

A new upper bound for 3-SAT*

J. Díaz¹, L. Kirousis^{2,3}, D. Mitsche¹, X. Pérez-Giménez¹

¹ Llenguatges i Sistemes Informàtics, UPC
E-08034 Barcelona
{diaz,dmitsche,xperez}@lsi.upc.edu

² RA Computer Technology Institute
GR-26504 Rio, Patras

³ Computer Engineering and Informatics, University of Patras
GR-26504 Rio, Patras
kirousis@ceid.upatras.gr

ABSTRACT. We show that a randomly chosen 3-CNF formula over n variables with clauses-to-variables ratio at least 4.4898 is asymptotically almost surely unsatisfiable. The previous best such bound, due to Dubois in 1999, was 4.506. The first such bound, independently discovered by many groups of researchers since 1983, was 5.19. Several decreasing values between 5.19 and 4.506 were published in the years between. The probabilistic techniques we use for the proof are, we believe, of independent interest.

1 Introduction

Satisfiability of Boolean formulas is a problem universally believed to be hard. Determining the source of this hardness will lead, as is often stressed, to applications in domains even outside the realm of mathematics or computer science; moreover, and perhaps more importantly, it will enhance our understanding of the foundations of computing.

In the beginning of the 90's several groups of experimentalists chose to examine the source of this hardness from the following viewpoint: consider a *random* 3-CNF formula with a given clauses-to-variables ratio, which is known as the *density* of the formula. What is the probability of it being satisfied and how does this probability depend on the density? Their simulation results led to the conclusion that if the density is fixed and below a number approximately equal to 4.27, then for large n , a randomly chosen formula is almost always satisfiable, whereas if the density is fixed and above 4.27, a randomly chosen formula is, for large n , almost always unsatisfiable. More importantly, around 4.27 the complexity of several well known complete algorithms for checking satisfiability reaches a steep peak (see e.g. [10, 15]). So, in a certain sense, 4.27 is the point where from an empirical, statistical viewpoint the "hard" instances of SAT are to be found. Similar results were obtained for other combinatorial problems, and also for k -SAT for values of $k > 3$.

*Partially supported by the and the Spanish CYCIT: TIN2007-66523 (FORMALISM). The first and second authors were also partially supported by *La distinció per a la promoció de la recerca* de la Generalitat de Catalunya, 2002, and by the EU within the 7th Framework Programme under contract 215270 (FRONTS).

These experimental results were followed by an intense activity to provide “rigorous results” (the expression often used in this context to refer to theorems). Perhaps the most important advance is due to Friedgut: in [7] he proved that there is a sequence of reals $(c_n)_n$ such that for any $\epsilon > 0$ the probability of a randomly chosen 3-CNF-formula with density $c_n - \epsilon$ being satisfiable approaches 1 (as $n \rightarrow \infty$), whereas for density $c_n + \epsilon$, it approaches 0. Intuitively, this means that the transition from satisfiability to unsatisfiability is sharp, however it is still not known if $(c_n)_n$ converges.

Despite the fact that the convergence of $(c_n)_n$ is still an open problem, increasingly improved upper and lower bounds on its terms have been computed in a rigorous way by many groups of researchers. The currently best lower bound is 3.52 [9, 2].

With respect to upper bounds, which is the subject of this work, the progress was slower but better, in the sense that the experimentally established threshold is more closely bounded from above, rather than from below. A naïve application of the first moment method yields an upper bound of 5.191 (see e.g., [6]). An important advance was made in [8], where the upper bound was improved to 4.76. In the sequel, the work of several groups of researchers, based on more refined variants of the first moment method, culminated in the value of 4.571 [4, 11] (see the nice surveys [12, 3] for a complete sequence of the events). The core idea in these works was to use the first moment method by computing the expected number of not all satisfying truth assignments, but only of those among them that are local maxima in the sense of a lexicographic ordering, within a degree of locality determined by the Hamming distance between truth assignments (considered as binary sequences). For degree of locality 1, this amounts to computing the expected number of satisfying assignments that become unsatisfying assignments by flipping any of their “false” values (value 0) to “true” (value 1). Such assignments are sometimes referred to as *single-flip* satisfying assignments.

The next big step was taken by Dubois et al. [5], who showed that 4.506 is an upper bound. Instead of considering further variations of satisfiability, they limited the domain of computations to formulas that have a typical *syntactic* characteristic. Namely, they considered formulas where the cardinality of variables with given numbers of occurrences as positive and negative literals, respectively, approaches a two dimensional Poisson distribution. Asymptotically almost all formulas have this typical property (we say that such formulas have a *Poisson 2D degree sequence*). It turns out that the expectation of the number of single-flip satisfying assignments is exponentially reduced when computed for such formulas. To get the afore mentioned upper bound, Dubois et al. further limited the domain of computations to formulas that are *positively unbalanced*, i.e. formulas where every variable has at least as many occurrences as a positive literal as it has as a negated one.

A completely different direction was recently taken in [13]. Their work was motivated by results on the geometry of satisfying assignments, and especially the way they form clusters (components where one can move from one satisfying assignment to another by hops of small Hamming distance). Most of these results were originally based on analytical, but non-rigorous, techniques of Statistical Physics; lately however important rigorous advances were made [1, 14]. The value of the upper bound obtained by Maneva and Sinclair (see [13]) was 4.453, far below any other upper bound presently known (including the one in this paper). However it was proved assuming a conjecture on the geometry of the satisfying truth

assignments which is presently proved only for k -SAT for $k \geq 8$ in [1].

In this paper, we show that 4.4898 is an upper bound. Our approach builds upon previous work. It makes use (i) of single-flip satisfying truth assignments, (ii) of formulas with a Poisson 2D degree sequence and (iii) of positively unbalanced formulas.

We add to these previously known techniques two novel elements that when combined further reduce the expectation computed. Our approach is rigorous: although we make use of computer programs, the outputs we use are formally justified. What is interesting is not that much the numerical value we get, although it constitutes a further improvement to a long series of results. The main interest lies, we believe, on one hand in the new techniques themselves and on the other in the fact that putting together so many disparate techniques necessitates a delicately balanced proof structure.

First, we start by recursively eliminating one-by-one the occurrences of pure literals from the random formula, until we get its *impure core*, i.e. the largest sub-formula with no pure literals (a pure literal is one that has at least one occurrence in the formula but whose negation has none). Obviously this elimination has no effect on the satisfiability of the formula. Since we consider random formulas with a given 2D degree sequence, we first have to determine what is the 2D degree sequence of the impure core. For this, we use the differential equation method. The setting of the differential equations is more conveniently carried out in the so called *configuration* model, where the random formula is constructed by starting with as many labelled copies of each literal as its occurrences and then by considering random 3D matchings of these copies. The matchings define the clauses. The change of models from the standard one to the configuration model with a Poisson 2D degree sequence is formalized in Lemma 2. We also take care of the fact that the configuration model allows formulas with (i) multiple clauses and (ii) multiple occurrences of the same variable in a clause, whereas we are interested in simple formulas, i.e. formulas where neither (i) nor (ii) holds. For our purposes, it is enough to bound from below the probability of getting a simple formula in the configuration model by $e^{-\Theta(n^{1/3} \log n)}$, see Lemma 3. The differential equations are then analytically solved, and we thus obtain the 2D degree sequence of the core, see Proposition 4.

