Satisfiability Threshold for Power Law Random 2-SAT in Configuration Model*

Oleksii Omelchenko[1] and Andrei A. Bulatov[1]

Simon Fraser University {oomelche,abulatov}@sfu.ca

Abstract. The Random Satisfiability problem has been intensively studied for decades. For a number of reasons the focus of this study has mostly been on the model, in which instances are sampled uniformly at random from a set of formulas satisfying some clear conditions, such as fixed density or the probability of a clause to occur. However, some non-uniform distributions are also of considerable interest. In this paper we consider Random 2-SAT problems, in which instances are sampled from a wide range of non-uniform distributions.
The model of random SAT we choose is the so-called configuration model, given by a distribution ξ for the degree (or the number of occurrences) of each variable. Then to generate a formula the degree of each variable is sampled from ξ, generating several *clones* of the variable. Then 2-clauses are created by choosing a random paritioning into 2-element sets on the set of clones and assigning the polarity of literals at random.
Here we consider the random 2-SAT problem in the configuration model for power-law-like distributions ξ. More precisely, we assume that ξ is such that its right tail $F_\xi(x)$ satisfies the conditions $W\ell^{-\alpha} \leq F_\xi(\ell) \leq V\ell^{-\alpha}$ for some constants V, W. The main goal is to study the satisfiability threshold phenomenon depending on the parameters α, V, W. We show that a satisfiability threshold exists and is determined by a simple relation between the first and second moments of ξ.

Keywords: Satisfiability, power law, phase transition

1 Introduction

The Random Satisfiability problem (Random SAT) and its special cases Random k-SAT as a model of 'typical case' instances of SAT has been intensively studied for decades. Apart from algorithmic questions related to the Random SAT, much attention has been paid to such problems as satisfiability thresholds and the structure of the solution space. The most widely studied model of the Random k-SAT is the uniform one parametrized by the (expected) density or clause-to-variable ratio ϱ of input formulas. Friedgut in [26] proved that depending on the parameter ϱ (and possibly the number of variables) Random k-SAT exhibits a sharp satisfiability threshold: a formula of density less than a certain value ϱ_0 (or possibly $\varrho_0(n)$) is satisfiable with high probability, and if the density

* This work was supported by an NSERC Discovery grant

is greater than ϱ_0, it is unsatisfiable with high probability. Moreover, a recent work of Friedrich and Rothenberger [28], which may be regarded as an extension of Fridgut's result to non-uniform random SAT instances, shows that if a distribution of variable's occurence in random formulas satisfies some criteria, then such formulas must undergo a sharp satisfiability threshold.

Hence, an impressive line of research aims at locating the satisfiability threshold for each random generating model. This includes more and more sophisticated methods of algorithms analysis [1, 17, 18, 22] and applications of the second moment method [2] to find lower bounds, and a variety of probabilistic and proof complexity tools to obtain upper bounds [23, 33, 24]. In the case of sufficiently large k the exact location of the satisfiability threshold was identified by Ding, Sly and Sun [22]. The satisfiability threshold and the structure of random k-CNFs received special attention for small values of k, see [15, 40, 30] for $k = 2$, and [21, 31, 33] for $k = 3$.

The satisfiability threshold phenomenon turned out to be closely connected with algorithmic properties of the Random SAT, as well as with the structure of its solution space. Experimental and theoretical results [39, 19] demonstrate that finding a solution or proving unsatisfiability is hardest around the satisfiability threshold. The geometry of the solution space also exhibits phase transitions not far from the satisfiability threshold, related to various clustering properties [35]. This phenomenon has been exploited by applications of methods from statistical physics that resulted in some of the most efficient algorithms for Random SAT with densities around the satisfiability threshold [36, 14].

Random k-SAT can be formulated using one of the three models whose statistical properties are very similar. In the model with fixed density ϱ, one fixes n distinct propositional variables $v_1, \dots, v_n$ and then chooses ϱn k-clauses uniformly at random [25, 39]. Alternatively, for selected variables every possible k-clause is included with probability tuned up so that the expected number of clauses equals ϱn. Finally, Kim [32] showed that one can also use the configuration model, which he called Poisson Cloning model. In this model for each variable v_i we first select a positive integer d_i accordingly to the Poisson distribution with expectation $k\varrho$, the degree of the variable. Then we create d_i *clones* of variable v_i, and choose $(d_1 + \cdots + d_n)/k$ random k-element subsets of the set of clones, then converting them into clauses randomly. The three models are largely equivalent and can be used whichever suits better to the task at hand.

The configuration model opens up a possibility for a wide range of different distributions of k-CNFs arising from different degree distributions. Starting with any random variable ξ that takes positive integer values one obtains a distribution $\Phi(\xi)$ on k-CNFs as above using ξ in place of the Poisson distribution. Note that ξ may depend on n, the number of variables, and even be different for different variables. One 'extreme' case of such a distribution is Poisson Cloning described above. Another case is studied by Cooper, Frieze, and Sorkin [20]. In their case each variable of a 2-SAT instance has a prescribed degree, which can be viewed as assigning a degree to every variable according to a random variable that only takes one value. We will be often returning to that paper, as our criterion for a satisfiability threshold is a generalization of that in [20]. Boufkhad et al. [13] considered another case of this kind — regular Random k-SAT.

In this paper we consider Random 2-SAT in the configuration model given by distribution $\Phi(\xi)$, where ξ is distributed according to the power law distribution in the following sense. Let $F_\xi(\ell) = \Pr[\xi \geq \ell]$ denote the tail function of a positive integer valued random variable ξ. We say that ξ is distributed according to the power law with parameter α if there exist constants V, W such that

$$W\ell^{-\alpha} \leq F_\xi(\ell) \leq V\ell^{-\alpha}. \tag{1}$$

Power law type distributions have received much attention. They have been widely observed in natural phenomena [37, 16], as well as in more artificial structures such as networks of various kinds [10]. Apart from the configuration model, graphs (and therefore 2-CNFs) whose degree sequences are distributed accordingly to a power law of some kind can also be generated in a number of ways. These include preferential attachment [3, 10, 12, 11], hyperbolic geometry [34], and others [5, 6]. Although the graphs resulting from all such processes satisfy the power law distributions of their degrees, other properties can be very different. We will encounter the same phenomenon in this paper.

The approach most closely related to this paper was suggested by Ansotegui et al. [5, 6]. Given the number of variables n, the number of clauses m, and a parameter β, the first step in their construction is to create m k-clauses without naming the variables. Then for every variable-place X in every clause, X is assigned to be one of the variables $v_1, \ldots, v_n$ according to the distribution

$$\Pr[X = v_i, \beta, n] = \frac{i^{-\beta}}{\sum_{j=1}^{n} j^{-\beta}}.$$

Ansotegui et al. argue that this model often well matches the experimental results on industrial instances, see also [4, 29, 8]. Interesting to note that although the model studied in these papers differ from the configuration model, it exhibits the same criterion of unsatisfiability $\mathbb{E}K^2 > 3\mathbb{E}K$, where K is the r.v. that governs the number of times a variable appears in 2-SAT formula ϕ [7].

