A$ to denote the substructure of $\mathbf{B}$ *induced by* A.

For every integer $n \geq 1$ and every σ-structure $\mathbf{A}$, the *n-power* of $\mathbf{A}$ is the σ-structure $\mathbf{A}^n$ with domain A^n and interpretations $R^{\mathbf{A}^n} = \{(\mathbf{t}_1, \ldots, \mathbf{t}_r) \in (A^n)^r \mid (\mathbf{t}_1(i), \ldots, \mathbf{t}_r(i)) \in R^{\mathbf{A}}$ for all $i \in [n]\}$ for all $R \in \sigma$, where r is the arity of R. Note that $R^{\mathbf{A}^n}$ is *not* literally the same relation as the n-power $(R^{\mathbf{A}})^n$ of the relation $R^{\mathbf{A}}$, but the two relations are the same up to flattening the tuples as elements of A^{rn} and permuting the coordinates.

Homomorphisms, CSPs and PCSPs Let σ be a signature and let $\mathbf{A}$ and $\mathbf{B}$ be σ-structures with domains A and B. A map $f : A \to B$ is a *homomorphism* from $\mathbf{A}$ to $\mathbf{B}$ if, for all $R \in \sigma$ and $\mathbf{a} \in A^r$, where r is the arity of R, it holds $\mathbf{a} \in R^{\mathbf{A}}$ implies $f(\mathbf{a}) \in R^{\mathbf{B}}$. We shall use $\mathbf{A} \to \mathbf{B}$ to denote the statement that there exists an homomorphism from $\mathbf{A}$ to $\mathbf{B}$.

We write CSP($\mathbf{A}$) to denote the following computational problem: *given an input σ-structure* $\mathbf{I}$*, decide whether* $\mathbf{I} \to \mathbf{A}$ *or* $\mathbf{I} \not\to \mathbf{A}$. The structure $\mathbf{A}$ is the *CSP template* of the problem CSP($\mathbf{A}$). If $\mathbf{A} \to \mathbf{B}$, we write PCSP($\mathbf{A}, \mathbf{B}$) to denote the following computational promise problem: *given an input σ-structure* $\mathbf{I}$*, distinguish the case* $\mathbf{I} \to \mathbf{A}$ *from the case* $\mathbf{I} \not\to \mathbf{B}$*, provided one of these cases holds (if not, any answer is valid).* The pair of structures ($\mathbf{A}, \mathbf{B}$) is the *PCSP template* of the problem PCSP($\mathbf{A}, \mathbf{B}$). Since it is assumed that $\mathbf{A} \to \mathbf{B}$, note that either $\mathbf{A}$ is the empty structure, or $\mathbf{B}$ cannot be trivial in the sense that not all the relations of $\mathbf{B}$ can be empty. Note also that the definition of the problem PCSP($\mathbf{A}, \mathbf{B}$) assumes that the input structure $\mathbf{I}$ satisfies $\mathbf{I} \to \mathbf{A}$ or $\mathbf{I} \not\to \mathbf{B}$, and that these cases are disjoint since $\mathbf{A} \to \mathbf{B}$. Finally, observe that PCSP($\mathbf{A}, \mathbf{A}$) is precisely the same problem as CSP($\mathbf{A}$). To simplify matters we shall take the liberty to use $\mathbf{A}$ to denote the PCSP template ($\mathbf{A}, \mathbf{A}$).

Minors and minions Most of the terminology that follows comes from [8]. Let m and n be positive integers and let A and B be finite sets. An n-ary function $f : A^n \to B$ is called the *minor* of an m-ary function $g : A^m \to B$ given by the map $\pi : [m] \to [n]$ if the following identity holds

$$f(x_1, \ldots, x_n) \approx g(x_{\pi(1)}, \ldots, x_{\pi(m)}), \tag{2}$$

i.e., if the equality $f(x_1, \ldots, x_n) = g(x_{\pi(1)}, \ldots, x_{\pi(m)})$ holds for all $x_1, \ldots, x_n \in A$. Informally, one can say that f is a minor of g if it can be obtained from g by permuting variables, identifying variables, and introducing dummy variables. We shall use g_π to denote the minor of g given by the map π. A *minion* on (A, B) is any non-empty subset of $\{f : A^n \to B \mid n \geq 1\}$ that is closed under minors. Let $\mathscr{M}$ and $\mathscr{N}$ be minions (not necessarily on the same pair of sets). A *minion homomorphism* from $\mathscr{M}$ to $\mathscr{N}$ is any mapping $\xi : \mathscr{M} \to \mathscr{N}$ satisfying the following conditions:

1. it preserves arities, i.e., for every g in $\mathscr{M}$, its image $\xi(g)$ has the same arity as g,

2. it preserves taking minors, i.e., for all integers m and n, all maps $\pi : [m] \to [n]$ and all m-ary functions g in $\mathscr{M}$, we have $\xi(g)_\pi = \xi(g_\pi)$.

Let σ be a signature, let $\mathbf{A}$ be a σ-structure, and let $\mathscr{M}$ be a minion (not necessarily related to $\mathbf{A}$). The *free structure* of $\mathbf{A}$ generated by $\mathscr{M}$ is the σ-structure $\mathbf{F}(\mathbf{A};\mathscr{M})$ defined as follows. Let $n = |A|$, regard the domain $A = \{x_1, \ldots, x_n\}$ of $\mathbf{A}$ as a collection of variables, and define the universe of $\mathbf{F}(\mathbf{A};\mathscr{M})$ to be the set of n-ary functions in $\mathscr{M}$. Let $R \in \sigma$ be any relation symbol, let r be its arity, and let $\mathbf{t}_1, \ldots, \mathbf{t}_m$ be an arbitrary ordering of the tuples in $R^{\mathbf{A}}$. For each $i \in [r]$, let $\pi_i : [m] \to [n]$ be the map defined by $\pi_i(j) = \mathbf{t}_j(i)$, for $j \in [m]$. Then, $R^{\mathbf{F}(\mathbf{A};\mathscr{M})}$ is defined to contain, for every m-ary function $g \in \mathscr{M}$, the tuple $(f_1, \ldots, f_r)$, where $f_i(x_1, \ldots, x_n)$ is the n-ary minor of g given by the map $\pi_i : [m] \to [n]$; i.e., the function f_i is the unique n-ary map that satisfies the identity

$$f_i(x_1, \ldots, x_n) \approx g(\mathbf{t}_1(i), \ldots, \mathbf{t}_m(i)) \tag{3}$$

over the base domain of g. We note here that since every minion $\mathscr{M}$ is closed under permuting the variables of a function, the structure $\mathbf{F}(\mathbf{A};\mathscr{M})$ is well defined in the sense that it does not depend on the choice of ordering of the elements in the domain A of $\mathbf{A}$, or on the choice for the ordering of the tuples of its relations.

Example 1. Let $m \geq 3$ be an integer and let $\mathbf{A}$ be the structure with domain $\{x, y\}$ with a single m-ary relation $R^{\mathbf{A}} = \{(y, x, \ldots, x), (x, y, \ldots, x), \ldots, (x, x, \ldots, y)\}$; the set of all tuples with exactly one occurrence of y. By construction, for every minion $\mathscr{M}$ and every m-ary function $g \in \mathscr{M}$, the relation $R^{\mathbf{F}(\mathbf{A};\mathscr{M})}$ contains the tuple $(f_1, \ldots, f_m)$ where f_i is the binary operation defined by the identity

$$f_i(x, y) \approx g(x, \ldots, x, y, x, \ldots, x), \tag{4}$$

where y appears in position i in the tuple $(x, \ldots, x, y, x, \ldots, x)$ on the right-hand side of (4). Now, if g satisfies the identities

$$g(y, x, x, \ldots, x) \approx g(x, y, x, \ldots, x) \approx \cdots \approx g(x, \ldots, x, x, y), \tag{5}$$

then $f_1 = \cdots = f_m$ and $R^{\mathbf{F}(\mathbf{A};\mathscr{M})}$ contains the reflexive tuple $(f, f, \ldots, f)$ where $f := f_1 = \cdots = f_m$. Conversely, if $R^{\mathbf{F}(\mathbf{A};\mathscr{M})}$ contains a reflexive tuple, then there exists some r-ary function $g \in \mathscr{M}$ satisfying (5). Any function of arity m satisfying (5) is called an m-ary *weak near unanimity* (WNU).

Observation 1. Let $\mathbf{A}$ be a σ-structure and $\mathscr{M}$ be a minion. Then $\mathbf{A} \to \mathbf{F}(\mathbf{A};\mathscr{M})$

Proof. Assume that the universe of $\mathbf{A}$ is $\{x_1, \ldots, x_n\}$ as usual. Since minions are nonempty and closed under identification of variables it follows that $\mathscr{M}$ contains a unary operation g. Hence, it must contain also, for every $i \in [n]$, the n-ary operation f_i defined by the identity $f_i(x_1, \ldots, x_n) \approx g(x_i)$. Define the mapping $\varphi : A \to F(\mathbf{A};\mathscr{M})$ as $\varphi(x_i) = f_i$, for $i \in [n]$. We shall show that φ defines an homomorphism from $\mathbf{A}$ to $\mathbf{F}(\mathbf{A};\mathscr{M})$. Let $R \in \sigma$ and let $\mathbf{t}_1, \ldots, \mathbf{t}_m$ be the ordering of the tuples in $R^{\mathbf{A}}$ used in the definition of $R^{\mathbf{F}(\mathbf{A};\mathscr{M})}$. It remains to be shown that $\varphi(\mathbf{t}_i) \in R^{\mathbf{F}(\mathbf{A};\mathscr{M})}$ for every $i \in [m]$. Indeed, for every $i \in [m]$, let g_i^m be the m-ary function defined by the identity

$$g_i^m(y_1, \ldots, y_m) \approx g(y_i).$$

Note that g_i^m is in $\mathscr{M}$ as it is a minor of g. To complete the proof it suffices to observe that the tuple included in $R^{\mathbf{F}(\mathbf{A};\mathscr{M})}$ due to g_i^m is precisely $\varphi(\mathbf{t}_i)$. □

Polymorphisms Let σ be a signature, let $(\mathbf{A}, \mathbf{B})$ be a PCSP template of signature σ, and let n be a positive integer. An n-ary *polymorphism* of $(\mathbf{A}, \mathbf{B})$ is a homomorphism from $\mathbf{A}^n$ to $\mathbf{B}$; that is, unfolding the definitions, a mapping $f : A^n \to B$ such that, for all $R \in \sigma$ and $\mathbf{t}_1, \ldots, \mathbf{t}_n \in R^{\mathbf{A}}$, it holds that $f(\mathbf{t}_1, \ldots, \mathbf{t}_n) \in R^{\mathbf{B}}$, where

$$f(\mathbf{t}_1, \ldots, \mathbf{t}_n) = (f(\mathbf{t}_1(1), \ldots, \mathbf{t}_n(1)), \ldots, f(\mathbf{t}_1(r), \ldots, \mathbf{t}_n(r))) \tag{6}$$

and r is the arity of R. We denote the set of all polymorphisms of $(\mathbf{A}, \mathbf{B})$ by $\mathrm{Pol}(\mathbf{A}, \mathbf{B})$. As usual we shall use $\mathrm{Pol}(\mathbf{A})$ as a shorthand for $\mathrm{Pol}(\mathbf{A}, \mathbf{A})$. It follows from the definitions that the collection $\mathrm{Pol}(\mathbf{A}, \mathbf{B})$ of all polymorphisms of $(\mathbf{A}, \mathbf{B})$ is a minion.

The following result, pointed out in [8], will be useful.

Observation 2. [8] For every minion $\mathscr{M}$ and every structure $\mathbf{A}$ there exists a minor homomorphism ξ from $\mathscr{M}$ to $\mathrm{Pol}(\mathbf{A}, \mathbf{F}(\mathbf{A}; \mathscr{M}))$.

Proof. To simplify notation assume again that $A = \{x_1, \ldots, x_n\}$. Then ξ maps every operation $g \in \mathscr{M}$ with arity, say, m, to the operation $\xi(g) \in \mathrm{Pol}(\mathbf{A}, \mathbf{F}(\mathbf{A}; \mathscr{M}))$ defined by $\xi(g)(x_{\pi(1)}, \ldots, x_{\pi(m)}) = g_\pi$ for every $\pi : [m] \to [n]$, where g_π is the n-ary minor of g given by map π, i.e., g_π satisfies the identity

$$g_\pi(x_1, \ldots, x_n) \approx g(x_{\pi(1)}, \ldots, x_{\pi(m)})$$

over the base domain of g. It can be readily verified that ξ thus defined defines a minor homomorphism from $\mathscr{M}$ to $\mathrm{Pol}(\mathbf{A}, \mathbf{F}(\mathbf{A}; \mathscr{M}))$. □

3 Relaxations for (P)CSP

A common heuristic method to solve an instance $\mathbf{I}$ of $\mathrm{PCSP}(\mathbf{S}, \mathbf{T})$ consists in solving a polynomial-time solvable relaxation of $\mathbf{I} \to \mathbf{S}$. If the relaxation turns out to be not feasible, then we can infer for sure that $\mathbf{I} \not\to \mathbf{S}$. Although the converse is not necessarily true, for some templates $(\mathbf{S}, \mathbf{T})$ it can be guaranteed that, if the relaxation is feasible, then $\mathbf{I} \to \mathbf{T}$. For such templates, the heurisitic method is a valid polynomial-time algorithm for $\mathrm{PCSP}(\mathbf{S}, \mathbf{T})$. In the present paper we will focus mostly on relaxations based on local consistency.

3.1 Local consistency algorithm

Fix a signature σ. Let $\mathbf{I}$ be an instance with domain I, let $\mathbf{S}$ be a constraint language with domain S, and let k be a positive integer. A mapping $f : X \to S$ with $X \subseteq I$ is a *partial* homomorphism from $\mathbf{I}$ to $\mathbf{S}$ if it is an homomorphism from $\mathbf{I}|_X$ to $\mathbf{S}$. A *k-strategy on* $\mathbf{I}$ *and* $\mathbf{S}$ [38] is any nonempty collection $\mathscr{H}$ of partial homomorphisms from $\mathbf{I}$ to $\mathbf{S}$ such that:

1. the family $\mathscr{H}$ it is closed under restrictions, i.e., for every h in $\mathscr{H}$ and every $X \subseteq \operatorname{dom}(h)$, the restriction $h|_X$ of h to X is in $\mathscr{H}$,
2. the family $\mathscr{H}$ has the extension property up to k, i.e., for every h in $\mathscr{H}$ with $|\operatorname{dom}(h)| < k$ and every $x \in I \setminus \operatorname{dom}(h)$, there exists f in $\mathscr{H}$ such that $h \subseteq f$ and $\operatorname{dom}(f) = \operatorname{dom}(h) \cup \{x\}$.

We write $\mathbf{I} \leq_k \mathbf{S}$ to denote the statement that there exists a k-strategy on $\mathbf{I}$ and $\mathbf{S}$. It follows directly from the definitions that, if $\mathbf{I} \to \mathbf{S}$, then $\mathbf{I} \leq_k \mathbf{S}$. Indeed, if h is a homomorphism from $\mathbf{I}$ to $\mathbf{S}$, then the collection $\mathscr{H} = \{h|_X \mid X \subseteq I, |X| \leq k\}$ is a k-strategy on $\mathbf{I}$ and $\mathbf{S}$.