Second, we require that not only the 2D degree sequence is Poisson, but also that the numbers of clauses with none, one, two and three positive literals, respectively, are close to the expected numbers. Notice that these expected numbers have to reflect the fact that we consider positively unbalanced formulas. This is formalized in Lemma 5.

The expectation of the number of satisfying assignments, in the framework determined by all the restrictions above, is computed in Lemma 6. This expectation turns out to be given by a sum of polynomially many terms of functions that are exponential in n . We estimate this sum by its maximum term, using a standard technique. However in this case, finding the maximum term entails maximizing a function of many variables whose number depends on n . To avoid a maximization that depends on n we prove a truncation result which allows us to consider formulas that have a Poisson 2D degree sequence only for *light* variables, i.e. variables whose number of occurrences, either as positive or negated literals, is at most a constant independent of n .

Then we carry out the maximization. The technique we use is the standard one by Lagrange multipliers. We get a complex 3×3 system which can be solved numerically. We

formally prove that the system does not maximize on the boundary of the system and we make a sweep over the domain which confirms the results of the numerical solution.

Due to lack of space, all proofs are omitted or just sketched in this extended abstract. As usual, *asymptotically almost surely* (a.a.s.) will mean with probability tending to 1 as $n \rightarrow \infty$. All asymptotic expressions as $1 - o(1)$ are always with respect to n . Our main result in the paper is the following:

THEOREM 1. *Let $\gamma = 4.4898$ and $m = \lfloor \gamma n \rfloor$. A random 3-CNF formula in $\mathcal{F}_{n,m}$ (i.e. with n variables and m clauses, no repetition of clauses and no repetition of variables in a clause) is not satisfiable a.a.s.*

2 Background and Technical highlights.

Consider a given set of n Boolean variables, and let $m = \lfloor \gamma n \rfloor$. Let $\mathcal{F}_{n,m}$ be the set of 3-CNF formulas with n variables and m clauses, where repetition of clauses or repetition of variables in a clause is not allowed. We also denote by $\mathcal{F}_{n,m}$ the probability space of formulas in $\mathcal{F}_{n,m}$ drawn with uniform probability. Throughout the paper, we fix the value $\gamma = 4.4898$ and prove that for that value a random 3-CNF formula is not satisfiable with high probability.

Throughout the paper, *scaled* will always mean divided by n , and a *scaled natural* will be a member of $\frac{1}{n}\mathbb{N}$. Given a formula $\phi \in \mathcal{F}_{n,m}$, we define the following parameters which depend on ϕ : For any $i, j \in \mathbb{N}$, let $d_{i,j}$ be the scaled number of variables with i positive occurrences and j negative occurrences in ϕ . Then,

$$\sum_{i,j \in \mathbb{N}} d_{i,j} = 1. \quad (1)$$

The sequence $\mathbf{d} = (d_{i,j})_{i,j \in \mathbb{N}}$ is called the *degree sequence* of ϕ . The scaled number of clauses of ϕ is denoted by c , and can be expressed by

$$c(\mathbf{d}) = \frac{1}{3} \sum_{i,j \in \mathbb{N}} (i+j)d_{i,j}. \quad (2)$$

Note that if $\phi \in \mathcal{F}_{n,m}$, then c must additionally satisfy $c = \lfloor \gamma n \rfloor / n$.

Given $\epsilon_1 > 0$ and any sequence $\boldsymbol{\xi} = (\xi_{i,j})_{i,j \in \mathbb{N}}$ of nonnegative reals with $\sum_{i,j \in \mathbb{N}} \xi_{i,j} = 1$, define

$$\mathcal{N}(n, \boldsymbol{\xi}, \epsilon_1) = \left\{ \mathbf{d} = (d_{i,j})_{i,j \in \mathbb{N}} : \sum_{i,j \in \mathbb{N}} d_{i,j} = 1, \frac{n}{3} \sum_{i,j \in \mathbb{N}} (i+j)d_{i,j} \in \mathbb{N}, \forall i, j \in \mathbb{N} \quad d_{i,j}n \in \mathbb{N}, \right. \\ \left. \text{and } |d_{i,j} - \xi_{i,j}| \leq \epsilon_1, \text{ and if } i > n^{1/6} \text{ or } j > n^{1/6} \text{ then } d_{i,j} = 0 \right\}.$$

Intuitively $\mathcal{N}(n, \boldsymbol{\xi}, \epsilon_1)$ can be interpreted as the set of degree sequences \mathbf{d} which are close to the ideal sequence $\boldsymbol{\xi}$, which in general is not a degree sequence since its entries $\xi_{i,j}$ need not be scaled naturals. However, if n is large enough, then $\mathcal{N}(n, \boldsymbol{\xi}, \epsilon_1) \neq \emptyset$. Now we consider the 2D Poisson ideal sequence $\boldsymbol{\delta}$ defined by $\delta_{i,j} = e^{-3\gamma} (3\gamma/2)^{i+j} / (i!j!)$. The following lemma reflects the fact that almost all $\phi \in \mathcal{F}_{n,m}$ have a degree sequence \mathbf{d} which is close to $\boldsymbol{\delta}$. A proof of an analogous result can be found in [5].

LEMMA 2. *Let \mathbf{d} be the degree sequence of a random $\phi \in \mathcal{F}_{n,m}$. For any $\epsilon_1 > 0$, we have that $\Pr_{\mathcal{F}_{n,m}}(\mathbf{d} \in \mathcal{N}(n, \delta, \epsilon_1)) = 1 - o(1)$.*

Given a fixed degree sequence $\mathbf{d} = (d_{i,j})_{i,j \in \mathbb{N}}$ satisfying (1) and such that $c = c(\mathbf{d})$ defined by (2) is also a scaled natural, we wish to generate 3-CNF formulas with that particular degree sequence \mathbf{d} . A natural approach to this is to use the *configuration model*. A *configuration* φ with degree sequence $\mathbf{d} = (d_{i,j})_{i,j \in \mathbb{N}}$ is constructed as follows: consider n variables and the corresponding $2n$ literals $x_1, \bar{x}_1, \dots, x_n, \bar{x}_n$; each literal has a certain number of distinct labelled *copies* in a way that the scaled number of variables with i positive copies and j negative copies is $d_{i,j}$; then partition the set of copies into sets of size 3, which we call clauses of φ . Let $\mathcal{C}_{n,\mathbf{d}}$ be the set of all configurations with degree sequence \mathbf{d} , and we also denote by $\mathcal{C}_{n,\mathbf{d}}$ the probability space on the set $\mathcal{C}_{n,\mathbf{d}}$ with the uniform distribution.

