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. We study Random 2-SAT problem, in which 2-CNFs are sampled from a wide range of non-uniform distributions, including heavy-tailed using random model that is quite different from the standard uniform Random 2-SAT.

This model of random SAT is the so-called configuration model, given by a distribution ξ for the degree (or the number of occurrences) of each variable. 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 partitioning 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, and we show that the 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 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 occurrence 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 to algorithmic properties of the Random SAT, as well as to 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 $v_1, \ldots, v_n$ 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 with support on positive integers, 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 case of such a distribution is Poisson Cloning described above. Another case is studied by Cooper, Frieze, and Sorkin [20]. In their paper 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 $\mathbb{E}\xi^2 > 3\mathbb{E}\xi$, where ξ is the r.v. that governs the number of times a variable appears in 2-SAT formula ϕ [7].

The satisfiability threshold of the model of Ansotegui et al. has been analytically 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 for random 2-CNF in the configuration model, in which the underlying degree random variable is distributed according to the power law. A formal statement of the result is given by Theorem 1 in Section 2.4.

We show that ϕ is unsatisfiable with high probability, when $0 < \alpha < 2$, 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 $\alpha < \frac{k}{k-1}$.

However, when $\alpha > 2$, then ξ has the first and second moments, and we show that the satisfiability of ϕ in this regime depends on the relation between the two moments. We then 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, while by $V(\phi)$ and $L(\phi)$ we denote the sets of variables and literals of ϕ respectively. Let d_i^+ denote the number of occurrences of v_i as a positive literal (or the number of the literal v_i), and let d_i^- denote the number of occurrences of the literal $\bar{v}_i$.

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 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 is associated with a literal if we need to emphasize the polarity of the clone.

We will denote a random formula ϕ obtained from $\mathbb{C}_n^k(\xi)$ by $\phi \sim \mathbb{C}_n^k(\xi)$. By $\mathbb{C}_n^k(\xi)$ we also denote the probability distribution implied by the Algorithm 1.

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** $S_n := \sum_{i=1}^n d_i$ is not a multiple of k **then**
4:         discard the sequence, and go to step 2
5:     **end if**
6:     Otherwise, introduce multi-set $S \leftarrow \bigcup_{i=1}^n \underbrace{\{v_i, v_i, \ldots, v_i\}}_{d_i \text{ times}}$
7:     Let $\phi \leftarrow \emptyset$
8:     **while** $S \neq \emptyset$ **do**
9:         Pick u.a.r. k elements $\{v_1, v_2, \ldots, v_k\}$ from S without replacement
10:        Let $C \leftarrow \{v_1, v_2, \ldots, v_k\}$
11:        $S \leftarrow S - C$
12:        Negate each element in C u.a.r with probability $1/2$
13:        $\phi \leftarrow \phi \cup C$
14:     **end while**
15:     **return** ϕ
16: **end procedure**

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 $\xi \sim \mathcal{P}(\alpha)$ depends only on α.

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

Proof. From the definition of m-th moment of a strictly positive r.v. ξ, we have

$$
\begin{aligned}
\mathbb{E}\xi^m &= \sum_{d \geq 1} d^m \Pr\left[\xi = d\right] \\
&= \sum_{d \geq 1} d^m \left(\Pr\left[\xi \geq d\right] - \Pr\left[\xi \geq d+1\right] \right) \\
&= \Pr\left[\xi \geq 1\right] + \sum_{d \geq 2} d^m \Pr\left[\xi \geq d\right] - \sum_{d \geq 2}(d-1)^m \Pr\left[\xi \geq d\right] \\
&= \sum_{d \geq 1} \left(d^m - (d-1)^m \right) \Pr\left[\xi \geq d\right]
\end{aligned}
$$

$$= \sum_{d \geq 1} \left(d^m - \sum_{i=0}^{m} \binom{m}{i} d^{m-i}(-1)^i \right) \Pr\left[\xi \geq d\right]$$

$$= \sum_{d \geq 1} \sum_{i=1}^{m} \binom{m}{i} d^{m-i}(-1)^{i+1} \Pr\left[\xi \geq d\right]$$

$$= \sum_{i=1}^{m} \binom{m}{i} (-1)^{i+1} \sum_{d \geq 1} d^{m-i} \Pr\left[\xi \geq d\right].$$

However, since $\xi \sim \mathcal{P}(\alpha)$, then for any $d \geq 1$

$$\Pr[\xi \geq d] = F_\xi(d) \leq V \, d^{-\alpha}.$$

Therefore,

$$\mathbb{E}\xi^m = \sum_{i=1}^{m} \binom{m}{i} (-1)^{i+1} \sum_{d \geq 1} d^{m-i} \Pr\left[\xi \geq \ell\right]$$

$$\leq V \sum_{i=1}^{m} \binom{m}{i} (-1)^{i+1} \sum_{d \geq 1} d^{m-i-\alpha},$$

which is finite iff $0 < m < \alpha$, and the result follows.   □

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 ξ_i's are independent copies of a r.v. $\xi \sim \mathcal{P}(\alpha)$.

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.

Lemma 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*

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

where $V > 0$ *and* $0 < \alpha \leq 1$ *are constants. Then w.h.p.* $S_n = O\left(n^{\frac{1}{\alpha}}\right)$.

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.

Lemma 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*

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

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 strictly 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 4. *Let $\Delta = \max\left(\xi_1, \xi_2, \cdots, \xi_n\right)$, where ξ_i's are independent copies of a r.v. $\xi \sim \mathcal{P}\left(\alpha\right)$ with $\alpha > 0$. Then w.h.p. $\Delta = O\left(n^{1/\alpha}\right)$.*

Proof. Simple calculation shows that

$$\Pr\left[\Delta \geq x\right] = \Pr\left[\max\left(\xi_1, \xi_2, \cdots, \xi_n\right) \geq x\right] = \Pr\left[\bigcup_{i=1}^{n}\{\xi_i \geq x\}\right]$$

$$\leq \sum_{i=1}^{n} \Pr\left[\xi_i \geq x\right], \text{ by Union bound}$$

$$= nF_\xi(x), \text{ since } \xi_i \stackrel{d}{=} \xi.$$

Since $\xi \sim \mathcal{P}\left(\alpha\right)$, then $F_\xi(\ell) \leq V\,\ell^{-\alpha}$. Hence,

$$\Pr\left[\Delta \geq x\right] \leq nF_\xi(x) \leq nVx^{-\alpha}.$$

Then if $\Delta \geq n^{1/\alpha + \epsilon}$ for any $\epsilon > 0$, we obtain

$$\Pr\left[\Delta \geq n^{1/\alpha + \epsilon}\right] \leq V\,n^{-\alpha\epsilon} = o(1),$$

hence, we do not expect to see variables with such large degrees. Thus, we conclude that $\Delta = O\left(n^{1/\alpha}\right)$ holds w.h.p., and the lemma follows. $\qquad\square$

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

$$d_i^+ \sim Bin\left(\deg(v_i), 1/2\right) \quad and \quad d_i^- = \deg(v_i) - d_i^+,$$

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

Lemma 5. *Let ξ be some positive integer-valued r.v., and let $d^+ \sim Bin\left(\xi, 1/2\right)$, while $d^- = \xi - d^+$. Then*

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

Proof. Observe that the event $\{d^+ \geq \ell\}$ implies $\{\xi \geq \ell\}$, therefore, $\{d^+ \geq \ell\} \subseteq \{\xi \geq \ell\}$, and as a result

$$F_{d^+}(\ell) = \Pr\left[d^+ \geq \ell\right] \leq \Pr\left[\xi \geq \ell\right] = F_\xi(\ell).$$

Finally, since $d^+ \overset{d}{=} d^-$ (by $R_1 \overset{d}{=} R_2$ we denote the fact that the two r.vs. R_1 and R_2 have identical distributions), the first result follows. As for the number of pairs of complementary clones, we have

$$\begin{aligned}
F_{d^+d^-}(\ell) &= \Pr\left[d^+d^- \geq \ell\right] \\
&= \Pr\left[d^+d^- \geq \ell \,|\, d^+ \geq \ell^{1/2}\right] \Pr\left[d^+ \geq \ell^{1/2}\right] \\
&\quad + \Pr\left[d^+d^- \geq \ell \,|\, d^+ < \ell^{1/2}\right] \Pr\left[d^+ < \ell^{1/2}\right] \\
&\leq \Pr\left[d^+ \geq \ell^{1/2}\right] + \Pr\left[d^- \geq \ell^{1/2} \,|\, d^+ < \ell^{1/2}\right]. \qquad (2)
\end{aligned}$$

