The Satisfiability Threshold for Non-Uniform Random 2-SAT

Tobias Friedrich
Algorithm Engineering Group, Hasso Plattner Institute, University of Potsdam, Germany
tobias.friedrich@hpi.de

Ralf Rothenberger
Algorithm Engineering Group, Hasso Plattner Institute, University of Potsdam, Germany
ralf.rothenberger@hpi.de

Abstract

Propositional satisfiability (SAT) is one of the most fundamental problems in computer science. Its worst-case hardness lies at the core of computational complexity theory, for example in the form of NP-hardness and the (Strong) Exponential Time Hypothesis. In practice however, SAT instances can often be solved efficiently. This contradicting behavior has spawned interest in the average-case analysis of SAT and has triggered the development of sophisticated rigorous and non-rigorous techniques for analyzing random structures.

Despite a long line of research and substantial progress, most theoretical work on random SAT assumes a *uniform* distribution on the variables. In contrast, real-world instances often exhibit large fluctuations in variable occurrence. This can be modeled by a *non-uniform* distribution of the variables, which can result in distributions closer to industrial SAT instances.

We study satisfiability thresholds of non-uniform random 2-SAT with n variables and m clauses and with an arbitrary probability distribution $(p_i)_{i\in[n]}$ with $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_n > 0$ over the n variables. We show for $p_1^2 = \Theta\left(\sum_{i=1}^n p_i^2\right)$ that the asymptotic satisfiability threshold is at $m = \Theta\left(\left(1-\sum_{i=1}^n p_i^2\right) / \left(p_1 \cdot \left(\sum_{i=2}^n p_i^2\right)^{1/2}\right)\right)$ and that it is coarse. For $p_1^2 = o\left(\sum_{i=1}^n p_i^2\right)$ we show that there is a sharp satisfiability threshold at $m = \left(\sum_{i=1}^n p_i^2\right)^{-1}$. This result generalizes the seminal works by Chvatal and Reed [FOCS 1992] and by Goerdt [JCSS 1996].

2012 ACM Subject Classification Theory of computation → Generating random combinatorial structures

Keywords and phrases random SAT, satisfiability threshold, sharpness, non-uniform distribution, 2-SAT

Digital Object Identifier 10.4230/LIPIcs.ICALP.2019.61

Category Track A: Algorithms, Complexity and Games

Related Version A full version of the paper is available at `https://arxiv.org/abs/1904.02027`.

Funding This paper is partially funded by the project *Skalenfreie Erfüllbarkeit* (project no. 416061626) of the German Research Foundation (DFG).

1 Introduction

Satisfiability of Propositional Formulas (SAT) is one of the most thoroughly researched topics in theoretical computer science. It was one of the first problems shown to be NP-complete by Cook [15] and, independently, by Levin [30]. Today SAT stands at the core of many techniques in modern complexity theory, for example NP-completeness proofs [29] or running time lower bounds assuming the (Strong) Exponential Time Hypothesis [10, 17, 26, 27].

In addition to its importance for theoretical research, Propositional Satisfiability is also famously applied in practice. Despite the theoretical hardness of SAT, many problems arising in practice can be transformed to SAT instances and then solved efficiently with

46th International Colloquium on Automata, Languages, and Programming (ICALP 2019).
Editors: Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi;
Article No. 61; pp. 61:1–61:14
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

state-of-the-art solvers. Problems like hard- and software verification, automated planning, and circuit design are often transformed into SAT instances. Such formulas arising from practical and industrial problems are therefore referred to as *industrial SAT instances.* The efficiency of SAT solvers on these instances suggests that they have a structure that makes them easier to solve than the theoretical worst-case.

1.1 Uniform Random k-SAT and the satisfiability threshold conjecture:

Random k-SAT is used to study the average-case complexity of Boolean Satisfiability. In the model, a random formula Φ with n variables, m clauses, and k literals per clause is generated in conjunctive normal form. Each of these formulas has the same uniform probability to be generated. Therefore, we also refer to this model as *uniform random k-SAT*.

One of the most prominent questions related to studying uniform random k-SAT is trying to prove the *satisfiability threshold conjecture.* The conjecture states that for a uniform random k-SAT formula Φ with n variables and m clauses there is a real number r_k such that

$$\lim_{n\to\infty} \Pr\left(\Phi \text{ is satisfiable}\right) = \begin{cases} 1 & m/n < r_k; \\ 0 & m/n > r_k. \end{cases}$$

Chvatal and Reed [11] and, independently, Goerdt [24] proved the conjecture for $k = 2$ and showed that $r_2 = 1$. For larger values of k upper and lower bounds have been established, e. g., $3.52 \leqslant r_3 \leqslant 4.4898$ [18, 25, 28]. Methods from statistical mechanics [32] were used to derive a numerical estimate of $r_3 \approx 4.26$. Coja-Oghlan and Panagiotou [12, 13] showed a bound (up to lower order terms) for $k \geqslant 3$ with $r_k = 2^k \log 2 - \frac{1}{2}(1 + \log 2) \pm o_k(1)$. Finally, Ding, Sly, and Sun [19] proved the exact position of the threshold for sufficiently large values of k. Still, for k between 3 and the values determined by Ding, Sly, and Sun the conjecture remains open.

The satisfiability threshold is also connected to the average hardness of solving instances. For uniform random k-SAT for example, the on average hardest instances are concentrated around the threshold [33].

1.2 Non-Uniform Random SAT

There is a large body of work which considers other random SAT models, e. g. regular random k-SAT [7, 8, 14, 38], random geometric k-SAT [9] and $2 + p$-SAT [1, 35, 34, 36]. However, most of these are not motivated by modeling the properties of industrial instances. One such property is community structure [6], i. e. some variables have a bias towards appearing together in clauses. It is clear by definition that such a bias does not exist in uniform random k-SAT. Therefore, Giráldez-Cru and Levy [23] proposed the Community Attachment Model, which creates random formulas with clear community structure. However, the work of Mull et al. [37] shows that instances generated by this model have exponentially long resolution proofs with high probability, making them hard on average for solvers based on conflict-driven clause learning.

Another important property of industrial instances is their degree distribution. The degree distribution of a formula Φ is a function $f\colon \mathbb{N} \to \mathbb{N}$, where $f(x)$ denotes the fraction of different Boolean variables that appear x times in Φ (negated or unnegated). Instances created with the uniform random k-SAT model have a binomial distribution, while some families of industrial instances appear to follow a power-law distribution [4], i. e. $f(x) \sim x^{-\beta}$, where β is a constant intrinsic to the instance. Therefore, Ansótegui et al. [5] proposed a random k-SAT model with a power-law degree distribution. Empirical studies by the same authors [2, 3, 4, 5] found that this distribution is beneficial for the runtime of SAT solvers specialized in industrial instances. However, it looks like instances generated with their model can be solved faster than uniform instances, but not as fast as industrial ones: median runtimes around the threshold still seem to scale exponentially for several state-of-the-art solvers [21].