There is an algorithm that, given $(\mathbf{I}, \mathbf{S}, k)$ as input, decides whether there exists a k-strategy and does so in time polynomial in $(|I| + |S|)^k$. This algorithm, usually called the $(k-1)$-consistency algorithm, starts by placing in $\mathscr{H}$ all partial homomorphisms from $\mathbf{I}$ to $\mathbf{S}$ with domain size at most k and repeatedly removes from $\mathscr{H}$ those h that falsify any one of the two conditions in the definition of k-strategy. The algorithm stops when it reaches a fixed-point. If the fixed-point obtained is non-empty, then it necessarily is a k-strategy. Otherwise, it can be safely concluded that no k-strategy exists. Since, as seen above, $\mathbf{I} \to \mathbf{S}$ implies $\mathbf{I} \leq_k \mathbf{S}$, one could use the consistency algorithm as a partial-check to decide whether $\mathbf{I}$ is homomorphic to $\mathbf{S}$. This is the basis of the *width heuristic* for CSP which, following [8], we now extend to PCSPs.

Let $(\mathbf{S}, \mathbf{T})$ be a PCSP template of signature σ; i.e., $\mathbf{S}$ and $\mathbf{T}$ are σ-structures such that $\mathbf{S} \to \mathbf{T}$. Let $k = k(n)$ be an integer function. We say that $\operatorname{PCSP}(\mathbf{S}, \mathbf{T})$ *is solvable in width* $k(n)$, or that the template $(\mathbf{S}, \mathbf{T})$ *has width* $k(n)$, if for every σ-structure instance $\mathbf{I}$ with n elements it holds that $\mathbf{I} \leq_{k(n)} \mathbf{S}$ implies $\mathbf{I} \to \mathbf{T}$. Note that this generalizes the definition of width for CSPs since, whenever $\operatorname{PCSP}(\mathbf{S}, \mathbf{S}) = \operatorname{CSP}(\mathbf{S})$ is solvable in width $k(n)$, the condition $\mathbf{I} \leq_{k(n)} \mathbf{S}$ is necessary *and* sufficient for $\mathbf{I} \to \mathbf{S}$. We note that every $\operatorname{PCSP}(\mathbf{S}, \mathbf{T})$ is solvable in width at most $k(n)$ for $k(n) = n$. Whenever it is solvable in width $k(n)$ for some $k(n) = O(1)$, we say that $\operatorname{PCSP}(\mathbf{S}, \mathbf{T})$ is solvable *in bounded width*, or that $(\mathbf{S}, \mathbf{T})$ *has* bounded width. Whenever it is solvable in width $k(n)$ for some $k(n) = o(n)$, we say that $\operatorname{PCSP}(\mathbf{S}, \mathbf{T})$ is solvable in *sublinear width*, or that $(\mathbf{S}, \mathbf{T})$ *has* sublinear width.

In the particular case of CSPs, the strength of bounded width is well understood. The breakthrough of Barto and Kozik [9], in combination with [5], yields the following characterization of CSP templates of bounded width:

Theorem 1. [9, 5] *For every structure* $\mathbf{T}$ *the following are equivalent:*

1. $\mathbf{T}$ *has bounded width,*
2. $\mathbf{T}$ *has width* $\max(r, 3)$ *where* r *is the maximum arity in* $\mathbf{T}$*,*

3. Pol($\mathbf{T}$) *contains an* m*-ary WNU for every integer* $m \geq 3$.

We illustrate Theorem 1 with an example that will become useful in Section 6.

Example 2. Consider the problem of deciding whether a graph $\mathbf{G}$ is 2-colorable, i.e, CSP($\mathbf{K}_2$). If $\mathbf{G}$ contains a cycle $a_0, a_1, \ldots, a_n = a_0$ of odd length n, then there is no k-strategy on $\mathbf{G}$ and $\mathbf{K}_2$: an easy inductive argument shows that every such strategy would need to contain for every $j \geq 1$ a partial homomorphism f with domain $\{a_0, a_j\}$ satisfying $f(a_0) = f(a_j)$ if j is even and $f(a_0) \neq f(a_j)$ if j is odd, which is not possible because n is odd and $a_n = a_0$. Consequently, $\mathbf{K}_2$ has width 3.

One can alternatively look at this algebraically and note that, for every integer $m \geq 3$, the set Pol($\mathbf{K}_2$) of polymorphisms of $\mathbf{K}_2$ contains the function $\varphi : [2]^m \to [2]$ that returns the majority of its arguments. These are WNUs and, therefore, by virtue of Theorem 1, it follows also that $\mathbf{K}_2$ has width 3. Indeed, it was already observed in [28] that a structure $\mathbf{T}$ has bounded width whenever Pol($\mathbf{T}$) contains a majority operation f, i.e, an operation $f : S^3 \to S$ satisfying

$$f(y, x, x) \approx f(x, y, x) \approx f(x, x, y) \approx x \tag{7}$$

Note that the first two identities in (7) alone are the WNU identities in (5) for arity 3.

Let us now include sublinear width into the picture. It has been known that Theorem 1 can be strengthened to also include *sublinear width* in the characterization. Concretely, items *1–3* are also equivalent to item *0* below:

0. $\mathbf{T}$ *has sublinear width.*

On one hand, clearly *1* implies *0*. On the other hand, it is known that, if *3* fails, then $\mathbf{T}$ is able to *simulate* the constraint language corresponding to systems of equations over a non-trivial finite Abelian group, for which linear width lower bounds are known (see, e.g., [3]). Thus, *0* implies *3*. As it turns out, this implication will also follow from our main result in Section 4.

As in the CSP world, whether a fixed-template PCSP is solvable in bounded width is controlled by the polymorphisms of the template [8].

Theorem 2. [8] *Let* $(\mathbf{S}_1, \mathbf{T}_1)$ *and* $(\mathbf{S}_2, \mathbf{T}_2)$ *be two PCSP templates such that there exists a minor homomorphism from* Pol($\mathbf{S}_1, \mathbf{T}_1$) *to* Pol($\mathbf{S}_2, \mathbf{T}_2$). *Then, the following statements hold:*

1. *If* $(\mathbf{S}_1, \mathbf{T}_1)$ *has bounded width, then so does* $(\mathbf{S}_2, \mathbf{T}_2)$.
2. *If* $(\mathbf{S}_1, \mathbf{T}_1)$ *has sublinear width, then so does* $(\mathbf{S}_2, \mathbf{T}_2)$.

We note here that, although not explicitly addressed in [8], the proof for the bounded width statement in Theorem 2 works also for sublinear width; i.e., the proof of Item *1* in Theorem 2 also proves Item *2*.

3.2 Linear programming and Sherali-Adams hierarchy

In this subsection we shall introduce a more general family of relaxations based on linear programming. Despite the fact that, in general, these relaxations are seemingly more powerful, in the realm of fixed-template CSPs they turn out to be not more powerful than local consistency. However, it will follow from our results that for fixed-template PCSPs they are, indeed, strictly more powerful.

Fix a signature σ. Let $\mathbf{I}$ be an instance with domain $I = [n]$, and let $\mathbf{S}$ be a constraint language with domain $S = [p]$, where n and p are positive integers. For each $R \in \sigma$ of arity r, let $Z_R \subseteq \{0,1\}^{rp}$ denote the set of characteristic vectors $v_{\mathbf{t}}$ as $\mathbf{t}$ ranges over the set of r-tuples in $R^{\mathbf{S}}$, seen as maps from $[r]$ to $[p]$; formally,

$$v_{\mathbf{t}} = (v_{\mathbf{t},1,1}, \ldots, v_{\mathbf{t},1,p}, \ldots, v_{\mathbf{t},r,1}, \ldots, v_{\mathbf{t},r,p}), \tag{8}$$

where $v_{\mathbf{t},i,j} = 1$ if $\mathbf{t}(i) = j$ and $v_{\mathbf{t},i,j} = 0$ if $\mathbf{t}(i) \neq j$, for all $i \in [r]$ and $j \in [p]$. Let $P_R \subseteq \mathbb{R}^{rp}$ denote the convex hull of $Z_R \subseteq \{0,1\}^{rp}$. The polytope P_R can be described by a linear program of the form $A_R x \leq b_R$, where x is a vector of rp many variables, A_R is an integer matrix with at most 2^{rp} many rows, and b_R is an integer vector of dimension at most 2^{rp}. Let $\mathrm{IP}(\mathbf{I}, \mathbf{S})$ denote the 0-1 linear program that has one variable $x_{u\mapsto a}$ for each $u \in I$ and $a \in S$ and the following defining equalities and inequalities:

$$\begin{aligned}
&x_{u\mapsto 1} + \cdots + x_{u\mapsto p} = 1 && \text{for } u \in I, && \text{(I1)}\\
&A_R x_{\mathbf{u}} \leq b_R && \text{for } R \in \sigma \text{ and } \mathbf{u} \in R^{\mathbf{I}}, && \text{(I2)}\\
&x_{u\mapsto a} \in \{0,1\} && \text{for } u \in I \text{ and } a \in S, && \text{(I3)}
\end{aligned}$$

where, for $\mathbf{u} = (u_1, \ldots, u_r)$, we define

$$x_{\mathbf{u}} := (x_{u_1\mapsto 1}, \ldots, x_{u_1\mapsto p}, \ldots, x_{u_r\mapsto 1}, \ldots, x_{u_r\mapsto p}). \tag{9}$$

The *direct LP relaxation* of $\mathrm{IP}(\mathbf{I}, \mathbf{S})$, denoted by $\mathrm{LP}(\mathbf{I}, \mathbf{S})$, is the linear program that is obtained by replacing the 0-1 constraints (I3) by their relaxation $0 \leq x_{u\mapsto a} \leq 1$. We note that $\mathrm{LP}(\mathbf{I}, \mathbf{S})$ is equivalent to the LP obtained by taking the so-called *basic linear programming* relaxation introduced in the optimization variants MAX-CSP of CSP (see, e.g., [15, 21, 25, 26, 31, 40, 41]), and turning the objective function into a family of inequalities. This is a weak relaxation. In order to obtain stronger LP relaxations we apply the method of Sherali and Adams [45] to $\mathrm{IP}(\mathbf{I}, \mathbf{S})$.

Fix an integer $k \geq 1$. Recall that $M_k(I, S)$ is used to denote the set of maps from a subset of I of cardinality at most k into S. Let $\mathrm{SA}^k(\mathbf{I}, \mathbf{S})$ denote the linear program that has one variable x_f for each $f \in M_k(I, S)$ and the following defining equalities and inequalities:

$$\begin{aligned}
&x_\emptyset = 1 && && \text{(T1)}\\
&x_{f\cup\{u\mapsto 1\}} + \cdots + x_{f\cup\{u\mapsto p\}} = x_f && \text{for } f \in M_{k-1}(I, S) \text{ and } u \in I \setminus \mathrm{dom}(f), && \text{(T2)}\\
&A_R x_{f,\mathbf{u}} \leq b_R x_f && \text{for } f \in M_{k-1}(I, S),\ R \in \sigma \text{ and } \mathbf{u} \in R^{\mathbf{I}}, && \text{(T3)}\\
&0 \leq x_f \leq 1 && \text{for } f \in M_k(I, S), && \text{(T4)}
\end{aligned}$$

where, for $f \in M_{k-1}(I,S)$ and $\mathbf{u} = (u_1, \ldots, u_r)$, we define

$$x_{f,\mathbf{u}} := (x_{f\cup\{u_1\mapsto 1\}}, \ldots, x_{f\cup\{u_1\mapsto p\}}, \ldots, x_{f\cup\{u_r\mapsto 1\}}, \ldots, x_{f\cup\{u_r\mapsto p\}}). \tag{10}$$

It is obvious that $\mathrm{SA}^1(\mathbf{I},\mathbf{S})$ is basically the same linear program as $\mathrm{LP}(\mathbf{I},\mathbf{S})$: in case $k=1$, the set $M_{k-1}(I,S)$ contains only the empty map $f=\emptyset$, and $x_\emptyset = 1$ by (T1) in SA^1. Furthermore, since $M_k(I,S) \subseteq M_{k+1}(I,S)$, it is clear that each $\mathrm{SA}^{k+1}(\mathbf{I},\mathbf{S})$ is at least as strong as $\mathrm{SA}^k(\mathbf{I},\mathbf{S})$. To understand the relaxation, think of each new inequality of $\mathrm{SA}^{k+1}(\mathbf{I},\mathbf{S})$ as obtained from multiplying an inequality in $\mathrm{SA}^k(\mathbf{I},\mathbf{S})$ by a variable $x_{u\mapsto a}$, simplifying the results by the rules $x^2_{u\mapsto a} = x_{u\mapsto a}$ and $x_{u\mapsto a}x_{u\mapsto b} = 0$ whenever $a \neq b$ (these relations are true in all solutions of $\mathrm{IP}(\mathbf{I},\mathbf{S})$), and finally linearizing all the quadratic terms by introducing new variables for the products.

It is worth noting that, according to the Sherali-Adams method as defined in [45], it would seem that our definition of $\mathrm{SA}^{k+1}(\mathbf{I},\mathbf{S})$ is missing the inequalities that can be obtained from those in $\mathrm{SA}^k(\mathbf{I},\mathbf{S})$ from multiplication by $1 - x_{u\mapsto a}$. However, we note that these multiplications are redundant: the equalities of type (T1) and (T2) imply that $1 - x_{u\mapsto a} = \sum_{b\in[p]\setminus\{a\}} x_{u\mapsto b}$, which means that a multiplication by $1 - x_{u\mapsto a}$ can be obtained as a positive linear combination of multiplications by $x_{u\mapsto b}$ for $b \neq a$.

We write $\mathbf{I} \leq_{\mathrm{SA}^k} \mathbf{S}$ if the linear program $\mathrm{SA}^k(\mathbf{I},\mathbf{S})$ is feasible. It follows directly from the definitions that if $\mathbf{I} \to \mathbf{S}$, then $\mathbf{I} \leq_{\mathrm{SA}^k} \mathbf{S}$. Indeed, if h is a homomorphism from $\mathbf{I}$ to $\mathbf{S}$, then the assignment that sets $x_f = 1$ if $f \subseteq h$, and $x_f = 0$ if $f \not\subseteq h$, is a feasible solution to the linear program. Furthermore, it also follows directly from the definitions that if $\mathbf{I} \leq_{\mathrm{SA}^k} \mathbf{S}$, then $\mathbf{I} \leq_k \mathbf{S}$; indeed, if $(x_f \mid f \in M_k(I,S))$ is a feasible solution for $\mathrm{SA}^k(\mathbf{I},\mathbf{S})$, then the collection $\mathscr{H} = \{f \in M_k(I,S) \mid x_f \neq 0\}$ is a k-strategy for $\mathbf{I}$ and $\mathbf{S}$; the non-emptiness condition follows from (T1), the closure under restrictions and the extension property up to k follow from (T2), and the condition that every map in $\mathscr{H}$ is a partial homomorphism follows from (T3) and the choice of the polytope P_R.

