Available online at www.sciencedirect.com
SCIENCE DIRECT®
Discrete Applied Mathematics 153 (2005) 48–57
DISCRETE APPLIED MATHEMATICS
www.elsevier.com/locate/dam

A sharp threshold for the renameable-Horn and the q-Horn properties

Nadia Creignou[a], Hervé Daudé[b], John Franco[c]

[a]LIF, UMR CNRS 6166, Université de la Méditerranée, 163, avenue de Luminy, 13 288 Marseille, France
[b]LATP, UMR CNRS 6632, Université de Provence, 39, rue Joliot-Curie, 13 453 Marseille, France
[c]ECECS, University of Cincinnati, Cincinnati, OH 45221-0030, USA

Received 12 December 2003; received in revised form 6 July 2004; accepted 6 May 2005
Available online 15 August 2005

Abstract

The sharp Satisfiability threshold is well known for random k-SAT formulas and is due to certain minimality and monotonic properties mentioned in this manuscript and reported in Chandru and Hooker [J. Assoc. Comput. Mach. 38 (1991) 205–221]. Whereas the Satisfiability threshold is on the probability that a satisfying assignment exists, we find that sharp thresholds also may be determined for certain formula structures, for example, the probability that a particular kind of cycle exists in a random formula. Such structures often have a direct relationship on the hardness of a formula because it is often the case that the presence of such a structure disallows a formula from a known, easily solved class of Satisfiability problems. We develop tools that should assist in determining threshold sharpness for a variety of applications. We use the tools to show a sharp threshold for the q-Horn and renameable-Horn properties.

Keywords: Sharp Satisfiability threshold; q-Horn; Renameable-Horn; Random formula

1. Introduction

The Satisfiability problem (SAT) is the problem of determining whether there exists an assignment of values to the variables of a given Boolean formula (an instance) which causes

E-mail addresses: creignou@lif.univ-mrs.fr (N. Creignou), daude@gyptis.univ-mrs.fr (H. Daudé), franco@gauss.ececs.uc.edu (J. Franco).

0166-218X/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.dam.2005.05.005

it to evaluate to *true* (a solution). The problem appears in numerous engineering, scientific, and operation research applications such as VLSI testing, design and verification, artificial intelligence, and decision analysis, to name a few. Unfortunately SAT is NP-complete. But, as a result of many advancements over the last decade, many instances previously considered prohibitively difficult are now being solved in a reasonable amount of time. In fact, so much progress has been made that it is sometimes better to translate an NP-complete problem to SAT and solve it in that domain.

However, there remain many hard instances of SAT, for example, in the area of bounded model checking [2,11]. Modern SAT solvers achieve success when they are able to detect and exploit structure in a given instance. But, in bounded model checking, the nature of the instances tends to blur the distinction between variables. In other words, such instances, which are often hard, tend to look like random formulae.

Hence, in order to better understand the nature of instances that are hard for current SAT solvers, it seems reasonable to study the relationship between hardness and random formulae. There has been much work on this subject in recent years (see, for example, [17] for a bibliography), mostly focusing on random instances of k-SAT: a conjunction of m disjunctions (clauses) of width k, chosen uniformly at random among the $2^k\binom{n}{k}$ k-clauses on n Boolean variables and their negations. As m and n tend to infinity with limiting ratio $m/n \to \alpha$, average case analysis and experimental results have provided evidence for the existence of a phase transition at some value r_k of the parameter α. Friedgut [20], with an appendix by Bourgain, proved that k-SAT exhibits a sharp threshold for $k \geqslant 2$ but without specifying its location. While the associated critical ratio has been identified for $k = 2$ ($r_2 = 1$, [10,23]), specifying it for $k \geqslant 3$ remains a challenging problem. For $k = 3$, the best upper bound is $r_3 \leqslant 4.506$ [16] and the best lower bound is $r_3 \geqslant 3.41$ [27]. For every $k \geqslant 3$, $r_k \geqslant 2^{k-2}/k$ [10] and it is known that $r_k \sim_{k\to\infty} 2^k/\ln 2$ [1].

It has also been observed that random instances become harder for SAT solvers when generated with values of m and n, where the ratio m/n is close to r_k and easier when m/n is distant from r_k: the more distant being easier.