Therefore, we want to consider a generalization of the model by Ansótegui et al. [4]. Our model allows instances with *any* given ensemble of variable distributions instead of only power laws: We draw m clauses of length k at random. For each clause the k variables are drawn with a probability proportional to the n-th distribution in the ensemble, then they are negated independently with a probability of $1/2$ each. This means, the probability ensemble is part of the model, but the number of variables n determines which distribution from the ensemble we actually use. We call this model *non-uniform random k-SAT* and denote it by $\mathcal{D}\left(n, k, (\vec{p}_x)_{x\in\mathbb{N}}, m\right)$. Although $\mathcal{D}\left(n, k, (\vec{p}_x)_{x\in\mathbb{N}}, m\right)$ cannot capture all properties of industrial instances, e.g. community structure, it can help us to investigate the influence of the degree distribution on the structure and on the computational complexity of such instances in an average-case scenario.

As one of the steps in analyzing this connection, we would like to find out for which ensembles of variable probability distributions an equivalent of the satisfiability threshold conjecture holds in non-uniform random k-SAT. In previous works we already proved upper and lower bounds on the threshold position [20] and showed sufficient conditions on sharpness [22]. In this work we are interested in actually determining the satisfiability threshold for $k = 2$. This is helpful in determining bounds on the satisfiability threshold for higher values of k, since 2-SAT instances appear as parts of k-SAT instances. We already successfully used this approach in [20] to derive lower bounds on the satisfiability threshold for non-uniform random k-SAT with a power-law distribution on the variables.

It has to be noted that Cooper et al. [16] and Levy [31] already studied thresholds in a similar random 2-SAT model. The difference is that in their models the degrees are fixed and the random instances determined in a configuration-model-like fashion, while in our model we only have a sequence of expected degrees from which the actual degrees might deviate. Note that it is not clear if the satisfiability thresholds in these two models coincide if they use the same sequence of (expected) degrees. Cooper et al. derive the position of the satisfiability threshold in their model if the maximum degree is sufficiently bounded. Levy only shows necessary conditions on unsatisfiability in the model of Cooper et al. This is not enough to derive the actual threshold position. In contrast, our result allows us to determine the position of the satisfiability threshold for *any* probability ensemble in the model we consider.

1.3 Our Results

We investigate the position and behavior of the satisfiability threshold for non-uniform random 2-SAT. That is, we fix the number of variables n and the variable distribution $\vec{p}_n$ from the ensemble and vary the number of clauses $m(n)$. To this end, we use the following definition and say that a function $m^*(n)$ is an *asymptotic satisfiability threshold* if

$$\Pr_{\Phi\sim\mathcal{D}(n,k,(\vec{p}_x)_{x\in\mathbb{N}},m)}(\Phi \text{ satisfiable}) = \begin{cases} 1-o(1) & \text{if } m(n) = o(m^*(n)) \\ o(1) & \text{if } m(n) = \omega(m^*(n)). \end{cases}$$

We also say that an asymptotic satisfiability threshold $m^*(n)$ is *sharp* if for all $\varepsilon > 0$

$$\Pr_{\Phi\sim\mathcal{D}(n,k,(\vec{p}_x)_{x\in\mathbb{N}},m)}(\Phi \text{ satisfiable}) = \begin{cases} 1-o(1) & \text{if } m(n) \leqslant (1-\varepsilon)\cdot m^*(n) \\ o(1) & \text{if } m(n) \geqslant (1+\varepsilon)\cdot m^*(n). \end{cases}$$

If an asymptotic threshold is not sharp, we call it *coarse*.

Let $\vec{p}_n = (p_1, p_2, \ldots, p_n)$ be the variable probability distribution we use. W. l. o. g. we assume $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_n$. We are going to show that there are three cases depending on $\vec{p}_n$:

1. If $p_1^2 = \Theta\left(\sum_{i=1}^n p_i^2\right)$ and $p_2^2 = \Theta\left(\sum_{i=2}^n p_i^2\right)$, then we can show that the asymptotic satisfiability threshold is at $m = \Theta\left(q_{\max}^{-1}\right)$, where $q_{\max} = \Theta\left((p_1 \cdot p_2) / \left(1 - \sum_{i=1}^n p_i^2\right)\right)$ is the maximum clause probability. We can also show that this threshold is coarse. The coarseness stems from the emergence of an unsatisfiable sub-formula of size 4, which contains only the two most probable variables.
2. If $p_1^2 = \Theta\left(\sum_{i=1}^n p_i^2\right)$ and $p_2^2 = o\left(\sum_{i=2}^n p_i^2\right)$, then the asymptotic threshold is at $m = \Theta\left(\left(1 - \sum_{i=1}^n p_i^2\right) / \left(p_1 \cdot \left(\sum_{i=2}^n p_i^2\right)^{1/2}\right)\right)$ and it is again coarse. This time the coarseness stems from the emergence of an unsatisfiable sub-formula with 3 variables and 4 clauses.
3. If $p_1^2 = o\left(\sum_{i=1}^n p_i^2\right)$, then there is a sharp threshold at exactly $m = 1/\left(\sum_{i=1}^n p_i^2\right)$.

Note that these three cases give us a complete dichotomy of coarseness and sharpness for the satisfiability threshold of non-uniform random 2-SAT. This result generalizes the seminal works by Chvatal and Reed [11] and by Goerdt [24] to arbitrary variable probability distributions and includes their findings as a special case (c. f. Section 6). We summarize our findings in the following theorem.

▶ **Theorem 1.1.** *Let $\mathcal{D}(n, 2, (\vec{p}_x)_{x\in\mathbb{N}}, m)$ be the non-uniform random 2-SAT model with n variables, m clauses, and an ensemble of probability distributions $(\vec{p}_x)_{x\in\mathbb{N}}$. Let $\vec{p}_n = (p_1, p_2, \ldots, p_n)$ be the n-th distribution from the ensemble. W. l. o. g. let $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_n$. If $p_1^2 = o\left(\sum_{i=1}^n p_i^2\right)$, then $\mathcal{D}(n, 2, (\vec{p}_x)_{x\in\mathbb{N}}, m)$ has a sharp satisfiability threshold at $m = \left(\sum_{i=1}^n p_i^2\right)^{-1}$. Otherwise, $\mathcal{D}(n, 2, (\vec{p}_x)_{x\in\mathbb{N}}, m)$ has a coarse satisfiability threshold at $m = \Theta\left(\left(1 - \sum_{i=1}^n p_i^2\right) / \left(p_1 \cdot \left(\sum_{i=2}^n p_i^2\right)^{1/2}\right)\right)$.*

1.4 Techniques

For the sharp threshold result, we only show the conditions on sharpness. These also imply the existence of an asymptotic threshold. For the coarse threshold results, however, we first have to show the existence of an asymptotic threshold at some number of clauses $m^*(n)$. Then, we have to show that for some range of constants $\varepsilon \in [\varepsilon_1, \varepsilon_2]$ the probability to generate a satisfiable instance at $\varepsilon \cdot m^*(n)$ is a constant bounded away from zero and one.

We extend and generalize the proof ideas of Chvatal and Reed [11]. In order to show a lower bound on the threshold, we investigate the existence of bicycles. Bicycles were introduced by Chvatal and Reed. They are sub-formulas which appear in every unsatisfiable formula. We can show with a first moment argument, that these do not appear below a certain number of clauses, thus making formulas satisfiable.