Since the feasibility problem for linear programs is solvable in polynomial time, there is an algorithm that, given $(\mathbf{I},\mathbf{S},k)$ as input, decides whether $\mathbf{I} \leq_{\mathrm{SA}^k} \mathbf{S}$ and does so in time polynomial in $(|I|+|S|)^k$. This algorithm is the *width-k Sherali-Adams heuristic* for CSPs which, as was the case for the width-k consistency algorithm, can also be used as a heuristic for PCSPs. If $(\mathbf{S},\mathbf{T})$ is a PCSP template and $k(n)$ is an integer function, then we say that $\mathrm{PCSP}(\mathbf{S},\mathbf{T})$ *is solvable in SA-width* $k(n)$, or that the template $(\mathbf{S},\mathbf{T})$ has *SA-width* $k(n)$, if for every σ-structure instance $\mathbf{I}$ with n elements it holds that $\mathbf{I} \leq_{\mathrm{SA}^{k(n)}} \mathbf{S}$ implies $\mathbf{I} \to \mathbf{T}$. It follows from the fact that it is always the case that $\mathbf{I} \leq_n \mathbf{S}$, where n is the number of elements of $\mathbf{I}$, that every $\mathrm{PCSP}(\mathbf{S},\mathbf{T})$ is solvable in SA-width at most $k(n)$ for $k(n) = n$. Whenever it is solvable in SA-width $k(n)$ for some $k(n) = O(1)$, we say that $\mathrm{PCSP}(\mathbf{S},\mathbf{T})$ is solvable in *bounded SA-width*, or that $(\mathbf{S},\mathbf{T})$ has bounded SA-width.

The SA relaxations have been intensively used in optimization versions of CSP [22, 30, 46, 50]. Our setup is, again, slightly different because the constraints of the instance must be encoded in the polytope-defining inequalities, rather than in the objective function, as is done in the optimization variants. Still, the family of SA relaxations commonly used for optimization can be readily adapted to our setting by turning the objective function into a set of inequalities (see [20]). The LP family obtained in this way, although not equivalent to ours, has the same power up to constants. Indeed, the notion of bounded SA-width remains unaltered.

The strength of SA-width k for $k = 1$ is well understood. Indeed, it is shown in [8] (Theorem 7.9) that $(\mathbf{S}, \mathbf{T})$ has SA-width 1 if and only if $\mathrm{Pol}(\mathbf{S}, \mathbf{T})$ has symmetric functions of all arities, where an operation is symmetric if its output is independent of the order of its input elements. The picture is also clear if $\mathbf{S} = \mathbf{T}$ (i.e., in the case of CSPs). In particular, it follows from combining the results in [1] and [2] with those in [9] that if $\mathrm{CSP}(\mathbf{T})$ is solvable in bounded SA-width, then it is solvable in bounded width as well. Furthermore, combining the results in [9] and [5] it can be concluded that, additionally, $\mathrm{CSP}(\mathbf{T})$ is solvable in *relational width 2*, which, in turn, can be shown to imply that it is solvable in SA-width 2.

4 Main result and applications

In this section we state the main structural result about PCSP templates of bounded width, derive from it the algebraic consequence about the presence of WNUs of arities three and more, and apply it to compare the relative power of bounded width and bounded SA-width.

4.1 Structure of PCSP templates of bounded width

If $R \subseteq A^r$ is a relation of arity r over the set A, then we write $\mathscr{G}(R)$ to denote the set of all binary projections of R; i.e., $\mathscr{G}(R) = \{\mathrm{pr}_{i,j} R \mid i, j \in [r],\ i \neq j\}$. If σ is a signature and $\mathbf{A}$ is a σ-structure, then we define $\mathscr{G}(\mathbf{A}) = \mathscr{G}(\prod_{\sigma \in R} R^{\mathbf{A}})$. The main technical result of this paper is the following.

Theorem 3. *Let $(\mathbf{S}, \mathbf{T})$ be a PCSP template that has sublinear width. If for all $U, V \in \mathscr{G}(\mathbf{S})$ we have $U \circ V = S^2$, then $\mathbf{T}$ is reflexive; i.e., there exists $a \in T$ such that each relation in $\mathbf{T}$ contains the reflexive tuple $(a, a, \ldots, a)$. In particular, this holds if $(\mathbf{S}, \mathbf{T})$ has bounded width.*

An example of a PCSP template that satisfies the condition of Theorem 3 is $(\mathbf{K}_2, \mathbf{H})$ for any graph $\mathbf{H}$ having at least one edge. Indeed, the composition of the edge relation of $\mathbf{K}_2$ with itself is the equality relation on $[2]$, which is not the full binary relation $[2] \times [2]$. This is consistent with the fact that $\mathbf{K}_2$, and hence $(\mathbf{K}_2, \mathbf{H})$, has width three (see Example 2). An example of a PCSP template that does *not* satisfy the condition of Theorem 3 is $(\mathbf{K}_p, \mathbf{H})$ for any integer $p \geq 3$ and any self-loop free graph $\mathbf{H}$ such that $\mathbf{K}_p \to \mathbf{H}$. Indeed, as it is easy to check, if $p \geq 3$ then the composition of the edge relation of $\mathbf{K}_p$ with itself is the full binary relation $[p] \times [p]$, but $\mathbf{H}$ is not reflexive (since it is self-loop free). We revisit these examples in Section 6.

The proof of Theorem 3 will be given in Section 5. We devote the rest of this section to derive some applications.

4.2 Algebraic implications

We first derive some algebraic implications. Recall the definition of WNU polymorphism from Example 1 in Section 2. Recall also that, by Theorem 1, if $\mathrm{CSP}(\mathbf{T})$ is solvable in bounded width, then $\mathrm{Pol}(\mathbf{T})$ must contain a WNU of every arity $m \geq 3$. It follows from Theorem 3 that this also holds for Promise CSPs.

Corollary 2. *Let* $(\mathbf{S}, \mathbf{T})$ *be a PCSP template of sublinear or bounded width. Then,* $\mathrm{Pol}(\mathbf{S}, \mathbf{T})$ *contains a WNU of arity* m *for every* $m \geq 3$.

Proof. Let $m \geq 3$, let $\mathscr{M} = \mathrm{Pol}(\mathbf{S}, \mathbf{T})$, and let $\mathbf{S}' = (\{x, y\}, R^{\mathbf{S}'})$ be the structure with a 2-element domain $\{x, y\}$ where $R^{\mathbf{S}'}$ is the m-ary relation that contains precisely all tuples in $\{x, y\}^m$ in which y appears exactly once (as in Example 1). As pointed out in Observation 1, there is a homomorphism from $\mathbf{S}'$ to $\mathbf{F}(\mathbf{S}'; \mathscr{M})$, and, hence, $(\mathbf{S}', \mathbf{F}(\mathbf{S}'; \mathscr{M}))$ is a legit PCSP template. Also, by Observation 2 there is a minor homomorphism from $\mathscr{M}$ to $\mathrm{Pol}(\mathbf{S}', \mathbf{F}(\mathbf{S}'; \mathscr{M}))$. Therefore, by Theorem 2, the PCSP template $(\mathbf{S}', \mathbf{F}(\mathbf{S}'; \mathscr{M}))$ has sublinear width. Then, it follows from Theorem 3 that $R^{\mathbf{F}(\mathbf{A}'; \mathscr{M})}$ contains a reflexive tuple, which implies, by Example 1, that $\mathscr{M}$ contains an m-ary WNU. □

Next we argue that the converse to Corollary 2 does not hold. For positive integers s and r, let s-IN-r-SAT be the structure with domain $\{0, 1\}$ and a single relation $R^{s\text{-IN-}r\text{-SAT}}$ of arity r containing all tuples with exactly s many 1's. Let r-NAE-SAT be the structure with domain $\{0, 1\}$ and a single relation $R^{r\text{-NAE-SAT}}$ of arity r containing all tuples in $\{0, 1\}^r$ except the two reflexive tuples $(0, 0, \ldots, 0)$ and $(1, 1, \ldots, 1)$. Clearly, if $0 < s < r$, then there is a homomorphism from s-IN-r-SAT to r-NAE-SAT, so (s-IN-r-SAT, r-NAE-SAT) is a PCSP template.

Lemma 1. *There is a PCSP template* $(\mathbf{S}, \mathbf{T})$ *that does not have sublinear or bounded width such that* $\mathrm{Pol}(\mathbf{S}, \mathbf{T})$ *contains a WNU of arity* m *for every* $m \geq 3$. *Concretely, setting* $\mathbf{S} =$ 2-IN-4-SAT *and* $\mathbf{T} =$ 4-NAE-SAT *gives such an example.*

Proof. Let $\mathbf{S}$ and $\mathbf{T}$ be set as in the second part of the lemma. It follows directly from Theorem 3 that $(\mathbf{S}, \mathbf{T})$ does not have sublinear width. However, $\mathrm{Pol}(\mathbf{S}, \mathbf{T})$ contains, for every $m \geq 3$, the WNU h or arity m, that returns the majority element in the input if exists and else (that is, in case of ties) it returns the first element. That is, h is the following function:

$$h(x_1, \ldots, x_m) = \begin{cases} a & \text{if } |\{x_i \mid x_i = a\}| > m/2 \\ x_1 & \text{otherwise} \end{cases}$$

To see that this is a WNU, use the assumption that $m \geq 3$. To see that this a polymorphism of $(\mathbf{S}, \mathbf{T})$, we argue by double counting. Any $m \times m$ matrix that has all its rows in $\mathbf{T}$ has the same number of 0's and 1's. Therefore, either all columns also have the same number of 0's and 1's, in which case h returns the first row of the matrix, which is not reflexive, or some column has more 0's than 1's and some column has more 1's than 0's, in which case h returns also a non-reflexive tuple. □

4.3 Separation of width from SA-width

Another application of Theorem 3 is the separation of bounded width and bounded SA-width for PCSPs. This separation is another evidence that the family of PCSPs is a more rich family of problems than classical CSPs from an algorithmic point of view. This separation is obtained combining our result with the following lemma, which in turn, builds upon [12].

Lemma 2. *Let* $\mathbf{S} = s$-IN-r-SAT *and* $\mathbf{T} = r$-NAE-SAT *where* s *and* r *are integers such that* $0 < s < r$. *Then* $(\mathbf{S}, \mathbf{T})$ *has SA-width* 2.

Proof. We start by noticing that a simple double counting argument shows that $\mathrm{Pol}(\mathbf{S}, \mathbf{T})$ contains, for every odd $m \geq 3$, the *alternating threshold* operation $f(x_1, \ldots, x_m)$ defined to be 1 if $\sum_{i \in [m]} (-1)^{i-1} x_i > 0$ and 0 otherwise. Indeed, let $\mathbf{t}_1, \ldots, \mathbf{t}_m$ be any sequence of tuples in $R^{\mathbf{S}}$ and let O and E be, respectively, the odd and even numbers in $[m]$. It we use $|\,\mathbf{t}\,|_1$ to denote the total number of 1's occurring in tuple $\mathbf{t}$, we observe that

$$\sum_{i \in O} |\,\mathbf{t}_i\,|_1 = s + \sum_{i \in E} |\,\mathbf{t}_i\,|_1 \tag{11}$$

It follows that there exist $j_0, j_1 \in [r]$ satisfying

$$\sum_{i \in O} \mathbf{t}_i(j_0) \geq \sum_{i \in E} \mathbf{t}_i(j_0) \quad \text{and} \quad \sum_{i \in O} \mathbf{t}_i(j_1) \leq \sum_{i \in E} \mathbf{t}_i(j_1). \tag{12}$$

We note that for the existence of j_1 we use $s < r$. Consequently, $f(\mathbf{t}_1, \ldots, \mathbf{t}_m)$ contains at least a 0 and a 1 (in coordinates j_0 and j_1 respectively). Then $f(\mathbf{t}_1, \ldots, \mathbf{t}_m) \in R^{\mathbf{T}}$ as desired.

It then follows from ([12], Section 3.2) that any instance $\mathbf{I}$ of $\mathrm{PCSP}(\mathbf{S}, \mathbf{T})$ is homomorphic to $\mathbf{T}$ whenever the following *augmented* SA^1 *feasibility* condition holds: for each variable $v \in I$ there exists a $b_v \in \{0, 1\}$ such that there is a feasible solution of $\mathrm{SA}^1(\mathbf{I}, \mathbf{S})$ that sets $x_{v \mapsto b_v}$ to 1. We are left to show that if $\mathrm{SA}^2(\mathbf{I}, \mathbf{S})$ has a feasible solution, then the augmented SA^1 feasibility condition holds.

To avoid confusion we use x'_f to refer to the variables in $\mathrm{SA}^2(\mathbf{I}, \mathbf{S})$ and x_f to refer to the variables in $\mathrm{SA}^1(\mathbf{I}, \mathbf{S})$. Fix $v \in I$. It follows from the first three types of constraint in the definition of SA^2 that there exists some $b \in \{0, 1\}$ that $x'_{v \mapsto b} > 0$; let $d := x'_{v \mapsto b}$. Now construct a solution of $\mathrm{SA}^1(\mathbf{I}, \mathbf{S})$ as follows: for every $u \in I$ and $a \in B$, set $x_{u \mapsto a}$ to $x'_{u \mapsto a, v \mapsto b}/d$. By choice of d we have $x_{v \mapsto b} = 1$ as desired. It is easy to verify that the assignment thus defined is a feasible solution of $\mathrm{SA}^1(\mathbf{I}, \mathbf{S})$. □

Hence, from Theorem 3 and Lemma 2 we have:

Corollary 3. *There is a PCSP template* $(\mathbf{S}, \mathbf{T})$ *that has SA-width* 2, *and hence bounded SA-width, but does not have sublinear or bounded width. Concretely, setting* $\mathbf{S} = s$-IN-r-SAT *and* $\mathbf{T} = r$-NAE-SAT *where* s *and* r *are integers such that* $0 < s < r$ *and* $r > 2$ *gives such an example.*

It should be pointed out that the condition that $r > 2$ in Corollary 3 is necessary. Indeed, the template 1-IN-2-SAT is isomorphic to $\mathbf{K}_2$, and we already know that $\mathbf{K}_2$, and hence (1-IN-2-SAT, 2-NAE-SAT), has width three (see Example 2). It is also easy to see that, consistently with this, (1-IN-2-SAT, 2-NAE-SAT) satisfies the condition of Theorem 3: the composition of the binary relation of 1-IN-2-SAT with itself is the equality relation on $\{0, 1\}$, which is not the full binary relation $\{0, 1\}^2$.