These results and observations have suggested a relationship between hardness and threshold. Further investigation has identified long "backbones," or chains of inferences, to be a good candidate for the underlying cause of the sharp thresholds and poor algorithm performance near the thresholds since it appears to be the high density of well-separated "almost solutions" induced by the backbones that lead to thrashing in search algorithms [9]. In [32] and other articles it has been suggested that there is a strong connection between the "order" of threshold sharpness and hardness.

But thresholds can exist for other properties. Two significant ones are: (1) the property that a particular polynomial-time incomplete algorithm finds a satisfying assignment; and (2) the property that a random k-CNF formula is a member of a well-known class of polynomial-time solvable formulas. Little has been done to explain the impact of such thresholds or even to find them. It is known that a sharp threshold exists at $m/n \approx 1.63$ for an incomplete algorithm that applies the pure literal rule to near exhaustion on 3-CNF formulas [7,30]. But there are coarse transitions for non-backtrack DPLL variants, spanning, for example, approximately $2.1 < m/n < 3.7$ when unit clauses are always satisfied and variables are otherwise picked according to the Johnson Heuristic [22,26]. In fact, all studied variants using the unit clause rule seem to have coarse transitions, even when the pure literal rule is added.

For succinctly defined classes of k-CNF that are solved in polynomial time even less is known. Notable examples are Horn [15,25], renameable-Horn [28], q-Horn [5,6], extended Horn [8], SLUR [33], balanced [12], and matched [19], to name a few. These classes have been studied partly in the belief that they will yield some distinction between hard and easy problems. For example, in [5] a satisfiability index is presented such that a class with index greater than $1 + \varepsilon$, for any positive constant ε, is NP-complete but the q-Horn class has satisfiability index 1. Thus, it seems that q-Horn is situated right at the point delineating hard and easy satisfiability problems. This hypothesis has been tested somewhat using m/n as a scale for determining the boundaries, in a probabilistic sense, of q-Horn and other classes; it has been found that a random k-CNF formula is q-Horn with probability tending to 0 if $m/n > 2/k(k-1)$ and that the probability that a random k-CNF formula is q-Horn is bounded away from 0 if $m/n < 1/k(k-1)$ [19]. Similar results have been obtained for other polynomial-time solvable classes. They illuminate the fact that most instances of such classes are satisfiable since their extent on the m/n scale is far below the r_k satisfiability threshold. Since their boundaries, in a probabilistic sense, are so distant from the threshold, all the polynomial-time classes mentioned above may be considered *extremely* easy, especially when compared to the good probabilistic performance shown for polynomial-time incomplete algorithms in the range $m/n < \frac{3}{8}(2^k/k)$ [10]. Why are so many succinctly defined polynomial-time solvable classes so weak and do there exist polynomial-time classes that are more of a challenge (that is, harder or having probabilistic boundaries closer to the satisfiability threshold) and are good candidates for revealing the distinction between hard and easy problems?

Surprisingly, this question seems to have a connection to thresholds of the second kind mentioned above. The classes above, including q-Horn, are “vulnerable” to cyclic clause structures, any one of which prevents a formula containing such a structure from being a member of the class. These structures have the recently discovered minimality and monotonic properties which are necessary for sharp thresholds and are defined in [13] and again in this manuscript. So, it seems to find challenging polynomial-time solvable classes it is advisable to look for classes which are not so vulnerable: that is, those for which formulas cannot be excluded by adding certain minimal monotonic structural components. The tools presented in this manuscript represent the beginning of a collection that may assist in doing so as they make the investigation of thresholds easier. Although the results here are derived specifically for the q-Horn class, similar results undoubtedly may be obtained for other classes as well.

In Section 2, we show how the class of q-Horn formulas can be seen as a non-Boolean CSP. Random CSPs have already been studied by various authors (see [13,31]) and q-Horn appears as a challenging property for proving sharpness. In Section 3, we recall a sharpness criterion [14] deduced from the well-known Friedgut’s one [20] and well suited for random CSPs of fixed arity. In Section 4, in using this criterion and a nice result on supersaturated hypergraphs [18] (which has already been used for proving sharpness of threshold for Ramsey properties on random graphs [21]) we prove that q-Horn exhibits a sharp transition.

2. The renameable-Horn and the q-Horn properties

Let $k \geqslant 3$. We consider k-CNF formulas, $F = \bigwedge_{i=1}^{L} C_i$ over the set of variables $V = \{x_1, \ldots, x_n\}$, where each clause C_i is a disjunction of k literals. The satisfiability problem k-SAT is to decide whether such a formula is satisfiable, that is, whether there exists a truth assignment to the variables that evaluates F true. The k-SAT problem is the prototypical NP-complete problem. Here, we recall two well-known classes of formulas for which satisfiability can be decided in polynomial time, namely renameable-Horn and q-Horn formulas.