A 3-CNF *multi-formula* is a formula with possible repetition of variables in one clause and/or possible repetition of clauses. A *simple formula* is a formula in $\mathcal{F}_{n,m}$. Let π be the projection from $\mathcal{C}_{n,\mathbf{d}}$ to 3-CNF multi-formulas obtained by unlabelling the copies of each literal. A configuration $\varphi \in \mathcal{C}_{n,\mathbf{d}}$ is satisfiable if $\phi = \pi(\varphi)$ is satisfiable. A configuration $\varphi \in \mathcal{C}_{n,\mathbf{d}}$ is *simple* iff $\phi = \pi(\varphi)$ is a simple formula, i.e. does not have repetition of variables or clauses. Notice that the number of anti-images of a simple formula ϕ with degree sequence \mathbf{d} under π does not depend on the particular choice of ϕ . Hence,

$$\Pr_{\mathcal{F}_{n,m}}(\phi \text{ is SAT} \mid \mathbf{d}) = \Pr_{\mathcal{C}_{n,\mathbf{d}}}(\varphi \text{ is SAT} \mid \text{SIMPLE}). \quad (3)$$

We need a lower bound on the probability that a configuration is simple. The following result gives a weak bound which is enough for our purposes.

LEMMA 3.

Let $\epsilon_1 > 0$ and $\mathbf{d} \in \mathcal{N}(n, \delta, \epsilon_1)$. Then

$$\Pr_{\mathcal{C}_{n,\mathbf{d}}}(\text{SIMPLE}) \geq e^{-\Theta(n^{1/3} \log n)},$$

where the $e^{-\Theta(n^{1/3} \log n)}$ bound is uniform for all $\mathbf{d} \in \mathcal{N}(n, \delta, \epsilon_1)$.

Given $\varphi \in \mathcal{C}_{n,\mathbf{d}}$, a *pure* variable of φ is a variable which has a non-zero number of occurrences which are either all syntactically positive or all syntactically negative. The only literal occurring in φ and all its copies are also called *pure*. If φ is satisfiable and x is a pure variable of φ , then there exists some satisfying truth assignment of φ which satisfies all copies of x in φ . Hence, in order to study the satisfiability of a $\varphi \in \mathcal{C}_{n,\mathbf{d}}$, we can satisfy each pure variable in φ and remove all clauses containing a copy of that variable. For each $\varphi \in \mathcal{C}_{n,\mathbf{d}}$, let $\tilde{\varphi}$ be the configuration obtained by greedily removing all pure variables and their corresponding clauses from φ . This $\tilde{\varphi}$ is independent of the particular elimination order of pure literals and is called the *impure core* of φ . In fact, in our analysis we will eliminate only one clause containing one copy of a pure literal at a time (the $\tilde{\varphi}$ obtained still remains the same). Note that φ is satisfiable iff $\tilde{\varphi}$ is satisfiable. Moreover, if φ is simple then $\tilde{\varphi}$ is also simple (but the converse is not necessarily true).

Furthermore, let $\hat{\varphi}$ be the configuration obtained from $\tilde{\varphi}$ by positively unbalancing all variables, i.e. switching the syntactic sign of those variables having initially more negative than positive occurrences in $\tilde{\varphi}$. Let $\hat{\mathcal{C}}_{n,\mathbf{d}}$ denote the probability space of configurations $\hat{\varphi}$,

where φ was chosen from $\mathcal{C}_{n,d}$ with uniform probability. Note that the probability distribution in $\widehat{\mathcal{C}}_{n,d}$ is not necessarily uniform. Since the simplicity and the satisfiability of a configuration are not affected by positively unbalancing the variables, we have

$$\Pr_{\mathcal{C}_{n,d}}(\varphi \text{ is SAT} \wedge \text{SIMPLE}) \leq \Pr_{\widehat{\mathcal{C}}_{n,d}}(\widehat{\varphi} \text{ is SAT} \wedge \text{SIMPLE}). \quad (4)$$

Let the random variable $\widehat{\mathbf{d}}$ be the degree sequence of a random configuration in $\widehat{\mathcal{C}}_{n,d}$. We prove in the following result that if the original \mathbf{d} is close to the ideal sequence δ , then with high probability $\widehat{\mathbf{d}}$ must be close to the ideal sequence $\widehat{\delta} = (\widehat{\delta}_{i,j})_{i,j \in \mathbb{N}}$ defined by

$$\widehat{\delta}_{i,j} = \begin{cases} 2e^{-3\gamma b} \frac{(3\gamma b/2)^{i+j}}{i!j!}, & \text{if } i > j, \\ e^{-3\gamma b} \frac{(3\gamma b/2)^{i+j}}{i!j!}, & \text{if } i = j, \\ 0, & \text{if } i < j, \end{cases}$$

where $b = (1 - t_{\mathcal{D}}/\gamma)^{2/3}$ and $t_{\mathcal{D}}$ is the scaled number of steps in the pure literal elimination algorithm.

PROPOSITION 4. *Given $\epsilon_2 > 0$, there exists $\epsilon_1 > 0$ and $0 < \beta < 1$ such that for any $\mathbf{d} \in \mathcal{N}(n, \delta, \epsilon_1)$*

$$\Pr_{\widehat{\mathcal{C}}_{n,d}}(\widehat{\mathbf{d}} \in \mathcal{N}(n, \widehat{\delta}, \epsilon_2)) = 1 - O(\beta^{n^{1/2}}).$$

Moreover, for each $\widehat{\mathbf{d}} \in \mathcal{N}(n, \widehat{\delta}, \epsilon_2)$, the probability space $\widehat{\mathcal{C}}_{n,d}$ conditional upon having degree sequence $\widehat{\mathbf{d}}$ has the uniform distribution (i.e. $\widehat{\mathcal{C}}_{n,d}$ conditional upon a fixed $\widehat{\mathbf{d}}$ behaves exactly as $\mathcal{C}_{n,\widehat{\mathbf{d}}}$).