The satisfiability threshold of this model has been studied by Friedrich et al. in [27]. Since the model has two parameters, β and $r = m/n$, the resulting picture is complicated. Friedrich et al. proved that a random CNF is unsatisfiable with high probability if r is large enough (although constant), and if $\beta < \frac{2k-1}{k-1}$. If $\beta \geq \frac{2k-1}{k-1}$, the formula is satisfiable with high probability provided r is smaller than a certain constant. The unsatisfiability results in [27] are mostly proved using the local structure of a formula.

In this paper we aim at a similar result for Random 2-SAT in the configuration model. Although the configuration model has only one parameter, the overall picture is somewhat more intricate, because there are more reasons for unsatisfiability than just the local structure of a formula. We show that for 2-SAT the parameter α from the tail condition (1) is what decides the satisfiability of such CNF. The main result of this paper is a satisfiability threshold given by the following

Theorem 1. *Let ϕ be a random 2-CNF in the configuration model, such that the number of occurrences of each variable in ϕ is an independent copy of the random variable ξ, satisfying the tail condition (1) for some α. Then for $n \to \infty$*

$$\Pr[\,\phi \text{ is satisfiable}\,] = \begin{cases} 0, & \text{when} \quad 0 < \alpha < 2 \\ 0, & \text{when} \quad \alpha = 2 \text{ or } \mathbb{E}\xi^2 > 3\mathbb{E}\xi, \\ 1, & \text{when} \quad \mathbb{E}\xi^2 < 3\mathbb{E}\xi. \end{cases}$$

In the first case of Theorem 1 we show that ϕ is unsatisfiable with high probability due to very local structure of the formula, such as the existence of variables of sufficiently high degree. Moreover, same structures persist with high probability in k-CNF formulas for *any* $k \geq 2$ obtained from the configuration model, when $\alpha < \frac{k}{k-1}$.

In the remaining cases we apply the approach of Cooper, Frieze, and Sorkin [20]. It makes use of the structural characterization of unsatisfiable 2-CNFs: a 2-CNF is unsatisfiable if and only if it contains so-called *contradictory paths*. If $\mathbb{E}\xi^2 < 3\mathbb{E}\xi$ we prove that w.h.p. formula ϕ does not have long paths, and contradictory paths are unlikely to form. If $\mathbb{E}\xi^2 > 3\mathbb{E}\xi$, we use the analysis of the dynamics of the growth of ϕ to show that contradictory paths appear w.h.p. However, the original method by Cooper et al. only works with strong restrictions on the maximal degree of variables that are not affordable in our case, and so it requires substantial modifications.

2 Notation and preliminaries

We use the standard terminology and notation of variables, positive and negative literals, clauses and 2-CNFs, and degrees of variables. The degree of variable v will be denoted by $deg(v)$, or when our CNF contains only variables $v_1, \ldots, v_n$, we use $d_i = deg(v_i)$. By $C(\phi)$ we denote the set of clauses in ϕ.

2.1 Configuration model

We describe the configuration model for k-CNFs, but will only use it for $k = 2$, see also [32]. In the configuration model of k-CNFs with n variables $v_1, \ldots, v_n$ we are a given a positive integer-valued random variable (r.v.) ξ from which we sample independently n integers $\{d_i\}_{i=1}^n$. Then d_i is the degree of v_i, that is, the number of occurrences of v_i in the resulting formula ϕ. Each occurrence of v_i in ϕ we call a *clone* of v_i. Hence, d_i is the number of clones of v_i. Then we sample k-element sets of clones from the set of all clones without replacement. Finally, every such subset is converted into a clause by choosing the polarity of every clone in it uniformly at random. If the total number of clones is not a multiple of k, we discard the set and repeat the procedure. Algorithm 1 gives a more precise description of the process. We will sometimes say that a clone p is associated with variable v if p is a clone of v. In a similar sense we will say a clone associated with a literal if we need to emphasize the polarity of the clone.

Algorithm 1 Configuration Model $\mathbb{C}_n^k(\xi)$

```
1: procedure SampleCNF(n, k, ξ)
2:     Form a sequence of n numbers {d_i}_{i=1}^n each sampled independently from ξ
3:     if S_n := Σ_{i=1}^n d_i is not a multiple of k then
4:         discard the sequence, and go to step 2
5:     end if
6:     Otherwise, introduce multi-set S ← ⋃_{i=1}^n {v_i, v_i, ..., v_i} (d_i times)
7:     Let φ ← ∅
8:     while S ≠ ∅ do
9:         Pick u.a.r. k elements {v_1, v_2, ..., v_k} from S without replacement
10:        Let C ← {v_1, v_2, ..., v_k}
11:        S ← S − C
12:        Negate each element in C u.a.r with probability 1/2
13:        φ ← φ ∪ C
14:    end while
15:    return φ
16: end procedure
```

We will denote a random formula ϕ obtained from $\mathbb{C}_n^k(R)$ by $\phi \sim \mathbb{C}_n^k(R)$. Clearly, formulas $\phi \sim \mathbb{C}_n^k(R)$ are defined over a set of n Boolean variables, which we denote by $V(\phi)$. By $L(\phi)$ we denote the set of all literals in ϕ. Let d_i^+ denote the number of occurrences of v_i as a positive literal (or the number of positive clones of v_i), and let d_i^- denote the number of negative clones of v_i.

2.2 Power law distributions

We focus our attention on the configuration model $\mathbb{C}_n^k(\xi)$, in which every variable is an i.i.d. copy of the random variable ξ having power-law distribution. In this paper we define such distributions through the properties of their tail functions. If ξ is an integer-valued r.v., its tail function is defined to be $F_\xi(\ell) = \Pr[\xi \geq \ell]$, where $\ell \geq 1$.

Definition 1. *An integer-valued positive r.v. ξ has power-law probability distribution, if $F_\xi(\ell) = \Theta\left(\ell^{-\alpha}\right)$, where $\alpha > 0$. We denote this fact as $\xi \sim \mathcal{P}(\alpha)$.*

Clearly, if $\xi \sim \mathcal{P}(\alpha)$, then there exist constants $V, W > 0$, such that $W\,\ell^{-\alpha} \leq F_\xi(\ell) \leq V\,\ell^{-\alpha}$, for every $\ell \geq 1$.

The existence of the moments of $\xi \sim \mathcal{P}(\alpha)$ depend only on α.

Lemma 1. *Let $\xi \sim \mathcal{P}(\alpha)$. Then $\mathbb{E}\xi^m < \infty$ iff $0 < m < \alpha$.*

We will write $\mathbb{E}\xi^m = \infty$ when the m-th moment of some r.v. ξ is not finite or does not exist. We will have to deal with cases when the second or even first moment of ξ does not exist.

Nevertheless, we can obtain good bounds on useful quantities formed from such variables with a good level of confidence, despite the absence of expectation or variance. One such quantity is the sum of *independent* variables drawn from $\mathcal{P}(\alpha)$: $S_n = \sum_{i=1}^{n} \xi_i$, where $\xi_i \sim \mathcal{P}(\alpha)$. Note that ξ_i's are not required to be identically distributed. They can come from different distributions, as long as their right tail can be bounded with some power-law functions with exponent α. But we do require their independence.