Definition 2.1. A formula F is *Horn* if each clause of F has at most one posistive literal.

Horn formulas can be solved in linear time by unit resolution [15,34].

Definition 2.2. *Renaming a variable* x_i corresponds to mapping x_i into $\bar{x}_i$ and vice versa.

Definition 2.3 (*Lewis [28]*). A formula F is *renameable-Horn* if renaming each of some subset of variables of F yields to a Horn formula.

Renameable-Horn formulas can also be solved in linear time [24].

Observe that deciding whether a formula F is renameable-Horn can be seen as a Boolean constraint satisfaction problem. Indeed, for every truth assignment to the variables $\Phi : V \longrightarrow \{0, 1\}$, extended to literals by $\Phi(\bar{x}_i) = 1 - \Phi(x_i)$, and every clause $C = (l' \vee \cdots \vee l_k)$, let us set $\Phi(C) := (\Phi(l_1), \ldots, \Phi(l_k)) = 1$ if and only if at most one of the literals from C is assigned true by Φ. Let us denote $S_n(C) = \{\Phi$ such that $\Phi(C) = 1\}$. Then, it is easy to see that F is renameable-Horn if and only if $\bigcap_{i=1}^{L} S_n(C_i) \neq \emptyset$. A certificate that F is renameable-Horn (or a "satisfying assignment" with respect to the property of being renameable-Horn) is given by a truth assignment Φ such that for every i, $\Phi(C_i) = 1$. Intuitively this assignment identifies the variables that have to be renamed, namely renaming each of the subset of variables $\{x_i \in V / \Phi(x_i) = 0\}$ yields to a Horn formula. Such an assignment Φ is called a *renameable-Horn-certificate* for F.

Note that in the terminology of [14] the renameable-Horn property corresponds to the symmetric Boolean CSP generated by the constraint function f defined by $f(a_1, \ldots, a_k) = 1$ if and only if at most one of the a_i's is equal to 1.

The class q-Horn was developed by Boros et al. [5,6]. Recognition of q-Horn formulas can be done in linear time and satisfiability of q-Horn formulas can be decided in linear time. The q-Horn property can be defined as follows [19, Lemma 3.1, p. 8, 5].

The letter D stands for decomposition, E for east, W for west, W^+ for west with positive polarization (in the matrix representation when the column is multiplied by $+1$), W^- for west with negative polarization (in the matrix representation when the column is multiplied by -1).

We say that F is q-Horn if there exists a decomposition $D : \{x_1, \ldots, x_n\} \longrightarrow \{W^+, W^-, E\}$ which extends to literals by $D(\bar{x}) = E$ if $D(x) = E$, $D(\bar{x}) = W^+$ if $D(x) = W^-$ and $D(\bar{x}) = W^-$ if $D(x) = W^+$, such that for each clause $C_i = (l_1 \vee \cdots \vee l_k)$:

1. either none of the literals l_i is assigned value E, and in this case at most one of them is assigned value W^+,
2. or one or two of the literals are assigned value E, and then all of the others are assigned value W^-.

This formulation has the advantage that in this way the q-Horn property appears as a satisfiability property. Each clause can be seen as a constraint, a satisfying assignment for the formula (a set of constraints) is a decomposition D as described above. Thus, the q-Horn property appears as a constraint satisfaction problem over the three-element domain $\{W^+, W^-, E\}$. For instance, deciding whether the formula $(w \vee x \vee y) \wedge (\bar{w} \vee \bar{x} \vee \bar{y}) \wedge (\bar{x} \vee y \vee z) \wedge (\bar{w} \vee \bar{y} \vee z)$ is q-Horn comes down to deciding whether the following collection of constraints $f_0(w, x, y) \wedge f_3(w, x, y) \wedge f_1(x, y, z) \wedge f_2(w, y, z)$ is satisfiable, where the f_i's are constraint functions over the domain $\{W^+, W^-, E\}$ such that $f_0(a, b, c) = 1$ if and only if $\{a, b, c\} \in \{\{W^-, W^-, W^-\}, \{W^+, W^-, W^-\}, \{E, W^-, W^-\}, \{E, E, W^-\}\}$, and $f_1(x_1, x_2, x_3)$(respectively, $f_2(x_1, x_2, x_3)$, $f_3(x_1, x_2, x_3)$) encodes the constraint f_0 $(\bar{x}_1, x_2, x_3)$ (respectively, $f_0(\bar{x}_1, \bar{x}_2, x_3)$, $f_3(\bar{x}_1, \bar{x}_2, \bar{x}_3)$).