Let $\widehat{\mathbf{d}} \in \mathcal{N}(n, \widehat{\delta}, \epsilon_2)$. Then, each $\varphi \in \mathcal{C}_{n,\widehat{\mathbf{d}}}$ has a scaled number of clauses of $\widehat{c} = c(\widehat{\mathbf{d}})$ (see (2)). Moreover, let ℓ_p and ℓ_n be the scaled number of copies in φ of positive and of negative literals respectively. Then

$$\ell_p(\widehat{\mathbf{d}}) = \sum_{i,j \in \mathbb{N}} i \widehat{d}_{i,j}, \quad \ell_n(\widehat{\mathbf{d}}) = \sum_{i,j \in \mathbb{N}} j \widehat{d}_{i,j}. \quad (5)$$

Given any fixed $\varphi \in \mathcal{C}_{n,\widehat{\mathbf{d}}}$ and for $k \in \{0, \dots, 3\}$, let \widehat{c}_k be the scaled number of clauses in φ containing exactly k positive copies (clauses of *syntactic type* k). We call $\widehat{c} = (\widehat{c}_0, \dots, \widehat{c}_3)$ the *clause-type sequence* of φ . By definition

$$\widehat{c}_1 + 2\widehat{c}_2 + 3\widehat{c}_3 = \ell_p, \quad 3\widehat{c}_0 + 2\widehat{c}_1 + \widehat{c}_2 = \ell_n, \quad (6)$$

and by adding the equations in (6), $\widehat{c}_0 + \dots + \widehat{c}_3 = \widehat{c}$. The $\widehat{c}_0, \dots, \widehat{c}_3$ are random variables in $\mathcal{C}_{n,\widehat{\mathbf{d}}}$, but the next result shows that if $\widehat{\mathbf{d}}$ is close enough to $\widehat{\delta}$, then $\widehat{c}_0, \dots, \widehat{c}_3$ as well as their sum $\widehat{c}_0 + \dots + \widehat{c}_3 = \widehat{c}$ are concentrated with high probability. In order to see this, we need to define $\widehat{\gamma} = c(\widehat{\delta})$, $\lambda_p = \ell_p(\widehat{\delta})$ and $\lambda_n = \ell_n(\widehat{\delta})$ (see (2) and (5)), which can be interpreted as the limit of \widehat{c} , ℓ_p and ℓ_n respectively when $\widehat{\mathbf{d}}$ approaches $\widehat{\delta}$. In terms of these numbers, we thus define for all $k \in \{0, \dots, 3\}$

$$\widehat{\gamma}_k = \binom{3}{k} \frac{\lambda_p^k \lambda_n^{3-k}}{(\lambda_p + \lambda_n)^3} \widehat{\gamma} \quad (7)$$

and also $\widehat{\gamma} = (\widehat{\gamma}_0, \dots, \widehat{\gamma}_3)$. Then we have $\widehat{\gamma}_1 + 2\widehat{\gamma}_2 + 3\widehat{\gamma}_3 = \lambda_p$, $3\widehat{\gamma}_0 + 2\widehat{\gamma}_1 + \widehat{\gamma}_2 = \lambda_n$ and $\widehat{\gamma}_0 + \widehat{\gamma}_1 + \widehat{\gamma}_2 + \widehat{\gamma}_3 = \widehat{\gamma}$.

The next result shows that when $\widehat{\mathbf{d}}$ is close enough to $\widehat{\delta}$, then each \widehat{c}_k is close to the corresponding $\widehat{\gamma}_k$. Indeed, given $\epsilon > 0$ and for any $\widehat{\mathbf{d}} \in \mathcal{N}(n, \widehat{\delta}, \epsilon_2)$, let $\mathcal{C}_{n, \widehat{\mathbf{d}}}^\epsilon$ be the set of all $\varphi \in \mathcal{C}_{n, \widehat{\mathbf{d}}}$ such that for $k \in \{0, \dots, 3\}$, $|\widehat{c}_k - \widehat{\gamma}_k| \leq \epsilon$. We also denote by $\mathcal{C}_{n, \widehat{\mathbf{d}}}^\epsilon$ the corresponding uniform probability space.

LEMMA 5. *Given $\epsilon > 0$, there is $\epsilon_2 > 0$ and $0 < \beta < 1$ such that for any $\widehat{\mathbf{d}} \in \mathcal{N}(n, \widehat{\delta}, \epsilon_2)$,*

$$\Pr_{\mathcal{C}_{n, \widehat{\mathbf{d}}}^\epsilon}(\mathcal{C}_{n, \widehat{\mathbf{d}}}^\epsilon) = 1 - O(\beta^n).$$

All the previous lemmata establish a connection between the uniform probability spaces $\mathcal{F}_{n, m}$ and $\mathcal{C}_{n, \widehat{\mathbf{d}}}^\epsilon$. In order to prove Theorem 1, it remains to bound the probability that a configuration $\varphi \in \mathcal{C}_{n, \widehat{\mathbf{d}}}^\epsilon$ is simple and satisfiable, as it is done in the following result.

LEMMA 6. *There exists $\epsilon > 0$ and $0 < \beta < 1$ such that for any $\widehat{\mathbf{d}} \in \mathcal{N}(n, \widehat{\delta}, \epsilon)$,*

$$\Pr_{\mathcal{C}_{n, \widehat{\mathbf{d}}}^\epsilon}(\text{SAT} \wedge \text{SIMPLE}) = O(\beta^n).$$

The proof of Lemma 6 is sketched in Section 3 below. The proof of Theorem 1 then follows from all the previous lemmata (see the full version for the proof).

3 Proof of Lemma 6

Let $\mathcal{N}(n, \widehat{\delta}, \widehat{\gamma}, \epsilon)$ be the set of tuples $(\widehat{\mathbf{d}}, \widehat{\mathbf{c}})$ such that $\widehat{\mathbf{d}} \in \mathcal{N}(n, \widehat{\delta}, \epsilon)$ and $\widehat{\mathbf{c}} = (\widehat{c}_k)_{0 \leq k \leq 3}$ is a tuple of scaled naturals satisfying (6) (recall also from (5) the definition of ℓ_p and ℓ_n), and moreover $|\widehat{c}_k - \widehat{\gamma}_k| \leq \epsilon$. For each $(\widehat{\mathbf{d}}, \widehat{\mathbf{c}}) \in \mathcal{N}(n, \widehat{\delta}, \widehat{\gamma}, \epsilon)$, we define $\mathcal{C}_{n, \widehat{\mathbf{d}}, \widehat{\mathbf{c}}}$ to be the uniform probability space of all configurations with degree sequence $\widehat{\mathbf{d}}$ and clause-type sequence $\widehat{\mathbf{c}}$. In order to prove the lemma, it suffices to show that for any $(\widehat{\mathbf{d}}, \widehat{\mathbf{c}}) \in \mathcal{N}(n, \widehat{\delta}, \widehat{\gamma}, \epsilon)$ we have $\Pr_{\mathcal{C}_{n, \widehat{\mathbf{d}}, \widehat{\mathbf{c}}}}(\text{SAT} \wedge \text{SIMPLE}) = O(\beta^n)$. Hence, we consider $\widehat{\mathbf{d}}, \widehat{\mathbf{c}}$ and the probability space $\mathcal{C}_{n, \widehat{\mathbf{d}}, \widehat{\mathbf{c}}}$ to be fixed throughout this section, and we try to find a suitable bound for $\Pr(\text{SAT} \wedge \text{SIMPLE})$.