The next two theorems provide bounds on the values of S_n, depending on α in a slightly more general case of r.vs. admitting negative values.

Theorem 2 (Corollary 1 from [38]). *Let $S_n = \sum_{i=1}^{n} \xi_i$, where ξ_i's are independent integer-valued random variables, with*

$$\Pr[\xi_i \geq \ell] \leq V\,\ell^{-\alpha}, \quad \textit{and} \quad \Pr[\xi_i \leq -\ell] \leq V\,\ell^{-\alpha},$$

where $V > 0$ and $0 < \alpha \leq 1$ are constants. Then w.h.p. $S_n \leq C\,n^{\frac{1}{\alpha}}$, where $C > 0$ is some constant.

As for the second theorem, we deal with a similar sum of random variables, but each variable's tail can be majorized with a power-law function with exponent $\alpha > 1$. Then, as it follows from Lemma 1, such variables have finite expectation, and due to the linearity of expectation, the sum itself has well defined mean value.

Theorem 3 (Corollary 5 from [38]). *Let $S_n = \sum_{i=1}^{n} \xi_i$, where ξ_i's are independent integer-valued random variables, with*

$$\Pr[\xi_i \geq \ell] \leq V\,\ell^{-\alpha}, \quad \textit{and} \quad \Pr[\xi_i \leq -\ell] \leq V\,\ell^{-\alpha},$$

where $V > 0$ and $\alpha > 1$ are constants. Then w.h.p. $S_n = \sum_{i=1}^{n} \mathbb{E}\xi_i + o(n)$.

Hence, as the theorem states, when ξ_i's are independent r.vs. with power-law boundable tails with tail exponent $\alpha > 1$, then the sum of such variables does not deviate much from its expected value.

Note that from now on we will deal with sctrictly positive power-law r.vs. ξ_i's, hence, their expectation (given that it exists) is a positive constant. Then when $\alpha > 1$, we have $S_n = \sum_{i=1}^{n} \mathbb{E}\xi_i + o(n) = (1 + o(1)) \sum_{i=1}^{n} \mathbb{E}\xi_i$.

Another important quantity we need is the maximum, Δ, of the sequence of n independent random variables (or the maximum degree of a CNF in our case).

Lemma 2. *Let $\Delta = \max(\xi_1, \xi_2, \cdots, \xi_n)$, where ξ_i's are independent copies of an r.v. $\xi \sim \mathcal{P}(\alpha)$ with $\alpha > 0$. Then w.h.p. $\Delta \leq C\,n^{1/\alpha}$, where $C > 0$ is some constant.*

We will also need some bounds on the number of pairs of complementary clones of a variable v_i, that is, the value $d_i^+ d_i^-$. By the definition of the configuration model

$$d_i^+ \sim Bin\Big(\deg(v_i), 1/2\Big) \quad \text{and} \quad d_i^- = \deg(v_i) - d_i^+,$$

where $Bin(n, p)$ is the binomial distribution with n trials and the success probability p.

Lemma 3. *Let ξ be some positive integer-valued r.v., and let $d^+ \sim Bin\Big(\xi, 1/2\Big)$, while $d^- = \xi - d^+$. Then*

$$F_{d^+}(\ell) = F_{d^-}(\ell) \le F_\xi(\ell), \quad \textit{and} \quad F_{d^+d^-}(\ell) \le 2F_\xi\left(\ell^{1/2}\right).$$

Hence, for $\xi \sim \mathcal{P}(\alpha)$, we have

Corollary 1. *Let $\xi \sim \mathcal{P}(\alpha)$, where $\alpha > 0$, be some positive integer-valued r.v., and let $d^+ \sim Bin\Big(\xi, 1/2\Big)$, while $d^- = \xi - d^+$. Then*

$$F_{d^+}(\ell) = F_{d^-}(\ell) \le V\,\ell^{-\alpha}, \tag{2}$$

$$F_{d^+d^-}(\ell) \le 2V\,\ell^{-\alpha/2}. \tag{3}$$

The expectations of d^+ and d^- are easy to find: $\mathbb{E}d^+ = \mathbb{E}d^- = \frac{\mathbb{E}\xi}{2}$. However, the expected value of d^+d^- requires a little more effort.

Lemma 4. *Let ξ be some positive integer-valued r.v., and let $d^+ \sim Bin\Big(\xi, 1/2\Big)$, while $d^- = \xi - d^+$. Then $\mathbb{E}\left[d^+d^-\right] = \frac{\mathbb{E}\xi^2 - \mathbb{E}\xi}{4}$.*

We use $T_n = \sum_{i=1}^n d_i^+ d_i^-$ to denote the total number of pairs of complementary clones, i.e. the sum of unordered pairs of complementary clones over all n variables,.

Note, that when $\alpha > 2$, the r.v. $d_i^+ d_i^-$ has finite expectation due to Lemma 1. Then by Theorem 3 w.h.p.

$$T_n = (1 + o(1)) \sum_{i=1}^n \mathbb{E}\left[d_i^+ d_i^-\right].$$

We finish this subsection with *Azuma-like inequality* first appeared in [20], which will be used in the proofs. Informally, the inequality states that a discrete-time random walk $X = \sum_{i=1}^n X_i$ with positive drift, consisting of not necessary independent steps, each having a right tail, which can be bounded by a power function with exponent at least 1, is very unlikely to drop much below the expected level, given n is large enough. Although, the original proof was relying on the rather artificial step of introducing a sequence of uniformly distributed random numbers, we figured out that the same result can be obtained by exploiting the tower property of expectation.

Lemma 5 (Azuma-like inequality). *Let $X = X_0 + \sum_{i=1}^t X_i$ be some random walk, such that $X_0 \ge 0$ is constant initial value of the process, $X_i \ge -a$, where $a > 0$ is constant, are bounded from below random variables, not necessary independent, and such that $\mathbb{E}[X_i \,|\, X_1, \ldots, X_{i-1}] \ge \mu > 0$ (μ is constant) and $\Pr[X_i \ge \ell \,|\, X_1, \ldots, X_{i-1}] \le V\,\ell^{-\alpha}$ for every $\ell \ge 1$ and constants $V > 0$, $\alpha > 1$. Then for any $0 < \varepsilon < \frac{1}{2}$, the following inequality holds*

$$\Pr\left[X \le \varepsilon \mu t\right] \le \exp\left(-\frac{t + X_0}{4\log^2 t}\mu^2\left(\frac{1}{2} - \varepsilon\right)^2\right).$$

2.3 Contradictory paths and bicycles

Unlike k-CNFs for larger values of k, 2-CNFs have a clear structural feature that indicates whether or not the formula is satisfiable. Let ϕ be a 2-CNF on variables $v_1, \ldots, v_n$. A sequence of clauses $(l_1, l_2), (\bar{l}_2, l_3), \ldots, (\bar{l}_{s_1}, l_s)$ is said to be a *path* from literal l_1 to literal l_s. As is easily seen, if there are variables u, v, w in ϕ such that there are paths from u to v and $\bar{v}$, and from $\bar{u}$ to w and $\bar{w}$, then ϕ is unsatisfiable, see also [9]. Such a collection of paths is sometimes called *contradictory paths.*

On the other hand, if ϕ is unsatisfiable, it has to contain a *bicycle*, see [15]. A bicycle of length s is a path $(u, l_1), (\bar{l}_1, l_2), \ldots, (\bar{l}_s, v)$, where the variables associated with literals $l_1, l_2, \ldots, l_s$ are distinct, and $u, v \in \{l_1, \bar{l}_1, l_2, \bar{l}_2, \ldots, l_s, \bar{l}_s\}$.