In order to show an upper bound on the threshold, we investigate the existence of snakes. Snakes are unsatisfiable sub-formulas and have also been introduced by Chvatal and Reed. We can show with a second-moment argument that snakes of certain sizes do appear above a certain number of clauses, thus making formulas unsatisfiable. However, we need to be careful and distinguish more possibilities of partially mapping snakes onto each other than in the uniform case. Unfortunately, this method does not work if $p_1^2 = \Theta\left(\sum_{i=1}^n p_i^2\right)$ and $p_2^2 = \Theta\left(\sum_{i=2}^n p_i^2\right)$. In that case we lower-bound the probability that an unsatisfiable sub-formula containing only the two most-probable variables exists. This can be done with a simple inclusion-exclusion argument and the resulting lemma also work for $k \geqslant 3$.

2 Preliminaries

We analyze non-uniform random k-SAT on n variables and m clauses. We denote by $X_1, \ldots, X_n$ the Boolean variables. A clause is a disjunction of k literals $\ell_1 \vee \ldots \vee \ell_k$, where each literal assumes a (possibly negated) variable. For a literal ℓ_i let $|\ell_i|$ denote the variable of the literal. A formula Φ in conjunctive normal form is a conjunction of clauses $c_1 \wedge \ldots \wedge c_m$. We conveniently interpret a clause c both as a Boolean formula and as a set of literals. We say that Φ is satisfiable if there exists an assignment of variables $X_1, \ldots, X_n$ such that the formula evaluates to 1. Now let $(\vec{p}_n)_{n\in\mathbb{N}}$ be an ensemble of probability distributions, where $\vec{p}_n = (p_{n,1}, p_{n,2}, \ldots, p_{n,n})$ is a probability distribution over n variables with $\Pr(X = X_i) = p_{n,i} =: p_n(X_i)$.

▶ **Definition 2.1** (Clause-Drawing Non-Uniform Random k-SAT)**.** *Let m, n, k be given, and consider any ensemble of probability distributions $(\vec{p}_n)_{n\in\mathbb{N}}$, where $\vec{p}_n = (p_{n,1}, p_{n,2}, \ldots, p_{n,n})$ is a probability distribution over n variables with $\sum_{i=1}^{n} p_{n,i} = 1$. The* clause-drawing non-uniform random k-SAT (non-uniform random k-SAT) *model $\mathcal{D}(n, k, (\vec{p}_x)_{x\in\mathbb{N}}, m)$ constructs a random SAT formula Φ by sampling m clauses independently at random. Each clause is sampled as follows:*

1. *Select k variables independently at random from the distribution $\vec{p}_n$. Repeat until no variables coincide.*
2. *Negate each of the k variables independently at random with probability $1/2$.*

For the sake of simplicity and since we will always only consider one distribution from the ensemble, we will omit the index n throughout the paper, e. g. the probability distribution $\vec{p}_n$ will be denoted as $(p_1, p_2, \ldots, p_n)$. W. l. o. g. we will assume $p_1 \geqslant p_2 \geqslant \ldots p_n$.

The clause-drawing non-uniform random k-SAT model is equivalent to drawing each clause independently at random from the set of all k-clauses which contain no variable more than once. The probability to draw a clause c over n variables is then

$$q_c := \frac{\prod_{\ell\in c} p(|\ell|)}{2^k \sum_{J\in\mathcal{P}_k(\{1,2,\ldots,n\})} \prod_{j\in J} p_j}, \tag{2.1}$$

where $\mathcal{P}_k(\cdot)$ denotes the set of cardinality-k elements of the power set. The factor 2^k in the denominator comes from the different possibilities to negate variables. Note that $k! \sum_{J\in\mathcal{P}_k(\{1,2,\ldots,n\})} \prod_{j\in J} p_{n,j}$ is the probability of choosing a k-clause that contains no variable more than once. We can now write

$$q_c = C \frac{k!}{2^k} \prod_{X\in S} p_n(X), \tag{2.2}$$

where we define $C := 1/\left(k! \cdot \sum_{J\in\mathcal{P}_k(\{1,2,\ldots,n\})} \prod_{j\in J} p_{n,j}\right)$. For $k = 2$ it holds that $C = 1/\left(1 - \left(\sum_{i=1}^{n} p_i^2\right)\right)$. Hiding this factor in C makes clause probabilities easier to handle. Throughout the paper we let $q_{\max}$ denote the maximum clause probability as defined in equation (2.2). In Section 3 and Section 4 we will assume $q_{\max} = o(1)$. The case $q_{\max} = \Theta(1)$ will be handled in Section 5. Note that this case can only happen for $p_1^2 = \Theta\left(\sum_{i=1}^{n} p_i^2\right)$ and $p_2^2 = \Theta\left(\sum_{i=2}^{n} p_i^2\right)$.

3 Bi-Cycles and a Lower Bound on the Satisfiability Threshold

Chvatal and Reed [11] define the following sub-structure of 2-SAT formulas and show that every unsatisfiable 2-CNF contains this substructure.

▶ **Definition 3.1** (bi-cycle). *We define a bicycle of length t to be a sequence of $t+1$ clauses of the form*

$$(u, w_1), (\overline{w}_1, w_2), \ldots, (\overline{w}_{t-1}, w_t), (\overline{w}_t, v),$$

where $w_1, \ldots, w_t$ are literals of distinct variables and $u, v \in \{w_1, \ldots, w_t, \overline{w}_1, \ldots, \overline{w}_t\}$.

To lower-bound the probability for a random 2-CNF to be satisfiable it is therefore sufficient to upper-bound the probability that such a formula contains a bicycle. This is done in the following two lemmas. Their proofs are oriented along the lines of the proof of Theorem 3 from [11].

▶ **Lemma 3.1.** *Consider a non-uniform random 2-SAT formula Φ with $p_1^2 = o\left(\sum_{i=1}^n p_i^2\right)$. Then, Φ is satisfiable with probability at least $1 - o(1)$ for a number of clauses $m < (1-\varepsilon)\left(\sum_{i=1}^n p_i^2\right)^{-1}$, where $\varepsilon > 0$ is a constant.*

Proof. To show this result, we show that the expected number of bicycles is $o(1)$ for the setting we consider. The result then follows by Markov's inequality.

First, we fix a set $S \subseteq [n]$ of variables to appear in a bicycle with $|S| = t \geqslant 2$. The probability that a *specific* bicycle B with these variables appears in Φ is

$$\Pr(B \text{ in } \Phi) = \underbrace{\binom{m}{t+1}(t+1)!}_{\text{positions of } B \text{ in } \Phi} \cdot \Pr(u \vee w_1) \cdot \Pr(\overline{w}_t \vee v) \prod_{h=1}^{t-1} \Pr(\overline{w}_h \vee w_{h+1}).$$