We need some definitions. Let us fix any given configuration $\varphi \in \mathcal{C}_{n, \widehat{\mathbf{d}}, \widehat{\mathbf{c}}}$. A *light* variable of φ is a variable with $i \leq M$ positive occurrences and $j \leq M$ negative occurrences in φ (we use in the numerical calculations the value $M = 23$). The other variables are called *heavy*. We consider a weaker notion of satisfiability in which heavy variables are treated as jokers and are always satisfied regardless of their sign in the formula and their assigned value. Given a configuration $\varphi \in \mathcal{C}_{n, \widehat{\mathbf{d}}, \widehat{\mathbf{c}}}$ and a truth assignment A , we say that $A \models^b \varphi$ iff each clause of φ contains at least one heavy variable or at least one satisfied occurrence of a light variable. Let SAT^b be the set of configurations $\varphi \in \mathcal{C}_{n, \widehat{\mathbf{d}}, \widehat{\mathbf{c}}}$ for which there exists at least one truth assignment A such that $A \models^b \varphi$. Clearly, if $A \models \varphi$, then also $A \models^b \varphi$, and hence $\text{SAT} \subset \text{SAT}^b$. We still introduce a further restriction to satisfiability in a way similar to [11] and [4], in order to decrease the number of satisfying truth assignments of each configuration without altering the set of satisfiable configurations (at least without altering this set for simple configurations). Given a configuration $\varphi \in \mathcal{C}_{n, \widehat{\mathbf{d}}, \widehat{\mathbf{c}}}$ and a truth assignment A , we say that

$A \models^{b'} \varphi$ iff $A \models^b \varphi$ and moreover each light variable which is assigned the value zero by A appears at least once as the only satisfied literal of a *blocking clause* (i.e. a clause with one satisfied negative literal and two unsatisfied ones). Let $\text{SAT}^{b'}$ be the set of configurations which are satisfiable according to this latter notion. Notice that if $\varphi \in \text{SIMPLE}$, then $\varphi \in \text{SAT}^{b'}$ iff $\varphi \in \text{SAT}^b$ (by an argument analogous to the one in [11] and [4]). Therefore, we have $\Pr(\text{SAT} \wedge \text{SIMPLE}) \leq \Pr(\text{SAT}^b \wedge \text{SIMPLE}) = \Pr(\text{SAT}^{b'} \wedge \text{SIMPLE}) \leq \Pr(\text{SAT}^{b'})$. Let X be the random variable counting the number of satisfying truth assignments of a randomly chosen configuration $\varphi \in \mathcal{C}_{n, \hat{a}, \hat{c}}$ in the $\text{SAT}^{b'}$ sense. We need to bound

$$\Pr(\text{SAT}^{b'}) = \Pr(X > 0) \leq \mathbf{EX} = \frac{|\{(\varphi, A) : \varphi \in \mathcal{C}_{n, \hat{a}, \hat{c}}, A \models^{b'} \varphi\}|}{|\mathcal{C}_{n, \hat{a}, \hat{c}}|}. \quad (8)$$

In the following subsection, we obtain an exact but complicated expression for \mathbf{EX} by a counting argument, and then we give a simple asymptotic bound which depends on the maximization of a particular continuous function over a bounded polytope. The next subsection contains the maximization of that function.

3.1 Asymptotic bound on \mathbf{EX}

First, we compute the denominator of the rightmost member in (8).

$$\begin{aligned} |\mathcal{C}_{n, \hat{a}, \hat{c}}| &= \binom{n}{(\hat{d}_{i,j}n)_{i,j}} \binom{\ell_p n}{\hat{c}_1 n, 2\hat{c}_2 n, 3\hat{c}_3 n} \binom{\ell_n n}{3\hat{c}_0 n, 2\hat{c}_1 n, \hat{c}_2 n} \frac{(3\hat{c}_0 n)!}{(\hat{c}_0 n)! 6^{\hat{c}_0 n}} \frac{(2\hat{c}_1 n)!}{2^{\hat{c}_1 n}} \frac{(2\hat{c}_2 n)!}{2^{\hat{c}_2 n}} \frac{(3\hat{c}_3 n)!}{(\hat{c}_3 n)! 6^{\hat{c}_3 n}} \\ &= \frac{n!}{\prod_{i,j} (\hat{d}_{i,j} n)!} \frac{(\ell_p n)! (\ell_n n)!}{2^{\hat{c}_2 n} 3^{(\hat{c}_0 + \hat{c}_3)n} (\hat{c}_0 n)! (\hat{c}_1 n)! (\hat{c}_2 n)! (\hat{c}_3 n)!} \end{aligned}$$

In order to deal with the numerator in (8), we need some definitions. Let us consider any fixed $\varphi \in \mathcal{C}_{n, \hat{a}, \hat{c}}$ and any assignment A such that $A \models^{b'} \varphi$. We will classify the variables, the clauses and the copies of literals in φ into several types, and define parameters counting the scaled number of items of each type. Variables are classified according to their degree. A variable is said to have degree (i, j) if it appears i times positively and j times negatively in φ . Let \mathcal{L} and \mathcal{H} , respectively, be the set of possible degrees for light and heavy variables, i.e. $\mathcal{L} = \{(i, j) \in \mathbb{N}^2 : 0 \leq i, j \leq M\}$, $\mathcal{H} = \{(i, j) \in \mathbb{N}^2 : i > M \text{ or } j > M\}$. We also consider an extended notion of degree for light variables which are assigned 0 by A . One of such variables has extended degree (i, j, k) if it has degree (i, j) and among its j negative occurrences k appear in a blocking clause (being the only satisfied literal of the clause). Let $\mathcal{L}' = \{(i, j, k) \in \mathbb{N}^3 : 0 \leq i \leq M, 1 \leq k \leq j \leq M\}$, be the set of possible extended degrees for these light 0-variables. For each $(i, j) \in \mathcal{L}$, let $t_{i,j}$ be the scaled number of light variables assigned 1 by A with degree (i, j) in φ . For each $(i, j, k) \in \mathcal{L}'$, let $f_{i,j,k}$ be the scaled number of light variables assigned 0 by A with extended degree (i, j, k) in φ . We must have

$$t_{i,j} + \sum_{k=1}^j f_{i,j,k} = \hat{d}_{i,j}, \quad \forall (i, j) \in \mathcal{L}. \quad (9)$$