2.4 The main result

Now we are ready to state our main result:

Theorem 4. *Let $\phi \sim \mathbb{C}_n^2(\xi)$, where $\xi \sim \mathcal{P}(\alpha)$. Then for $n \to \infty$*

$$\Pr[\,\phi \textit{ is SAT}\,] = \begin{cases} 0, & \textit{when} \quad 0 < \alpha < 2, \\ 0, & \textit{when} \quad \alpha = 2 \textit{ or } \mathbb{E}\xi^2 > 3\mathbb{E}\xi, \\ 1, & \textit{when} \quad \mathbb{E}\xi^2 < 3\mathbb{E}\xi. \end{cases}$$

If the r.v. ξ is distributed according to the zeta distribution, that is, $\Pr[\xi = \ell] = \frac{\ell^{-\beta}}{\zeta(\beta)}$ for some $\beta > 1$ and where $\zeta(\beta) = \sum_{d \geq 1} d^{-\beta}$ is the Riemann zeta function (note that in this case $\xi \sim \mathcal{P}(\beta - 1)$), then the satisfiabitliy threshold is given by a certain value of β.

Corollary 2. *Let $\phi \sim \mathbb{C}_n^2(\xi)$, where the pdf of ξ is $\Pr[\xi = \ell] = \frac{\ell^{-\beta}}{\zeta(\beta)}$ for some $\beta > 1$ and all $\ell \geq 1$. Then there exists β_0 such that for $n \to \infty$*

$$\Pr[\,\phi \textit{ is SAT}\,] = \begin{cases} 0, & \textit{when} \quad 1 < \beta < \beta_0, \\ 1, & \textit{when} \quad \beta > \beta_0. \end{cases}$$

The value β_0 is the positive solution of the equation $\mathbb{E}\xi^2 = 3\mathbb{E}\xi$, and $\beta_0 \approx 3.26$.

A proof of this theorem constitutes the rest of the paper. We consider each case separately, and the first case is proved in Proposition 1, while the other two cases are examined in Propositions 2 and 3.

3 Satisfiability of $\mathbb{C}_n^2(\xi)$, when $\xi \sim \mathcal{P}(\alpha)$ and $0 < \alpha < 2$

This case is the easiest to analyze. Moreover, we show that the same result holds for any $\phi \sim \mathbb{C}_n^k(\xi)$, where $k \geq 2$, when $\alpha < \frac{k}{k-1}$. Hence, the case $0 < \alpha < 2$ for unsatisfiable 2-CNFs follows. In other words, if $\alpha < \frac{k}{k-1}$, then *any* k-CNF formulas from $\mathbb{C}_n^k(\xi)$ will be unsatisfiable w.h.p.

What happens here, is that we expect many variables to have degree $\gg S_n^{(k-1)/k}$. Let us fix k such variables. Then, as it is shown in the proof, the formula ϕ contains $(k-1)!\log^k n$ clauses that are formed only from literals of these k variables. However, one of the possible subformulas, which is formed from only k variables, that renders the whole k-CNF formula unsatisfiable consists of only 2^k clauses.

The next proposition establishes a lower bound of satisfiability threshold for any power-law distributed k-CNF from configuration model.

Proposition 1. *Let $\phi \sim \mathbb{C}_n^k(\xi)$, where $\xi \sim \mathcal{P}(\alpha)$, $k \geq 2$ and $0 < \alpha < \frac{k}{k-1}$. Then w.h.p. ϕ is unsatisfiable.*

After proving the above proposition, result for 2-CNF from $\mathbb{C}_n^2(\xi)$ naturally follows.

Corollary 3. *Let $\phi \sim \mathbb{C}_n^2(\xi)$, where $\xi \sim \mathcal{P}(\alpha)$, such that $0 < \alpha < 2$. Then w.h.p. ϕ is unsatisfiable.*

4 Satisfiability of $\mathbb{C}_n^2(\xi)$, when $\xi \sim \mathcal{P}(\alpha)$ and $\alpha = 2$ or $\mathbb{E}\xi^2 > 3\mathbb{E}\xi$

4.1 The inequality $\mathbb{E}\xi^2 > 3\mathbb{E}\xi$

Analysis of this and subsequent cases mainly follows the approach suggested in [20], where they deal with random 2-SAT instances having prescribed literal degrees. In other words, the assumption in [20] is that the degree sequences $d_1^+, \dots, d_n^+$ and $d_1^-, \dots, d_n^-$ are fixed, and a random 2-CNF is generated as in the configuration model. Then two quantities play a very important role. The first one is the sum of all degrees $S_n = \sum_{i=1}^n (d_i^+ + d_i^-)$ (we use our notation) and the second one is the number of pairs of complementary clones $T_n = \sum_{i=1}^n d_i^+ d_i^-$. It is then proved that a 2-CNF with a given degree sequence is satisfiable w.h.p. if and only if $2T_n < (1-\varepsilon)S_n$ for some $\varepsilon > 0$. We will quickly show that the conditions $\alpha = 2$ and $\mathbb{E}\xi^2 > 3\mathbb{E}\xi$ imply the inequality $2T_n > (1+\varepsilon)S_n$ w.h.p., see Lemma 6, and therefore a random 2-CNF in this case should be unsatisfiable w.h.p. The problem however is that Cooper et al. only prove their result under a significant restrictions on the maximal degree of literals, $\Delta < n^{1/11}$. By Lemma 2 the maximal degree of literals in our case tends to be much higher, and we cannot directly utilize the result from [20]. Therefore we follow the main steps of the argument in [20] changing parameters, calculations, and in a number of cases giving a completely new proofs.

Lemma 6. *Let $\phi \sim \mathbb{C}_n^2(\xi)$, where $\xi \sim \mathcal{P}(\alpha)$ and $\alpha = 2$ or $\mathbb{E}\xi^2 > 3\mathbb{E}\xi$. Let also $S_n = \sum_{i=1}^n d_i$ and $T_n = \sum_{i=1}^n d_i^+ d_i^-$. Then w.h.p. $2T_n > (1+\varepsilon)S_n$.*

Proof. Let us first consider the case, when $\alpha > 2$ and $\mathbb{E}\xi^2 > 3\mathbb{E}\xi$. Then by Lemma 1 and Theorem 3, we have that w.h.p.

$$S_n = \sum_{i=1}^n d_i = (1+o(1)) \sum_{i=1}^n \mathbb{E}d_i = (1+o(1))\, n\mathbb{E}\xi,$$

since $d_i \stackrel{d}{=} \xi$. Likewise, since $\alpha > 2$, we also have that w.h.p.

$$T_n = \sum_{i=1}^{n} d_i^+ d_i^- = (1+o(1)) \sum_{i=1}^{n} \mathbb{E}\left[d_i^+ d_i^-\right] = (1+o(1))\, n \frac{\mathbb{E}\xi^2 - \mathbb{E}\xi}{4},$$

where the last equality follows from Lemma 4.