We have already established that $\Pr\left[d^+ \geq \ell^{1/2}\right] \leq F_\xi\left(\ell^{1/2}\right)$. As for the second probability in (2), it is bounded in a similar manner. Since the event $\{d^- \geq \ell^{1/2} \,|\, d^+ < \ell^{1/2}\}$ implies $\{\xi \geq \ell^{1/2}\}$, it follows

$$\Pr\left[d^- \geq \ell^{1/2} \,|\, d^+ < \ell^{1/2}\right] \leq \Pr\left[\xi \geq \ell^{1/2}\right] = F_\xi\left(\ell^{1/2}\right).$$

Hence, after combining the two probabilities together, we obtain

$$F_{d^+d^-}(\ell) \leq \Pr\left[d^+ \geq \ell^{1/2}\right] + \Pr\left[d^- \geq \ell^{1/2} \,|\, d^+ < \ell^{1/2}\right] \leq 2F_\xi\left(\ell^{1/2}\right),$$

and the lemma follows. $\qquad\square$

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

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

$$F_{d^+}(\ell) = F_{d^-}(\ell) \leq V\,\ell^{-\alpha}, \qquad (3)$$
$$F_{d^+d^-}(\ell) \leq 2V\,\ell^{-\alpha/2}. \qquad (4)$$

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

Proof. From the definition of the expected value, and the way quantities d^+ and d^- are calculated, it follows that

$$\mathbb{E}\left[d^+ d^-\right] = \sum_{\ell=1}^{\infty}\sum_{d=0}^{\ell} d\left(\ell - d\right)\Pr\left[d^+ = d \mid \xi = \ell\right]\Pr\left[\xi = \ell\right]$$

$$= \sum_{\ell=1}^{\infty}\sum_{d=0}^{\ell} d\left(\ell - d\right)\Pr\left[Bin(\ell, 1/2) = d \mid \xi = \ell\right]\Pr\left[\xi = \ell\right]$$

$$= \sum_{\ell=1}^{\infty}\frac{\ell\Pr\left[\xi = \ell\right]}{2^\ell}\sum_{d=0}^{\ell} d\binom{\ell}{d} - \sum_{\ell=1}^{\infty}\frac{\Pr\left[\xi = \ell\right]}{2^\ell}\sum_{d=0}^{\ell} d^2\binom{\ell}{d}.$$

Next, we apply two well-known relations

$$\sum_{j=0}^{n} j\binom{n}{j} = n2^{n-1} \quad \text{and} \quad \sum_{j=0}^{n} j^2\binom{n}{j} = (n + n^2)2^{n-2},$$

to get

$$\mathbb{E}\left[d^+ d^-\right] = \sum_{\ell=1}^{\infty}\frac{\ell\Pr\left[\xi = \ell\right]}{2^\ell}\sum_{d=0}^{\ell} d\binom{\ell}{d} - \sum_{\ell=1}^{\infty}\frac{\Pr\left[\xi = \ell\right]}{2^\ell}\sum_{d=0}^{\ell} d^2\binom{\ell}{d} = \frac{\mathbb{E}\xi^2 - \mathbb{E}\xi}{4},$$

and the proof is finished.   □

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 Lemma 3 w.h.p. it holds $T_n = (1 + o(1))\sum_{i=1}^{n}\mathbb{E}\left[d_i^+ d_i^-\right]$.

We finish this subsection with our version of the *Azuma-like inequality* first appeared in [20], which will be used in the proofs. Intuitively, our version states that a discrete-time random walk $X = X_0 + \sum_{i=1}^{n} X_i$ with initial value X_0 and positive drift, consisting of not necessarily 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 a 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. Moreover, we exploit the tail inequality of each step in the random walk in order to overcome the rather restrictive bounds on the maximal variable degree in the original paper.

Lemma 7 (Azuma-like inequality). *Let $X = X_0 + \sum_{i=1}^{t} X_i$ be some random walk, such that $X_0 \geq 0$ is constant initial value of the process, $X_i \geq -a$, where*

$a > 0$ *is constant, are bounded from below random variables, not necessary independent, and such that* $\mathbb{E}[X_i \,|\, X_1, \ldots, X_{i-1}] \geq \mu > 0$ *(μ is constant) and* $\Pr[X_i \geq \ell \,|\, X_1, \ldots, X_{i-1}] \leq V \, \ell^{-\alpha}$ *for every* $\ell \geq 1$ *and constants* $V > 0$, $\alpha > 1$. *Then for any* $0 < \varepsilon < \frac{1}{2}$, *the following inequality holds*

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

Proof. First, let us introduce "truncated" at $\delta := \lfloor \log n \rfloor$ versions of the variables X_i, i.e.

$$Y_i = X_i \cdot \mathbb{1}_{X_i \leq \delta}.$$

Then the conditional expectation for the new variables is

$$\begin{aligned}
\mathbb{E}\left[Y_i \,|\, X_1, \ldots, X_{i-1}\right] &= \mathbb{E}\left[X_i \cdot \mathbb{1}_{X_i \leq \delta} \,|\, X_1, \ldots, X_{i-1}\right] \\
&\geq \mathbb{E}\left[X_i \,|\, X_1, \ldots, X_{i-1}\right] - \mathbb{E}\left[X_i \cdot \mathbb{1}_{X_i \geq \delta} \,|\, X_1, \ldots, X_{i-1}\right] \\
&\geq \mu - \mathbb{E}\left[X_i \cdot \mathbb{1}_{X_i \geq \delta} \,|\, X_1, \ldots, X_{i-1}\right],
\end{aligned}$$

since $\mathbb{E}\left[X_i \,|\, X_1, \ldots, X_{i-1}\right] \geq \mu$. Let us denote by $\mathcal{P}_i\left[A\right]$, where A is some event, the following function

$$\mathcal{P}_i\left[A\right] := \Pr\left[A \,|\, X_1, \ldots, X_{i-1}\right].$$

Then

$$\begin{aligned}
\mathbb{E}\left[Y_i \,|\, X_1, \ldots, X_{i-1}\right] &\geq \mu - \mathbb{E}\left[X_i \cdot \mathbb{1}_{X_i \geq \delta} \,|\, X_1, \ldots, X_{i-1}\right] \\
&= \mu - \sum_{\ell=\delta}^{\infty} \ell \Pr\left[X_i = \ell \,|\, X_1, \ldots, X_{i-1}\right] \\
&= \mu - \left(\sum_{\ell=\delta}^{\infty} \ell \, \mathcal{P}_i\left[X_i \geq \ell\right] - \sum_{\ell=\delta+1}^{\infty} (\ell - 1) \, \mathcal{P}_i\left[X_i \geq \ell\right]\right) \\
&= \mu - \delta \, \mathcal{P}_i\left[X_i \geq \delta\right] - \sum_{\ell=\delta+1}^{\infty} \mathcal{P}_i\left[X_i \geq \ell\right].
\end{aligned}$$

Now recall that $\mathcal{P}_i\left[X_i \geq \delta\right] = \Pr\left[X_i \geq \delta \,|\, X_1, \ldots, X_{i-1}\right] \leq V \, \delta^{-\alpha}$. Hence,

$$\begin{aligned}
\mathbb{E}\left[Y_i \,|\, X_1, \ldots, X_{i-1}\right] &\geq \mu - \delta \, \mathcal{P}_i\left[X_i \geq \delta\right] - \sum_{\ell=\delta+1}^{\infty} \mathcal{P}_i\left[X_i \geq \ell\right] \\
&\geq \mu - V \, \delta^{1-\alpha} - V \sum_{\ell=\delta+1}^{\infty} \ell^{-\alpha} \\
&\geq \mu - C \, \delta^{1-\alpha},
\end{aligned}$$

where $C > 0$ is constant.

Since $\delta = \lfloor \log n \rfloor$ and $\alpha > 1$, we obtain

$$\mathbb{E}\left[Y_i \,|\, X_1, \ldots, X_{i-1}\right] \geq \mu - C\,\delta^{1-\alpha} = \mu - C\,\left(\lfloor \log n \rfloor\right)^{1-\alpha} \geq \mu - o(1) \geq \frac{\mu}{2}.$$

Next, notice that the above lower bound of the conditional expectation does not depend on realizations of the r.vs. $X_1, \ldots, X_{i-1}$. Thus, for *any* trajectory of the random walk $\{x_j\}_{j=1}^{i-1}$, where x_j is an element from the domain of r.v. X_j, $\mathbb{E}\left[Y_i \,|\, x_1, \ldots, x_{i-1}\right] \geq \frac{\mu}{2}$. Hence, since domain of the r.v. Y_i is a subset of the domain of r.v. X_i, it follows that

$$\mathbb{E}\left[Y_i \,|\, Y_1, \ldots, Y_{i-1}\right] \geq \frac{\mu}{2}. \tag{5}$$

As for the second moment, since $Y_i \leq \delta$, we have an obvious upper bound

$$\mathbb{E}\left[Y_i^2 \,|\, Y_1, \ldots, Y_{i-1}\right] \leq \delta^2 \leq \log^2 n. \tag{6}$$