Notation. In the sequel, in order to deal with monotone increasing properties, we denote by $\mathscr{R}$ (resp., $\mathscr{H}$) the property for a k-CNF formula of NOT being renameable-Horn (of NOT being q-Horn). If a formula F is not in $\mathscr{R}$, i.e. if F is renameable-Horn, then F has a *renameable-Horn-certificate* as defined above. If a formula F is not in $\mathscr{H}$, i.e. if F is q-Horn, then F has a *q-Horn-certificate* $D : Var(F) \longrightarrow \{W^+, W^-, E\}$ that verifies the conditions described above.

3. Probabilistic tools

As we noted in the previous section, the number, N, of k-clauses one can build from n variables, and that are of interest in our study is: $N = 2^k\binom{n}{k}$. The properties $\mathscr{R}$ (resp., $\mathscr{H}$) are monotone increasing in the sense that if s is a set of clauses verifying such a property, then so does any set s' of clauses containing s.

When each k-clause appears independently with probability p, the probability for a set of clauses to verify $\mathscr{R}$ (resp., $\mathscr{H}$) can be nicely evaluated in a probabilistic model analogous to the well-known $G_n(p)$ model for random graphs. For any p in [0, 1], and all subset A of k-CNF-formulas we will denote: $\mu_p(A) = \sum_{s \in A} (1-p)^{N-w(s)} p^{w(s)}$, where $w(s)$, the size of s, is the number of clauses in s. In this model, the average size of a set of clauses is $p \cdot N$. Then greater is p, greater is the probability $\mu_p(\mathscr{R})$ (resp., $\mu_p(\mathscr{H})$), which evaluates the probability for a set of clauses of verifying $\mathscr{R}$ (resp., $\mathscr{H}$)

In this paper, we will establish a sharp transition, in the sense of Friedgut–Bourgain [20]: $\mu_p(\mathscr{H})$ (resp., $\mu_p(\mathscr{R})$) increases in a small interval from near 0 to near 1. More precisely, for each of these properties and for any $c \in (0, 1)$, let $p_c(N)$ be defined by $\mu_{p_c}(\mathscr{H}) = c$ (resp., $\mu_{p_c}(\mathscr{R}) = c$). Thus, we will show that for any $\varepsilon \in]0, 1/2]$ the ratio $(p_{1-\varepsilon}(N) - p_\varepsilon(N))/p_{1/2}(N)$ tends to 0 as N tends to infinity. For this we will use a criterion for sharpness, given in [14], deduced from Friedgut–Bourgain's one [20] and dedicated to random CSPs of fixed arity k over a finite domain Dom.

In the previous section, we have shown that $\mathscr{R}$ and $\mathscr{H}$ can be seen as constraint satisfaction problems, with, respectively, $\mathsf{Dom} = \{0, 1\}$ and $\mathsf{Dom} = \{W^+, W^-, E\}$. Thus $\mathscr{R}$, (resp., $\mathscr{H}$) falls in the scope of application of Creignou–Daudé's criterion [14, Theorem 3.4], which tells us that the three following conditions are sufficient to prove sharpness.

(D0) For each $c \in (0, 1)$, $p_c(n) = \mathrm{O}(n^{1-k})$.
(D1) For every m minimal for $\mathscr{R}$ (resp., for $\mathscr{H}$), $\#\mathit{Var}(m) \leqslant (k-1)w(m) - 1$.
(D2) For each $c \in (0, 1)$, for each t, for all $\delta = (\delta_1, \ldots, \delta_t) \in \mathsf{Dom}^t$, and all $\gamma > 0$

$$\mu_{p_c(n)}(s \notin Q_\delta, \#\mathscr{A}_\delta(s) \geqslant \gamma \cdot n^{k-1}) = \mathrm{o}(1),$$

Q_δ denoting the property for a set of clauses s of having no renameable-Horn-certificate (resp., no q-Horn-certificate) with $x_1 = \delta_1, \ldots, x_t = \delta_t$,
$\mathscr{A}_\delta(s)$ denoting, for $s \notin Q_\delta$, the set of clauses C having at least one variable in $\{x_1, \ldots, x_t\}$ and such that $s \cup \{C\} \in Q_\delta$.