Hence, when $\mathbb{E}\xi^2 > 3\mathbb{E}\xi$, we have that w.h.p.

$$\frac{2T_n}{S_n} = (1 \pm o(1)) \frac{\mathbb{E}\xi^2 - \mathbb{E}\xi}{2\mathbb{E}\xi} = (1 \pm o(1)) \left(\frac{\mathbb{E}\xi^2}{2\mathbb{E}\xi} - \frac{1}{2} \right) > 1.$$

Now we consider the case $\alpha = 2$. Unfortunately, then $\mathbb{E}\left[d_i^+ d_i^-\right] = \infty$ for any $i \in [1 \ldots n]$, and so we cannot claim that T_n is concentrated around its mean. Nevertheless, the quantity $\frac{2T_n}{S_n}$ is still greater than 1 in this case.

Since $\xi \sim \mathcal{P}(2)$, there are constants V, W such that $W\,\ell^{-2} \le F_\xi(\ell) \le V\,\ell^{-2}$. We construct auxiliary random variables $\xi_\varepsilon \sim \mathcal{P}(2+\varepsilon)$ for $\varepsilon > 0$. Later we will argue that ξ_ε can be chosen such that $\mathbb{E}\xi_\varepsilon^2 > 3\mathbb{E}\xi_\varepsilon$. Specifically, let ξ_ε be such that $F_{\xi_\varepsilon}(1) = 1$ and $F_{\xi_\varepsilon}(\ell) = W\,\ell^{-2-\varepsilon}$ for $\ell > 1$.

Let T_n^ε be the number of pairs of complementary clones in formula $\phi_0 \sim \mathbb{C}_n^2(\xi_\varepsilon)$. Since $\Pr\left[\xi_\varepsilon \ge \ell\right] \le \Pr\left[\xi \ge \ell\right]$ for any $\ell \ge 1$, we have that

$$\Pr\left[2T_n > S_n\right] \ge \Pr\left[2T_n^\varepsilon > S_n\right], \tag{4}$$

due to the stochastic dominance of the r.v. T_n over T_n^ε. As is easily seen, for sufficiently small ε we have $\mathbb{E}\xi_\varepsilon^2 > 3\mathbb{E}\xi_\varepsilon$. Therefore, by the first part of the proof $2T_n^\varepsilon > S_n$ w.h.p. The result follows.

Thus, in either case we obtain that for some $\mu > 0$ w.h.p. $\frac{2T_n}{S_n} = 1 + \mu$.

In what follows, we will always assume that $\alpha > 2$.

4.2 TSPAN

The process of generating a random 2-CNF in the configuration model can be viewed as follows. After creating a pool of clones, we assign each clone a polarity, making it a clone of a positive or negative literal. Then we choose a random paritioning of the set of clones into 2-element sets. The important point here is that in the process of selection of a random matching we pair clones up one after another, and it does not matter in which order a clone to match is selected, as long as it is paired with a random unpaired clone.

Our goal is to show that our random 2-CNF ϕ contains contradictory paths. In order to achieve this we exploit the property above as follows. Starting from a random literal p we will grow a set *span*(p) of literals reachable from p in the sense of paths introduced in Section 2.3. This is done by trying to iteratively extend *span*(p) by pairing one the unpaired clones of the negation of a literal from *span*(p). The details of the process will be described later. The hope is that at some point *span*(p) contains a pair of literals of the form $v, \bar{v}$, and therefore

ϕ contains a part of the required contradictory paths. To obtain the remaining part we run the same process starting from $\bar{p}$.

To show that this approach works we need to prove three key facts:

– that *span*(p) grows to a certain size with reasonable probability (Lemma 10),

– that if *span*(p) has grown to the required size, it contains a pair $v, \bar{v}$ w.h.p. (Lemma 11), and

– that the processes initiated at p and $\bar{p}$ do not interact too much w.h.p. (Lemma 8).

Since the probability that *span*(p) grows to the required size is not very high, most likely this process will have to be repeated multiple times. It is therefore important that the probabilities above are estimated when some clones are already paired up, and that all the quantities involved are carefully chosen.

We now fill in some details. The basic "growing" algorithm is TSPAN (short for *truncated span*), see Algorithm 2. Take a literal and pick a clone p associated with it. Then partition the set $\mathcal{S}$ of all clones into 3 subsets: the set $\mathcal{L}(p)$ of "live" clones from which we can grow the span, the set $\mathcal{C}$ of paired (or "connected") clones, and the set $\mathcal{U}$ of "untouched" yet clones. We start with $\mathcal{L}(p) = \{p\}$, $\mathcal{U} = \mathcal{S} - \{p\}$, and empty set $\mathcal{C}$.

TSPAN works as follows: while the set of live clones is not empty, pick u.a.r. clone c_1 from the live set, and pair it u.a.r. with any non-paired clone $c_2 \in \mathcal{U} \cup \mathcal{L}(p) \setminus \{c_1\}$. Since clones c_1 and c_2 are paired now, we move them into the set of paired clones $\mathcal{C}$, while removing them from both sets $\mathcal{L}(p)$ and $\mathcal{U}$ to preserve the property that the sets $\mathcal{C}$,$\mathcal{U}$, and $\mathcal{L}(p)$ form a partition of $\mathcal{S}$.

Next, we identify the literal l which clone c_2 is associated with, and we move all the complementary clones of $\bar{l}$ from the set of untouched clones $\mathcal{U}$ into $\mathcal{L}(p)$. The idea of this step is, when we add an edge (c_1, c_2), where c_2 is one the l's clones, to grow the span further we will need to add another directed edge $(c_3, \cdot)$, where c_3 is one of the clones belonging to $\bar{l}$. Hence, we make all clones of $\bar{l}$ live, making them available to pick as a starting point during next iterations of TSPAN. This way we can grow a span, starting from the clone p, and then the set

$$span(p) = \{c \in \mathcal{S} \,|\, c \text{ is reachable from } p\},$$

contains all the clones, which are reachable from the clone p (or literal that is associated with p) at a certain iteration of TSPAN. We call this set a *p-span*.

The version of TSPAN given in Algorithm 2 takes as input sets $\mathcal{C}, \mathcal{L}, \mathcal{U}$ (which therefore do not have to be empty in the beginning of execution of the procedure), a maximal number of iterations τ, and a maximal size of the set of live clones. It starts by using the given sets, $\mathcal{C}, \mathcal{L}, \mathcal{U}$, stops after at most τ iterations or when $\mathcal{L}$ reaches size σ.