For literals w_i over variables x_i it holds that

$$\Pr(w_j \vee w_i) = \frac{C}{2} p_i \cdot p_j,$$

where $1 \leqslant C = \left(1 - \sum_{i=1}^n p_i^2\right)^{-1} = 1 + o(1)$, since $\sum_{i=1}^n p_i^2 = o(1)$ due to the requirement $p_1^2 = o\left(\sum_{i=1}^n p_i^2\right)$. There are at most $t!$ possibilities to arrange the t variables in a bicycle and 2^t possibilities to choose literals from the t variables. For the probability that *any* bicycle with the variables from S appears in Φ it now holds that

$$\Pr(S\text{-bicycle in } \Phi) \leqslant m^{t+1} \cdot t! \cdot 2^t \cdot \left(\frac{C}{2}\right)^{t+1} \cdot \prod_{i \in S} p_i^2 \left(2 \cdot \sum_{i \in S} p_i\right)^2$$

where the last factor accounts for the possibilities to choose u and v. It now holds that

$$\begin{aligned}
\Pr(\Phi \text{ contains a bicycle}) &\leqslant \sum_{t=2}^{n} \sum_{S \subseteq \mathcal{P}_t(V)} m^{t+1} \cdot t! \cdot 2^t \cdot \left(\frac{C}{2}\right)^{t+1} 2^2 \cdot \prod_{i \in S} p_i^2 \left(\sum_{i \in S} p_i\right)^2 \\
&\leqslant 2 \cdot \sum_{t=2}^{n} (C \cdot m)^{t+1} \cdot t! \cdot t^2 \cdot p_1^2 \cdot \sum_{S \subseteq \mathcal{P}_t(V)} \prod_{i \in S} p_i^2 \\
&\leqslant 2 \cdot \sum_{t=2}^{n} (C \cdot m)^{t+1} \cdot t^2 \cdot p_1^2 \cdot \left(\sum_{i \in S} p_i^2\right)^t \\
&= o\left(2 \cdot \sum_{t=2}^{n} \left(C \cdot m \left(\sum_{i \in S} p_i^2\right)\right)^{t+1} \cdot t^2\right),
\end{aligned}$$

where we used $\sum_{i\in S} p_i \leqslant t \cdot p_1$ in the second, $\sum_{S\subseteq \mathcal{P}_t(V)} \prod_{i\in S} p_i^2 \leqslant \frac{1}{t!} \cdot \left(\sum_{i\in S} p_i^2\right)^t$ in the third line, and the requirement $p_1^2 = o\left(\sum_{i=1}^n p_i^2\right)$ in the fourth line. It is obvious that this probability is o(1) as soon as the sum becomes a constant. This holds for $m < (1-\varepsilon)\left(\sum_{i=1}^n p_i^2\right)^{-1} < \left(C \cdot \sum_{i=1}^n p_i^2\right)^{-1}$, where $\varepsilon > 0$ is a constant. ◀

It has to be noted that in the former lemma we ignored the factor C in our bound. We can do this, since for $p_1^2 = o\left(\sum_{i=1}^n p_i^2\right)$ it always is $1 + o(1)$ and does not make a difference for sharpness due to our definition. In the case of $p_1^2 = \Theta\left(\sum_{i=1}^n p_i^2\right)$, we can show the following result with a similar proof, but now we have to take C into account, since it might become super-constant. Also we have to do a case distinction between the terms with $p_1 \in S$ and $p_1 \notin S$ to get $\sum_{S\subseteq \mathcal{P}_t(V)} \left(\prod_{i\in S} p_i^2\right) \cdot \left(\sum_{i\in S} p_i\right)^2 = \mathcal{O}\left(t^3 \cdot p_1^4 \cdot \frac{1}{t!} \cdot \left(\sum_{i=2}^n p_i^2\right)^{t-1}\right)$. See the full version of the paper for the whole proof.

▶ **Lemma 3.2.** *Consider a non-uniform random 2-SAT formula Φ with $p_1^2 = \Theta\left(\sum_{i=1}^n p_i^2\right)$ and $q_{\max} = o(1)$. Then, Φ is satisfiable with probability at least $1 - o(1)$ for a number of clauses $m = o\left(\left(C \cdot p_1 \cdot \left(\sum_{i=2}^n p_i^2\right)^{1/2}\right)^{-1}\right)$. Also, there is a constant $\varepsilon \in (0,1)$ such that Φ is satisfiable with a positive constant probability for a number of clauses $m \leqslant (1-\varepsilon)\left(C \cdot p_1 \cdot \left(\sum_{i=2}^n p_i^2\right)^{1/2}\right)^{-1}$.*

Note that this lemma captures both cases for $p_1^2 = \Theta\left(\sum_{i=1}^n p_i^2\right)$. If also $p_2^2 = \Theta\left(\sum_{i=2}^n p_i^2\right)$, then $\left(C \cdot p_1 \cdot \left(\sum_{i=2}^n p_i^2\right)^{1/2}\right)^{-1} = \Theta\left(q_{\max}^{-1}\right)$ is the asymptotic threshold as we stated in the introduction. The case $q_{\max} = \Theta(1)$ has to be excluded, since for that case the asymptotic threshold is a constant. The above lemma might then give us a value so small that the ranges where we can lower- and upper-bound satisfiability to constants away from zero resp. one do not overlap. Thus, this case is handled separately in Section 5.

4 Snakes and an Upper Bound on the Satisfiability Threshold

The two lemmas from the previous section provided a lower bound on the satisfiability threshold for non-uniform random 2-SAT. By using the second moment method, we can also derive an upper bound on the threshold. Again, this proof is inspired by Chvatal and Reed [11, Theorem 4], who provide us with the following definition.

▶ **Definition 4.1** (snake). *A* snake *of size t is a sequence of literals $w_1, w_2, \ldots, w_{2t-1}$ over distinct variables. Each snake A is associated with a set F_A of $2t$ clauses $(\overline{w}_i, w_{i+1})$, $0 \leqslant i \leqslant 2t-1$, such that $w_0 = w_{2t} = \overline{w}_t$.*

We will also call the variable $|w_t|$ of a snake its *central* variable. Note that the set of clauses F_A defined by a snake A is unsatisfiable. Also, the snakes $(w_1, \ldots, w_{t-1}, w_t, w_{t+1}, \ldots, w_s)$, $(\overline{w_{t-1}}, \overline{w_{t-2}} \ldots, \overline{w_1}, w_t, w_{t+1}, \ldots, w_s)$, $(w_1, \ldots, w_{t-1}, w_t, \overline{w_s}, \overline{w_{s-1}} \ldots, \overline{w_{t+1}})$ and $(\overline{w_{t-1}}, \overline{w_{t-2}} \ldots, \overline{w_1}, w_t \overline{w_s}, \overline{w_{s-1}} \ldots, \overline{w_{t+1}})$ create the same set of formulas.

The *variable-variable incidence graph (VIG)* for a formula Φ is a simple graph $G_\Phi = (V_\Phi, E_\Phi)$ with V_Φ consisting of all variables appearing in Φ and two variables being connected by an edge if they appear together in at least one clause of Φ. An example for a snake's VIG can be seen in Figure 1. This representation will come in handy later in the proofs of Lemmas 4.4 and 4.6.

In order to show our upper bounds, we will prove that snakes of a certain length t appear with sufficiently high probability in a random formula $\Phi \sim \mathcal{D}\left(n, k, (\vec{p}_x)_{x\in\mathbb{N}}, m\right)$ To this end we utilize the second moment method: If $X \geqslant 0$ is a random variable with finite variance, then

$\Pr(X > 0) \geqslant \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}$.

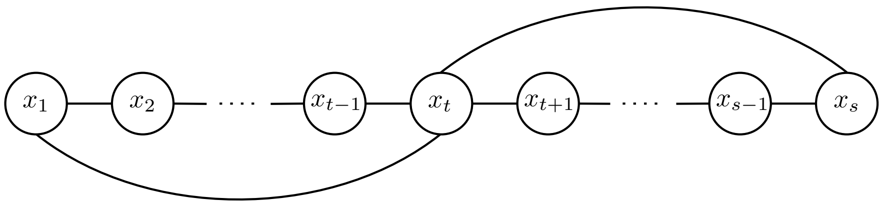

Figure 1 Variable-variable-incidence graph of a snake $w_1, w_2, \ldots, w_s$ where $|w_i| = x_i$ (the variable of the literal w_i) for $1 \leqslant i \leqslant s = 2t - 1$.