4. Sharp threshold results

It turns out that the sharpness of the transition associated to $\mathscr{R}$ can be proved in using the classification theorem on the nature of the threshold for symmetric Boolean CSPs established in [14].

Theorem 4.1. *The property $\mathscr{R}$ exhibits a sharp threshold and the scale for the transition is of order n.*

Proof. In Section 2, we noticed that $\mathscr{R} = \mathsf{UNSAT}(\{f\})$ with

$$f^{-1}(1) = \{(0, \ldots, 0), (0, \ldots, 0, 1), \ldots, (0, 1, \ldots, 0), (1, 0, \ldots, 0)\},$$

in the symmetric model of [14]. It is clear that the function f has no unary clause as an implicate nor a 2-XOR-clause as an implicate (since for any $\varepsilon = 0$ or 1, $f(a_1, \ldots, a_k) = 1$ does not imply $a_i = \varepsilon$ for any $1 \leqslant i \leqslant k$, nor $a_i \oplus a_j = \varepsilon$ for any $1 \leqslant i \neq j \leqslant k$). Therefore, the result follows from the application of the classification theorem given in [14]. □

The sharpness for the transition of property $\mathscr{H}$ is a more challenging task. This property deals with a non-Boolean CSP and therefore does not fall into the scope of application of the classification theorem in [14], moreover it does not verify the sufficient condition for sharpness of random CSPs identified in [31]. We will prove the sharpness in using the criterion recalled in Section 3 and a nice combinatorial tool coming from supersaturated hypergraphs theory [18].

Theorem 4.2. *The property $\mathscr{H}$ exhibits a sharp threshold and the scale for the transition is of order n.*

As we have seen in Section 2 the q-Horn property can be seen as a CSP, $\mathscr{H} = \mathsf{UNSAT}(\{f_0, f_1, f_2, f_3\})$ in the non-symmetric model defined in [13]. Thus, according to the previous section, the proof of Theorem 4.2 follows from the three following propositions.

Proposition 4.3.

1. *For every* $r > 2/(k(k-1))$, *if* $Np \geqslant rn$, *then* $\mu_p(\mathscr{H}) \longrightarrow 1$, *in particular* $p_c(n) = \mathrm{O}(n^{1-k})$.
2. *For every* $r < 1/(k(k-1))$, *if* $Np \leqslant rn$, *then* $\mu_p(\mathscr{H}) \longrightarrow 0$, *in particular* $p_c(n) = \Omega(n^{1-k})$.

Proof. In a slightly different probabilistic model, Franco and Gelder [19] obtained upper and lower bounds for the scale at which the transition occurs for $\mathscr{H}$. These bounds correspond to those given here modulo a change of probability model analogous to the one from $G(n, M)$ to $G_n(p)$ in random graph theory. □

Observe that the first assertion shows that (D0) holds, and that the two bounds together make precise the scale of the transition, which occurs when $Np_c(n) = \Theta(n)$.

Proposition 4.4. *For every k-CNF formula m minimal for* $\mathscr{H}$ *we have*

$$\#Var(m) \leqslant (k-1)w(m) - 1.$$

Proof. The proof is similar to the one used in [13, Proposition 3.6] for minimal unsatisfiable formulas. Observe that a minimal non-q-Horn formula *m* cannot have any *free* clause, that is a clause with $(k-1)$ variables occurring only once. Indeed, by contradiction suppose that *m* contains a free clause *C*. Then, let us consider the formula m' obtained from *m* by removing *C*. By minimality of *m*, m' is q-Horn. Let Φ be a q-Horn certificate for m'. One can extend Φ in assigning the literals from *C* occurring only once to W^-, thus, obtaining a q-Horn certificate for *m*, a contradiction. This is sufficient to prove that $\#Var(m) \leqslant (k-1)w(m)$. Now, let us consider formulas *m* verifying $\#Var(m) = (k-1)w(m)$. Then, either $w(m) = 2$, or $w(m) > 2$ and *m* can be described as a cycle of the form $(l_1 \vee \cdots \vee l_2) \wedge (l_2 \vee \cdots \vee l_3) \wedge \cdots \wedge (l_{w(m)} \vee \cdots \vee l_1)$, where literals with distinct indices refer to distinct variables and the literals not specified correspond to variables occurring only once. But such formulas are always q-Horn since it suffices to assign the literals occurring twice to *E* and the ones occurring only once to W^-. This concludes the proof. □