5 Proof of Theorem 3

The proof is structured in five parts. The first part sets the stage. In the second part we define a probability distribution on instances and prove that a random instance $\mathbf{I}$ sampled from this distribution has, with high probability, two key properties: (a) it is *dense enough* to guarantee that $\mathbf{I} \not\to \mathbf{T}$, and (b) it is also *sparse enough*, in a sense compatible with (a), to guarantee that $\mathbf{I} \leq_k \mathbf{S}$, as will be proved in the next part. In the third part we show that any instance $\mathbf{I}$ that satisfies the sparsity condition indeed satisfies $\mathbf{I} \leq_k \mathbf{S}$. In the fourth part we discuss the setting of parameters that satisfies all the required conditions to prove the theorem. Finally, the fifth part of the proof focuses on the special case of the theorem that applies to digraphs (structures with a single binary relation). For this special case we are able to slightly improve the parameters; this special case for (di)graphs will be used in Section 6.

5.1 Setting the stage

Let σ be a fixed signature and let $(\mathbf{S}, \mathbf{T})$ be a PCSP template of signature σ that has sublinear width. Let $\mathbf{S}' := (S, \prod_{R\in\sigma} R^{\mathbf{S}})$ and $\mathbf{T}' := (T, \prod_{R\in\sigma} R^{\mathbf{T}})$ and note that $\mathbf{S} \to \mathbf{T}$ implies $\mathbf{S}' \to \mathbf{T}'$. Since the signature σ is finite and fixed, the template $(\mathbf{S}, \mathbf{T})$ has sublinear width if and only if the template $(\mathbf{S}', \mathbf{T}')$ has sublinear width. Also, it holds that $\mathscr{G}(\mathbf{S}) = \mathscr{G}(\mathbf{S}')$, and that $\mathbf{T}$ is reflexive if and only if $\mathbf{T}'$ is reflexive. Therefore, it suffices to prove the theorem when the signature σ consists of a single relation symbol R. Furthermore, since any unary relation is either reflexive or empty, we may assume that the arity r of R is at least 2; i.e., $r \geq 2$.

Let $(\mathbf{S}, \mathbf{T})$ satisfy the following assumptions:

A1: for all $U, V \in \mathscr{G}(\mathbf{S})$ we have $U \circ V = S^2$,
A2: for all $a \in T$, the reflexive tuple $(a, a, \ldots, a)$ is not in $R^{\mathbf{T}}$.

Let $p = |S|$ and $q = |T|$ be the cardinalities of the domains of $\mathbf{S}$ and $\mathbf{T}$, respectively, and let $k = k(n)$ be an integer function such that $k(n) = o(n)$. Our goal is to show that there exist arbitrarily large instances $\mathbf{I}$ that witness that the PCSP template $(\mathbf{S}, \mathbf{T})$ does not have width k; i.e., the instance $\mathbf{I}$ is such that $\mathbf{I} \leq_k \mathbf{S}$ and $\mathbf{I} \not\to \mathbf{T}$. We show that such an instance $\mathbf{I}$ exists by the probabilistic method.

In anticipation for the proof, in addition to the data r, k, p, q, we fix some real parameters $\delta, \beta, \alpha, c, d$, as well as an integer parameter n. These parameters are required to satisfy the following conditions:

C1: $0 < \delta \leq 1/((r+1)(3r+1))$,
C2: $0 < \beta \leq (1+\delta)/(r-1)$,
C3: $0 < \alpha \leq (\beta/d)^{1/(r-1)}(r/e)^{r/(r-1)}$,
C4: $c \geq kp/\delta$,
C5: $n \geq \max\{c/(\alpha\beta), q\}$,
C6: $1 \leq d \leq n^{r-1}$,

C7: $p_1(r,d,n,q) + p_2(r,d,n,\alpha,\beta) < 1$,

where

$$p_1(r,d,n,q) := q^n \exp\left(-dn/(r^r q^{r-1}\right), \tag{13}$$
$$p_2(r,d,n,\alpha,\beta) := \sum_{v=1}^{\lfloor \alpha n \rfloor} \left((n/v)^{1-(r-1)\beta} d^{\beta} e^{1+(r+1)\beta} r^{-r\beta} \beta^{-\beta}\right)^v. \tag{14}$$

We have separated the first six conditions C1–C6 from the last one C7 because the first six are easily feasible by themselves; fulfilling Condition C7 simultaneously is more delicate. At the end of the proof we discuss a settings of the parameters $\delta, \beta, \alpha, c, d$ and n that satisfy Conditions C1–C7. For now, we assume that the conditions are feasible.

5.2 Probabilistic construction

Let H denote a random Erdős-Rényi r-uniform hypergraph with n vertices and edge probability d/n^{r-1}; i.e., $V(H) = [n]$, and each r-element subset $C \subseteq [n]$ is or is not an edge in $E(H)$, independently, with probability d/n^{r-1}. Note that d/n^{r-1} is a proper probability by Condition C6. Let $\mathbf{I} = \mathbf{I}(H)$ be the (random) instance with domain $[n]$ that has one r-tuple $\mathbf{v}_C$ in $R^{\mathbf{I}}$ for each $C \in E(H)$, where $\mathbf{v}_C$ is obtained by ordering the elements in C is some arbitrary way.

Lemma 3. *The probability that* $\mathbf{I}$ *is homomorphic to* $\mathbf{T}$ *is at most* $p_1(r,d,n,q)$.

Proof. Let h be any mapping from the domain I of $\mathbf{I}$ to the domain T of $\mathbf{T}$. Let $S_1, \ldots, S_q$ be the partition induced by h; i.e., $S_i = h^{-1}(i)$ for $i = 1, \ldots, q$. Let $n_i := |S_i|$. By Assumption A2, the relation $R^{\mathbf{T}}$ does not contain a reflexive tuple. Therefore, for all $i \in [q]$ and $C \subseteq S_i$ we have $h(\mathbf{v}_C) \notin R^{\mathbf{T}}$. By independence, the probability that $h(\mathbf{v}_C)$ belongs to $R^{\mathbf{T}}$ for every $\mathbf{v}_C \in R^{\mathbf{I}}$ is at most $(1-d/n^{r-1})^N$, where $N = \sum_{i=1}^{q} \binom{n_i}{r}$. Now note that $\binom{n_i}{r} \geq (n_i/r)^r$, and that, subject to the constraints $x_1, \ldots, x_q \geq 0$ and $\sum_{i=1}^{q} x_i = n$, the sum $\sum_{i=1}^{q}(x_i/r)^r$ is minimized at $x_1 = \cdots = x_q = n/q$. It follows that $N \geq q(n/(rq))^r = n^r/(r^r q^{r-1})$. Therefore, $(1-d/n^{r-1})^N \leq \exp(-dn^r/(n^{r-1}r^r q^{r-1})) \leq \exp(-dn/(r^r q^{r-1}))$. By the union bound the probability that there is a homomorphism from $\mathbf{I}$ to $\mathbf{T}$ is at most $q^n \exp(-dn/(r^r q^{r-1})$, which equals $p_1(r,d,n,q)$. $\square$

We say that $\mathbf{I}$ is (α, β)-*sparse* if every substructure $\mathbf{J}$ of $\mathbf{I}$ with v elements satisfying $1 \leq v \leq \alpha n$ has less than βv many tuples. Equivalently, $\mathbf{I}$ is (α, β)-sparse if every substructure $\mathbf{J}$ of $\mathbf{I}$ with at most αn elements and at least βv many tuples has more than v elements. The argument for the lemma below is almost the same as the one for Lemma 1 in [23] except that we need to apply it to our probability model. Also as in [23], we use the following Chernoff-like bound

$$\sum_{i=\lceil tm \rceil}^{m} \binom{m}{i} \gamma^i (1-\gamma)^{m-i} \leq (e\gamma/t)^{tm}, \tag{15}$$

which holds for every integer $m \geq 1$, every real $\gamma \in [0, 1]$, and every real $t \in (\gamma, 1]$.

Lemma 4. *The probability that* $\mathbf{I}$ *is not* (α, β)*-sparse is at most* $p_2(r, d, n, \alpha, \beta)$*.*

Proof. For each integer v such that $1 \leq v \leq \alpha n$, set $m_v = \binom{v}{r}$. Setting $p_0 = d/n^{r-1}$, the probability that $\mathbf{I}$ is not (α, β)-sparse is bounded by

$$\sum_{v=r}^{\lfloor \alpha n \rfloor} \binom{n}{v} \sum_{i=\lceil \beta v \rceil}^{m_v} \binom{m_v}{i} p_0^i (1-p_0)^{m_v - i}. \tag{16}$$

Now set $t_v = \beta v / m_v$. Using $\binom{v}{r} < (ve/r)^r$ note that $t_v > \beta r^r/(e^r v^{r-1}) \geq d/n^{r-1} = p_0$ for $v \leq \alpha n$ since α satisfies Condition C3. If $t_v > 1$, then the inner sum in (16) is zero, while if $t_v \leq 1$, then we have $t_v \in (p_0, 1]$, so (15) applies to bound (16) by

$$\sum_{v=r}^{\lfloor \alpha n \rfloor} \binom{n}{v} (ep_0/t_v)^{t_v m_v}. \tag{17}$$

Using $\binom{n}{v} \leq (ne/v)^v$ and $\binom{v}{r} \leq (ve/r)^r$, we bound this further by

$$\sum_{v=r}^{\lfloor \alpha n \rfloor} \left((n/v)^{1-(r-1)\beta} d^\beta e^{1+(r+1)\beta} r^{-r\beta} \beta^{-\beta} \right)^v, \tag{18}$$

which is bounded by $p_2(r, d, n, \alpha, \beta)$. ☐

Lemma 5. *There exists a structure* $\mathbf{I}$ *with* n *elements that is not homomorphic to* $\mathbf{T}$ *and is* (α, β)*-sparse. In particular,* $\mathbf{I}$ *is such that for every substructure* $\mathbf{J}$ *of* $\mathbf{I}$ *and every integer* $m \geq 0$*, if* $\mathbf{J}$ *has* m *many tuples and* $m \leq c$*, then* $\mathbf{J}$ *has more than* $(r-1)m/(1+\delta)$ *elements.*

Proof. The existence of $\mathbf{I}$ follows directly from the assumption that Condition C7 holds and Lemmas 3 and 4. For the second part of the statement, assume that $\mathbf{J}$ is a substructure of $\mathbf{I}$ that has v many elements and m many tuples, with $m \leq c$ and $v \leq (r-1)m/(1+\delta)$. In particular $v \leq m/\beta \leq c/\beta \leq \alpha n$ since β satisfies Condition C2 and n satisfies Condition C5. But $\mathbf{I}$ is (α, β)-sparse and it follows that $\mathbf{J}$ has less than $\beta v \leq m$ many tuples; a contradiction. ☐

5.3 Proof of consistency

Still assuming that the setting of parameters satisfies the conditions C1–C7, let $\mathbf{I}$ be a structure as in the second part of Lemma 5. We shall now prove that $\mathbf{I} \leq_k \mathbf{S}$.

The proof adapts the notion of *boundary* from [44]; this was used to prove size lower bounds for the Resolution proof complexity of random CSPs. Let $\mathbf{J}$ be any substructure of $\mathbf{I}$. A set $D \subseteq J$ is said to be a *boundary set of* $\mathbf{J}$ if it is non-empty and every homomorphism from $\mathbf{J}|_{J\setminus D}$ to $\mathbf{S}$ extends to a homomorphism from $\mathbf{J}$ to $\mathbf{S}$. We introduce two types of subsets $D \subseteq J$ of $\mathbf{J}$ and show that any set of any of these types is a boundary set of $\mathbf{J}$.

We define the *degree* of an element in $\mathbf{J}$ to be the number of tuples of $R^{\mathbf{J}}$ in which it appears. We say that $D \subseteq J$ is of type (1) if $D = \{x_1, \ldots, x_{r-1}\}$, where $x_1, \ldots, x_{r-1}$ are all distinct and have degree one in $\mathbf{J}$, and there exists $\mathbf{v}_C \in R^{\mathbf{J}}$ with $D \subseteq C$. In this case we say that the set C *witnesses* that D is of type (1). We say that $D \subseteq J$ is of type (2) if $D = \{x_1, y_1, \ldots, x_{r-2}, y_{r-2}, z\}$, where $x_1, y_1, \ldots, x_{r-2}, y_{r-2}$ are all distinct and have degree one in $\mathbf{J}$, also z is distinct from the rest of elements in D and has degree two in $\mathbf{J}$, and there exist two different tuples $\mathbf{v}_{C_1}, \mathbf{v}_{C_2} \in R^{\mathbf{J}}$ such that $\{x_1, \ldots, x_{r-2}, z\} \subseteq C_1$ and $\{y_1, \ldots, y_{r-2}, z\} \subseteq C_2$. In this case we say that the sets C_1, C_2 *witness* that D is of type (2). We check below that, since $\mathscr{G}(R^{\mathbf{S}})$ satisfies Assumption A1, every set of these two types is a boundary set.

Lemma 6. *For every substructure $\mathbf{J}$ of $\mathbf{I}$ and every $D \subseteq J$, if D is of type (1) or (2) in $\mathbf{J}$, then D is a boundary set of $\mathbf{J}$.*

Proof. Fix $\mathbf{J}$ and D as in the hypothesis and let h be a homomomorphism from $\mathbf{J}|_{J\setminus D}$ to $\mathbf{S}$.

Firstly, assume that $D = \{x_1, \ldots, x_{r-1}\}$ is of type (1), and let C witness so; i.e., $\mathbf{v}_C \in R^{\mathbf{J}}$ and $D \subseteq C$. Let $i_0 \in [r]$ be such that $\mathbf{v}_C(i_0) \in J \setminus D$. By Assumption A1 on $\mathscr{G}(R^{\mathbf{S}})$ we have that

$$\mathrm{pr}_{i_0} R^{\mathbf{S}} = S. \tag{19}$$

Consequently, there is a tuple $\mathbf{a} \in R^{\mathbf{S}}$ such that $\mathbf{a}(i_0) = h(\mathbf{v}_C(i_0))$. We extend h to map every element of J by setting $h(\mathbf{v}_C(i)) = \mathbf{a}(i)$ for every $i \in [r] \setminus \{i_0\}$. Since each $x \in D$ appears in $\mathbf{v}_C$ but in no other tuple of $\mathbf{J}$, the result is a homomorphism from $\mathbf{J}$ to $\mathbf{S}$.