Now, using the truncated variables, we introduce an auxiliary random process

$$Y = X_0 + \sum_{i=1}^{t} Y_i.$$

Then the probability that the original random walk X will drop below the $\epsilon t \mu$ level, where $0 < \epsilon < 1/2$, is at most

$$\begin{aligned}
\Pr\left[X \leq \epsilon t \mu\right] &\leq \Pr\left[Y \leq \epsilon t \mu\right], \text{ since } Y \leq X \\
&= \Pr\left[e^{-\lambda Y} \geq e^{-\lambda \epsilon t \mu}\right], \text{ for any } \lambda > 0 \\
&\leq e^{\lambda \epsilon t \mu} \mathbb{E} e^{-\lambda Y}, \text{ by Markov's inequality} \\
&= e^{\lambda \epsilon t \mu - \lambda X_0} \mathbb{E}\left[\prod_{i=1}^{t} e^{-\lambda Y_i}\right].
\end{aligned}$$

Next, we apply the tower property of expectation, that is for any two random variables A and B defined over the same probability space and $\mathbb{E}|A| < \infty$, the following holds (subscript indicates over which variable calculation of expectation is performed)

$$\mathbb{E}A = \mathbb{E}_A\left[A\right] = \mathbb{E}_B\left[\mathbb{E}_A\left[A \,|\, B\right]\right].$$

Thus, we have

$$\begin{aligned}
\Pr\left[X \leq \epsilon t \mu\right] &\leq e^{\lambda \epsilon t \mu - \lambda X_0} \mathbb{E}\left[\prod_{i=1}^{t} e^{-\lambda Y_i}\right] \\
&= e^{\lambda \epsilon t \mu - \lambda X_0} \mathbb{E}_{Y_1, \ldots, Y_{t-1}}\left[\mathbb{E}_{Y_t}\left[\prod_{i=1}^{t} e^{-\lambda Y_i} \,|\, Y_1, \ldots, Y_{i-1}\right]\right].
\end{aligned}$$

Consider the innermost expectation. Since we condition it over variables $Y_1, \ldots, Y_{t-1}$, we consider such variables as given (or constant). Hence,

$$
\begin{aligned}
\Pr\left[X \le \epsilon t\mu\right] &\le e^{\lambda \epsilon t\mu - \lambda X_0} \mathbb{E}_{Y_1,\ldots,Y_{t-1}}\left[\mathbb{E}_{Y_t}\left[\prod_{i=1}^{t} e^{-\lambda Y_i} \mid Y_1, \ldots, Y_{i-1}\right]\right] \\
&= e^{\lambda \epsilon t\mu - \lambda X_0} \mathbb{E}_{Y_1,\ldots,Y_{t-1}}\left[\prod_{i=1}^{t-1} e^{-\lambda Y_i} \cdot \mathbb{E}_{Y_t}\left[e^{-\lambda Y_t} \mid Y_1, \ldots, Y_{i-1}\right]\right].
\end{aligned}
\tag{7}
$$

Now, since $Y_i \ge -a$ just like the original variables X_i's, we can upper bound the inner expectation by applying the well-known inequality

$$
e^{-x} \le 1 - x + x^2,
$$

which is valid for every $x \ge -1$. Hence, by restricting $0 < \lambda \le 1/a$, we have

$$
\begin{aligned}
\mathbb{E}_{Y_t}\left[e^{-\mu Y_t} \mid Y_1, \ldots, Y_{t-1}\right] &\le \mathbb{E}_{Y_t}\left[1 - \lambda Y_t + \lambda^2 Y_t^2 \mid Y_1, \ldots, Y_{t-1}\right] \\
&= 1 - \lambda \mathbb{E}_{Y_t}\left[Y_t \mid Y_1, \ldots, Y_{t-1}\right] + \lambda^2 \mathbb{E}\left[Y_t^2 \mid Y_1, \ldots, Y_{t-1}\right] \\
&\le 1 - \lambda \frac{\mu}{2} + \lambda^2 \log^2 t,
\end{aligned}
\tag{8}
$$

since $\mathbb{E}\left[Y_i \mid Y_1, \ldots, Y_{i-1}\right] \ge \frac{\mu}{2}$ by (5) and $\mathbb{E}\left[Y_i^2 \mid Y_1, \ldots, Y_{i-1}\right] \le \log^2 t$ by (6). Therefore, we obtain

$$
\mathbb{E}_{Y_t}\left[e^{-\mu Y_t} \mid Y_1, \ldots, Y_{t-1}\right] \le 1 - \lambda \frac{\mu}{2} + \lambda^2 \log^2 t \le \exp\left(-\lambda \frac{\mu}{2} + \lambda^2 \log^2 t\right).
$$

Thus, the probability in (7) is upper bounded as

$$
\begin{aligned}
\Pr\left[X \le \epsilon t\mu\right] &\le e^{\lambda \epsilon t\mu - \lambda X_0} \mathbb{E}_{Y_1,\ldots,Y_{t-1}}\left[\prod_{i=1}^{t-1} e^{-\lambda Y_i} \cdot \mathbb{E}_{Y_t}\left[e^{-\lambda Y_t} \mid Y_1, \ldots, Y_{t-1}\right]\right] \\
&\le e^{\lambda \epsilon t\mu - \lambda X_0} \cdot \exp\left(-\lambda \frac{\mu}{2} + \lambda^2 \log^2 t\right) \mathbb{E}_{Y_1,\ldots,Y_{t-1}}\left[\prod_{i=1}^{t-1} e^{-\lambda Y_i}\right].
\end{aligned}
$$

Repeating the same process of isolation of conditional expectation of each step Y_i and bounding it with (8) for another $t-1$ times, we obtain

$$
\begin{aligned}
\Pr\left[X \le \epsilon t\mu\right] &\le e^{\lambda \epsilon t\mu - \lambda X_0} \cdot \exp\left(-\lambda \frac{\mu}{2} + \lambda^2 \log^2 t\right) \mathbb{E}_{Y_1,\ldots,Y_{t-1}}\left[\prod_{i=1}^{t-1} e^{-\lambda Y_i}\right] \\
&= \exp\left(\lambda \epsilon t\mu - \lambda X_0 - \lambda t \frac{\mu}{2} + \lambda^2 t \log^2 t\right).
\end{aligned}
$$

Fix $\lambda = \frac{\mu}{2 \log^2 t}\left(\frac{1}{2} - \epsilon\right)$. Note, that when $t \to \infty$ and $\mu = const$, then $0 < \lambda < 1/a$. Hence, we have

$$\Pr\left[X \le \epsilon t\mu\right] \le \exp\left(\lambda\epsilon t\mu - \lambda X_0 - \lambda t\frac{\mu}{2} + \lambda^2 t \log^2 t\right)$$

$$= \exp\left(-\frac{t + X_0}{4\log^2 t}\mu^2\left(\frac{1}{2} - \epsilon\right)^2\right).$$

And the lemma follows.   □

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

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

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 $k \ge 2$, 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 with 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 the 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.*

Proof. First, recall quantity S_n that denotes as the total number of clones

$$S_n = \sum_{i=1}^{n} \deg(v_i).$$

Since each $\deg(v_i)$ is an independent realization of the r.v. ξ in $\mathbb{C}_n^k(\xi)$, it follows that $\deg(v_i) \overset{d}{=} \xi$.

Next, let us estimate how many variables v_i's in ϕ have degree at least $S_n^{(k-1)/k} \log n$:

$$\mathbb{E}\left[\sum_{i=1}^{n} \mathbb{1}_{\deg(v_i) \geq S_n^{(k-1)/k} \log n}\right] = \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{1}_{\deg(v_i) \geq S_n^{(k-1)/k} \log n}\right]$$
$$= nF_\xi\left(S_n^{(k-1)/k} \log n\right).$$

However, since $\xi \sim \mathcal{P}(\alpha)$, then $F_\xi(x) \geq W x^{-\alpha}$ for some $W > 0$ and all $x \geq 1$, hence,

$$\mathbb{E}\left[\sum_{i=1}^{n} \mathbb{1}_{\deg(v_i) \geq S_n^{(k-1)/k} \log n}\right] = nF_\xi\left(S_n^{(k-1)/k} \log n\right) \geq W\, nS_n^{-\alpha(k-1)/k} \log^{-\alpha} n. \tag{9}$$

Prior to moving to the next steps of the proof, we note that the actual number of variables with degrees at least $S_n^{(k-1)/k} \log n$ is distributed according to Binomial distribution, hence, it is concentrated around its mean.