Proposition 4.5. *For each* $c \in (0, 1)$, *for each t, for all* $(\delta_1, \ldots, \delta_t) \in \{W^+, W^-, E\}^t$, *and all* $\gamma > 0$

$$\mu_{p_c(n)}(s \notin Q_\delta, \#\mathscr{A}_\delta(s) \geqslant \gamma n^{k-1}) = \mathrm{o}(1),$$

Q_δ *denoting the property for a k-CNF formula s of having no q-Horn certificate with* $x_1 = \delta_1, \ldots, x_t = \delta_t$,

$\mathscr{A}_\delta(s)$ *denoting, for* $s \notin Q_\delta$, *the set of clauses C having at least one variable in* $\{x_1, \ldots, x_t\}$ *and such that* $s \wedge C \in Q_\delta$.

Proof. For more readability we will perform the proof in the special case $k = 3$, it will be clear that it is extendable to any $k \geqslant 3$.

For $s \notin Q_\delta$, $\#\mathscr{A}_\delta(s)$ is the number of ways one can reach the property Q_δ from *s* by adding a clause having at least one variable in $\{x_1, \ldots, x_t\}$. Observe that there are $\Theta(n^{k-1})$ such clauses. Therefore, the proposition says that for $s \notin Q_\delta$, $\mathscr{A}_\delta(s)$ is negligible.

The strategy will be as follows. First, for $s \notin Q_\delta$, let us consider the following set:

$$\mathscr{B}_\delta(s) = \{(l' \vee l'' \vee l''') \wedge (l^{iv} \vee l^{v} \vee l^{vi}) \text{ such that } s \wedge (l' \vee l'' \vee l''') \wedge (l^{iv} \vee l^{v} \vee l^{vi}) \in Q_\delta\}.$$

We know that the probability that $\mathscr{B}_\delta(s)$ is dense in the set of conjunctions of two clauses is negligible (see [14, Lemma 5.2]):

$$\text{For all } v > 0, \quad \mu_{p_c(n)}(s \notin Q_\delta, \#\mathscr{B}_\delta(s) \geqslant vn^6) = o(1).$$

Therefore, in order to prove our proposition we will prove that there exists some $v > 0$ such that for all γ:

$$\mu_{p_c(n)}(s \notin Q_\delta, \#\mathscr{A}_\delta(s) \geqslant \gamma n^2) \leqslant \mu_{p_c(n)}(s \notin Q_\delta, \#\mathscr{B}_\delta(s) \geqslant vn^6), \tag{1}$$

thus proving the proposition.

Hence, the trick is to provide a relationship between the cardinality of $\mathscr{A}_\delta(s)$ and the one of $\mathscr{B}_\delta(s)$. So, for $s \notin Q_\delta$, suppose that there exists $\gamma > 0$ such that $\#\mathscr{A}_\delta(s) \geqslant \gamma n^2$

$$\mathscr{A}_\delta(s) = \{C = (l_{i_1} \vee l_{i_2} \vee l_{i_3}) / l_{i_1} \in \{x_1, \ldots, x_t\} \cup \{\bar{x}_1, \ldots, \bar{x}_t\} \text{ such that } s \wedge C \text{ has no q-Horn-certificate with } x_1 = \delta_1, \ldots, x_t = \delta_t\}$$