4.3 Searching for contradictory paths

The procedure TSPAN is used to find contradictory paths as follows:

Algorithm 2 Procedure TSPAN

procedure TSPAN($\mathcal{C}$,$\mathcal{L}$,$\mathcal{U}$, σ, τ)
  while $0 < |\mathcal{L}| \leq \sigma$ **and** less than τ pairings performed **do**
    Pick u.a.r. a live clone $c_1 \in \mathcal{L}$
    Pick u.a.r. an unpaired clone $c_2 \in \mathcal{U} \cup \mathcal{L} \setminus \{c_1\}$
    Pair clones c_1 and c_2, i.e.
      $\mathcal{C} \leftarrow \mathcal{C} \cup \{c_1, c_2\}$
      $\mathcal{L} \leftarrow \mathcal{L} \setminus \{c_1, c_2\}$
      $\mathcal{U} \leftarrow \mathcal{U} \setminus \{c_1, c_2\}$
    Let w be the literal associated with c_2
    Make live the clones associated with $\bar{w}$, i.e
    Let
      $\kappa(\bar{w}) = \{c \in \mathcal{S} \mid c \text{ is associated with } \bar{w}\}$
      $L \leftarrow L \cup (U \cap \kappa(\bar{w}))$
      $U \leftarrow U \setminus \kappa(\bar{w})$
  end while
end procedure

Step 1. Pick a variable and a pair of its complementary clones p, q.

Step 2. Run TSPAN starting from p for at most $s_1 = n^{\frac{\alpha+4}{6(\alpha+1)}}$ steps. If $\mathcal{L}(p)$ becomes empty during the process, or if q gets included into *span*(p), or if in the end $|\mathcal{L}(p)| < \sigma = s_1\mu/6$ (μ is determined by the value $2T_n/S_n$, see Lemma 7), declare failure.

Step 3. Run TSPAN starting from q and the current set $\mathcal{C}$ of paired clones for at most $s_1 = n^{\frac{\alpha+4}{6(\alpha+1)}}$ steps. If $\mathcal{L}(q)$ becomes empty during the process, or if $|\mathcal{L}(q) \cap \mathcal{L}(q)| = \Theta(s_1)$, or if in the end $|\mathcal{L}(q)| < \sigma$, declare failure.

Step 4. Run TSPAN starting from $\mathcal{L}(p)$ and the current set $\mathcal{C}$ of paired clones for at most $s_2 = n^{\frac{11\alpha^2+3\alpha-2}{12\alpha(\alpha+1)}}$ steps. If $\mathcal{L}(p)$ becomes empty during the process, declare failure.

Step 5. Similarly, run TSPAN starting from $\mathcal{L}(q)$ and the current set $\mathcal{C}$ of paired clones for at most $s_2 = n^{\frac{11\alpha^2+3\alpha-2}{12\alpha(\alpha+1)}}$ steps. If $\mathcal{L}(q)$ becomes empty during the process, declare failure.

If a failure is declared at any step, we abandon the current pair p, q and pick another variable and a pair of clones keeping the current set $\mathcal{C}$ of paired clones that will be used in the next round. Also, even if all the Steps are successful, but the constructed span does not contain contradictory paths, we also declare a failure. It is important that the set $\mathcal{C}$ never grows too large, that is, it remains of size $|\mathcal{C}| = o(n)$. This implies that the number of restarts does not exceed $K = n^{\frac{7\alpha+10}{12(\alpha+1)}}$.

The next lemma shows how we choose the value of μ. It also shows that we expect a positive change in the size of the live set when $\frac{2T_n}{S_n} > 1$. However, first, we need to introduce several variables. Let $\mathcal{L}_i, \mathcal{U}_i$, and $\mathcal{C}_i$ are the live, untouched, and connected sets respectively after the i-th iteration of some execution of TSPAN. Additionally we have $L_i = |\mathcal{L}_i|$, $C_i = |\mathcal{C}_i|$, $U_i = |\mathcal{U}_i|$. Also let X_i indicate the change in the number of live clones after performing the ith iteration, i.e. $X_i = L_i - L_{i-1}$.

Lemma 7. *Let* $\frac{2T_n}{S_n} = 1 + \mu$, *where* $\mu > 0$. *Then for any* $t \leq |\mathcal{C}| = o(n)$, *we have*

$$\mathbb{E}[X_t \mid X_1, \ldots, X_{t-1}] \geq \mu/2.$$

Next, we bound the probability of failure in each of Steps 2–5. We start with Step 2 assuming that the number of paired clones is $o(n)$.

Lemma 8 (Step 2). *(1) Let* $s_1 = n^{\frac{\alpha+4}{6(\alpha+1)}}$. *If* TSPAN *starts with a live set containing only a single point* $L_0 = 1$, *time bound* $\tau = s_1$, *the live set size bound* $\sigma = s_1\mu/6$, *and the number of already paired clones* $|\mathcal{C}| = o(n)$, *then with probability at least* $\frac{1}{2s_1}$ TSPAN *terminates with the live set of size at least* σ.

(2) For any fixed clone q, *the probability it will be paired in* $s_1 = n^{\frac{\alpha+4}{6(\alpha+1)}} \leq t = o(n)$ *steps of the algorithm, is at most* $o\left(\frac{1}{s_1}\right)$.

Note that in Lemma 8(1) the size of $\mathcal{L}(p)$ can be slightly greater than σ, as it may increase by more than 1 in the last iteration. Also, in Lemma 8(2) the bound on the probability is only useful when s_1 is sufficiently large.

Proof. We prove item (1) here. The TSPAN procedure may terminate at the moment $i < \tau$ due to one of two reasons: first, when L_i hits 0, and second, when $L_i = \sigma$. To simplify analysis of the lemma, instead of dealing with conditional probabilities that the live set hasn't paired all its clones, we suggest to use a slightly modified version of TSPAN, which *always* runs for τ steps.

The modified version works exactly as the original TSPAN procedure when the live set has at least one clone. But if at some moment, the live set has no clones to pick, we perform a "restart": we restore the sets $\mathcal{L}, \mathcal{C}$, and $\mathcal{D}$ to the states they'd been before the first iteration of TSPAN procedure occurred. After that we continue the normal pairing process. Although during restarts we reset the values of the sets, the counter that tracks the number of iterations the TSPAN has performed is never reset, and keeps increasing with every iteration until the procedure has performed pairings τ times, or the live set was able to grow up to size σ, and only then the TSPAN terminates.

Now, let $r_i = 1$ represents a "successfull" restart that started at i iteration, meaning during this restart the live set accumulated σ clones, while $r_i = 0$ means there was no restart or the live set became empty. What we are looking for $\Pr[r_1 = 1]$, since this probability is identical to the probability that the original TSPAN was able to grow the live set to the desired size. Next, we can have at most τ restarts, and, since the very first restart has the most time and we expect the live set to grow in the long run, it follows that it stochastically dominates over other r_i's. Thus,

$$\Pr[L_{s_1} \geq s_1\mu/6] \leq \Pr\left[\sum_{i=1}^{s_1} r_i \geq 1\right] \leq \mathbb{E}\sum_{i=1}^{s_1} r_i \leq s_1 \mathbb{E} r_1 = s_1 \Pr[r_1 = 1]$$

from which we obtain the probability that the TSPAN terminates with large enough live set from the very first try:

$$P := \Pr[r_1 = 1] \geq \frac{\Pr[L_{s_1} \geq s_1\mu/6]}{s_1}. \tag{5}$$

Now what is left is to obtain bounds on the right-hand side probability. We have a random process

$$L_{s_1} = \sum_{i=1}^{s_1} (L_i - L_{i-1}) = \sum_{i=1}^{s_1} X_i,$$

which consists of steps X_i, each of which can be proved to have the right tail bounded by $\Pr[X_i \geq \ell \,|\, X_1, \ldots, X_{i-1}] \leq V\,\ell^{-\alpha}$, and positive expectation (Lemma 7) $\mathbb{E}[X_i \,|\, X_1, \ldots, X_{i-1}] \geq \frac{\mu}{2}$.