We define the following indicator variables for each snake A of size t

$$X_A = \begin{cases} 1 & \text{if } F_A \text{ appears exactly once in } \Phi \\ 0 & \text{otherwise} \end{cases}$$

and their sum $X_t = \sum_A X_A$. For carefully chosen t we will show $\mathbb{E}\left[X_t^2 \right] = \mathcal{O}(\mathbb{E}\left[X_t \right]^2)$ to show a coarse and $\mathbb{E}\left[X_t^2 \right] = (1 + o(1)) \cdot (\mathbb{E}\left[X_t \right]^2)$ to show a sharp threshold. This implies a constant resp. $1 - o(1)$ probability to be unsatisfiable due to the second moment method. In the case of $p_1^2 = o(\sum_{i=1}^n p_i^2)$, we will chose $t = \Theta\left(\log^2 f(n)\right)$, where we define $f(n) = \left(\sum_{i=1}^n p_i^2\right)/p_1^2$. For $p_1^2 = \Theta(\sum_{i=1}^n p_i^2)$ and $p_2^2 = o(\sum_{i=2}^n p_i^2)$ we choose $t = 2$. We only want to use the method for these two cases. The third case with $p_1 = \Theta\left(\sum_{i=1}^n p_i^2\right)$ and $p_2 = \Theta\left(\sum_{i=2}^n p_i^2\right)$ will be handled with the more general Lemma 4.7.

Now, if we want to use the second moment method, we first have to ensure that the expected number of snakes of a certain size is large enough. The following lemma provides a lower bound on this expected number. Due to space limitations, its proof can be found in the full version.

▶ **Lemma 4.1.** *Let X_t be the number of snakes of size $s + 1 = 2t$ whose associated formulas appear exactly once in a non-uniform random 2-SAT formula. Then it holds that*

$$\mathbb{E}\left[X_t \right] \geqslant \frac{1}{2}(m-2t)^{2t} \cdot C^{2t} \cdot e^{-(m-2t)\frac{2t \cdot q_{\max}}{1-2t \cdot q_{\max}}} \cdot \left(\sum_{i=1}^{n} p_i^4 \right) \cdot \left(\sum_{i=2}^{n} p_i^2 - (2t-2) \cdot p_2^2 \right)^{2t-2}.$$

In order to use the second moment method we have to show that this expected value is at least a constant if we want to show a coarse threshold and asymptotically bigger than a constant if we want to show a sharp threshold. Hence, the following lemmas give lower bounds on $\mathbb{E}\left[X_t \right]$ for the first two cases and the respective ranges of t we consider. Again, their proofs can be found in the full version of this paper.

▶ **Lemma 4.2.** *Let X_t be the number of snakes of size t that appear exactly once in a non-uniform random 2-SAT formula with $p_1^2 = o(\sum_{i=1}^n p_i^2)$ and $m = (1+\varepsilon)\left(\sum_{i=1}^n p_i^2\right)^{-1}$ for some $\varepsilon > 0$. Then it holds that*

$$\mathbb{E}\left[X_t \right] \geqslant (1 - o(1)) \cdot m^{2t} \left(\sum_{i=1}^{n} p_i^4 \right) \cdot \left(\sum_{i=1}^{n} p_i^2 \right)^{2t-2} = \omega(1)$$

if $t = o\left(\sqrt{f(n)}\right) \cap \omega\left(\log f(n)\right)$, where $f(n) = \left(\sum_{i=1}^n p_i^2\right)/p_1^2$.

▶ **Lemma 4.3.** *Let X_t be the number of snakes of size t that appear exactly once in a non-uniform random 2-SAT formula with $p_1^2 = \Theta(\sum_{i=1}^n p_i^2)$ and $p_2^2 = o(\sum_{i=2}^n p_i^2)$. For $t = 2$ and $m = \Omega\left(\left(C \cdot p_1 \cdot \left(\sum_{i=2}^n p_i^2\right)^{1/2}\right)^{-1}\right) \cap o\left(q_{\max}^{-1}\right)$ it holds that*

$$\mathbb{E}\left[\, X_2 \,\right] \geqslant (1 - o(1)) \cdot m^4 \cdot C^4 \cdot p_1^4 \cdot \left(\sum_{i=1}^n p_i^2\right)^2 .$$

Furthermore,

$$\mathbb{E}\left[\, X_2 \,\right] = \begin{cases} \Omega(1) & , m = \Theta\left(\left(C \cdot p_1 \left(\sum_{i=2}^n p_i^2\right)^{1/2}\right)^{-1}\right) \text{ and} \\ \omega(1) & , m = \omega\left(\left(C \cdot p_1 \left(\sum_{i=2}^n p_i^2\right)^{1/2}\right)^{-1}\right) \cap o\left((q_{\max})^{-1}\right). \end{cases}$$

Now we are ready to prove an upper bound on the non-uniform random 2-SAT threshold. To get to know the proof technique, we start with the much simpler case $p_1^2 = \Theta\left(\sum_{i=1}^n p_i^2\right)$ and $p_2^2 = o\left(\sum_{i=2}^n p_i^2\right)$. The proof contains a small case distinction depending on how the shared clauses of two snakes A and B influence $\Pr\left(X_A \wedge X_B\right)$. The next lemma establishes that there is a regime of m where random formulas are unsatisfiable with a positive constant probability. Its proof is in the full version.

▶ **Lemma 4.4.** *Consider a non-uniform random 2-SAT formula Φ with $p_1^2 = \Theta\left(\sum_{i=1}^n p_i^2\right)$ and $p_2^2 = o\left(\sum_{i=2}^n p_i^2\right)$. Then Φ is unsatisfiable with positive constant probability for $m = \Theta\left(\left(C \cdot p_1 \left(\sum_{i=2}^n p_i^2\right)^{1/2}\right)^{-1}\right)$.*

The following lemma complements the former one, showing that above that regime of m random formulas are unsatisfiable with probability $1 - o(1)$. Its proof is very similar and is therefore omitted. It can be found in the full version of this work.

▶ **Lemma 4.5.** *Consider a non-uniform random 2-SAT formula Φ with $p_1^2 = \Theta\left(\sum_{i=1}^n p_i^2\right)$ and $p_2^2 = o\left(\sum_{i=2}^n p_i^2\right)$. Then Φ is unsatisfiable with probability $1 - o(1)$ for $m = \omega\left(\left(C \cdot p_1 \left(\sum_{i=2}^n p_i^2\right)^{1/2}\right)^{-1}\right)$.*

The former two lemmas together with Lemma 3.2 establish that in the case of $p_1^2 = \Theta\left(\sum_{i=1}^n p_i^2\right)$ and $p_2^2 = o\left(\sum_{i=2}^n p_i^2\right)$ the asymptotic threshold is at $m = \Theta\left(\left(C \cdot p_1 \left(\sum_{i=2}^n p_i^2\right)^{1/2}\right)^{-1}\right)$ and that it is coarse.