Secondly, assume that $D = \{x_1, y_1, \ldots, x_{r-2}, y_{r-2}, z\}$ is of type (2), and let C_1, C_2 witness so; i.e., $\mathbf{v}_{C_1}, \mathbf{v}_{C_2} \in R^{\mathbf{J}}$ are distinct, and $\{x_1, \ldots, x_{r-2}, z\} \subseteq C_1$ and $\{y_1, \ldots, y_{r-2}, z\} \subseteq C_2$. For $j = 1, 2$, let $i_j \in [r]$ be such that $\mathbf{v}_{C_j}(i_j) \in J \setminus D$ and let $k_j \in [r] \setminus \{i_j\}$ be such that $\mathbf{v}_{C_j}(k_j) = z$. Again, by Assumption A1 on $\mathscr{G}(R^{\mathbf{S}})$, we have that

$$\mathrm{pr}_{i_1,k_1} R^{\mathbf{S}} \circ \mathrm{pr}_{k_2,i_2} R^{\mathbf{S}} = S^2. \tag{20}$$

Consequently, there exist tuples $\mathbf{a}_1, \mathbf{a}_2 \in R^{\mathbf{S}}$ such that $\mathbf{a}_j(i_j) = h(\mathbf{v}_{C_j}(i_j))$ for $j = 1, 2$ and $\mathbf{a}_1(k_1) = \mathbf{a}_2(k_2)$. We extend h to map every element of J by setting $h(\mathbf{v}_{C_j}(i)) = \mathbf{a}_j(i)$ for $j = 1, 2$ and every $i \in [r] \setminus i_j$. Since each $x \in D$ appears in $\mathbf{v}_{C_1}$ or in $\mathbf{v}_{C_2}$ (or in both, in case of z) but in no other tuple of $\mathbf{J}$, the result is a homomorphism from $\mathbf{J}$ to $\mathbf{S}$. $\square$

Lemma 7. *For every substructure $\mathbf{J}$ of $\mathbf{I}$ and every integer $m \geq 0$, if $\mathbf{J}$ has m many tuples and $m \leq c$, then $\mathbf{J}$ has at least δm many pairwise disjoint boundary sets.*

Proof. We can assume that **J** does not contain elements of degree zero since every boundary set of any substructure obtained by removing elements of degree 0 from **J** is also a boundary set of **J**. We shall show that the total number of boundary sets in **J** that are of type (1) or (2) is at least δm. Since any two distinct boundary sets of these types must be disjoint, the claim will follow.

Let $\mathscr{C} = \{C \mid \mathbf{v}_C \in R^{\mathbf{J}}\}$. For $C = \{x_1, \ldots, x_r\} \in \mathscr{C}$, define

$$\mathrm{sdr}(C) := \textstyle\sum_{i=1}^{r} 1/d_i \tag{21}$$

where d_i is the degree of x_i in **J** (*sdr* stands for sum of degree reciprocals). Since **J** does not have elements of degree zero, it is easy to see that $\sum_{C\in\mathscr{C}} \mathrm{sdr}(C)$ is equal to the number v of elements in **J**. The idea of the proof is the following. Since, by Lemma 5 we have that $v \geq (r-1)m/(1+\delta)$, there is a large number of sets $C \in \mathscr{C}$ with $\mathrm{sdr}(C) \geq (r-1)/(1+\delta)$. Since δ is small enough each one of these sets must have at least $r-2$ vertices of degree one, and one vertex of degree at most two. From this large pool of sets it is not difficult to find a large number of witnesses for boundary sets of type (1) or (2). We formalize this below.

Let $\mathscr{D}$ be a collection of boundary sets of types (1) and (2) with the largest possible cardinality and assume towards a contradiction that $|\mathscr{D}| < \delta m$. We partition $\mathscr{C}$ in three sets $\mathscr{C}_1$, $\mathscr{C}_2$, $\mathscr{C}_3$. The set $\mathscr{C}_1$ contains the witness $C \in \mathscr{C}$ of every boundary set of type (1) in $\mathscr{D}$, and exactly one among the two sets $C_1, C_2 \in \mathscr{C}$ that witness some boundary set of type (2) in $\mathscr{D}$. Note that $|\mathscr{C}_1| = |\mathscr{D}| < \delta m$. The set $\mathscr{C}_2$ contains all $C \in \mathscr{C} \setminus \mathscr{C}_1$ such that $\mathrm{sdr}(C) > r - 4/3$. The set $\mathscr{C}_3$ contains the rest, i.e., all $C \in \mathscr{C} \setminus (\mathscr{C}_1 \cup \mathscr{C}_2)$. Note that every $C \in \mathscr{C} \setminus \mathscr{C}_1$ must contain at least two elements of degree larger than one: otherwise it would be the witness of a boundary set of type (1), against the maximality of $\mathscr{D}$. In particular, for any $C \in \mathscr{C} \setminus \mathscr{C}_1$, we have $\mathrm{sdr}(C) \leq r - 1$. Note also that $\mathrm{sdr}(C) \leq r$ for every $C \in \mathscr{C}$, and $\mathrm{sdr}(C) \leq r - 4/3$ for every $C \in \mathscr{C}_3$.

Let α_0 and β_0 be reals in $[0,1]$ such that $|\mathscr{C}_1| = \alpha_0 m$, $|\mathscr{C}_2| = (1 - \alpha_0 - \beta_0)m$, and $|\mathscr{C}_3| = \beta_0 m$. Since each boundary set in $\mathscr{D}$ contributes exactly one set to $\mathscr{C}_1$ and $|\mathscr{D}| < \delta m$, we have $\alpha_0 < \delta$. We shall prove that $\beta_0 \leq 3r\delta$. Assume otherwise; i.e., $\beta_0 > 3r\delta$. Recall that $v = \sum_{C\in\mathscr{C}} \mathrm{sdr}(C)$ and, hence,

$$\begin{aligned}
v &\leq |\mathscr{C}_1| r + |\mathscr{C}_2|(r-1) + |\mathscr{C}_3|(r - 4/3) && (22)\\
&= \alpha_0 m r + (1 - \alpha_0 - \beta_0) m (r-1) + \beta_0 m (r - 4/3) && (23)\\
&= m(r - 1 + \alpha_0 - \beta_0/3) && (24)\\
&< m(r-1)(1-\delta), && (25)
\end{aligned}$$

where the first inequality follows from the fact that $\mathscr{C}_1, \mathscr{C}_2, \mathscr{C}_3$ is a partition of $\mathscr{C}$ and the already noted bounds on $\mathrm{sdr}(C)$ for the C in these sets, the first equality follows from the choices of α_0 and β_0, the second equality follows from plain arithmetic, and the strict inequality follows from $\alpha_0 < \delta$ and the assumption that $\beta_0 > 3r\delta$. On the other hand, by Lemma 5 we have that $v \geq (r-1)m/(1+\delta)$. Combined with (22)-(25), this means that $1 - \delta \geq 1/(1+\delta)$, which is impossible since $\delta > 0$ by Condition C1.

It follows that $|\mathscr{C}_2| \geq (1 - (1 + 3r)\delta)m$. Let $C \in \mathscr{C}_2$. Since $r - 3/2 < r - 4/3 < \mathrm{sdr}(C)$, the set C must contain at least $r - 2$ elements of degree one. We also know that the remaining two elements in C must have degree at least two since, otherwise, C would witness a boundary set of type (1), against the maximality of $\mathscr{D}$. Again from $r - 4/3 < \mathrm{sdr}(C)$ it follows that at least one of the two remaining elements must have degree exactly two. Two sets in $\mathscr{C}_2$ cannot share an element of degree two since otherwise they would witness a boundary set of type (2), against the maximality of $\mathscr{D}$ again. Consequently, each one of the $(1-(1+3r)\delta)m$ sets in $\mathscr{C}_2$ contains an element of degree two which must also appear in some set of $\mathscr{C}_1 \cup \mathscr{C}_3$. Since $|\mathscr{C}_1 \cup \mathscr{C}_3| = (\alpha_0 + \beta_0)m \leq (1+3r)\delta m$, the total number of elements that can appear in $\mathscr{C}_1 \cup \mathscr{C}_3$ is at most $r(1+3r)\delta m$, which yields $1-(1+3r)\delta < r(1+3r)\delta$, against Condition C1. □

Lemma 8. *For every substructure* $\mathbf{J}$ *of* $\mathbf{I}$*, if* $\mathbf{J}$ *has at most* c *many tuples, then* $\mathbf{J}$ *is homomorphic to* $\mathbf{S}$*.*

Proof. Assume, towards a contradiction, that there is a substructure $\mathbf{J}$ of $\mathbf{I}$ with $m \leq c$ many tuples that is not homomorphic to $\mathbf{S}$. We can further assume that $m \geq 1$, and that $\mathbf{J}$ is minimal in the sense that any of its proper substructures is homomorphic to $\mathbf{S}$. By Lemma 7, the substructure $\mathbf{J}$ has at least $\delta m > 0$ boundary sets. Let $D \subseteq J$ be any boundary set of $\mathbf{J}$, which is non-empty by definition. By the minimality of $\mathbf{J}$, there is an homomorphism f from $\mathbf{J}|_{J \setminus D}$ to $\mathbf{S}$. By the definition of boundary set, f can be extended to a homomorphism from $\mathbf{J}$ to $\mathbf{S}$; a contradiction. □

Let $\mathbf{J}$ be a substructure of $\mathbf{I}$. A mapping $h : X \to S$ with $X \subseteq I$ is said to be *consistent with* $\mathbf{J}$ if there is an homomorphism g from $\mathbf{J}$ to $\mathbf{S}$ such that h and g agree on the intersection $\mathrm{dom}(h) \cap \mathrm{dom}(g)$. For the next lemma, recall that $p = |S|$.

Lemma 9. *For every partial homomorphism* h *from* $\mathbf{I}$ *to* $\mathbf{S}$ *with* $\mathrm{dom}(h) \leq k$*, if* h *is consistent with every substructure of* $\mathbf{I}$ *with a most* c/p *many tuples, then* h *is also consistent with every substructure of* $\mathbf{I}$ *with at most* c *many tuples.*

Proof. Fix h as in the hypothesis and assume, towards a contradiction, that h is not consistent with some substructure $\mathbf{J}$ of $\mathbf{I}$ with $m \leq c$ many tuples. We can further assume that $\mathbf{J}$ is minimal in the sense that h is consistent with any proper substructure of $\mathbf{J}$. By assumption we have $c/p < m \leq c$, where $p = |S|$. It follows from Lemma 7 that $\mathbf{J}$ has a collection $\mathscr{D}$ of at least $\delta m > \delta c/p$ many pairwise disjoint boundary sets. By the minimality of $\mathbf{J}$, for each boundary set $D \in \mathscr{D}$ of $\mathbf{J}$, which is non-empty, we have that h is consistent with the substructure $\mathbf{J}|_{J \setminus D}$. Therefore, there exists a homomorphism g_D from $\mathbf{J}|_{J \setminus D}$ to $\mathbf{S}$ that agrees with h. Since D is a boundary set of $\mathbf{J}$, the homomorphism g_D extends to a homomorphism h_D from $\mathbf{J}$ to $\mathbf{S}$. But h is not consistent with $\mathbf{J}$ which means that h_D and h disagree somewhere in D; in particular, $D \cap \mathrm{dom}(h) \neq \emptyset$. Since the boundary sets in $\mathscr{D}$ are pairwise disjoint, we get $|\mathrm{dom}(h)| > \delta c/p$, which is at least k by Condition C4; a contradiction. □

Lemma 10. *There is a* k*-strategy on* $\mathbf{I}$ *and* $\mathbf{S}$*; i.e.,* $\mathbf{I} \leq_k \mathbf{S}$*.*

Proof. Define $\mathscr{H}$ to be set of all partial mappings from I to S with domain size at most k that are consistent with every substructure $\mathbf{J}$ of $\mathbf{I}$ with at most c many tuples. We show that $\mathscr{H}$ is a k-strategy.

First, note that every $h \in \mathscr{H}$ is a partial homomorphism from $\mathbf{I}$ to $\mathbf{S}$. To see this, take any tuple $\mathbf{v}_C$ in $\mathbf{I}|_{\mathrm{dom}(h)}$ and we show that $h(\mathbf{v}_C)$ is in $R^{\mathbf{S}}$. Let $\mathbf{J}$ be the substructure of $\mathbf{I}$ that contains only $\mathbf{v}_C$. Since $\mathbf{J}$ has only one tuple and $c \geq 1$, by the definition of $\mathscr{H}$ and the fact that h is in $\mathscr{H}$ we have that h is consistent with $\mathbf{J}$; that is, there is a homomorphism g from $\mathbf{J}$ to $\mathbf{S}$ that agrees with h on the intersection $C = \mathrm{dom}(h) \cap \mathrm{dom}(g)$; i.e., $h(\mathbf{v}_C) = g(\mathbf{v}_C) \in R^{\mathbf{S}}$. Also it follows directly from Lemma 8 that $\mathscr{H}$ is non-empty as it contains the partial mapping h with $\mathrm{dom}(h) = \emptyset$.

It remains to be shown that $\mathscr{H}$ is closed under restrictions and satisfies the extension property up to k. The closure under restrictions follows directly from the definitions. Let us then verify that $\mathscr{H}$ satisfies the extension property up to k. Fix $h \in \mathscr{H}$ with $|\mathrm{dom}(h)| < k$, and fix $x \in I \backslash \mathrm{dom}(h)$. For every $a \in S$, let h_a be the extension of h with domain $\mathrm{dom}(h) \cup \{x\}$ that maps x to a. We claim that there exists some $a \in S$ such that h_a is consistent with all substructures of $\mathbf{I}$ with at most c/p many tuples. Once the claim is proved we can conclude from Lemma 9 that h_a belongs to $\mathscr{H}$ and the proof will be complete. To prove the claim, assume that for each $a \in S$ there is some substructure $\mathbf{J}_a$ of $\mathbf{I}$ that falsifies it. Then, h is not consistent with $\bigcup_{a \in S} \mathbf{J}_a$, which is a substructure of $\mathbf{I}$ that has at most c many tuples, since $|J_a| \leq c/p$ and $p = |S|$; a contradiction. $\square$

5.4 Settings of parameters

The data r, p, q are fixed and independent of n, but k is set to ϵn for a small enough positive constant $\epsilon > 0$ to be determined next (in Equations (26) and (28) below). We set δ and β to their upper bounds in Conditions C1 and C2. In particular, δ and β are constants independent of n and $\delta = (r-1)\beta - 1$. Set $d = r^r q^{r-1} \ln(2q)$, so d is also a constant independent of n. Set $\alpha = \epsilon p/(\delta\beta)$. If ϵ is small enough, namely, if

$$0 < \epsilon < (\delta\beta/p)(\beta/d)^{1/(r-1)}(r/e)^{r/(r-1)}, \tag{26}$$

then Condition C3 is satisfied. Set $c = kp/\delta$, so that Condition C4 is satisfied. Now choose n is large enough so that $n \geq q$; observe that the choices of α and c are made so that $c/(\alpha\beta) = n$, so Condition C5 is satisfied. Since d is a constant independent of n, Condition C6 is also satisfied, if n is large enough. We still need to check that Condition C7 holds. By the choice of d we have that $p_1(r, d, n, q) = 1/2^n$, which approaches 0 as n grows to infinity. Set

$$\begin{aligned}
\rho_1(n) &:= (1/\sqrt{n})^{\delta} d^{\beta} e^{1+(r+1)\beta} r^{-r\beta} \beta^{-\beta}, \\
\rho_2(n) &:= (\epsilon p/(\delta\beta))^{\delta} d^{\beta} e^{1+(r+1)\beta} r^{-r\beta} \beta^{-\beta}.
\end{aligned}$$

Splitting the sum that defines $p_2(r, d, n, \alpha, \beta)$ into $v \leq \lfloor\sqrt{n}\rfloor$ and $v \geq \lfloor\sqrt{n}\rfloor + 1$ we get

$$p_2(r, d, n, \alpha, \beta) \leq \textstyle\sum_{v=1}^{\lfloor\sqrt{n}\rfloor} \rho_1(n)^v + \sum_{v=\lfloor\sqrt{n}\rfloor+1}^{\lfloor\alpha n\rfloor} \rho_2(n)^v. \tag{27}$$

Now note that $\rho_1(n) < 1/3$ for large enough n because δ, d, r, β, p are constants independent of n. Also, if ϵ is positive but small enough, namely, if

$$0 < \epsilon < (\delta\beta/(3p))d^{-\beta/\delta}e^{-1/\delta-(r+1)\beta/\delta}r^{r\beta/\delta}\beta^{\beta/\delta}, \quad (28)$$

then also $\rho_2(n) < 1/3$ again because δ, d, r, β, p are constants independent of n and of ϵ. Thus, the sum in (27) is strictly less than $\sum_{v=1}^{\infty}(1/3)^v = 1/2$. Therefore, $p_1 + p_2 < 1/2^n + 1/2 < 1$ for large enough n, which proves that Condition C7 is satisfied, as was to be shown.