with $\mathscr{A}_\delta(s)$ we associate a graph $G_\delta(s)$: the set of vertices is $\{x_1, \ldots, x_n\} \cup \{\bar{x}_1, \ldots, \bar{x}_n\}$, and for each clause $(l_{i_1} \vee l_{i_2} \vee l_{i_3}) \in \mathscr{A}_\delta(s)$ we create the edge $\{l_{i_2}, l_{i_3}\}$. By assumption $G_\delta(s)$ is dense, i.e. its number of edges is greater than or equal to $\gamma' n^2$ for some $\gamma' > 0$. Following the result from Erdös and Simonovits on supersaturated graphs [18, Corollary 2, p. 184], there exists $v > 0$ such that $G_\delta(s)$ contains at least vn^6 copies of the complete bipartite graph $K_{3,3}$. Consider such a copy whose bipartition is $\{l', l'', l'''\} \cup \{l^{iv}, l^{v}, l^{vi}\}$. Then, we claim that $(l' \vee l'' \vee l''') \wedge (l^{iv} \vee l^{v} \vee l^{vi}) \in \mathscr{B}_\delta(s)$.Indeed, in order to get a contradiction suppose that $s' = s \wedge (l' \vee l'' \vee l''') \wedge (l^{iv} \vee l^{v} \vee l^{vi}) \notin Q_\delta$. Then, s' has a q-Horn-certificate D with $x_1 = \delta_1, \ldots, x_t = \delta_t$. By definition of a q-Horn-certificate, D assigns at least one of the literals from the clause $(l' \vee l'' \vee l''')$ to W^-, w.l.o.g let us suppose that $D(l') = W^-$. In the same way we can suppose that $D(l^{iv}) = W^-$. Thus, for every literal l, the decomposition D is also a q-Horn-certificate for the formula $s \wedge (l \vee l' \vee l^{iv})$. But by assumption $\{l', l^{iv}\}$ is an edge of $G_\delta(s)$, which means that there exists some literal $l \in \{x_1, \ldots, x_t\} \cup \{\bar{x}_1, \ldots, \bar{x}_t\}$ such that $s \wedge (l \vee l' \vee l^{iv})$ has no q-Horn-certificate with $x_1 = \delta_1, \ldots, x_t = \delta_t$, a contradiction.

The one-to-one correspondence we have established between the copies of $K_{3,3}$ in $G_\delta(s)$ and $\mathscr{B}_\delta(s)$ proves that if $\#\mathscr{A}_\delta(s) \geqslant \gamma n^2$, then $\#\mathscr{B}_\delta(s) \geqslant vn^6$. Therefore, we have proved (1), the desired inequality.

The proof can be extended to any $k \geqslant 3$. In the general case, $\mathscr{B}_\delta(s)$ is formed with conjunctions of $(k-1)$ k-clauses. The graph $G_\delta(s)$ will be a $(k-1)$-uniform hypergraph which contains at least γn^{k-1} hyperedges. The central result from Erdös and Simonovits actually holds for such supersaturated $(k-1)$-uniform hypergraphs and says that there exists $v > 0$ such that $G_\delta(s)$ contains at least $vn^{(k-1)k}$ copies of the generalization of the complete $(k-1)$-partite graph $K_{k-1}^{(k-1)}(k, \ldots, k)$ (see [18, p. 184]), thus concluding the proof. □

5. Conclusion

We have found a sharp threshold for the q-Horn property which defines, in some probabilistic sense, the boundary of the q-Horn class on the m/n scale. The sharp threshold is due to minimality and monotonic properties which characterize quite a few polynomial-time solvable classes of CNF satisfiability problems and significantly limit their extent.

The sharpness of the well-studied satisfiability threshold is thought to be related to problem hardness. The sharpness of the q-Horn threshold similarly seems to keep many formulas out of the q-Horn class and, therefore, perhaps harder than otherwise.

The results here emphasize the need to look for polynomial-time solvable classes that do not have the minimality and monotonic properties discussed here. We expect such classes are good candidates for delineating easy and hard satisfiability problems.

References

[1] D. Achlioptas, C. Moore, The asymptotic order of the random k-SAT threshold, in: Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002, pp. 779–788.

[2] A. Biere, A. Cimatti, E.M. Clarke, Y. Zhu, Symbolic model checking without BDDs, in: Proceedings of Tools and Algorithms for the Construction and Analysis of Systems (TACAS'99), Part of the European Conferences on Theory and Practice of Software, ETAPS'99, Lecture Notes in Computer Science, vol. 1579, Springer, Amsterdam, 1999, pp. 193–207.

[5] E. Boros, Y. Crama, P.L. Hammer, M. Saks, A complexity index for satisfiability problems, SIAM J. Comput. 23 (1994) 45–49.

[6] E. Boros, P.L. Hammer, X. Sun, Recognition of q-Horn formulae in linear time, Discrete Appl. Math. 55 (1994) 1–13.

[7] A.Z. Broder, A.M. Frieze, E. Upfal, On the satisfiability and maximum satisfiability of random 3-CNF formulas, in: Proceedings of the Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, 1993, pp. 322–330.

[8] V. Chandru, J.N. Hooker, Extended Horn sets in propositional logic, J. Assoc. Comput. Mach. 38 (1991) 205–221.