We now turn to the case $p_1^2 = \Theta\left(\sum_{i=1}^n p_i^2\right)$. Again, we have to consider different possibilities for the shared clauses of snakes A and B to influence $\Pr\left(X_A \wedge X_B\right)$. In the proofs of the former case this was rather easy, since we only considered the smallest possible snakes of size 3. Now the distinction becomes a bit more difficult. We will distinguish several cases: If the number of shared clauses is at least $t - 1$ then $\Pr\left(X_A \wedge X_B\right)$ is by roughly a factor of $(1 + \varepsilon)^t$ smaller than $\mathbb{E}\left[\, X_t \,\right]^2$. If the shared clauses form at least two connected sub-formulas, then there are enough variable appearances pre-defined for B to make $\Pr\left(X_A \wedge X_B\right)$ sufficiently small. The last case is that there is only one connected sub-formula, which is a lot smaller than $t - 1$. In that case we have to carefully consider what happens to the central variable from B, since this variable appears most times in B and the many appearances take degrees of freedom away from other variables, therefore making $\Pr\left(X_A \wedge X_B\right)$ small. The whole proof of the lemma can be found in the full version.

▶ **Lemma 4.6.** *Consider a non-uniform random 2-SAT formula Φ with $p_1^2 = o\left(\sum_{i=1}^n p_i^2\right)$. Then Φ is unsatisfiable with probability $1 - o(1)$ for $m > (1+\varepsilon) \cdot \left(\sum_{i=1}^n p_i^2\right)^{-1}$, where $\varepsilon > 0$ is a constant.*

Lemma 4.6 and Lemma 3.1 now establish the existence of a sharp threshold at $m = \left(\sum_{i=1}^n p_i^2\right)^{-1}$.

Now we still have to consider the case $p_1^2 = \Theta\left(\sum_{i=1}^n p_i^2\right)$ and $p_2^2 = \Theta\left(\sum_{i=2}^n p_i^2\right)$. In the following lemma, we give a lower bound on the probability to be unsatisfiable by showing the existence of an unsatisfiable sub-formula consisting only of the two most-probable variables. The lemma generally holds for $k \geqslant 2$, but it especially serves our purpose of considering the remaining case. The proof uses a simple inlcusion-exclusion argument and can be found in the full version.

▶ **Lemma 4.7.** *Consider a non-uniform random k-SAT formula Φ with $q_{\max} = o(1)$. Then Φ is unsatisfiable with probability at least*

$$\sum_{l=0}^{2^k} \left(\binom{2^k}{l} (-1)^l \left(1 - l \cdot q_{\max}\right)^m \right)$$
$$\geqslant \left(1 - \exp\left(-q_{\max} \cdot m\right)\right)^{2^k} - q_{\max}^2 \cdot 2^{2k} \cdot m \cdot \left(1 + \exp\left(-q_{\max} \cdot m\right)\right)^{2^k} .$$

The former lemma now yields the following corollary.

▶ **Corollary 4.1.** *Consider a non-uniform random k-SAT formula Φ with $q_{\max} = o(1)$. Then*

1. $\Pr(\Phi$ *unsatisfiable*$) = \Omega(1)$ *for* $m = \Theta\left(q_{\max}^{-1}\right)$ *and*
2. $\Pr(\Phi$ *unsatisfiable*$) = 1 - o(1)$ *for* $m = \omega(q_{\max}^{-1})$.

In the second case the result follows from Lemma 4.7 for $m = \omega(q_{\max}^{-1}) \cap o(q_{\max}^{-2})$ and by monotonicity of the satisfiability probability in m. This corollary together with Lemma 3.2 establishes the existence of a coarse threshold at $m = \Theta\left(\left(C \cdot p_1 \left(\sum_{i=2}^n p_i^2\right)^{1/2}\right)^{-1}\right) = \Theta\left(q_{\max}^{-1}\right)$ for non-uniform random 2-SAT with $p_1^2 = \Theta\left(\sum_{i=1}^n p_i^2\right)$, $p_2^2 = \Theta\left(\sum_{i=2}^n p_i^2\right)$.

5 Constant Clause Probabilities

We assumed $q_{\max} = o(1)$ throughout the paper. For the sake of completeness we still have to take care of the case $q_{\max} = \Theta(1)$. It is easy to see that for $\Phi \sim \mathcal{D}(n, 2, (\vec{p}_x)_{x\in\mathbb{N}}, m)$ and a constant $m \geqslant 4$ it holds that $\Pr(\Phi \text{ unsatisfiable}) \geqslant q_{\max}^m$, since this is the probability of an unsatisfiable instance, where the most probable clause appears with all four combinations of signs and then one of these clauses appears an additional $m - 4$ times. Similarly, $\Pr(\Phi \text{ satisfiable}) \geqslant q_{\max}^m$, as this is the probability of a satisfiable instance, where the same most probable clause appears m times with the same sign. Since $0 < q_{\max} \leqslant 1/4$ is a constant, the probability is a constant bounded away from zero and one. It remains to show that Φ is unsatisfiable with probability $1 - o(1)$ for $m = \omega(1)$. The following lemma establishes this. Again, this lemma also holds for $k \geqslant 2$ in general.

▶ **Lemma 5.1.** *Consider a non-uniform random k-SAT formula Φ. Then Φ is unsatisfiable with probability at least*

$$2 - \left(1 + \exp\left(-q_{\max} \cdot m\right)\right)^{2^k} .$$

Proof. As in Lemma 4.7, it holds that

$$\Pr\left(\Phi \text{ unsat}\right) \geqslant \sum_{l=0}^{2^k} \left(\binom{2^k}{l} (-1)^l \left(1 - l \cdot q_{\max}\right)^m \right).$$

We can now estimate

$$\begin{aligned}
\sum_{l=0}^{2^k} \left(\binom{2^k}{l} (-1)^l \left(1 - l \cdot q_{\max}\right)^m \right) &\geqslant 1 - \sum_{l=1}^{2^k} \left(\binom{2^k}{l} \left(1 - l \cdot q_{\max}\right)^m \right) \\
&\geqslant 1 - \sum_{l=1}^{2^k} \left(\binom{2^k}{l} \exp\left(-m \cdot \frac{l \cdot q_{\max}}{1 - l \cdot q_{\max}}\right)^m \right) \\
&\geqslant 1 - \sum_{l=1}^{2^k} \left(\binom{2^k}{l} \exp\left(-m \cdot l \cdot q_{\max}\right) \right) \\
&= 2 - \left(1 + \exp\left(-m \cdot q_{\max}\right)\right)^{2^k}
\end{aligned}$$ ◀

For $q_{\max} = \Theta(1)$ and $m = \omega\left(q_{\max}^{-1}\right)$ this lemma implies $\Pr\left(\Phi \text{ unsatisfiable}\right) \geqslant 1 - o(1)$. All lemmas together now imply our main theorem.

▶ **Theorem 1.1.** *Let $\mathcal{D}\left(n, 2, (\vec{p}_x)_{x\in\mathbb{N}}, m\right)$ be the non-uniform random 2-SAT model with n variables, m clauses, and an ensemble of probability distributions $(\vec{p}_x)_{x\in\mathbb{N}}$. Let $\vec{p}_n = (p_1, p_2, \ldots, p_n)$ be the n-th distribution from the ensemble. W. l. o. g. let $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_n$. If $p_1^2 = o\left(\sum_{i=1}^n p_i^2\right)$, then $\mathcal{D}\left(n, 2, (\vec{p}_x)_{x\in\mathbb{N}}, m\right)$ has a sharp satisfiability threshold at $m = \left(\sum_{i=1}^n p_i^2\right)^{-1}$. Otherwise, $\mathcal{D}\left(n, 2, (\vec{p}_x)_{x\in\mathbb{N}}, m\right)$ has a coarse satisfiability threshold at $m = \Theta\left(\left(1 - \sum_{i=1}^n p_i^2\right) / \left(p_1 \cdot \left(\sum_{i=2}^n p_i^2\right)^{1/2}\right)\right)$.*

6 Example Applications of our Theorem

We will now show on some examples how our main theorem can be applied.