Next, we consider 2 cases. The first one is when $0 < \alpha \leq 1$, and the second case is when $1 < \alpha < \frac{k}{k-1}$. Hence, for the first case, by Lemma 2, w.h.p. $S_n = O\left(n^{1/\alpha}\right)$. Therefore,

$$\mathbb{E}\left[\sum_{i=1}^{n} \mathbb{1}_{\deg(v_i) \geq S_n^{(k-1)/k} \log n}\right] \geq W\, nS_n^{-\alpha(k-1)/k} \log^{-\alpha} n$$
$$= \Omega\left(n^{1-(k-1)/k} \log^{-\alpha} n\right).$$

Hence, we expect polynomially many variables to have large degrees. The same holds in the case, when $1 < \alpha < \frac{k}{k-1}$. Then $S_n = (1 + o(1))\, n\mathbb{E}\xi$ w.h.p. by Lemma 3, and so the expected number of variables with degree at least $S_n^{(k-1)/k} \log n$ is:

$$\mathbb{E}\left[\sum_{i=1}^{n} \mathbb{1}_{\deg(v_i) \geq S_n^{(k-1)/k} \log n}\right] \geq W\, n S_n^{-\alpha(k-1)/k} \log^{-\alpha} n$$

$$= \Omega\left(n^{1-\alpha(k-1)/k} \log^{-\alpha} n\right).$$

And so in both cases we note that we expect many variables to have degree at least $S_n^{(k-1)/k} \log n$. Let us fix k such variables $v_1, v_2, \ldots, v_k \in V(\phi)$ such that $\deg(v_i) \geq S_n^{(k-1)/k} \log n$ for all $i \in [1 \ldots k]$. Next we introduce indicator r.v. I_c, which is equal to 1 iff clause $c \in C(\phi)$ consists solely of clones of variables $v_1, v_2, \ldots, v_k$. Then

$$H = \sum_{c \in C(\phi)} I_c$$

is the total number of clauses formed only from clones of variables $v_1, v_2, \ldots, v_k$.

It is easy to show that H is the sum of weakly correlated binary r.vs., since the covariance of I_{c_0} and I_{c_1} for some fixed clauses $c_0, c_1 \in C(\phi)$ is

$$Cov(I_{c_0}, I_{c_1}) = o\left(\Pr\left[I_{c_0} = 1\right]^2\right) = o\left(\Pr\left[I_{c_1} = 1\right]^2\right).$$

Therefore, due to Chebyshev's inequality, it follows that H is concentrated around its expectation, i.e. the expected value $\mathbb{E}[H]$ serves as a good approximation to the actual value of the r.v. H.

Finally, the expected number of clauses formed only from clones of variables $v_1, v_2, \ldots, v_k \in V(\phi)$ is

$$\mathbb{E}[H] = \mathbb{E}\left[\sum_{c \in C(\phi)} I_c\right] = \sum_{c \in C(\phi)} \Pr[I_c = 1]$$

$$= \sum_{c \in C(\phi)} k! \frac{\deg(v_1)}{S_n} \times \frac{\deg(v_2)}{S_n - 1} \times \cdots \times \frac{\deg(v_k)}{S_n - k + 1}$$

$$\geq (1 + o(1)) \sum_{c \in C(\phi)} k! \left(\frac{S_n^{1-\frac{1}{k}} \log n}{S_n}\right)^k, \text{ since } \deg(v_i) \geq S_n^{1-\frac{1}{k}} \log n$$

$$= (1 + o(1)) |C(\phi)| k! \frac{\log^k n}{S_n}$$

$$= (1 + o(1)) (k - 1)! \log^k n,$$

since $|C(\phi)| = \frac{S_n}{k}$. Hence, from the facts that w.h.p. $H = (1 + o(1))\mathbb{E}[H]$ and $\mathbb{E}[H] \geq (1 + o(1))(k - 1)! \log^k n$, it follows that the number of clauses formed solely of clones of the fixed k variables grows together with n. However, as it was pointed earlier, we need only a 2^k clauses subformula to make ϕ unsatisfiable. Thus, ϕ is *UNSAT* w.h.p. when $0 < \alpha < \frac{k}{k-1}$. $\qquad\square$

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

Corollary 2. *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 $d_1^+, \ldots, d_n^+$ and $d_1^-, \ldots, d_n^-$ 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 8, 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 4 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 8. *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 Lemma 3, it holds 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 $d_i \overset{d}{=} \xi$. Likewise, since $\alpha > 2$, then 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 6.

Hence, when $\mathbb{E}\xi^2 > 3\mathbb{E}\xi$, then

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

holds w.h.p.

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 $W\,\ell^{-2} \leq F_\xi(\ell) \leq V\,\ell^{-2}$. We construct auxiliary random variables $\xi_\varepsilon \sim \mathcal{P}(2 + \varepsilon)$ for $\varepsilon > 0$. Later we will argue that 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 $\Pr[\xi_\varepsilon \geq \ell] \leq \Pr[\xi \geq \ell]$ for any $\ell \geq 1$, we notice that

$$\Pr[2T_n > S_n] \geq \Pr[2T_n^\varepsilon > S_n], \tag{10}$$

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 observe that for some $\mu > 0$ it holds 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 partitioning of the set of clones into 2-element sets. The important point here is that in the process of 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 of 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 15),

– that if $span(p)$ has grown to the required size, it contains a pair $v, \bar{v}$ w.h.p. (Lemma 17), and

– that the processes initiated at p and $\bar{p}$ do not interact too much w.h.p. (Lemma 12).

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 and [20]. 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 of 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 σ 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:

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 9), 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 in the end $|\mathcal{L}(q)| < \sigma$, declare failure.

Algorithm 2 Procedure TSPAN

1: **procedure** TSPAN$(\mathcal{C}, \mathcal{L}, \mathcal{U}, \sigma, \tau)$
2:     **while** $0 < |\mathcal{L}| \leq \sigma$ **and** less than τ pairings performed **do**
3:         Pick u.a.r. a live clone $c_1 \in \mathcal{L}$
4:         Pick u.a.r. an unpaired clone $c_2 \in \mathcal{U} \cup \mathcal{L} \setminus \{c_1\}$
5:         Pair clones c_1 and c_2, i.e.
6:           $\mathcal{C} \leftarrow \mathcal{C} \cup \{c_1, c_2\}$
7:           $\mathcal{L} \leftarrow \mathcal{L} \setminus \{c_1, c_2\}$
8:           $\mathcal{U} \leftarrow \mathcal{U} \setminus \{c_1, c_2\}$
9:         Let l be the literal associated with c_2
10:        Make live the clones associated with $\bar{l}$, i.e
11:          $\kappa(\bar{l}) = \{c \in \mathcal{S} \mid c \text{ is a clone of } \bar{l}\}$
12:          $\mathcal{L} \leftarrow \mathcal{L} \cup (\mathcal{U} \cap \kappa(\bar{l}))$
13:          $\mathcal{U} \leftarrow \mathcal{U} \setminus \kappa(\bar{l})$
14:     **end while**
15: **end procedure**

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 pair of complementary 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 $|\mathcal{C}| = o(n)$. We will show later that the probability of failure in Steps 4,5 is very low, and therefore we are mainly concern with restarts occurring during Steps 2,3. 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 exploit the value of μ, since it acts as an approximation to the number of newly added live clone into the set of live clones 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 9. *Let* $\frac{2T_n}{S_n} = 1 + \mu$, *where* $\mu > 0$. *Then for any* $t \leq |\mathcal{C}| = o(n)$, *we have*

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

Before proving Lemma 9 we prove three useful auxiliary results regarding X_i's and L_i's:

Lemma 10. *There exists constant $V > 0$, such that for any $d \geq 1$ and $i \leq |\mathcal{C}| = o(n)$*

$$\Pr\left[X_i \geq d \,|\, X_1, X_2, \cdots, X_{i-1}\right] \leq V \, d^{1-\alpha}. \tag{11}$$

Proof. Notice that $X_i \geq 0$ indicates that at the i-th iteration, the TSPAN picked a clone c_i to pair with from the set of untouched clones $\mathcal{U}_{i-1}$. Assume that c_i is associated with literal l_i. Then $X_i = -1 + \deg(\bar{l})$ (according to the TSPAN algorithm, $\mathcal{L}_{i-1}$ always loses one clone and gets $\deg(\bar{l})$ new clones).

Hence, for the live set to gain d clones, c_i must be one of the clones associated with literals whose complements have exactly $d+1$ clones. Thus,

$$
\begin{aligned}
\Pr\left[X_i \geq d \,|\, X_1, X_2, \cdots, X_{i-1}\right] &\leq \frac{1}{S_n - 2i - 1} \mathbb{E}\left[\sum_{l \in L(\phi)} \deg(l) \cdot \mathbb{1}_{\deg(\bar{l}) \geq d+1}\right] \\
&= \frac{1}{S_n - 2i - 1} \sum_{l \in L(\phi)} \mathbb{E}\left[\deg(l) \cdot \mathbb{1}_{\deg(\bar{l}) \geq d+1}\right].
\end{aligned}
\tag{12}
$$

In the above estimate of the probability we claim that the total number of clones of the literals with degree at least $d+1$ is well-approximated by its mean. To see why this is true, note that the tail function of the r.v. $\deg(l) \cdot \mathbb{1}_{\deg(\bar{l}) \geq d+1}$ is upper bounded by the right tail function of the r.v. $\deg(l) = d^+ = d^-$. Hence, for $\alpha > 2$, we expect that

$$\sum_{l \in L(\phi)} \deg(l) \cdot \mathbb{1}_{\deg(\bar{l}) \geq d+1} = (1 + o(1)) \sum_{l \in L(\phi)} \mathbb{E}\left[\deg(l) \cdot \mathbb{1}_{\deg(\bar{l}) \geq d+1}\right],$$

due to Lemma 3.