Therefore, according to *Azuma-like inequality* (Lemma 5), we obtain that

$$\Pr[L_{s_1} \leq s_1\mu/6] = \Pr\left[L_{s_1} \leq \left(s_1\frac{\mu}{2}\right)\frac{1}{3}\right] \leq \exp\left(-\frac{s_1}{4\log^2 s_1}\frac{\mu^2}{576}\right)$$

Fixing $s_1 = n^{\frac{\alpha+4}{6(\alpha+1)}}$, we have for some constant $C > 0$

$$\Pr[L_{s_1} \leq s_1\mu/6] \leq \exp\left(-C\frac{n^{\frac{\alpha+4}{6(\alpha+1)}}}{\log^2 n}\right) = o(1) \leq 1/2.$$

Thus, from (5) follows $P \geq \frac{\Pr[L_{s_1} \geq s_1\mu/6]}{s_1} = \frac{1}{2s_1}$, which proves item (1) of the lemma. □

The probability that both runs of TSPAN for p and q are successful is given by the following

Lemma 9 (Step 3). *The probability that two specific clones p and q accumulate $s_1\mu/7$ clones in their corresponding live sets $\mathcal{L}$ during the execution of* Steps 2,3, *such that the span from clone p doesn't include q nor make it live, is at least $\frac{1}{5s_1^2}$.*

Next, we show that we can grow the spans for another s_2 steps, while keeping the size of the respective live sets of order at least $s_1\mu/8$.

Lemma 10 (Steps 4,5). *Assume that p- and q-spans were both able to accumulate at least $s_1\mu/8$ live clones after $s_1 = n^{\frac{\alpha+4}{6(\alpha+1)}}$ steps, and q is not in the p-span. Then with probability $1 - o(1)$,* TSPAN *will be able to perform another $s_2 = n^{\frac{11\alpha^2+3\alpha-2}{12\alpha(\alpha+1)}}$ iterations, and $L_{s_1+j} \geq s_1\mu/8$ for every $0 \leq j \leq s_2$ for each clone p and q.*

Finally, we show that w.h.p. the spans produced in Steps 1–5, provided no failure occurred, contain contradictory paths. In other words, we are looking for the probability that spans do not contain complement clones after growing them for $s_1 + s_2$ steps, i.e. at each step TSPAN was choosing only untouched clones from the set $\mathcal{U}$.

Lemma 11 (Contradiction paths). *If for a pair of complementary clones p, q Steps 1–5 are completed successfully, the probability that span(p) or span(q) contain no 2 complementary clones is less than* $\exp\left(-n^{\frac{\alpha^2-\alpha-2}{12\alpha(\alpha+1)}}\right)$.

This completes the proof in the case $\alpha = 2$ or $\mathbb{E}\xi^2 > 3\mathbb{E}\xi$, and the next proposition summarizes the result.

Proposition 2. *Let $\phi \sim \mathbb{C}_n^2(\xi)$, where $\xi \sim \mathcal{P}(\alpha)$ and $\alpha = 2$ or $\mathbb{E}\xi^2 > 3\mathbb{E}\xi$. Then w.h.p. ϕ is unsatisfiable.*

5 Satisfiability of $\mathbb{C}_n^2(\xi)$, when $\xi \sim \mathcal{P}(\alpha)$ and $\mathbb{E}\xi^2 < 3\mathbb{E}\xi$

Chvátal and Reed [15] argue that if *2-SAT* formula ϕ is unsatisfiable, then it contains a *bicycle*, see Section 2.3. Thus, the absence of *bicycles* may serve as a convenient witness of formula's satisfiability. The general idea of this section is to show that w.h.p. there are no bicycles in $\phi \sim \mathbb{C}_n^2(\xi)$, when $\xi \sim \mathcal{P}(\alpha)$ and $\mathbb{E}\xi^2 < 3\mathbb{E}\xi$.

Intuitively, when $\mathbb{E}\xi^2 < 3\mathbb{E}\xi$, then we expect $\frac{2T_n}{S_n} = 1 - \mu' > 0$, where $\mu' > 0$ is some small number. As it was showed in Lemma 7, the latter quantity approximates the number of newly added live clones, when running the TSPAN procedure. Since TSPAN always performs at least one iteration of growing the span, it may add at most Δ clones into the live set after constructing the very first span from the root. After that each subsequent iteration adds *on average* $\approx \frac{2T_n}{S_n}$ new live clones. So after running the TSPAN for j iterations, where $j \to \infty$, when $n \to \infty$, then we expect the live set to contain around

$$L_{t^*} = \Delta\left(\frac{2T_n}{S_n}\right)^j = \Delta(1-\mu')^j \leq \Delta e^{-j\mu'}$$

clones. Therefore, after $O(\log n)$ iterations, the live set becomes empty, and TSPAN terminates. Thus, we expect paths of length at most $O(\log n)$, which is not enough for bicycles to occur.

More formally we first show that in the case $\frac{2T_n}{S_n} = 1 - \mu'$ a random formula is unlikely to contain long paths.

Lemma 12. *If $\frac{2T_n}{S_n} = 1 - \mu' < 1$, then paths in ϕ are of length $O(\log n)$, w.h.p.*

Then we give a straightforward estimation of the number of 'short' bicycles

Lemma 13. *If $\frac{2T_n}{S_n} = 1 - \mu' < 1$, then for any k the probability that formula ϕ contains a bicycle of length r is at most $(1 + o(1))^r(1 - \mu')^r$.*

These two lemmas imply that ϕ contains no bicycles.

Corollary 4. *If $\frac{2T_n}{S_n} = 1 - \mu' < 1$, then ϕ contains no bicycles, w.h.p.*

It remains to argue that the inequality $\frac{2T_n}{S_n} = 1 - \mu' < 1$ holds w.h.p.