[9] P. Cheeseman, B. Kanefsky, W.M. Taylor, Where the really hard problems are, in: Proceedings of the 12th International Joint Conference on Artificial Intelligence, Sydney, Australia, 1991.

[10] V. Chvátal, B. Reed, Mick gets some (the odds are on his side), in: Proceedings of the 33rd IEEE Symposium on Foundations of Computer Science, 1992, pp. 620–627.

[11] E.M. Clarke, A. Biere, R. Raimi, Y. Zhu, Bounded model checking using satisfiability solving, Formal Methods System Design 19 (2001) 7–34.

[12] M. Conforti, G. Cornuéjols, A. Kapoor, K. Vušković, M.R. Rao, Balanced matrices, in: J.R. Birge, K.G. Murty (Eds.), Mathematical Programming: State of the Art, Braun-Brumfield, USA, produced in association with the 15th International Symposium on Mathematical Programming, University of Michigan, 1994.

[13] N. Creignou, H. Daudé, Generalized satisfiability problems: minimal elements and phase transitions, Theoret. Comput. Sci. 302 (1–3) (2003) 417–430.

[14] N. Creignou, H. Daudé, Combinatorial sharpness criterion and phase transition classification for random CSPs, Inform. and Comput. 190 (2) (2004) 220–238.

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

[16] O. Dubois, Y. Boufkhad, J. Mandler, Typical random 3-SAT formulae and the satisfiability threshold, in: Proceedings of the 11th ACM-SIAM Symposium on Discrete Algorithms, SODA'2000, 2000, pp. 124–126.

[17] O. Dubois, R. Monasson, B. Selman, R. Zecchina, Phase transition in combinatorial problems, Theoret. Comput. Sci. 265 (1–2) (2001).

[18] P. Erdös, M. Simonovits, Supersaturated graphs and hypergraphs, Combinatorica 3 (2) (1982) 181–192.

[19] J. Franco, A. Van Gelder, A perspective on certain polynomial time solvable classes of satisfiability, Discrete Appl. Math. 125 (2–3) (2003) 177–214.
[20] E. Friedgut, and an appendix by J. Bourgain, Sharp thresholds of graph properties, and the *k*-sat problem, J. Amer. Math. Soc. 12(4) (1999) 1017–1054.
[21] E. Friedgut, M. Krivelevich, Sharp thresholds for Ramsey properties of random graphs, Random Structures Algorithms 17 (1) (2000) 1–19.
[22] C. Giannella, An empirical study of non-threshold behavior in iterative 3-SAT algorithms, Master Thesis, University of Cincinnati, 1999.
[23] A. Goerdt, A threshold for unsatisfiability, J. Comput. System Sci. 53 (3) (1996) 469–486.
[24] J.-J. Hébrard, A linear algorithm for renaming a set of clauses as a Horn set, Theoret. Comput. Sci. 124 (1994) 343–350.
[25] A. Itai, J. Makowsky, On the complexity of Herbrand's theorem, Working paper 243, Department of Computer Science, Israel Institute of Technology, 1982.
[26] D.S. Johnson, Approximation algorithms for combinatorial problems, J. Comput. Systems Sci. 9 (1974) 256–278.
[27] A. Kaporis, L. Kirousis, E. Lalas, The probabilistic analysis of a greedy satisfiability algorithm, in: Proceedings of the Fifth International Symposium on Theory and Applications of Satisfiability Testing, SAT'2002, 2002, pp. 362–376.
[28] H.R. Lewis, Renaming a set of clauses as a Horn set, J. Assoc. Comput. Mach. 25 (1978) 134–135.
[30] M. Mitzenmacher, Tight thresholds for the pure literal rule, DEC/SRC Technical Note 1997-011, 1997.
[31] M. Molloy, Models for random constraint satisfaction problems, SIAM J. Comput. 32 (4) (2003) 935–949.
[32] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, L. Troyansky, Determining computational complexity from characteristic "phase transitions", Nature 400 (1999) 133–137.
[33] J.S. Schlipf, F. Annexstein, J. Franco, R. Swaminathan, On finding solutions for extended Horn formulas, Inform. Process. Lett. 54 (1995) 133–137.
[34] M.G. Scutella, A note on Dowling and Gallier's top-down algorithm for propositional Horn satisfiability, J. Logic Programming 8 (1990) 265–273.