Now, let us fix some specific literal l and let v be its corresponding variable. Then

$$
\begin{aligned}
\mathbb{E}\left[\deg(l) \cdot \mathbb{1}_{\deg(\bar{l}) \geq d+1}\right] &= \sum_{j=d+1}^{\infty} \sum_{j_0=d+1}^{j} (j - j_0) \Pr\left[\deg(\bar{l}) = j_0 \,|\, \deg(v) = j\right] \Pr\left[\deg(v) = j\right] \\
&\leq \sum_{j=d+1}^{\infty} \sum_{j_0=d+1}^{j} j \, \Pr\left[\deg(\bar{l}) = j_0 \,|\, \deg(v) = j\right] \Pr\left[\deg(v) = j\right] \\
&\leq \sum_{j=d+1}^{\infty} \frac{j}{2^j} \Pr\left[\deg(v) = j\right] \sum_{j_0=0}^{j} \binom{j}{j_0} \\
&\leq \sum_{j=d}^{\infty} j \, \Pr\left[\deg(v) = j\right].
\end{aligned}
$$

Next, we apply summation by parts

$$\mathbb{E}\left[\deg(l) \cdot \mathbb{1}_{\deg(\bar{l}) \geq d+1}\right] \leq \sum_{j=d}^{\infty} j \Pr\left[\deg(v) = j\right]$$

$$= \sum_{j=d}^{\infty} j \Big(\Pr\left[\deg(v) \geq j\right] - \Pr\left[\deg(v) \geq j+1\right] \Big)$$

$$= d \Pr\left[\deg(v) \geq j\right] + \sum_{j=d+1}^{\infty} \Pr\left[\deg(v) \geq j\right].$$

Recall that $\deg(v) \stackrel{d}{=} \xi$, while $\xi \sim \mathcal{P}(\alpha)$; hence, for a fixed l

$$\mathbb{E}\left[\deg(l) \cdot \mathbb{1}_{\deg(\bar{l}) \geq d+1}\right] \leq d \Pr\left[\deg(v) \geq d\right] + \sum_{j=d+1}^{\infty} \Pr\left[\deg(v) \geq j\right] \leq V_r \, d^{1-\alpha},$$

for some constant $V_r > 0$.

Therefore, (12) can be further simplified

$$\Pr\left[X_i \geq d \,|\, X_1, X_2, \cdots, X_{i-1}\right] \leq \frac{1}{S_n - 2i - 1} \sum_{l \in L(\phi)} \mathbb{E}\left[\deg(l) \cdot \mathbb{1}_{\deg(\bar{l}) \geq d+1}\right]$$

$$\leq \frac{1}{S_n - 2i - 1} \sum_{l \in L(\phi)} V_r \, d^{1-\alpha}$$

$$\leq \frac{2n}{S_n - 2i - 1} V_r \, d^{1-\alpha},$$

since $|L(\phi)| \leq 2n$. Recall that w.h.p. $S_n = (1 + o(1))\, n \mathbb{E}\xi$. Then for $i = o(n)$ we obtain

$$\Pr\left[X_i \geq d \,|\, X_1, X_2, \cdots, X_{i-1}\right] \leq \frac{2n}{S_n - 2i - 1} V_r \, d^{1-\alpha}$$

$$= \frac{2n}{(1 + o(1))\, n \mathbb{E}\xi - 2i - 1} V_r \, d^{1-\alpha}$$

$$\leq 2 V_r \, d^{1-\alpha},$$

since $\mathbb{E}\xi \geq 1$. Finally, denote by $V := 2V_r$, and the lemma follows with

$$\Pr\left[X_i \geq d \,|\, X_1, X_2, \cdots, X_{i-1}\right] \leq V \, d^{1-\alpha}.$$

$\square$

Next, we show that $L_i = o(n)$ as long as i is not too large. However, instead of showing this result directly, we will prove a somewhat more general statement from which the desired property naturally follows.

Lemma 11. *Let the sequence $\{\xi_i\}_{i=1}^{cn}$, where $c > 0$ is constant, contains cn independent copies of the r.v. ξ, which has the right tail for any $\ell \geq 1$*

$$F_\xi(\ell) = \Pr[\xi \geq \ell] \leq V\, \ell^{-\alpha},$$

with $V > 0$ and $\alpha > 1$. Let $A \subset \{\xi_i\}_{i=1}^{cn}$ be any subset of size $t = n^\beta$ ($0 < \beta < 1$). Then w.h.p.

$$\sum_{X \in A} X = O\left(n^{\beta + \frac{1-\beta}{\alpha}}\right).$$

Proof. Let us call a variable ξ_j "heavy", if $\xi_j \geq n^{\frac{1-\beta}{\alpha}}$. Then the expected sum of "heavy" variables in the original sequence $\{\xi_i\}_{i=1}^{cn}$ is

$$\mathbb{E}\left[\sum_{1 \leq i \leq cn} \xi_i\, \mathbb{1}_{\xi_i \geq n^{\frac{1-\beta}{\alpha}}}\right] = cn \sum_{d \geq n^{\frac{1-\beta}{\alpha}}} d \Pr[\xi = d]$$

$$= cn \sum_{d \geq n^{\frac{1-\beta}{\alpha}}} d\Big(\Pr[\xi \geq d] - \Pr[\xi \geq d+1]\Big)$$

$$= cn \sum_{d \geq n^{\frac{1-\beta}{\alpha}}} d\Pr[\xi \geq d] - cn \sum_{d \geq n^{\frac{1-\beta}{\alpha}}+1} (d-1)\Pr[\xi \geq d]$$

$$= cn^{1 + \frac{1-\beta}{\alpha}} \Pr\left[\xi \geq n^{\frac{1-\beta}{\alpha}}\right] + cn \sum_{d \geq n^{\frac{1-\beta}{\alpha}}+1} \Pr[\xi \geq d]$$

$$\leq cn^{1 + \frac{1-\beta}{\alpha}} F_\xi\left(n^{\frac{1-\beta}{\alpha}}\right) + cn \sum_{d \geq n^{\frac{1-\beta}{\alpha}}} F_\xi(d).$$

Next, recall that $F_\xi\left(n^{\frac{1-\beta}{\alpha}}\right) \leq V\, n^{\beta-1}$. Then

$$\mathbb{E}\left[\sum_{1 \leq i \leq n} \xi_i\, \mathbb{1}_{\xi_i \geq n^{\frac{1-\beta}{\alpha}}}\right] \leq cn^{1 + \frac{1-\beta}{\alpha}} F_\xi\left(n^{\frac{1-\beta}{\alpha}}\right) + cn \sum_{d \geq n^{\frac{1-\beta}{\alpha}}} F_\xi(d) = O\left(n^{\beta + \frac{1-\beta}{\alpha}}\right).$$

Since the right tail functions of the r.vs. $\xi_i\, \mathbb{1}_{\xi_i \geq n^{\frac{1-\beta}{\alpha}}}$ is upper bounded by the right tail function of the original non-truncated variable ξ, the sum of "heavy" variables is concentrated around its mean. Therefore, even if A contained only the "heavy" variables, the resulting sum would be at most $O\left(n^{\beta + \frac{1-\beta}{\alpha}}\right)$. However, if we pick "non-heavy" variables, meaning the ones that have degree at most $n^{\frac{1-\beta}{\alpha}}$, then again the sum would be at most $O\left(n^{\beta + \frac{1-\beta}{\alpha}}\right)$. And so no matter what variables the set A includes, the sum of its elements is $O\left(n^{\beta + \frac{1-\beta}{\alpha}}\right)$ w.h.p. $\square$

However, our concern was to bound the size of the live set L_i. And the next corollary shows that we expect $L_i = o(n)$ as long as i is not too large.

Corollary 3. *When $t \leq |\mathcal{C}| = o(n)$, then $L_t = \sum_{i=1}^{t} X_i = o(n)$ holds w.h.p.*

Proof. Proof is pretty straightforward now, given Lemma 11. We have at most $2n$ literals (it is not exactly $2n$, since some variables may produce only one literal). And the right tail function of the literals is $\Pr\left[\deg(l) \geq \ell\right] \leq V \ell^{-\alpha}$, due to (3). Since $\alpha > 2$, and after applying Lemma 11 we obtain $L_t = \sum_{i=1}^{t} X_i = o(n)$ holds w.h.p. $\qquad\square$

Now we are in a position to prove Lemma 9.