5.5 Special case of digraphs

In this section we assume that $r = 2$ and use it to improve the parameters. The improvement is that we can replace the 1/21 that would result from plugging $r = 2$ into the right-hand side of Condition C1 by 1/2. The price to pay for this is that the right-hand side of Condition C4 becomes slightly bigger.

We indicate the required changes in the previous proof. Conditions C1 and C4 are replaced by the following:

C1': $\quad 0 < \delta < 1/2,$
C4': $\quad c \geq kp/\delta'.$

where

$$\delta' := (1 - 2\delta)/(6(1 + \delta)) \quad (29)$$

Lemma 7 becomes the following:

Lemma 11. *For every substructure* $\mathbf{J}$ *of* $\mathbf{I}$ *and every integer* $m \geq 0$*, if* $r = 2$ *and* $\mathbf{J}$ *has* m *many tuples and* $m \leq c$*, then* $\mathbf{J}$ *has at least* $\delta' m$ *many pairwise disjoint boundary sets.*

Proof. We can assume that $\mathbf{J}$ does not contain elements of degree zero since every boundary set of any substructure obtained by removing elements of degree 0 from $\mathbf{J}$ is also a boundary set of $\mathbf{J}$. In the special case $r = 2$, the boundary sets of types (1) and (2) are in one-to-one correspondence with the vertices of degree one and the vertices of degree two, respectively. Thus, since $\mathbf{J}$ does not contain elements of degree zero, and since each boundary set involves two vertices, it suffices to show that $\mathbf{J}$ contains at least $2\delta' m$ many vertices of degree at most two; this will give at least $\delta' m$ many pairwise disjoint boundary sets.

Let $v = |J|$ and let X denote the random variable that equals the degree, in $\mathbf{J}$, of a uniformly chosen random element of J. The sum of the degrees of the elements in $\mathbf{J}$ is $2m$. Therefore, $\mathbb{E}[X] = 2m/v$. Now recall that, by Lemma 5, we have $v \geq (r-1)m/(1+\delta) = m/(1+\delta)$, since $m \leq c$ and $r = 2$. By Markov's inequality, the probability that $X \geq 3$ is bounded by $\mathbb{E}[X]/3 \leq (2m/v)/3 \leq 2(1+\delta)/3$. Therefore, at least a $1 - 2(1+\delta)/3$ fraction of the elements of $\mathbf{J}$ have degree strictly less than 3 in $\mathbf{J}$, which means that $\mathbf{J}$ has at least $(1 - 2(1+\delta)/3)v \geq (1 - 2(1+\delta)/3)m/(1+\delta) = 2\delta' m$ elements of degree at most two. The lemma is proved. □

In the proof of Lemma 8, we need to replace the occurrence of δ by δ', so we can call Lemma 11 (instead of calling Lemma 7), and use Condition C1' (instead of using Condition C1) to ensure that $\delta' m > 0$, as is required in the proof. In the proof of Lemma 9, we need to replace the three occurrences of δ by δ', so we can call Lemma 11 (instead of calling Lemma 7), and use Condition C4' (instead of using Condition C4) to ensure that $\delta' c/p \geq k$, as is required in the proof. Except for these changes, the structure of the proof is exactly the same: Lemma 5 states that there is an instance $\mathbf{I}$ that is (α, β)-sparse such that $\mathbf{I} \not\to \mathbf{S}$, and Lemmas 8 and 9 give Lemma 10; i.e., $\mathbf{I} \leq_k \mathbf{S}$.

6 Approximate Chromatic Number

For a graph $\mathbf{G}$ with n vertices, there is always an integer q satisfying $1 \leq q \leq n$ such that $\mathbf{G} \to \mathbf{K}_q$; the smallest such q is the *chromatic number* of $\mathbf{G}$. Any homomorphism from $\mathbf{G}$ to $\mathbf{K}_q$ is called a proper q-coloring of $\mathbf{G}$. The problem of properly coloring a graph with as few as possible number of colors has a long history. Finding the exact chromatic number is of course one of the classical NP-hard problems, straight from Karp's 21 list [34]. The exact computational complexity of the problem of approximating the chromatic number is much less understood, despite the important progress on the problem since the discovery of the PCP Theorem. In this section we study the width complexity of the problem.

6.1 Constant chromatic numbers

In the regime of constant chromatic numbers, it is conjectured that the problem of q-coloring p-colorable graphs is NP-hard for any two constants p and q such that $3 \leq p \leq q$. The larger the gap between q and p, the stronger the NP-hardness result. For $p = 3$, the current best result of this type, due to Barto, Bulín, Krokhin, and Opršal [8], is that it is NP-hard to 5-color 3-colorable graphs. For constant $p \geq 4$, the current best such result, due to Wrochna and Živný [49], is that it is NP-hard to $q(p)$-color p-colorable graphs, where $q(p) := \binom{p}{\lfloor p/2 \rfloor} - 1$. This improved over the previously known point of NP-hardness for the weaker $q(p) = \exp(\Omega(p^{1/3}))$, which holds for sufficiently large p [32], and for the even weaker $q(p) = 2p - 1$, which holds for all $p \geq 3$ [8]. The full conjecture stating that the problem of q-coloring p-colorable graphs is NP-hard for any two constants p and q such that $3 \leq p \leq q$ is known to follow from certain variants of the *Unique Games Conjecture* (UGC) [27].

These results predict that the promise problems $\mathrm{PCSP}(\mathbf{K}_3, \mathbf{K}_5)$ and $\mathrm{PCSP}(\mathbf{K}_p, \mathbf{K}_{q(p)})$ for $p \geq 4$ are not solvable in bounded width, unless P = NP. The UGC-based results predict that $\mathrm{PCSP}(\mathbf{K}_p, \mathbf{K}_q)$ is not solvable in bounded width for any two constants p and q such that $3 \leq p \leq q$, unless either the suitable variants of the UGC fail or P = NP. We show that our Main Theorem 3 confirms all these predictions, unconditionally, and in the stronger sense of ruling out, not only solvability in constant width, but solvability in sublinear width:

Theorem 4. *For any two integers p and q such that $1 \leq p \leq q$, the following statements are equivalent:*

(a) $\mathrm{PCSP}(\mathbf{K}_p, \mathbf{K}_q)$ *is solvable by the consistency algorithm in width* 3*,*

(b) $\mathrm{PCSP}(\mathbf{K}_p, \mathbf{K}_q)$ *is solvable by the consistency algorithm in bounded width,*

(c) $\mathrm{PCSP}(\mathbf{K}_p, \mathbf{K}_q)$ *is solvable by the consistency algorithm in sublinear width,*

(d) $p = 1$ *or* $p = 2$*.*

Proof. If $p = 1$, then the problem is trivial and solvable in width 2, and hence in width 3. If $p = 2$, then the problem is solvable in width 3 since, in this case, the problem amounts to detecting whether $\mathbf{G}$ is 2-colorable, which is solvable in width 3 (see Example 2). For p and q such that $3 \leq p \leq q$, we have $\mathscr{G}(\mathbf{K}_p) = E^{\mathbf{K}_p} = \{(i,j) \in [p]^2 \mid i \neq j\}$, and $E^{\mathbf{K}_p} \circ E^{\mathbf{K}_p} = [p]^2$, while $E^{\mathbf{K}_q}$ is clearly irreflexive. Therefore, Theorem 3 implies that $\mathrm{PCSP}(\mathbf{K}_p, \mathbf{K}_q)$ is not solvable in sublinear width, which completes the proof. □

6.2 Algorithms with sublinear guarantees

Turning to upper bounds, a well-known algorithm of Wigderson [47] shows that the problem of $O(\sqrt{n})$-coloring 3-colorable graphs is solvable in polynomial time. We observe that, in its decision variant, Wigderson's algorithm can be thought of as a *width four* algorithm:

Theorem 5. *For every graph* $\mathbf{G}$*, if* $\mathbf{G} \leq_4 \mathbf{K}_3$*, then* $\mathbf{G} \to \mathbf{K}_{3\lceil\sqrt{n}\rceil}$*, where* n *is the number of vertices of* $\mathbf{G}$*.*

Proof. Fix a graph $\mathbf{G}$ with vertex-set $V = [n]$. Let $t = \lceil\sqrt{n}\rceil$. Assume that $\mathbf{G} \leq_4 \mathbf{K}_3$. We prove, by induction on m, that for every $Y \subseteq V$ with $|Y| \leq mt$, we have $\mathbf{G}|_Y \to \mathbf{K}_{t+2m}$. Since $t + 2\lceil n/t\rceil \leq 3t$, the claim will follow by setting $m = \lceil n/t\rceil$ and $Y = V$.

Fix m and $Y \subseteq V$ with $|Y| \leq mt$. If $m = 1$ or $\mathbf{G}|_Y$ has all vertices of degree less than t, then $\mathbf{G}|_Y$ is t-colorable, so $\mathbf{G}|_Y \to \mathbf{K}_t \to \mathbf{K}_{t+2m}$ and the claim is proved. Assume then that $m > 1$ and $\mathbf{G}|_Y$ has some vertex $v \in Y$ of degree at least t. Let $N \subseteq Y \setminus \{v\}$ be the neighborhood of v in $\mathbf{G}|_Y$. Since $\mathbf{G} \leq_4 \mathbf{K}_3$, also $\mathbf{G}|_{N\cup\{v\}} \leq_4 \mathbf{K}_3$, implying $\mathbf{G}|_N \leq_3 \mathbf{K}_2$. But then (see Example 2) $\mathbf{G}|_N$ is 2-colorable and hence $\mathbf{G}|_{N\cup\{v\}}$ is 3-colorable. Let h_v be a 3-coloring $\mathbf{G}|_{N\cup\{v\}}$ using the colors $\{a, a+1, a+2\}$, where $a := t + 2m - 2$ and $h_v(v) = a$. Let $X = Y \setminus (N \cup \{v\})$. Then $|X| < |Y| - t \leq (m-1)t$ and hence, by induction hypothesis, $\mathbf{G}|_X \to \mathbf{K}_{t+2(m-1)}$. Let g be a homomorphism from $\mathbf{G}|_X$ to $\mathbf{K}_{t+2m-2}$ and note that $g \cup h_v$ is a homomorphism from $\mathbf{G}|_Y$ to $\mathbf{K}_{t+2m}$; to see this, observe that the only possible color in $\mathrm{img}(g) \cap \mathrm{img}(h_v)$ is $a = h_v(v)$, but v is not adjacent in $\mathbf{G}|_Y$ to any vertex in $\mathbf{G}|_X$, since $N \cap X = \emptyset$ by choice of X. The proof is complete. □

In the rest of section we study the optimal width $k = k(n)$ that guarantees that, for any graph $\mathbf{G}$ with n vertices it holds that $\mathbf{G} \leq_{k(n)} \mathbf{K}_3$ implies $\mathbf{G} \to \mathbf{K}_{O(n^\epsilon)}$ for arbitrary but fixed $\epsilon \in (0, 1/2)$.

As a first observation, it is straightforward to show that $k(n) \leq \lceil n^{1-\epsilon}\rceil$ suffices; i.e., for every graph $\mathbf{G}$ with n vertices, if $\mathbf{G} \leq_{\lceil n^{1-\epsilon}\rceil} \mathbf{K}_3$, then $\mathbf{G} \to \mathbf{K}_{3\lceil n^\epsilon\rceil}$. Indeed, if V denotes the set of vertices of $\mathbf{G}$, then one just selects a subset $X \subseteq V$ of $\lceil n^{1-\epsilon}\rceil$ many vertices, properly colors $\mathbf{G}|_X$ with three colors using the assumption that $\mathbf{G} \leq_{\lceil n^{1-\epsilon}\rceil} \mathbf{K}_3$, and proceeds to the rest of the graph $\mathbf{G}|_{V\setminus X}$ with new colors. Overall this uses at most $3\lceil n^\epsilon \rceil$ colors. We show that, by generalizing Wigderson's method, this width upper bound of $O(n^{1-\epsilon})$ can be improved to $O(n^{1-2\epsilon})$. This builds on some of the ideas of Blum [11], who showed that Wigderson's algorithm can be improved to $n^{3/8}$-color 3-colorable graphs in polynomial time. Note that we do not achieve polynomial time, only sublinear width, but we also show that this is necessary: for any $\gamma < 1 - 3\epsilon$, width $O(n^\gamma)$ does *not* suffice. Note the slight gap between the $1 - 2\epsilon$ in the upper bound and the $1 - 3\epsilon$ in the lower bound.

Theorem 6. *Fix a real $\epsilon \in (0, 1/2)$ and an integer function $q(n)$ such that $q(n) \geq 3$ holds for all integers $n \geq 1$, and $q(n) = \Theta(n^\epsilon)$. The following statements hold:*

1. *There is an integer function $k(n) = O(n^{1-2\epsilon})$ such that, for every graph $\mathbf{G}$, if $\mathbf{G} \leq_{k(n)} \mathbf{K}_3$ then $\mathbf{G} \to \mathbf{K}_{q(n)}$, where n is the number of vertices of $\mathbf{G}$.*
2. *For every real $\gamma < 1-3\epsilon$ and every integer function $k(n) = O(n^\gamma)$ there exist arbitrarily large graphs $\mathbf{G}$ such that $\mathbf{G} \leq_{k(n)} \mathbf{K}_3$ and $\mathbf{G} \not\to \mathbf{K}_{q(n)}$, where n is the number of vertices of $\mathbf{G}$.*

Proof of 1. Choose $k(n) := \max\{3 + \lceil C^2 n^{1-2\epsilon} \rceil, n_0\}$ for sufficiently large integers C and n_0 to be determined later; their choice will depend on the constants that are implicit in the assumption that $q(n) = \Theta(n^\epsilon)$. Note that $k(n)$ is indeed $O(n^{1-2\epsilon})$. Fix any integer n and let $\mathbf{G}$ be a graph with vertex-set V and edge-set E with $|V| = n$. Assume that $\mathbf{G} \leq_{k(n)} \mathbf{K}_3$ and let $\mathscr{H}$ be a $k(n)$-strategy on $\mathbf{G}$ and $\mathbf{K}_3$. We shall prove that for every $Y \subseteq V$ there exists a set $X \subseteq Y$ such that $\mathbf{G}|_X$ is 3-colorable and $|X| \geq \min\{m, Cm^{1-\epsilon}\}$ where $m = |Y|$.