On the other hand, we classify the copies of literals occurring in φ into five different types depending on their sign in φ , their assignment by A and whether they belong or not to a

blocking clause. Each copy receives a label from the set $\mathcal{S} = \{\text{ps}, \text{ns1}, \text{ns2}, \text{pu}, \text{nu}\}$, where the labels ps, pu, ns1, ns2 and nu denote positive-satisfied, positive-unsatisfied, negative-satisfied in a blocking clause, negative-satisfied in a non-blocking clause and negative-unsatisfied, respectively. It is useful to consider as well coarser classifications of the copies of literals in φ and thus we define the types p, n and ns which correspond to positive, negative and negative-satisfied copies, respectively. Also, let $\mathcal{S}' = \{\text{ps}, \text{ns}, \text{pu}, \text{nu}\}$ and $\mathcal{S}'' = \{\text{p}, \text{n}\}$. For each of the types $\sigma \in \mathcal{S} \cup \mathcal{S}' \cup \mathcal{S}''$ that we defined, let ℓ_σ be the scaled number of copies of type σ . Note that ℓ_p and ℓ_n were already defined (see (5) and (6)). Also, let h_σ be the scaled number of copies of type σ which come from heavy variables (recall that these copies are always satisfied by definition regardless of their sign). In view of $h_{\text{ps}} = \sum_{\mathcal{H}} i\widehat{d}_{i,j}$, $h_{\text{ns}} = \sum_{\mathcal{H}} j\widehat{d}_{i,j}$ and of (5) and (6), we observe that ℓ_p , ℓ_n , h_{ps} and h_{ns} are constants which do not depend on the particular choice of (φ, A) . The parameters h_{ns1} and h_{ns2} depend on the particular (φ, A) and satisfy

$$h_{\text{ns1}} + h_{\text{ns2}} = h_{\text{ns}}. \quad (10)$$

The parameters ℓ_{ps} , ℓ_{pu} , ℓ_{ns1} , ℓ_{ns2} and ℓ_{nu} also depend on (φ, A) and can be expressed as

$$\begin{aligned} \ell_{\text{ps}} &= \sum_{\mathcal{L}} it_{i,j} + h_{\text{ps}}, & \ell_{\text{pu}} &= \sum_{\mathcal{L}'} if_{i,j,k}, & \ell_{\text{ns1}} &= \sum_{\mathcal{L}'} kf_{i,j,k} + h_{\text{ns1}}, \\ \ell_{\text{ns2}} &= \sum_{\mathcal{L}'} (j-k)f_{i,j,k} + h_{\text{ns2}}, & \ell_{\text{nu}} &= \sum_{\mathcal{L}} jt_{i,j}. \end{aligned} \quad (11)$$

Finally, the clauses of φ are classified into 16 extended types (not to be mistaken with the four syntactic types defined immediately before (6)). Each type is represented by a 2×2 matrix from the set $\mathcal{A} = \left\{ \alpha = \begin{pmatrix} \text{ps}(\alpha) & \text{ns}(\alpha) \\ \text{pu}(\alpha) & \text{nu}(\alpha) \end{pmatrix} \in \mathbb{N}^4 : \sum_{\sigma \in \mathcal{S}'} \sigma(\alpha) = 3, \text{ps}(\alpha) + \text{ns}(\alpha) > 0 \right\}$.

A clause is said to be of extended type $\alpha = \begin{pmatrix} \text{ps}(\alpha) & \text{ns}(\alpha) \\ \text{pu}(\alpha) & \text{nu}(\alpha) \end{pmatrix}$ if for each $\sigma \in \mathcal{S}'$ the clause contains $\sigma(\alpha)$ copies of literals of type σ . Notice that all clauses of extended type α also contain the same number of copies of type σ for all other $\sigma \in \mathcal{S} \cup \mathcal{S}''$ and thus we can define $\sigma(\alpha)$ to be this number. For each $\alpha \in \mathcal{A}$, let c_α be the scaled number of clauses of extended type α (while \widehat{c}_k , $0 \leq k \leq 3$ is the number of clauses of syntactic type k , i.e. with k positive literals). We have

$$\sum_{\substack{\alpha \in \mathcal{A} \\ \text{p}(\alpha)=k}} c_\alpha = \widehat{c}_k. \quad (12)$$

The parameters ℓ_{ps} , ℓ_{pu} , ℓ_{ns1} , ℓ_{ns2} and ℓ_{nu} can also be expressed in terms of the c_α by

$$\ell_\sigma = \sum_{\alpha \in \mathcal{A}} \sigma(\alpha) c_\alpha, \quad \forall \sigma \in \mathcal{S}. \quad (13)$$

We now consider the following equations:

$$\ell_{\text{ps}} + \ell_{\text{pu}} = \ell_p \quad \ell_{\text{ns1}} + \ell_{\text{ns2}} + \ell_{\text{nu}} = \ell_n \quad (14)$$

$$\ell_{\text{ps}} = \sum_{\mathcal{L}} it_{i,j} + h_{\text{ps}} \quad \ell_{\text{ns1}} = \sum_{\mathcal{L}'} kf_{i,j,k} + h_{\text{ns1}} \quad \ell_{\text{ns2}} = \sum_{\mathcal{L}'} (j-k)f_{i,j,k} + h_{\text{ns2}} \quad (15)$$

$$\ell_{\text{ps}} = \sum_{\alpha \in \mathcal{A}} \text{ps}(\alpha) c_\alpha \quad \ell_{\text{ns1}} = \sum_{\alpha \in \mathcal{A}} \text{ns1}(\alpha) c_\alpha \quad \ell_{\text{ns2}} = \sum_{\alpha \in \mathcal{A}} \text{ns2}(\alpha) c_\alpha \quad (16)$$

In view of (5) and (6), the system of equations {(9), (10), (11), (12), (13)} is equivalent to {(9), (10), (12), (14), (15), (16)}.

So far we verified that the constraints {(9), (10), (12), (14), (15), (16)} express necessary conditions for the parameters of any particular (φ, A) , with $\varphi \in \mathcal{C}_{n, \hat{\mathbf{d}}, \hat{\mathbf{c}}}$ and $A \models^{b'} \varphi$. Now we will see that they are also sufficient, in the sense that for each tuple of parameters satisfying the above-mentioned constraints we will be able to construct pairs (φ, A) .