Proof. [Proof of Lemma 9] Assume that at the t-th iteration of TSPAN procedure we paired clones (p, p'). Clearly, according to the algorithm, $p \in \mathcal{L}_{t-1}$, thus, $\mathcal{L}_i$ loses at least one clone, and, with probability $\frac{L_{t-1}}{S_n - 2t + 1}$, p' can be also from $\mathcal{L}_{t-1}$, making the live set to lose another clone.

Next let us call variable v_i *undisturbed*, if none of its clones were paired or made live. Then, if p is paired with a positive clone of the undisturbed variable v_i (and we can do so in d_i^+ ways with uniform probability $\frac{1}{S_n - 2t - 1}$), then the live set will gain d_i^- new clones. Similarly, when p is paired with a negative clone of v_i (which can be done in d_i^- ways with uniform probability $\frac{1}{S_n - 2t - 1}$), then the live set gains d_i^+ new clones.

Note, it may happen that p is paired with clones of a "disturbed" variable, which may add some clones into $\mathcal{L}_t$, but since we are looking for a lower bound, we can ignore this case. Therefore,

$$
\begin{aligned}
\mathbb{E}\left[X_t \mid X_1, \ldots, X_{t-1}\right] & \\
&= \mathbb{E}\left[L_t - L_{t-1} \mid X_1, \ldots, X_{t-1}\right] \\
&\geq -1 - \frac{L_{t-1}}{S - 2t - 1} + \frac{1}{S - 2t - 1} \sum_{v_j \text{ undisturbed}} \left(d_j^+ d_j^- + d_j^- d_j^+\right) \\
&\geq -1 + \frac{2}{S_n - 2t - 1}\left(\sum_{j=1}^{n} d_j^+ d_j^- - \sum_{v_j \text{ disturbed}} d_j^+ d_j^- - \frac{L_{t-1}}{2}\right).
\end{aligned}
$$

Recall that $\sum_{j=1}^{n} d_j^+ d_j^- = T_n$, and w.h.p. $T_n = (1 + o(1))\, n \frac{\mathbb{E}\xi^2 - \mathbb{E}\xi}{4}$. Then

$$
\begin{aligned}
\mathbb{E}\left[X_t \mid X_1, \ldots, X_{t-1}\right] & \\
&\geq -1 + \frac{2}{S_n - 2t - 1}\left(\sum_{j=1}^{n} d_j^+ d_j^- - \sum_{v_j \text{ disturbed}} d_j^+ d_j^- - \frac{L_{t-1}}{2}\right) \\
&\geq -1 + \frac{2}{S_n - 2t - 1}\left(T_n - \sum_{v_j \text{ disturbed}} d_j^+ d_j^- - \frac{L_{t-1}}{2}\right).
\end{aligned}
$$

The sum $\sum_{v_j \text{ disturbed}} d_j^+ d_j^-$ is the sum of at most $t = o(n)$ random variables $d_j^+ d_j^-$. Note, though, that these variables $d_j^+ d_j^-$ are not independent, nor identically distributed in the aforementioned sum, since during the first iterations

the TSPAN procedure favours products of complementary literals having larger numbers of clones. Now, every element $d^+ d^- \in \{d_i^+ d_i^- \mid v_i \in V(\phi)\}$ has the right tail function $\Pr[d^+ d^- \geq \ell] \leq 2V \, \ell^{-\alpha/2}$ by Corollary 1 for some $V > 0$ and every $\ell \geq 1$. Thus, since $\alpha > 2$, and according to Lemma 11, we obtain

$$\sum_{v_j \text{ disturbed}} d_j^+ d_j^- = o(n).$$

As for the L_{t-1} term, we know from Corollary 3 that it remains $o(n)$ as well.
Hence,

$$
\begin{aligned}
\mathbb{E}\left[X_t \mid X_1, \ldots, X_{t-1}\right] &\geq -1 + \frac{2}{S_n - 2t - 1}\left(T_n - \sum_{j \text{ disturbed}} d_j^+ d_j^- - \frac{L_{t-1}}{2}\right) \\
&\geq -1 + \frac{2}{S_n - o(n)}\left(T_n - o(n)\right) \\
&\geq -1 + \frac{2T_n}{S_n}\left(1 - o(1)\right) \\
&\geq -1 + (1 + \mu)(1 - o(1)) \\
&\geq \mu/2.
\end{aligned}
$$

$\square$

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 12 (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 $o\left(\frac{1}{s_1}\right)$.

Proof. (1) 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 has not 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 had 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 "successful" 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 is $\Pr[r_1 = 1]$, 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,

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

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

$$P := \Pr[r_1 = 1] \geq \frac{\Pr[L_{s_1} \geq s_1\mu/6]}{s_1}. \tag{13}$$

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 having the right tail (Lemma 10)

$$\Pr\left[X_i \geq \ell \,|\, X_1, \ldots, X_{i-1}\right] \leq V \ell^{-\alpha},$$

and positive expectation (Lemma 9)

$$\mathbb{E}\left[X_i \,|\, X_1, \ldots, X_{i-1}\right] \geq \frac{\mu}{2}.$$

Therefore, according to *Azuma-like inequality* (Lemma 7), we obtain

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

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

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

Thus, from (13) it follows that $P \geq \frac{\Pr[L_{s_1} \geq s_1\mu/6]}{s_1} = \frac{1}{2s_1}$, which proves the first part of the lemma.

(2) To pair clone q, we must select it uniformly among $|\mathcal{S}| - |\mathcal{C}|$ non-paired clones. Hence, due to Union bound, we have

$$
\begin{aligned}
\Pr\left[q \notin \mathcal{U}_{s_1}\right] &\le \Pr\left[\bigcup_{i=1}^{s_1}\{q \in \mathcal{C}_i\}\right] \\
&\le \sum_{i=1}^{s_1}\Pr\left[q \in \mathcal{C}_i\right] \\
&= s_1 \frac{1}{|\mathcal{S}| - |\mathcal{C}|} \\
&= \frac{s_1}{S_n - o(n)}, \text{ since } |\mathcal{C}| = t = o(n), \text{ and } |\mathcal{S}| = S_n \\
&= (1 + o(1))\frac{n^{\frac{\alpha+4}{6(\alpha+1)}}}{n\mathbb{E}\xi}, \text{ since } S_n = (1 + o(1))\, n\mathbb{E}\xi \text{ w.h.p.} \\
&= o\left(\frac{1}{s_1}\right).
\end{aligned}
$$

$\square$

Note that in Lemma 12(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 12(2) the bound on the probability is only useful when s_1 is sufficiently large.

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

Lemma 13 (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 does not include q nor make it live, is at least $\frac{1}{5s_1^2}$.*

However, before proving this result, we start with an auxiliary lemma after which we provide the proof of Lemma 13.

Lemma 14. *The TSPAN procedure will pair at most $o(s_1)$ clones from the set of live clones $\mathcal{L}(p)$, while constructing the span from q, when $|\mathcal{C}| = o(n)$.*

Proof. Since the TSPAN from p and q runs for at most s_1 steps, and $|\mathcal{L}(p)| = o(n)$ w.h.p. due to Corollary 3, it follows that the expected number of paired clones, which belong to the set $\mathcal{L}(p)$ is at most

$$
\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{s_1}\mathbb{1}_{\text{u.a.r. picked clone } c \in \mathcal{L}(p)}\right] &= \sum_{i=1}^{s_1}\Pr\left[\text{u.a.r. picked clone } c \in \mathcal{L}(p)\right] \\
&= \sum_{i=1}^{s_1}\frac{|\mathcal{L}(p)|}{|\mathcal{S}| - |\mathcal{C}|} \\
&= \sum_{i=1}^{s_1}\frac{o(n)}{(1 + o(1))\, n\mathbb{E}\xi - o(n)},
\end{aligned}
$$

where the last step follows from $|\mathcal{S}| = S_n = (1 + o(1)) \, n\mathbb{E}\xi$ w.h.p. Hence,

$$\mathbb{E}\left[\sum_{i=1}^{s_1} \mathbb{1}_{\text{u.a.r. picked clone } c \in \mathcal{L}(p)}\right] = \sum_{i=1}^{s_1} \frac{o(n)}{(1 + o(1)) \, n\mathbb{E}\xi - o(n)} = o(s_1).$$

Moreover, since the process $\sum_{i=1}^{s_1} \mathbb{1}_{\text{u.a.r. picked clone } c \in \mathcal{L}(p)}$ forms a binomial trial, it follows that the actual number of paired clones does not deviate much from its expectation, and hence, w.h.p. $\sum_{i=1}^{s_1} \mathbb{1}_{\text{u.a.r. picked clone } c \in \mathcal{L}(p)} = o(s_1)$. $\square$

Thus, after growing span from clone p, there is a good probability that q is not paired. So we can run the same TSPAN algorithm for clone q now (STEP 2), and Lemma 12 suggests that we will be able to construct the second span of size $s_1\mu/6$ with probability at least $\frac{1}{2s_1}$.