Let $Y \subseteq V$. For every $S \subseteq Y$ we use $N(S)$ to denote the neighbourgood of S in the subgraph $\mathbf{G}|_Y$ induced by Y in $\mathbf{G}$, i.e., the set of vertices in Y that are adjacent in $\mathbf{G}|_Y$ to some vertex in S. If $|Y| \leq k(n)$, then, by the extension property up to $k(n)$, the strategy $\mathscr{H}$ contains a mapping h with domain Y. By definition, h is a partial homomorphism from $\mathbf{G}$ to $\mathbf{K}_3$, and hence a proper 3-coloring of $\mathbf{G}|_Y$. Thus, we can just set $X = Y$ and the claim is proved since $|X| = |Y| = m \geq \min\{m, Cm^{1-\epsilon}\}$. Assume then that $|Y| > k(n)$. Now, consider two cases: (a) there exists a subset $S \subseteq Y$ with $|S| = k(n)-3$ such that $|S \cup N(S)| > Cm^{1-\epsilon}$, and (b) such a set does not exist.

Case (a). We shall prove that in this case $\mathbf{G}|_X$ is 3-colorable for $X = S \cup N(S)$, which proves the claim since $|X| > Cm^{1-\epsilon}$. First note that, by the extension property up to $k(n)$, the strategy $\mathscr{H}$ contains a mapping $h : S \to [3]$ with domain S; this follows from the fact that $|S| = k(n) - 3 \leq k(n)$. Furthermore, h is a partial homomorphism, so it is a proper 3-coloring of $\mathbf{G}|_S$. We shall prove that h can be extended to a proper 3-coloring of $\mathbf{G}|_X$. For every node $v \in N(S)$, let $L_v \subseteq [3]$ be the set of *remaining* colors for v, i.e., the set $[3] \setminus \{h(u) \mid u \in S \cap N(\{v\})\}$. Note that $|L_v| \leq 2$ for each $v \in N(S)$. We want to show that there exists a *list* coloring of $\mathbf{G}|_{N(s)}$, i.e., a coloring g of $\mathbf{G}|_{N(s)}$ such that $g(v) \in L_v$ for every $v \in N(S)$. Consider the CSP instance $(\mathbf{A}, \mathbf{B})$ where $A = N(S)$ and $B = [3]$, whose signature σ contains a relation $R_{u,v}$ for every pair u, v of different elements in $N(S)$. In particular, for every $u, v \in N(S)$ with $u \neq v$ we have $R^{\mathbf{A}}_{u,v} = \{(u, v)\}$ and $R^{\mathbf{B}}_{u,v} = [3]^2 \cap (L_u \times L_v)$. It follows immediately from the definition of $\mathbf{A}$ and $\mathbf{B}$ that for every mapping $g : A \to B$, it holds that g is a homomomorphism from $\mathbf{A}$ to $\mathbf{B}$ if and only if g is a list-coloring of $\mathbf{G}|_{N(S)}$. Since $|L_v| \leq 2$ for all $v \in A$ then $\mathrm{Pol}(\mathbf{B})$ contains the function $\varphi : B^3 \to B$ defined as

$$\varphi(x, y, z) = \begin{cases} x & \text{if } x = y \\ z & \text{otherwise} \end{cases}$$

This is a majority operation (see Example 2) and, therefore, it follows that $\mathbf{B}$ has width 3. Consider the set $\mathscr{H}'$ of partial maps from $\mathbf{A}$ to $\mathbf{B}$ that contains $f|_{X \setminus S}$ for every $f \in \mathscr{H}$ with domain S and every set $X \subseteq V$ such that $S \subseteq X$ and $|X| \leq k(n)$. It is easy to see that $\mathcal{H}'$ is a 3-strategy on $\mathbf{A}$ and $\mathbf{B}$, hence $\mathbf{A} \leq_3 \mathbf{B}$. Since $\mathbf{B}$ has width 3 it follows that $\mathbf{A} \to \mathbf{B}$ and we are done.

Case (b). Let $\mathscr{S}$ be any maximal collection of subsets of Y of cardinality $k(n) - 3$ such that $(S \cup N(S)) \cap (S' \cup N(S')) = \emptyset$ for any two distinct $S, S' \in \mathscr{S}$. Note that, as in Case (a), we have that $\mathbf{G}|_S$ has a proper 3-coloring for every $S \in \mathscr{S}$. Then, if we let $X = \bigcup_{S \in \mathscr{S}} S$, then $\mathbf{G}|_X$ also has a proper 3-coloring. Since $|S \cup N(S)| \leq Cm^{1-\epsilon}$ for every $S \in \mathscr{S}$ and $|Y| = m$, we have $|\mathscr{S}| \geq m^\epsilon / C$. It then follows that X has cardinality at least $(m^\epsilon/C) \cdot (k(n) - 3) \geq Cm^{1-\epsilon}$, by the choice of $k(n)$ and the fact that $n \geq m$. This finishes the proof of the claim.

The rest of the proof is fairly standard. The following recursive algorithm produces a valid coloring for $\mathbf{G}|_Y$ for any $Y \subseteq V$. Let $m := |Y|$. If $m \leq Cm^{1-\epsilon}$, then color $\mathbf{G}|_Y$ with three colors, which is possible since in this case $\min\{m, Cm^{1-\epsilon}\} = m = |Y|$. Else, select an $X \subseteq Y$ with $|X| \geq Cm^{1-\epsilon}$ such that $G|_X$ has a proper 3-coloring g, which is again possible since in this case $\min\{m, Cm^{1-\epsilon}\} = Cm^{1-\epsilon}$. Recursively, obtain a proper coloring h of $\mathbf{G}|_{Y \setminus X}$. Renaming colors if necessary we can assume that g and h do not use any color in common. Return $g \cup h$. Note that $g \cup h$ uses at most $Q(m)$ different colors where $Q(m)$ is the solution to the following recurrence:

$$\begin{array}{ll} Q(m) = 3 + Q(m - \lceil Cm^{1-\epsilon} \rceil) & \text{if } m > Cm^{1-\epsilon} \\ Q(m) = 3 & \text{otherwise} \end{array}$$

We shall show that the bound $Q(n) \leq q(n)$ holds for every $n \geq n_0$ for sufficiently large n_0. The statement will follow since the choice of $k(n)$ guarantees $k(n) \geq n_0$ and, therefore, any graph $\mathbf{G}$ with $n < n_0$ many vertices that satisfies $\mathbf{G} \leq_{k(n)} \mathbf{K}_3$ is even 3-colorable. Recall that $q(n) \geq 3$ holds by assumption.

To prove our claim, we first note that, whenever $m \geq n/2$, we have $\lceil Cm^{1-\epsilon} \rceil \geq C(n/2)^{1-\epsilon}$, and therefore it takes at most $\lceil (n/2)/(C(n/2)^{1-\epsilon}) \rceil = \lceil (n/2)^\epsilon / C \rceil$ many iterations of the recurrence to get from $Q(n)$ down to $Q(\lfloor n/2 \rfloor)$. It follows that

$$Q(n) \leq 3\lceil (n/2)^\epsilon / C \rceil + Q(\lfloor n/2 \rfloor) \leq 3(n/2)^\epsilon / C + 3 + Q(\lfloor n/2 \rfloor). \tag{30}$$

Iterating this recurrence we get

$$Q(n) \leq (3/C) n^\epsilon \textstyle\sum_{i \geq 1} (1/2^\epsilon)^i + 3 \log_2(n) \leq (9/C) n^\epsilon + 3 \log_2(n) \leq q(n), \tag{31}$$

where the second inequality follows from the identity $\sum_{i\geq 1} r^i = r/(1-r)$ and the fact that $1/2^\epsilon \leq 1/\sqrt{2}$ holds since $\epsilon \leq 1/2$. The third inequality follows from the assumption that $q(n) = \Theta(n^\epsilon)$ and $n \geq n_0$, provided the constants C and n_0 are chosen large enough. □

Proof of 2. Fix $\gamma < 1 - 3\epsilon$ and $k(n) = O(n^\gamma)$. We analyze the probabilistic construction in the proof of Theorem 3 when $\mathbf{S} = \mathbf{K}_p$ and $\mathbf{T} = \mathbf{K}_q$ for $p = 3$ and $q = q(n) = \Theta(n^\epsilon)$. Note that q is now also a function of n and that the argument presented in Section 5 allows this generality. Since $r = 2$, we use the version of the proof in Section 5.5. Furthermore, as in the proof of Theorem 4, the left template $\mathbf{S} = \mathbf{K}_p$ satisfies Assumption A1, and it is obvious that the right template $\mathbf{T} = \mathbf{K}_q$ satisfies Assumption A2. Thus, we just need to set the parameters in the proof.

The given data is r, p, q, k, where $q = q(n)$ and $k = k(n)$ are functions of n, and we need to produce, for every large enough integer n, a choice of the real parameters $\delta, \beta, \alpha, c, d$, that may or may not be functions of n, in such a way that Conditions C1',C2,C3,C4',C5,C6,C7 hold, where Conditions C1' and C4' are stated in Section 5.5, and Conditions C2,C3,C5,C6,C7 are stated in Section 5.1. Once we achieve this, Lemma 5 will provide a graph $\mathbf{G}$ with n vertices that is (α, β)-sparse such that $\mathbf{G} \not\to \mathbf{K}_q$, holds. By Lemma 10 (derived as in Section 5.5), this $\mathbf{G}$ will also satisfy $\mathbf{G} \leq_k \mathbf{K}_p$, and since this will succeed for any large enough n, the statement will be proved.

Set $\delta_0 := \epsilon/(1 - \epsilon - \gamma)$ and note that the assumption $\gamma < 1 - 3\epsilon$ implies $\delta_0 < 1/2$. Set δ to be any positive real in the interval $(\delta_0, 1/2)$. The upper bound $\delta < 1/2$ means that Condition C1' holds. For later use, we note that the lower bound $\delta_0 < \delta$ implies

$$\gamma < 1 - (1 + \delta)\epsilon/\delta. \tag{32}$$

Set $\beta := 1 + \delta$, so Condition C2 is satisfied. Set $d = d(n) := 5q(n)\ln(q(n))$, so Condition C6 is satisfied for all large enough n. Set $\alpha = \alpha(n) := (C/d(n))^{(1+\delta)/\delta}$ where $C := (1 + \delta)4^{\delta/(1+\delta)}e^{(-4-3\delta)/(1+\delta)}$. Observe that $d(n)$ is an increasing function of n, while C, β, r are constants independent of n. In particular the rate at which $\alpha(n)$ approaches 0 is that of $(1/d(n))^{(1+\delta)/\delta}$, and the rate at which the right-hand side in Condition C3 approaches 0 is that of $(1/d(n))^{1/(r-1)}$. Since $1/(r-1) = 1 < (1+\delta)/\delta$, this means that Condition C3 holds for all large enough n. Finally, set $c = c(n) := (p/\delta')k(n)$ for $\delta' := (1 - 2\delta)/(6(1 + \delta))$, so Condition C4' is satisfied for all n. We need to argue that Conditions C5 and C7 hold for all large enough n

To argue that Condition C5 holds we need to show that $n \geq \max\{c(n)/(\alpha(n)\beta), q(n)\}$ for all large enough n. Clearly $n \geq q(n)$ for all large enough n since $q(n) = \Theta(n^\epsilon)$ and $\epsilon < 1/2$. Thus, it suffices to show that $n \geq c(n)/(\alpha(n)\beta)$ or, equivalently, that $k(n) \leq f(n) := (\beta\delta'/p)\alpha(n)n$, for all large enough n. First, recall that the parameters $\epsilon, p, \beta, C, \delta'$ are constants independent of n. Therefore, recalling that $d(n) = 5q(n)\ln(q(n)) = \Theta(\epsilon n^\epsilon \ln(n))$, the growth rate of $f(n)$ is that of $n^{1-(1+\delta)\epsilon/\delta}\ln(n)^{-(1+\delta)/\delta}$. On the other hand, the growth rate of $k(n)$ is bounded above by that of n^γ. Now, the choice of δ guarantees (32) and, therefore, we have $k(n) = o(f(n))$. It follows that, for all large enough n, it holds that $k(n) \leq f(n)$, which means that Condition C5 holds.

To argue that Condition C7 holds, observe that the choice of $d(n)$ ensures that $p_1(n) \to 0$ as $n \to \infty$, and the choice of $\alpha(n)$ (and the constant C) ensures that $p_2(n)$ is bounded by $\sum_{v \geq 1}(1/4)^v = 1/3$ for all n. Both facts together imply that $p_1(n) + p_2(n) < 1$ and Condition C7 holds for all large enough n. $\square$

Recall from Section 3.1 that there is an algorithm that, given a graph $\mathbf{G}$ and an integer k, decides whether $\mathbf{G} \leq_k \mathbf{K}_3$ in time polynomial in n^k, where n is the number of vertices of $\mathbf{G}$ and, if so, returns a strategy $\mathscr{H}$. The proof of the width upper bound in Theorem 6 gives the following:

Theorem 7. *Fix a real $\epsilon \in (0, 1/2)$ and an integer function $q(n)$ such that $q(n) \geq 3$ holds for all integers $n \geq 1$, and $q(n) = \Theta(n^\epsilon)$. Then, there is an algorithm that finds a proper $q(n)$-coloring of any given 3-colorable graph with n vertices in $2^{\Theta(n^{1-2\epsilon}\log(n))}$-time.*

Proof. We analyse the recursive algorithm given in the proof of the first part of Theorem 6. The algorithm starts by computing a strategy $\mathscr{H}$ that witnesses $\mathbf{G} \leq_{k(n)} \mathbf{K}_3$; such a strategy exists because, indeed, the assumption is that $\mathbf{G} \to \mathbf{K}_3$. The runtime of this step is polynomial in $n^{k(n)}$. Once $\mathscr{H}$ is computed, the algorithm proceeds recursively as described in the proof of Theorem 6 starting at $Y = V$, where V is the set of all vertices of $\mathbf{G}$. To find the required set $X \subseteq Y$ with $|X| \geq \min\{m, Cm^{1-\epsilon}\}$ for the Y of cardinality m in the current recursive call, we first need to tell whether Y falls in Case (a) or in Case (b). For this, it suffices to loop through all subsets $S \subseteq Y$ with $|S| = k(n) - 3$, and compute $|S \cup N(S)|$. The number of such sets is bounded by $n^{k(n)-3}$ and hence can be looped in time polynomial in $n^{k(n)}$. In Case (a), we color $\mathbf{G}|_X$ for $X = S \cup N(S)$ with 3 colors as follows: first find an $h \in \mathscr{H}$ with $\mathrm{dom}(h) = S$, and then extend h to a proper 3-coloring of $\mathbf{G}|_X$ by solving the CSP instance $(\mathbf{A}, \mathbf{B})$. We are using here the well-known fact (see [24]) that for CSPs an polynomial-time algorithm for the decision variant yields immediately and polynomial-time algorithm for the search version. In Case (b), we greedily find a maximally disjoint family $\mathscr{S}$ of sets of the form $S \cup N(S)$ with $S \subseteq Y$, and color $\mathbf{G}|_X$ for $X = \bigcup_{S \in \mathscr{S}}$ with three colors as $h = \bigcup_{S \in \mathscr{S}} h_S$, where $h_S \in \mathscr{H}$ with $\mathrm{dom}(h_S) = S$ is a suitably found proper 3-coloring of $\mathbf{G}|_S$ for each $S \in \mathscr{S}$. Each recursive call shrinks the size of the calling set Y from m to $m - \lceil Cm^{1-\epsilon} \rceil$, which means that the algorithm ends after a linear in n number of recursive calls, each of which takes time polynomial in $n^{k(n)}$. For $k(n) = \Theta(n^{1-2\epsilon})$, this is time complexity $2^{\Theta(n^{1-2\epsilon}\log(n))}$ overall, and the proof is complete. $\square$

6.3 Discussion and an open problem

Some discussion on the runtime of the algorithm in Theorem 7 is in order. On one hand, the simple observation we made just before the statement of Theorem 7 that width $\lceil n^{1-\epsilon} \rceil$ suffices already gives a very simple algorithm that properly $\Theta(n^\epsilon)$-colors 3-colorable graphs with n vertices in subexponential $2^{\Theta(n^{1-\epsilon})}$-time. The algorithm of Theorem 7 is only slightly more complicated and asymptotically beats this. On the other hand, using more sophisticated techniques, it was shown in [4] that, for any desired approximation factor f, there is a $2^{\tilde{\Theta}(n/(f\log(f)+f\log(f)^2))}$-time randomized algorithm that approximates the chromatic number of a graph with n vertices within a factor of f. For $f = \Theta(n^\epsilon)$, this gives a $2^{\Theta(n^{1-\epsilon-o(1)})}$-time randomized algorithm for the problem of $\Theta(n^\epsilon)$-coloring 3-colorable graphs with n vertices. Interestingly, the simple width-based algorithm from Theorem 7 also beats this, and is deterministic (but of course it applies only to our problem and not to the more general problem of approximating the chromatic number).