Proposition 3. *Let $\phi \sim \mathbb{C}_n^2(\xi)$, where $\xi \sim \mathcal{P}(\alpha)$ and $\mathbb{E}\xi^2 < 3\mathbb{E}\xi$. Then w.h.p. ϕ is satisfiable.*

References

1. Achlioptas, D.: Lower bounds for random 3-SAT via differential equations. Theor. Comput. Sci. **265**(1-2), 159–185 (2001)
2. Achlioptas, D., Moore, C.: Random k-SAT: Two moments suffice to cross a sharp threshold. SIAM J. Comput. **36**(3), 740–762 (2006)
3. Aiello, W., Graham, F.C., Lu, L.: A random graph model for power law graphs. Experimental Mathematics **10**(1), 53–66 (2001)
4. Ansótegui, C., Bonet, M.L., Giráldez-Cru, J., Levy, J.: Community structure in industrial SAT instances (2016)
5. Ansótegui, C., Bonet, M.L., Levy, J.: On the structure of industrial SAT instances. In: Principles and Practice of Constraint Programming - CP 2009, 15th International Conference, CP 2009, Lisbon, Portugal, September 20-24, 2009, Proceedings. pp. 127–141 (2009)
6. Ansótegui, C., Bonet, M.L., Levy, J.: Towards industrial-like random SAT instances. In: IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, California, USA, July 11-17, 2009. pp. 387–392 (2009)
7. Ansótegui, C., Bonet, M.L., Levy, J.: Scale-free random SAT instances (2017), available at `https://arxiv.org/abs/1708.06805v2`
8. Ansótegui, C., Bonet, M.L., Levy, J., Manyà, F.: Measuring the hardness of SAT instances. In: Proceedings of the 23rd National Conference on Artificial Intelligence - Volume 1. pp. 222–228. AAAI'08 (2008)
9. Aspvall, B., Plass, M.F., Tarjan, R.E.: A linear-time algorithm for testing the truth of certain quantified boolean formulas. Information Processing Letters **8**(3), 121 – 123 (1979)
10. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. Science **286**, 509–512 (1999)
11. Bollobás, B., Riordan, O.: Mathematical results on scale-free random graphs. In: Handbook of Graphs and Networks, pp. 1–34. Wiley-VCH (2002)
12. Bollobás, B., Riordan, O., Spencer, J., Tusnády, G.E.: The degree sequence of a scale-free random graph process. Random Struct. Algorithms **18**(3), 279–290 (2001)
13. Boufkhad, Y., Dubois, O., Interian, Y., Selman, B.: Regular random k-sat: Properties of balanced formulas. J. Autom. Reasoning **35**(1-3), 181–200 (2005)
14. Braunstein, A., Mézard, M., Zecchina, R.: Survey propagation: An algorithm for satisfiability. Random Struct. Algorithms **27**(2), 201–226 (2005)
15. Chvátal, V., Reed, B.A.: Mick gets some (the odds are on his side). In: 33rd Annual Symposium on Foundations of Computer Science, Pittsburgh, Pennsylvania, USA, 24-27 October 1992. pp. 620–627 (1992)
16. Clauset, A., Shalizi, C., Newman, M.: Power-law distributions in empirical data. SIAM Review **51**(4), 661–703 (2009)
17. Coja-Oghlan, A.: A better algorithm for random k-SAT. SIAM J. Comput. **39**(7), 2823–2864 (2010)
18. Coja-Oghlan, A., Panagiotou, K.: Going after the k-sat threshold. In: Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013. pp. 705–714 (2013)
19. Cook, S.A., Mitchell, D.G.: Finding hard instances of the satisfiability problem: A survey. In: Satisfiability Problem: Theory and Applications, Proceedings of a DIMACS Workshop, Piscataway, New Jersey, USA, March 11-13, 1996. pp. 1–18 (1996)

20. Cooper, C., Frieze, A., Sorkin, G.B.: Random 2-SAT with prescribed literal degrees. Algorithmica **48**(3), 249–265 (Jul 2007)
21. Díaz, J., Kirousis, L.M., Mitsche, D., Pérez-Giménez, X.: On the satisfiability threshold of formulas with three literals per clause. Theor. Comput. Sci. **410**(30-32), 2920–2934 (2009)
22. Ding, J., Sly, A., Sun, N.: Proof of the satisfiability conjecture for large k. In: Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015. pp. 59–68 (2015)
23. Dubios, O., Boufkhad, Y.: A general upper bound for the satisfiability threshold of random r-sat formulae. J. Algorithms **24**(2), 395–420 (1997)
24. Dubois, O., Boufkhad, Y., Mandler, J.: Typical random 3-SAT formulae and the satisfiability threshold. Electronic Colloquium on Computational Complexity (ECCC) **10**(007) (2003)
25. Franco, J., Paull, M.C.: Probabilistic analysis of the Davis Putnam procedure for solving the satisfiability problem. Discrete Applied Mathematics **5**(1), 77–87 (1983)
26. Friedgut, E.: Sharp thresholds of graph properties, and the k-SAT problem. J. ACM **12**(4), 1017–1054 (1999)
27. Friedrich, T., Krohmer, A., Rothenberger, R., Sauerwald, T., Sutton, A.M.: Bounds on the satisfiability threshold for power law distributed random SAT. In: 25th Annual European Symposium on Algorithms, ESA 2017, September 4-6, 2017, Vienna, Austria. pp. 37:1–37:15 (2017)
28. Friedrich, T., Rothenberger, R.: Sharpness of the satisfiability threshold for non-uniform random k-SAT. In: Beyersdorff, O., Wintersteiger, C.M. (eds.) Theory and Applications of Satisfiability Testing – SAT 2018. pp. 273–291. Springer International Publishing, Cham (2018)
29. Giráldez-Cru, J., Levy, J.: Generating SAT instances with community structure. Artificial Intelligence **238**(C), 119–134 (2016)
30. Goerdt, A.: A threshold for unsatisfiability. J. Comput. Syst. Sci. **53**(3), 469–486 (1996)
31. Kaporis, A.C., Kirousis, L.M., Lalas, E.G.: The probabilistic analysis of a greedy satisfiability algorithm. Random Struct. Algorithms **28**(4), 444–480 (2006)
32. Kim, J.H.: The poisson cloning model for random graphs, random directed graphs and random k-SAT problems. In: COCOON. p. 2 (2004)
33. Kirousis, L.M., Kranakis, E., Krizanc, D., Stamatiou, Y.C.: Approximating the unsatisfiability threshold of random formulas. Random Struct. Algorithms **12**(3), 253–269 (1998)
34. Krioukov, D.V., Papadopoulos, F., Kitsak, M., Vahdat, A., Boguñá, M.: Hyperbolic geometry of complex networks. CoRR **abs/1006.5169** (2010)
35. Krzakała, F., Montanari, A., Ricci-Tersenghi, F., Semerjian, G., Zdeborová, L.: Gibbs states and the set of solutions of random constraint satisfaction problems. PNAS **104**(25), 10318–10323 (2007)
36. M ézard, M., Parisi, G., Zecchina, R.: Analytic and algorithmic solution of random satisfiability problems. Science **297**(5582), 812–815 (2002)
37. Newman, M.: Power laws, Pareto distributions and Zipf's law. Contemporary Physics **46**(5), 323–351 (2005)
38. Omelchenko, O., Bulatov, A.: Concentration inequalities for sums of random variables, each having power bounded tails (2018), available at `https://arxiv.org/abs/1903.02529`
39. Selman, B., Mitchell, D.G., Levesque, H.J.: Generating hard satisfiability problems. Artif. Intell. **81**(1-2), 17–29 (1996)

40. de la Vega, W.F.: Random 2-SAT: results and problems. Theor. Comput. Sci. **265**(1-2), 131–146 (2001)