6.1 Uniform Distribution

The simplest distribution we can apply our theorem to is the uniform distribution, i. e. $\vec{p}_n = \left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$ for all $n \in \mathbb{N}$. It holds that $p_1^2 = \frac{1}{n^2}$ and $\sum_{i=1}^n p_i^2 = \frac{1}{n}$. Thus, Theorem 1.1 implies a sharp threshold at $m^*(n) = n$ for all $n \in \mathbb{N}$. This reproves [11, 24] the satisfiability threshold conjecture for $k = 2$, since these sharp threshold are all at $m^*(n) = n$.

6.2 Power Law Distribution

Another ensemble of distributions we can choose are power-law distributions, i.e. we consider the power law random 2-SAT model introduced by Ansótegui et al. [5]. Thus, for a constant $\beta > 2$ we choose $\vec{p}_n = (p_1, p_2, \ldots, p_n)$ with

$$p_i = \frac{(n/i)^{\frac{1}{\beta-1}}}{\left(\sum_{j=1}^n (n/j)^{\frac{1}{\beta-1}}\right)}.$$

It already holds that $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_n$. Now it is an easy exercise to show that

$$\left(\sum_{j=1}^n (n/j)^{\frac{1}{\beta-1}}\right) = (1 - o(1)) \cdot \frac{\beta - 1}{\beta - 2}.$$

Furthermore

$$p_1^2 = (1 \pm o(1)) \cdot \left(\frac{\beta-1}{\beta-2}\right)^2 \cdot n^{-2\frac{\beta-2}{\beta-1}}.$$

Finally, one can show that

$$\sum_{i=1}^{n} p_i^2 = \begin{cases} (1 \pm o(1)) \cdot \frac{(\beta-2)^2}{(\beta-3)\cdot(\beta-1)} \cdot n^{-2\frac{\beta-2}{\beta-1}} & \text{for } \beta < 3 \\ (1 \pm o(1)) \cdot \frac{1}{4} \cdot \frac{\ln n}{n} & \text{for } \beta = 3 \\ (1 \pm o(1)) \cdot \frac{(\beta-2)^2}{(\beta-3)\cdot(\beta-1)} \cdot n^{-1} & \text{for } \beta > 3. \end{cases}$$

Thus, applying our theorem we can see that for $\beta < 3$ there is a coarse threshold at $m = \Theta\left(n^{-2\frac{\beta-2}{\beta-1}}\right)$, since $p_1^2 = \Theta\left(\sum_{i=1}^n p_i^2\right) = \Theta\left(n^{2\frac{\beta-2}{\beta-1}}\right)$ and $C = 1 + o(1)$. For $\beta = 3$ there is a sharp threshold at $4 \cdot \frac{n}{\ln n}$, since $p_1^2 = \Theta\left(n^{-1}\right) = o\left(\frac{\ln n}{n}\right)$. Also, there is a sharp threshold at $\frac{(\beta-3)\cdot(\beta-1)}{(\beta-2)^2} \cdot n$ for $\beta > 3$, since $p_1^2 = \Theta\left(n^{-2\frac{\beta-2}{\beta-1}}\right) = o(n)$. We already observed the behavior for the latter case experimentally in previous works [21, 20]. Thus, we can say that for power-law random 2-SAT with a fixed power-law exponent $\beta \geqslant 3$ an equivalent of the satisfiability threshold conjecture holds, since the sharp thresholds converge to a function with the same leading constant factor.

6.3 Geometric Distribution

Ansótegui et al. [5] also considered an ensemble of geometric distributions with

$$p_i = \frac{b \cdot (1 - b^{-1/n})}{b-1} \cdot b^{-(i-1)/n}$$

for $i = 1, \ldots, n$ and for some constant $b > 1$. Again, it already holds that $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_n$. It holds that

$$p_1^2 = \frac{b^2 \cdot (1 - b^{-1/n})^2}{(b-1)^2}$$

and

$$\sum_{i=1}^{n} p_i^2 = \frac{b+1}{b-1} \cdot \frac{1 - b^{-1/n}}{1 + b^{-1/n}}.$$

One can show that $p_1^2 = o(\sum_{i=1}^n p_i^2)$. Theorem 1.1 now tells us that there is a sharp threshold at $\frac{b-1}{b+1} \cdot \frac{1+b^{-1/n}}{1-b^{-1/n}}$. This function grows as fast as $\frac{2\cdot(b-1)}{(b+1)\cdot\ln b} \cdot n$ in the limit. Again, we can say that an equivalent of the satisfiability threshold conjecture holds for geometric random 2-SAT with some fixed $b > 1$, since the sharp thresholds $m^*(n)$ converge to $\frac{2\cdot(b-1)}{(b+1)\cdot\ln b} \cdot n$.

7 Discussion and Future Work

We showed a dichotomy of coarse and sharp thresholds for the non-uniform random 2-SAT model depending on the variable probability distribution. In the case of a coarse threshold, the coarseness either stems from two variables being present in too many clauses and forming an unsatisfiable sub-formula of size 4 with constant probability or from a snake with three variables which emerges with constant probability. Furthermore we determined the exact position of the satisfiability threshold in the case of a sharp threshold. Hence, our result generalizes the seminal works by Chvatal and Reed [11] and by Goerdt [24] to arbitrary variable probability distributions. It allows us to prove or disprove an equivalent of the satisfiability threshold conjecture for non-uniform random 2-SAT. For example for power-law random 2-SAT, an equivalent of the conjecture holds for power law exponents $\beta \geqslant 3$ and the satisfiability threshold is at exactly $\frac{(\beta-3)\cdot(\beta-1)}{(\beta-2)^2} \cdot n$ for $\beta > 3$ and exactly at $4 \cdot \frac{n}{\ln n}$ for $\beta = 3$.

The grand goal of our works is to show similar results for higher values of k, where we already made a first step by showing sharpness for certain variable probability distributions [22]. Another direction we are interested in for $k \geqslant 3$ is proving bounds on the average computational hardness of formulas around the threshold, for example by showing resolution lower bounds like Mull et al. [37].