Let $\bar{t} = (t_{i,j})_{\mathcal{L}}$, $\bar{f} = (f_{i,j,k})_{\mathcal{L}'}$, $\bar{h} = (h_{\text{ns}1}, h_{\text{ns}2})$, $\bar{c} = (c_\alpha)_{\alpha \in \mathcal{A}}$, $\bar{\ell} = (\ell_\sigma)_{\sigma \in \mathcal{S}}$ and $K = |\mathcal{L}| + |\mathcal{L}'| + 2 + |\mathcal{A}| + |\mathcal{S}| = (M+1)^2(1+M/2) + 23$. We define the bounded polytope $\mathcal{P}(\hat{\mathbf{d}}, \hat{\mathbf{c}}) \subset \mathbb{R}^K$ as the set of tuples $\bar{x} = (\bar{t}, \bar{f}, \bar{h}, \bar{c}, \bar{\ell})$ of non-negative reals satisfying {(9), (10), (12), (14), (15), (16)}, and consider the following set of lattice points in $\mathcal{P}(\hat{\mathbf{d}}, \hat{\mathbf{c}})$: $\mathcal{I}(n, \hat{\mathbf{d}}, \hat{\mathbf{c}}) = \mathcal{P}(\hat{\mathbf{d}}, \hat{\mathbf{c}}) \cap (\frac{1}{n}\mathbb{N})^K$. For any tuple of parameters $\bar{x} \in \mathcal{I}(n, \hat{\mathbf{d}}, \hat{\mathbf{c}})$, we count the number of pairs (φ, A) , with $\varphi \in \mathcal{C}_{n, \hat{\mathbf{d}}, \hat{\mathbf{c}}}$ and $A \models^{b'} \varphi$, satisfying these parameters. We denote this number by $T(\bar{x}, n, \hat{\mathbf{d}}, \hat{\mathbf{c}})$. We obtain (see the full version for details)

$$T(\bar{x}, n, \hat{\mathbf{d}}, \hat{\mathbf{c}}) = 2^{\sum_{\mathcal{H}} \hat{d}_{i,j} n} \binom{n}{(t_{i,j} n)_{\mathcal{L}}, (f_{i,j,k} n)_{\mathcal{L}'}, (\hat{d}_{i,j} n)_{\mathcal{H}}} \\ \left(\prod_{\mathcal{L}'} \binom{j}{k}^{f_{i,j,k} n} \right) \binom{h_{\text{ns}} n}{h_{\text{ns}1} n, h_{\text{ns}2} n} \prod_{\sigma \in \mathcal{S}} \binom{\ell_\sigma n}{(\sigma(\alpha) c_\alpha n)_{\alpha \in \mathcal{A}}} \prod_{\alpha \in \mathcal{A}} W(\alpha),$$

where $W(\alpha) = \frac{(w(\alpha) c_\alpha n)! (c_\alpha n)!^{2-w(\alpha)}}{(w(\alpha)!)^{c_\alpha n}}$, and $w(\alpha)$ is the number of 0's in the matrix α . Hence $\text{EX} = \frac{1}{|\mathcal{C}_{n, \hat{\mathbf{d}}, \hat{\mathbf{c}}}|} \sum_{\bar{x} \in \mathcal{I}(n, \hat{\mathbf{d}}, \hat{\mathbf{c}})} T(\bar{x}, n, \hat{\mathbf{d}}, \hat{\mathbf{c}})$.

To characterize the asymptotic behaviour of $T(\bar{x}, n, \hat{\mathbf{d}}, \hat{\mathbf{c}}) / |\mathcal{C}_{n, \hat{\mathbf{d}}, \hat{\mathbf{c}}}|$ with respect to n , we define

$$F(\bar{x}) = \frac{\prod_{\sigma \in \mathcal{S}} \ell_\sigma^{\ell_\sigma}}{\prod_{\mathcal{L}} t_{i,j}^{t_{i,j}} \prod_{\mathcal{L}'} \left(f_{i,j,k} / \binom{j}{k} \right)^{f_{i,j,k}} h_{\text{ns}1}^{h_{\text{ns}1}} h_{\text{ns}2}^{h_{\text{ns}2}} \prod_{\alpha \in \mathcal{A}} ((w(\alpha)!/2) c_\alpha)^{c_\alpha}}$$

and

$$B(\hat{\mathbf{d}}, \hat{\mathbf{c}}) = 2^{\sum_{\mathcal{H}} \hat{d}_{i,j}} h_{\text{ns}}^{h_{\text{ns}}} \prod_{\mathcal{L}} \hat{d}_{i,j} \frac{3^{c_0+c_3} c_0^{c_0} c_1^{c_1} c_2^{c_2} c_3^{c_3}}{\ell_p^{\ell_p} \ell_n^{\ell_n}}.$$

By Stirling's inequality we obtain $\frac{T(\bar{x}, n, \hat{\mathbf{d}}, \hat{\mathbf{c}})}{|\mathcal{C}_{n, \hat{\mathbf{d}}, \hat{\mathbf{c}}}|} \leq \text{poly}_1(n) (B(\hat{\mathbf{d}}, \hat{\mathbf{c}}) F(\bar{x}))^n$, where $\text{poly}_1(n)$ is some fixed polynomial in n which can be chosen to be independent of \bar{x} , $\hat{\mathbf{d}}$ and $\hat{\mathbf{c}}$ (as long as $\bar{x} \in \mathcal{I}(n, \hat{\mathbf{d}}, \hat{\mathbf{c}})$ and $(\hat{\mathbf{d}}, \hat{\mathbf{c}}) \in \mathcal{N}(n, \hat{\delta}, \hat{\gamma}, \epsilon)$). Moreover, since the size of $\mathcal{I}(n, \hat{\mathbf{d}}, \hat{\mathbf{c}})$ is also polynomial in n , we can write

$$\text{EX} \leq \text{poly}_2(n) \left(B(\hat{\mathbf{d}}, \hat{\mathbf{c}}) \max_{\bar{x} \in \mathcal{I}(n, \hat{\mathbf{d}}, \hat{\mathbf{c}})} F(\bar{x}) \right)^n \leq \text{poly}_2(n) \left(B(\hat{\mathbf{d}}, \hat{\mathbf{c}}) \max_{\bar{x} \in \mathcal{P}(n, \hat{\mathbf{d}}, \hat{\mathbf{c}})} F(\bar{x}) \right)^n,$$

for some other fixed polynomial $\text{poly}_2(n)$. By continuity, if we choose ϵ to be small enough, we can guarantee that

$$\text{EX} \leq \left((1 + 10^{-7}) B \max_{\bar{x} \in \mathcal{P}(n, \hat{\delta}, \hat{\gamma})} F(\bar{x}) \right)^n, \quad (17)$$

where (recall the definition in (7))

$$\begin{aligned}
B = B(\widehat{\delta}, \widehat{\gamma}) &= 2^{\sum_{\mathcal{H}} \widehat{\delta}_{i,j}} \left(\sum_{\mathcal{H}} j \widehat{\delta}_{i,j} \right)^{\sum_{\mathcal{H}} j \widehat{\delta}_{i,j}} \prod_{\mathcal{L}} \widehat{\delta}_{i,j}^{\widehat{\delta}_{i,j}} \frac{3^{\widehat{\gamma}_0 + \widehat{\gamma}_3} \widehat{\gamma}_0^{\widehat{\gamma}_0} \widehat{\gamma}_1^{\widehat{\gamma}_1} \widehat{\gamma}_2^{\widehat{\gamma}_2} \widehat{\gamma}_3^{\widehat{\gamma}_3}}{\lambda_p^{\lambda_p} \lambda_n^{\lambda_n}} \\
&= 2^{\sum_{\mathcal{H}} \widehat{\delta}_{i,j}} \left(\sum_{\mathcal{H}} j \widehat{\delta}_{i,j} \right)^{\sum_{\mathcal{H}} j \widehat{\delta}_{i,j}} \frac{\prod_{\mathcal{L}} \widehat{\delta}_{i,j}^{\widehat{\delta}_{i,j}}}{(3\widehat{\gamma})^{2\widehat{\gamma}}}.
\end{aligned} \tag{18}$$