However, we need to make sure that while the TSPAN process constructs span from q, it does not pair too many clones that were marked as "live" and placed into $\mathcal{L}(p)$. But since $L_i = o(n)$ w.h.p. for $i \leq t = o(n)$, this "bad" event is unlikely to happen.

Proof (of Lemma 13). Let $\mathcal{L}_p^I$ and $\mathcal{L}_q^I$ be the state of the sets of live clones $\mathcal{L}(p)$ and $\mathcal{L}(q)$ respectively after finishing STEP 2 and STEP 3, while $\mathcal{U}^I$ be the state of the set $\mathcal{U}$ after finishing STEP 2. Then the probability that STEP 2 and STEP 3 do not fail is at least

$$\Pr\left[\{\mathcal{L}_p^I \geq s_1\mu/6\} \wedge \{\mathcal{L}_q^I \geq s_1\mu/6\} \wedge \{q \in \mathcal{U}^I\}\right] =$$
$$= \Pr\left[\mathcal{L}_p^I \geq s_1\mu/6 \,|\, \{\mathcal{L}_q^I \geq s_1\mu/6\} \wedge \{q \in \mathcal{U}^I\}\right] \times$$
$$\times \Pr\left[\{\mathcal{L}_q^I \geq s_1\mu/6\} \wedge \{q \in \mathcal{U}^I\}\right]$$
$$\geq \Pr\left[\mathcal{L}_p^I \geq s_1\mu/6 \,|\, \{\mathcal{L}_q^I \geq s_1\mu/6\} \wedge \{q \in \mathcal{U}^I\}\right] \times$$
$$\times \left(\Pr\left[\mathcal{L}_q^I \geq s_1\mu/6\right] - \Pr\left[q \notin \mathcal{U}^I\right]\right)$$
$$\geq \frac{1}{2s_1}\left(\frac{1}{2s_1} - \Pr\left[q \notin \mathcal{U}^q\right]\right),$$

since $\Pr\left[\mathcal{L}_p^I \geq s_1\mu/6\right] = \Pr\left[\mathcal{L}_q^I \geq s_1\mu/6\right] > \frac{1}{2s_1}$ by Lemma 12. Next recall that

$$\Pr\left[q \notin \mathcal{U}^I\right] = o\left(\frac{1}{s_1}\right)$$

from Lemma 12.2, and so it follows that

$$\Pr\left[\{\mathcal{L}_p^I \geq s_1\mu/6\} \wedge \{\mathcal{L}_q^I \geq s_1\mu/6\} \wedge \{q \in \mathcal{U}^I\}\right] \geq$$
$$\geq \frac{1}{2s_1}\left(\frac{1}{2s_1} - \Pr[q \notin \mathcal{U}^I]\right)$$
$$\geq \frac{1}{2s_1}\left(\frac{1}{2s_1} - o\left(\frac{1}{s_1}\right)\right)$$

$$\geq \frac{1}{5s_1^2},$$

and the lemma follows.   □

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

Lemma 15 (Steps 4,5). *Assume that p- and q-spans were both able to accumulate at least $s_1\mu/7$ 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, while maintaining the sizes of the sets $\mathcal{L}(p)$ and $\mathcal{L}(q)$ at the level of at least $s_1\mu/8$ live clones at each iteration.*

We start by showing that TSPAN in Step 4 does not decimate $\mathcal{L}(q)$.

Lemma 16. *Let Steps 2,3 succeed. Then the number of clones from $\mathcal{L}(q)$ that will be paired after finishing Step 4 is at most $o(s_1)$ w.h.p.*

Proof. Let $\mathcal{L}(p)$ and $\mathcal{L}(q)$ be the sets of live clones right after finishing Step 2 and Step 3. Then $|\mathcal{L}(p)| \geq s_1\mu/7$ and $|\mathcal{L}(q)| \geq s_1\mu/7$ by Lemma 13. Hence, the expected number of clones in $\mathcal{L}(q)$, which get also marked as "live" clones in $\mathcal{L}(p)$ by TSPAN, while it expands the p-span for another s_2 iterations, is

$$\mathbb{E}\left[\sum_{i=1}^{s_2} \mathbb{1}_{\text{u.a.r. picked clone } c \in \mathcal{L}(q)}\right] = \sum_{i=1}^{s_2} \Pr\left[\text{u.a.r. picked clone } c \in \mathcal{L}(q)\right]$$
$$= \sum_{i=1}^{s_2} \frac{|\mathcal{L}(q)|}{|\mathcal{S}| - |\mathcal{C}|}$$
$$= (1 + o(1)) \sum_{i=1}^{s_2} \frac{|\mathcal{L}(q)|}{n\mathbb{E}\xi},$$

since $|\mathcal{S}| - |\mathcal{C}| = S_n - o(n) = (1 + o(1))\,n\mathbb{E}\xi$ w.h.p. Now recall that $|\mathcal{L}(q)|$ is the sum of at most s_1 r.vs. $\{Z_i\}_{i=1}^{s_1}$, where each Z_i represents the number of clones of literals being added into $\mathcal{L}(q)$. Therefore, $\{Z_i\}_{i=1}^{s_1}$ is the subset of the multi-set of degrees of all literals in ϕ, i.e. $\{Z_i\}_{i=1}^{s_1} \subseteq D$, where $D := \{\deg(l)\,|\,l \in L(\phi)\}$ and $|D| \leq 2n$. However, each element in D is a realization of a r.v. with the right tail function $\Pr[\deg(l) \geq \ell] \leq V\,\ell^{-\alpha}$ by Corollary 1.

Then, according to Lemma 11, w.h.p.

$$|\mathcal{L}(q)| = O\left(n^{\frac{\alpha+4}{6(\alpha+1)} + \frac{1 - \frac{\alpha+4}{6(\alpha+1)}}{\alpha}}\right) = O\left(n^{\frac{\alpha^2+9\alpha+2}{6\alpha(\alpha+1)}}\right).$$

Thus, we obtain

$$
\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{s_2} \mathbb{1}_{\text{u.a.r. picked clone } c \in \mathcal{L}(q)}\right] &= (1 + o(1)) \sum_{i=1}^{s_2} \frac{|\mathcal{L}(q)|}{n\mathbb{E}\xi} \\
&= (1 + o(1))\, s_2 \frac{|\mathcal{L}(q)|}{n\mathbb{E}\xi} \\
&= (1 + o(1))\, s_2\, O\left(\frac{n^{\frac{\alpha^2 + 9\alpha + 2}{6\alpha(\alpha+1)}}}{n}\right) \\
&= o\left(n^{\frac{\alpha+4}{6(\alpha+1)}}\right), \quad \text{when } \alpha > 2 \\
&= o(s_1).
\end{aligned}
$$

And since $\sum_{i=1}^{s_2} \mathbb{1}_{\text{u.a.r. picked clone } c \in \mathcal{L}(q)}$ forms a binomial trial, it follows that the actual number of clones from $\mathcal{L}(q)$ that get paired by the p-span is concentrated around its expectation. Hence, we do not expect more than $o(s_1)$ clones from $\mathcal{L}(q)$ to be paired, while expanding the span from p. □

Therefore, we conclude that after finishing Steps 2,3 and given that the Step 4 succeeds, the size of the live set $\mathcal{L}(q)$ is at least $s_1\mu/7 - o(1) \geq s_1\mu/7.5$.

However, now we need to show that Steps 4,5 succeed w.h.p., assuming that Steps 1,2 did not fail.