References

1. Dimitris Achlioptas, Lefteris M. Kirousis, Evangelos Kranakis, and Danny Krizanc. Rigorous results for random (2+p)-SAT. *Theor. Comput. Sci.*, 265(1-2):109–129, 2001.
2. Carlos Ansótegui, Maria Luisa Bonet, Jesús Giráldez-Cru, and Jordi Levy. The Fractal Dimension of SAT Formulas. In *7th Intl. Joint Conf. Automated Reasoning (IJCAR)*, pages 107–121, 2014.
3. Carlos Ansótegui, Maria Luisa Bonet, Jesús Giráldez-Cru, and Jordi Levy. On the Classification of Industrial SAT Families. In *18th Intl. Conf. of the Catalan Association for Artificial Intelligence (CCIA)*, pages 163–172, 2015.
4. Carlos Ansótegui, Maria Luisa Bonet, and Jordi Levy. On the Structure of Industrial SAT Instances. In *15th Intl. Conf. Principles and Practice of Constraint Programming (CP)*, pages 127–141, 2009.
5. Carlos Ansótegui, Maria Luisa Bonet, and Jordi Levy. Towards Industrial-Like Random SAT Instances. In *21st Intl. Joint Conf. Artificial Intelligence (IJCAI)*, pages 387–392, 2009.
6. Carlos Ansótegui, Jesús Giráldez-Cru, and Jordi Levy. The Community Structure of SAT Formulas. In *15th Intl. Conf. Theory and Applications of Satisfiability Testing (SAT)*, pages 410–423, 2012.
7. Victor Bapst and Amin Coja-Oghlan. The Condensation Phase Transition in the Regular k-SAT Model. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2016*, pages 22:1–22:18, 2016.
8. Yacine Boufkhad, Olivier Dubois, Yannet Interian, and Bart Selman. Regular Random k-SAT: Properties of Balanced Formulas. *J. Autom. Reasoning*, 35(1-3):181–200, 2005.
9. Milan Bradonjic and Will Perkins. On Sharp Thresholds in Random Geometric Graphs. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2014*, pages 500–514, 2014.
10. Karl Bringmann. Why Walking the Dog Takes Time: Frechet Distance Has No Strongly Subquadratic Algorithms Unless SETH Fails. In *55th Symp. Foundations of Computer Science (FOCS)*, pages 661–670, 2014.
11. Václav Chvatal and Bruce Reed. Mick gets some (the odds are on his side). In *33rd Symp. Foundations of Computer Science (FOCS)*, pages 620–627, 1992.
12. Amin Coja-Oghlan. The Asymptotic k-SAT Threshold. In *46th Symp. Theory of Computing (STOC)*, pages 804–813, 2014.
13. Amin Coja-Oghlan and Konstantinos Panagiotou. The asymptotic k-SAT threshold. *Advances in Mathematics*, 288:985–1068, 2016.
14. Amin Coja-Oghlan and Nick Wormald. The Number of Satisfying Assignments of Random Regular k-SAT Formulas. *Combinatorics, Probability and Computing*, 27(4):496–530, 2018.
15. Stephen A. Cook. The Complexity of Theorem-Proving Procedures. In *3rd Symp. Theory of Computing (STOC)*, pages 151–158, 1971.
16. Colin Cooper, Alan Frieze, and Gregory Sorkin. Random 2SAT with Prescribed Literal Degrees. *Algorithmica*, 48:249–265, July 2007.
17. Marek Cygan, Jesper Nederlof, Marcin Pilipczuk, Michal Pilipczuk, Johan M. M. van Rooij, and Jakub Onufry Wojtaszczyk. Solving Connectivity Problems Parameterized by Treewidth in Single Exponential Time. In *52nd Symp. Foundations of Computer Science (FOCS)*, pages 150–159, 2011.

18 Josep Díaz, Lefteris M. Kirousis, Dieter Mitsche, and Xavier Pérez-Giménez. On the satisfiability threshold of formulas with three literals per clause. *Theoretical Computer Science*, 410(30-32):2920–2934, 2009.

19 Jian Ding, Allan Sly, and Nike Sun. Proof of the Satisfiability Conjecture for Large K. In *47th Symp. Theory of Computing (STOC)*, pages 59–68, 2015.

20 Tobias Friedrich, Anton Krohmer, Ralf Rothenberger, Thomas Sauerwald, and Andrew M. Sutton. Bounds on the Satisfiability Threshold for Power Law Distributed Random SAT. In *25th European Symposium on Algorithms (ESA)*, pages 37:1–37:15, 2017.

21 Tobias Friedrich, Anton Krohmer, Ralf Rothenberger, and Andrew M. Sutton. Phase Transitions for Scale-Free SAT Formulas. In *31st Conf. Artificial Intelligence (AAAI)*, pages 3893–3899, 2017.

22 Tobias Friedrich and Ralf Rothenberger. Sharpness of the Satisfiability Threshold for Non-uniform Random k-SAT. In *21st Intl. Conf. Theory and Applications of Satisfiability Testing (SAT)*, pages 273–291, 2018.

23 Jesús Giráldez-Cru and Jordi Levy. A Modularity-Based Random SAT Instances Generator. In *24 thIntl. Joint Conf. Artificial Intelligence (IJCAI)*, pages 1952–1958, 2015.

24 Andreas Goerdt. A Threshold for Unsatisfiability. *J. Comput. Syst. Sci.*, 53(3):469–486, 1996.

25 Mohammad Taghi Hajiaghayi and Gregory B. Sorkin. The Satisfiability Threshold of Random 3-SAT is at Least 3.52. Technical Report RC22942, IBM, October 2003.

26 Russell Impagliazzo and Ramamohan Paturi. On the Complexity of k-SAT. *J. Comput. Syst. Sci.*, 62(2):367–375, 2001.

27 Russell Impagliazzo, Ramamohan Paturi, and Francis Zane. Which Problems Have Strongly Exponential Complexity? In *39th Symp. Foundations of Computer Science (FOCS)*, pages 653–663, 1998.

28 Alexis C. Kaporis, Lefteris M. Kirousis, and Efthimios G. Lalas. The probabilistic analysis of a greedy satisfiability algorithm. *Random Struct. Algorithms*, 28(4):444–480, 2006.

29 Richard M. Karp. Reducibility Among Combinatorial Problems. In *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.*, pages 85–103, 1972.

30 Leonid A. Levin. Universal sorting problems. *Problems of Information Transmission*, 9:265–266, 1973.

31 Jordi Levy. Percolation and Phase Transition in SAT. *CoRR*, abs/1708.06805, 2017. `arXiv:1708.06805`.

32 Marc Mézard, Giorgio Parisi, and Riccardo Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297(5582):812–815, 2002.

33 David G. Mitchell, Bart Selman, and Hector J. Levesque. Hard and Easy Distributions of SAT Problems. In *10th Conf. Artificial Intelligence (AAAI)*, pages 459–465, 1992.

34 Rémi Monasson and Riccardo Zecchina. Statistical mechanics of the random K-satisfiability model. *Phys. Rev. E*, 56:1357–1370, August 1997.

35 Rémi Monasson, Riccardo Zecchina, Scott Kirkpatric, Bart Selman, and Lidror Troyansky. Phase transition and search cost in the 2+ p-sat problem. *4th Workshop on Physics and Computation, Boston, MA, 1996.*, 1996.

36 Rémi Monasson, Riccardo Zecchina, Scott Kirkpatrick, Bart Selman, and Lidror Troyansky. 2+p-SAT: Relation of typical-case complexity to the nature of the phase transition. *Random Struct. Algorithms*, 15(3-4):414–435, 1999.

37 Nathan Mull, Daniel J. Fremont, and Sanjit A. Seshia. On the Hardness of SAT with Community Structure. In *19th Intl. Conf. Theory and Applications of Satisfiability Testing (SAT)*, pages 141–159, 2016.

38 Vishwambhar Rathi, Erik Aurell, Lars K. Rasmussen, and Mikael Skoglund. Bounds on Threshold of Regular Random k-SAT. In *13th Intl. Conf. Theory and Applications of Satisfiability Testing (SAT)*, pages 264–277, 2010.