Whether the runtime $2^{\Theta(n^{1-2\epsilon}\log(n))}$ of Theorem 7 can be beaten is an interesting question left open by our work. Our width lower bound of $n^{1-3\epsilon}$ has as a consequence that $2^{\Omega(n^{1-3\epsilon})}$ appears to be a lower limit on the runtime of any width-based algorithm. The obstacle to improving the width lower bound from $n^{1-3\epsilon}$ to $n^{1-2\epsilon}$ is Condition C1', which is an improvement for the special case of graphs over Condition C1 of the general case. Ideally, Condition C1' should be improved further to Condition C1" defined as $0 < \delta < 1$. We do not know if this is possible; we leave it as an open problem.

References

[1] Albert Atserias, Andrei A. Bulatov, and Anuj Dawar. Affine systems of equations and counting infinitary logic. *Theor. Comput. Sci.*, 410(18):1666–1683, 2009.

[2] Albert Atserias and Elitza N. Maneva. Sherali-Adams Relaxations and Indistinguishability in Counting Logics. *SIAM J. Comput.*, 42(1):112–137, 2013.

[3] Albert Atserias and Joanna Ochremiak. Proof Complexity Meets Algebra. *ACM Trans. Comput. Log.*, 20(1):1:1–1:46, 2019.

[4] Nikhil Bansal, Parinya Chalermsook, Bundit Laekhanukit, Danupon Nanongkai, and Jesper Nederlof. New Tools and Connections for Exponential-Time Approximation. *Algorithmica*, 81(10):3993–4009, 2019.

[5] Libor Barto. The collapse of the bounded width hierarchy. *Journal of Logic and Computation*, 26(3):923–943, 2016.

[6] Libor Barto. Promises Make Finite (Constraint Satisfaction) Problems Infinitary. In *34th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2019, Vancouver, BC, Canada, June 24-27, 2019*, pages 1–8. IEEE, 2019.

[7] Libor Barto, Diego Battistelli, and Kevin M. Berg. Symmetric Promise Constraint Satisfaction Problems: Beyond the Boolean Case. In Markus Bläser and Benjamin Monmege, editors, *38th International Symposium on Theoretical Aspects of Computer Science, STACS 2021, March 16-19, 2021, Saarbrücken, Germany (Virtual Conference)*, volume 187 of *LIPIcs*, pages 10:1–10:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021.

[8] Libor Barto, Jakub Bulín, Andrei A. Krokhin, and Jakub Oprsal. Algebraic approach to promise constraint satisfaction. *To appear at J.ACM (also CoRR abs/1811.00970)*, 2018. Preliminary version at Proceedings of STOC 2019.

[9] Libor Barto and Marcin Kozik. Constraint Satisfaction Problems Solvable by Local Consistency Methods. *J. ACM*, 61(1):3:1–3:19, 2014.

[10] Paul Beame, Joseph C. Culberson, David G. Mitchell, and Cristopher Moore. The resolution complexity of random graph k-colorability. *Discret. Appl. Math.*, 153(1-3):25–47, 2005.

[11] Avrim Blum. New Approximation Algorithms for Graph Coloring. *J. ACM*, 41(3):470–516, 1994.

[12] Joshua Brakensiek and Venkatesan Guruswami. Promise Constraint Satisfaction: Algebraic Structure and a Symmetric Boolean Dichotomy. *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018 (Full version at CoRR abs/1704.01937)*, 2018.

[13] Joshua Brakensiek and Venkatesan Guruswami. An Algorithmic Blend of LPs and Ring Equations for Promise CSPs. In Timothy M. Chan, editor, *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 436–455. SIAM, 2019.

[14] Joshua Brakensiek, Venkatesan Guruswami, and Sai Sandeep. Conditional Dichotomy of Boolean Ordered Promise CSPs. In Nikhil Bansal, Emanuela Merelli, and James Worrell, editors, *48th International Colloquium on Automata, Languages, and Programming, ICALP 2021, July 12-16, 2021, Glasgow, Scotland (Virtual Conference)*, volume 198 of *LIPIcs*, pages 37:1–37:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021.

[15] Joshua Brakensiek, Venkatesan Guruswami, Marcin Wrochna, and Stanislav Zivný. The Power of the Combined Basic Linear Programming and Affine Relaxation for Promise Constraint Satisfaction Problems. *SIAM J. Comput.*, 49(6):1232–1248, 2020.

[16] Alex Brandts, Marcin Wrochna, and Stanislav Zivný. The Complexity of Promise SAT on Non-Boolean Domains. In Artur Czumaj, Anuj Dawar, and Emanuela Merelli, editors, *47th International Colloquium on Automata, Languages, and Programming, ICALP 2020, July 8-11, 2020, Saarbrücken, Germany (Virtual Conference)*, volume 168 of *LIPIcs*, pages 17:1–17:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

[17] Alex Brandts and Stanislav Zivný. Beyond PCSP(1-in-3, NAE). In Nikhil Bansal, Emanuela Merelli, and James Worrell, editors, *48th International Colloquium on Automata, Languages, and Programming, ICALP 2021, July 12-16, 2021, Glasgow, Scotland (Virtual Conference)*, volume 198 of *LIPIcs*, pages 121:1–121:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021.

[18] Andrei A. Bulatov. A Dichotomy Theorem for Nonuniform CSPs. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS*

2017, Berkeley, CA, USA, October 15-17, 2017, pages 319–330. IEEE Computer Society, 2017.

[19] Andrei A. Bulatov, Peter Jeavons, and Andrei A. Krokhin. Classifying the Complexity of Constraints Using Finite Algebras. *SIAM J. Comput.*, 34(3):720–742, 2005.

[20] Silvia Butti and Victor Dalmau. Fractional Homomorphism, Weisfeiler-Leman invariance, and the Sherali-Adams hierarchy for the Constraint Satisfaction Problem. In *46th International Symposium on Mathematical Foundations of Computer Science (MFCS'21), to appear*, 2021.

[21] Silvia Butti and Victor Dalmau. The Complexity of the Distributed Constraint Satisfaction Problem. In Markus Bläser and Benjamin Monmege, editors, *38th International Symposium on Theoretical Aspects of Computer Science, STACS 2021, March 16-19, 2021, Saarbrücken, Germany (Virtual Conference)*, volume 187 of *LIPIcs*, pages 20:1–20:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021.

[22] Siu On Chan, James R. Lee, Prasad Raghavendra, and David Steurer. Approximate Constraint Satisfaction Requires Large LP Relaxations. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 350–359. IEEE Computer Society, 2013.

[23] Vasek Chvátal and Endre Szemerédi. Many Hard Examples for Resolution. *J. ACM*, 35(4):759–768, 1988.

[24] David A. Cohen. Tractable Decision for a Constraint Language Implies Tractable Search. *Constraints An Int. J.*, 9(3):219–229, 2004.

[25] Víctor Dalmau and Andrei A. Krokhin. Robust Satisfiability for CSPs: Hardness and Algorithmic Results. *ACM Trans. Comput. Theory*, 5(4):15:1–15:25, 2013.

[26] Víctor Dalmau, Andrei A. Krokhin, and Rajsekar Manokaran. Towards a characterization of constant-factor approximable finite-valued CSPs. *J. Comput. Syst. Sci.*, 97:14–27, 2018.

[27] Irit Dinur, Elchanan Mossel, and Oded Regev. Conditional Hardness for Approximate Coloring. *SIAM J. Comput.*, 39(3):843–873, 2009.

[28] Tomás Feder and Moshe Y. Vardi. The Computational Structure of Monotone Monadic SNP and Constraint Satisfaction: A Study through Datalog and Group Theory. *SIAM J. Comput.*, 28(1):57–104, 1998.

[29] Miron Ficak, Marcin Kozik, Miroslav Olšák, and Szymon Stankiewicz. Dichotomy for Symmetric Boolean PCSPs. In Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi, editors, *46th International Colloquium on Automata, Languages, and Programming, ICALP 2019, July 9-12, 2019, Patras, Greece*, volume 132 of *LIPIcs*, pages 57:1–57:12. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019.

[30] Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. Optimal Sherali-Adams Gaps from Pairwise Independence. In Irit Dinur, Klaus Jansen, Joseph Naor, and José D. P. Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, 12th International Workshop, APPROX 2009, and 13th International Workshop, RANDOM 2009, Berkeley, CA, USA, August 21-23, 2009. Proceedings*, volume 5687 of *Lecture Notes in Computer Science*, pages 125–139. Springer, 2009.

[31] Mrinalkanti Ghosh and Madhur Tulsiani. From Weak to Strong Linear Programming Gaps for All Constraint Satisfaction Problems. *Theory Comput.*, 14(1):1–33, 2018.

[32] Sangxia Huang. Improved Hardness of Approximating Chromatic Number. *CoRR*, abs/1301.5216, 2013. Preliminary version in APPROX/RANDOM 2013.

[33] Peter Jeavons, David A. Cohen, and Marc Gyssens. Closure properties of constraints. *J. ACM*, 44(4):527–548, 1997.

[34] Richard M. Karp. Reducibility Among Combinatorial Problems. In Raymond E. Miller and James W. Thatcher, editors, *Proceedings of a symposium on the Complexity of Computer Computations, held March 20-22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, USA*, The IBM Research Symposia Series, pages 85–103. Plenum Press, New York, 1972.

[35] Sanjeev Khanna, Nathan Linial, and Shmuel Safra. On the Hardness of Approximating the Chromatic Number. *Comb.*, 20(3):393–415, 2000.

[36] Phokion G. Kolaitis and Moshe Y. Vardi. A Game-Theoretic Approach to Constraint Satisfaction. In Henry A. Kautz and Bruce W. Porter, editors, *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on on Innovative Applications of Artificial Intelligence, July 30 - August 3, 2000, Austin, Texas, USA*, pages 175–181. AAAI Press / The MIT Press, 2000.

[37] Phokion G. Kolaitis and Moshe Y. Vardi. Conjunctive-Query Containment and Constraint Satisfaction. *J. Comput. Syst. Sci.*, 61(2):302–332, 2000.

[38] Phokion G. Kolaitis and Moshe Y. Vardi. A Logical Approach to Constraint Satisfaction. In *Complexity of Constraints - An Overview of Current Research Themes [Result of a Dagstuhl Seminar]*, pages 125–155, 2008.

[39] Andrei A. Krokhin and Jakub Oprsal. The Complexity of 3-Colouring H-Colourable Graphs. In David Zuckerman, editor, *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 1227–1239. IEEE Computer Society, 2019.

[40] Amit Kumar, Rajsekar Manokaran, Madhur Tulsiani, and Nisheeth K. Vishnoi. On LP-Based Approximability for Strict CSPs. In Dana Randall, editor, *Proceedings of the*

Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2011, San Francisco, California, USA, January 23-25, 2011, pages 1560–1573. SIAM, 2011.

[41] Gabor Kun, Ryan O'Donnell, Suguru Tamaki, Yuichi Yoshida, and Yuan Zhou. Linear Programming, Width-1 CSPs, and Robust Satisfaction. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ITCS '12, page 484–495, New York, NY, USA, 2012. Association for Computing Machinery.

[42] Benoît Larose and László Zádori. Taylor Terms, Constraint Satisfaction and the Complexity of Polynomial Equations over Finite Algebras. *Int. J. Algebra Comput.*, 16(3):563–582, 2006.

[43] Carsten Lund and Mihalis Yannakakis. On the Hardness of Approximating Minimization Problems. *J. ACM*, 41(5):960–981, 1994.

[44] Michael Molloy and Mohammad R. Salavatipour. The Resolution Complexity of Random Constraint Satisfaction Problems. *SIAM J. Comput.*, 37(3):895—922, 2007.

[45] Hanif D. Sherali and Warren P. Adams. A Hierarchy of Relaxations Between the Continuous and Convex Hull Representations for Zero-One Programming Problems. *SIAM J. Discrete Math.*, 3(3):411–430, 1990.

[46] Johan Thapper and Stanislav Zivný. The Power of Sherali-Adams Relaxations for General-Valued CSPs. *SIAM J. Comput.*, 46(4):1241–1279, 2017.

[47] Avi Wigderson. Improving the Performance Guarantee for Approximate Graph Coloring. *J. ACM*, 30(4):729–735, 1983.

[48] Marcin Wrochna and Stanislav Zivný. Improved hardness for H-colourings of G-colourable graphs. In Shuchi Chawla, editor, *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 1426–1435. SIAM, 2020.

[49] Marcin Wrochna and Stanislav Zivný. Improved hardness for H-colourings of G-colourable graphs. In Shuchi Chawla, editor, *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 1426–1435. SIAM, 2020.

[50] Yuichi Yoshida and Yuan Zhou. Approximation schemes via Sherali-Adams hierarchy for dense constraint satisfaction problems and assignment problems. In Moni Naor, editor, *Innovations in Theoretical Computer Science, ITCS'14, Princeton, NJ, USA, January 12-14, 2014*, pages 423–438. ACM, 2014.

[51] Dmitriy Zhuk. A Proof of the CSP Dichotomy Conjecture. *J. ACM*, 67(5):30:1–30:78, 2020.