3.2 Maximization of $F(\bar{x})$

We wish to maximize F or equivalently $\log F$ over the domain $\mathcal{P}(n, \widehat{\delta}, \widehat{\gamma})$. We need the following lemma:

LEMMA 7. $F(\bar{x})$ does not maximize on the boundary of $\mathcal{P}(n, \widehat{\delta}, \widehat{\gamma})$.

Since $\log F$ does not maximize on the boundary of its domain, the maximum must be attained at a critical point of $\log F$ in the interior of $\mathcal{P}(n, \widehat{\delta}, \widehat{\gamma})$. We use the Lagrange multipliers technique and characterize each critical point of $\log F$ in terms of the solution of a 3×3 system. The system is numerically solved with the help of Maple, which finds just one solution. We express the maximum of F over $\mathcal{P}(n, \widehat{\delta}, \widehat{\gamma})$ in terms of this solution, and multiply it by B given in (18), and from (17) we obtain the bound

$$EX \leq ((1 + 10^{-7})0.9999998965)^n, \tag{19}$$

which concludes the proof of Lemma 6, since $(1 + 10^{-7})0.9999998965 < 1$.

Note that the validity of our approach relies on the assumption that the solution of the 3×3 system found by Maple is unique, which implies that the critical point of $\log F$ we found is indeed the global maximum (if an alternative solution exists it could happen that at the corresponding critical point the function F attains a value greater than the maximum obtained).

In order to be more certain about the correctness of (19) we performed the following alternative experiment: Let $\mathcal{P}_{\bar{l}}$ be the polytope obtained by restricting $\mathcal{P}(n, \widehat{\delta}, \widehat{\gamma})$ to the coordinates $l_{ps}, l_{pu}, l_{ns1}, l_{ns2}, l_{nu}$. Observe that this is a 3-dimensional polytope in \mathbb{R}^5 , since its elements are determined by the values of the coordinates l_{ps}, l_{ns1}, l_{ns2} . We performed a sweep over this polytope by considering a grid of 100 equispaced points in each of the three dimensions. For each of the 100^3 fixed tuples of $(l_{ps}, l_{ns1}, l_{ns2})$ which correspond to the points on the grid, we determine the remaining two coordinates of $\mathcal{P}_{\bar{l}}$, and maximize $\log F$ restricted to those fixed values of \bar{l} . Observe that in this case $\log F$ is strictly concave and thus has a unique maximum which can be efficiently found by any iterative Newton-like algorithm. We checked, again using Maple, that the value obtained for each fixed tuple of \bar{l} is below the maximum in (19).

References

- [1] D. Achlioptas and F. Ricci-Tersenghi. On the solution-space geometry of random constraint satisfaction problems. In *38th Annual ACM Symposium on Theory of Computing (STOC)*, pages 130–139, 2006.

- [2] D. Coppersmith, D. Gamarnik, M. T. Hajiaghayi, and G. Sorkin. Random Max SAT, Random Max Cut, and their phase transitions. *Random Structures and Algorithms*, 24:502–545, 2004.
- [3] O. Dubois. Upper bound on the satisfiability threshold. *Theoretical Computer Science*, 265:187–197, 2001.
- [4] O. Dubois and Y. Boufkhad. A general upper bound for the satisfiability threshold of random r -SAT formulae. *Journal of Algorithms*, 24:395–420, 1997.
- [5] O. Dubois, Y. Boufkhad, and J. Mandler. Typical random 3-SAT formulae and the satisfiability threshold. In *Proceedings of the 11th Symposium on Discrete Algorithms (SODA)*, pages 126–127, 2000.
- [6] J. Franco and M. Paull. Probabilistic analysis of the Davis-Putnam procedure for solving the satisfiability problem. *Discrete Applied Mathematics*, 5:77–87, 1983.
- [7] E. Friedgut and J. Bourgain. Sharp thresholds of graph properties, and the k -SAT problem. *Journal of the American Mathematical Society*, 12:1017–1054, 1999.
- [8] A. Kamath, R. Motwani, K. Palem, and P. Spirakis. Tail bounds for occupancy and the satisfiability threshold conjecture. *Random Structures and Algorithms*, 7:59–80, 1995.
- [9] A. C. Kaporis, L. M. Kirousis, and E. G. Lalas. The probabilistic analysis of a greedy satisfiability algorithm. *Random Structures and Algorithms*, 28:444–480, 2006.
- [10] S. Kirkpatrick and B. Selman. Critical behavior in the satisfiability of random Boolean expressions. *Science*, 264(5163):1297–1301, 1994.
- [11] L. M. Kirousis, E. Kranakis, D. Krizanc, and Y. C. Stamatiou. Approximating the unsatisfiability threshold of random formulas. *Random Structures and Algorithms*, 12:253–269, 1998.
- [12] L. M. Kirousis, Y. C. Stamatiou, and M. Zito. The satisfiability threshold conjecture: Techniques behind upper bound. In A. P. G. Istrate and C. Moore, editors, *Computational Complexity and Statistical Physics*. Oxford University Press, 2005.
- [13] E. Maneva and A. Sinclair. On the satisfiability threshold and clustering of solutions of random 3-SAT formulas. *Theoretical Computer Science*, 2008.
- [14] M. Mézard and R. Zecchina. Random k -satisfiability: from an analytic solution to a new efficient algorithm. *Physics Review*, E-66, 056126:1357–1361, 2002.
- [15] R. Monasson and R. Zecchina. Statistical mechanics of the random k -SAT problem. *Physics Review*, E-56:1357–1361, 1997.
- [16] N. C. Wormald. The differential equation method for random graph processes and greedy algorithms. In M. Karoński and H. J. Prömel, editors, *Lectures on approximation and randomized algorithms*, pages 73–155. PWN, Warsaw, 1999.