Proof (of Lemma 15). When we start from a live set with size of order at least $s_1\mu/7.5$, to be able to grow the span for another s_2 steps, we need to make sure that the size of the live set never drops to zero. First, recall that $Ks_1 + s_2 = o(n)$, where $K = n^{\frac{7\alpha+10}{12(\alpha+1)}}$ is the maximal number of tries to pick complementary clones from which TSPAN grows spans. Thus, according to Lemma 9, for any $j = o(n)$

$$
\mathbb{E}\left[X_{s_1+j} \mid X_1, \ldots, X_{s_1+j-1}\right] \geq \mu/2 > 0
$$

and for any $d \geq 1$

$$
\Pr\left[X_{s_1+j} \geq d \mid X_1, \ldots, X_{s_1+j-1}\right] \leq V\, d^{-\alpha}.
$$

Hence

$$
\begin{aligned}
\Pr\left[L_{s_1+j} \leq s_1 \frac{\mu}{8}\right] &\leq \Pr\left[L_{s_1+j} \leq \left(s_1 \frac{\mu}{2}\right) \frac{2}{7.5}\right] \\
&= \Pr\left[L_{s_1} + \sum_{i=1}^{j} X_{s_1+i} \leq \left(s_1 \frac{\mu}{2}\right) \frac{2}{7.5}\right],
\end{aligned}
$$

which, after applying the Azuma-like inequality (Lemma 7), leads to an upper bound

$$\Pr\left[L_{s_1+j} \le s_1\frac{\mu}{8}\right] \le \Pr\left[L_{s_1} + \sum_{i=1}^{j} X_{s_1+i} \le \left(s_1\frac{\mu}{2}\right)\frac{2}{7.5}\right]$$

$$\le \exp\left(-\frac{j+L_{s_1}}{4\log^2 s_2}\mu^2\frac{121}{900}\right)$$

$$\le \exp\left(-\frac{j+s_1\mu}{\log^2 s_2}\mu^2\frac{121}{6750}\right), \text{ since } L_{s_1} \ge s_1\mu/7.5$$

$$\le \exp\left(-\frac{s_1\mu}{\log^2 s_2}\frac{\mu^2}{6750}\right).$$

Recall that $s_1 = n^{\frac{\alpha+4}{6(\alpha+1)}}$ and $s_2 = n^{\frac{11\alpha^2+3\alpha-2}{12\alpha(\alpha+1)}}$. Hence, we obtain

$$\Pr\left[L_{s_1+j} \le s_1\frac{\mu}{8}\right] \le \exp\left(-\frac{s_1\mu}{\log^2 s_2}\frac{\mu^2}{6750}\right) \le \exp\left(-C\frac{n^{\frac{\alpha+4}{6(\alpha+1)}}}{\log^2 n}\right),$$

for *specific* $1 \le j \le s_2$ and for some constant $C > 0$. Then the probability that the live set will drop below the $s_1\mu/8$ level during *any* of the $1 \le j \le s_2$ steps is, by *Union bound*,

$$\Pr[\bigcup_{j=1}^{s_2}\{L_{s_1+j} \le s_1\frac{\mu}{8}\}] \le \sum_{j=1}^{s_2}\Pr\left[L_{s_1+j} \le s_1\frac{\mu}{8}\right]$$

$$\le s_2 \cdot \exp\left(-C\frac{n^{\frac{\alpha+4}{6(\alpha+1)}}}{\log^2 n}\right)$$

$$= o(1).$$

$\square$

Therefore, by Lemma 16, after the TSPAN finishes constructing the p-span in STEP 4, the size of the "live" set of q-span is at least

$$|\mathcal{L}(q)| \ge s_1\mu/7 - o(s1) \ge s_1\frac{\mu}{7.5}.$$

To keep our calculations as simple as possible, we will assume that both $|\mathcal{L}(p)|$ and $|\mathcal{L}(p)|$ are at least $s_1\frac{\mu}{8}$ during the execution of STEPS 4,5.

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

Proof. Let B ("bad") be the event that after performing $s_1 + s_2$ steps, there are no two complementary clones ever paired. This means, as was mentioned previously, that at each step TSPAN was choosing u.a.r. only untouched clones $c \in \mathcal{U}$ among $\mathcal{L} \cup \mathcal{U}$ clones.

Recall that sets $\mathcal{U}, \mathcal{L}$, and $\mathcal{C}$ form a partition of the set $\mathcal{S}$ of all clones. Then at any moment of time i we have $S_n = C_i + U_i + L_i = 2i + U_i + L_i$. Hence,

$$\Pr[B] = \prod_{i=1}^{s_1+s_2} \frac{U_i}{L_i + U_i - 1} = \prod_{i=1}^{s_1+s_2} \left(1 - \frac{L_i}{L_i + U_i - 1}\right) \leq \prod_{i=0}^{s_2} \left(1 - \frac{L_{s_1+i}}{S_n}\right).$$

Next, Lemma 15 implies, that $L_{s_1+i} \geq s_1\mu/8$ for $0 \leq i \leq s_2$, so

$$\Pr[B] = \prod_{i=0}^{s_2} \left(1 - \frac{L_{t+i}}{S_n}\right) \leq \prod_{i=0}^{s_2} \left(1 - \frac{s_1\mu}{8S_n}\right) \leq \exp\left(-\frac{s_1 s_2 \mu}{8S_n}\right).$$

Having $s_1 = n^{\frac{\alpha+4}{6(\alpha+1)}}$, $s_2 = n^{\frac{11\alpha^2+3\alpha-2}{12\alpha(\alpha+1)}}$, and $S_n = (1+o(1))\, n\mathbb{E}\xi$ w.h.p., we obtain

$$\Pr[B] \leq \exp\left(-\frac{s_1 s_2 \mu}{8S_n}\right) \leq \exp\left(-n^{\frac{\alpha^2-\alpha-2}{12\alpha(\alpha+1)}}\right),$$

which proves the lemma. $\square$

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.*

Proof. Recall that our algorithm grows spans for a sequence of pairs of complementary clones waiting for the first success, that is, contradictory paths. First, we estimate the probability of finding contradictory paths for some pair of complementary clones in this sequence. Suppose that at some point we picked clones p and q, and let $\{L_{s_1}(p) \geq s_1\mu/7\}$, $\{L_{2s_1}(q) \geq s_1\mu/7\}$, and $\{q \in \mathcal{U}_{s_1}\}$ be the events from Lemma 13. Then by Lemma 13 with probability at least $\frac{1}{5s_1^2}$ we will be able to accumulate in each of the live sets at least $s_1\mu/7$ elements, while constructing spans from the complementary clones p and q. Thus,

$$\Pr[A_1] := \Pr\left[\{L_{s_1}(p) \geq s_1\mu/7\} \wedge \{L_{2s_1}(q) \geq s_1\mu/7\} \wedge \{q \in \mathcal{U}_{s_1}\}\right] \geq \frac{1}{5s_1^2}.$$

Next, Lemma 15 implies that if event A_1 happens, then with probability at least $1 - o(1)$ we will be able to grow both spans for another s_2 iterations, such that sizes of the live sets never drop below $s_1\mu/8$ clones, i.e.

$$\Pr[A_2 \mid A_1] := \Pr\left[\bigcap_{j=1}^{s_2} \{L_{s_1+j}(p) \geq s_1\mu/8\} \wedge \bigcap_{j=1}^{s_2} \{L_{2s_1+s_2+j}(q) \geq s_1\mu/8\} \mid A_1\right]$$
$$\geq 1 - o(1).$$

If the events A_1 and A_2 happen, then by Lemma 17 with probability at least $1 - o(1)$ the corresponding span contains 2 complementary clones. Let us denote by $Comp(p)$ the event that there exist 2 complementary clones in the span originating at p.

Then, the probability of event $Comp(p, q)$ that a pair of *fixed* complementary clones p and q, after completing STEPS 1–5 form contradictory paths is at least

$$\Pr\left[Comp(p, q)\right] = \Pr[Comp(p) \wedge Comp(q) \wedge A_1 \wedge A_2]$$
$$\geq \Pr\left[Comp(p) \wedge Comp(q)\right] \Pr\left[A_1 \wedge A_2\right]$$
$$= \Pr[\bar{B}]^2 \Pr\left[A_2 \,|\, A_1\right] \Pr\left[A_1\right]$$
$$\geq \left(1 - \exp\left(-n^{\frac{\alpha^2 - \alpha - 2}{12\alpha(\alpha+1)}}\right)\right)^2 (1 - o(1)) \frac{1}{5s_1^2}$$
$$= (1 - o(1)) \frac{1}{5s_1^2} \quad \geq \frac{1}{6s_1^2} =: P_{final},$$

when $n \to \infty$. In other words, P_{final} is a lower bound of the probability that 2 *specific* complementary clones p and q form contradictory paths. However, recall that we have $T_n = \sum_{i=1}^{n} d_i^+ d_i^-$ different pairs of complementary clones, and, moreover, w.h.p. $T_n = (1 + o(1)) \, n \frac{\mathbb{E}\xi^2 - \mathbb{E}\xi}{4}$, when $\alpha > 2$. Since we repeat the TSPAN procedure for at most $K = n^{\frac{7\alpha+10}{12(\alpha+1)}} \ll T_n$ different pair of clones, the probability that none of the picked pairs form a contradiction path is at most

$$\Pr\left[\bigcap_{i=1}^{K} \overline{Comp(p_i, q_i)}\right] \leq \prod_{i=1}^{K} (1 - \Pr\left[Comp(p, q)\right])$$
$$= \prod_{i=1}^{K} \left(1 - \frac{1}{6s_1^2}\right) \quad \leq \exp\left(-\frac{K}{6s_1^2}\right)$$
$$= \exp\left(-\frac{1}{6} n^{\frac{7\alpha+10}{12(\alpha+1)}} \, n^{-2\frac{\alpha+4}{6(\alpha+1)}}\right), \text{ since } s_1 = n^{\frac{\alpha+4}{6(\alpha+1)}}$$
$$= \exp\left(-\frac{1}{6} n^{\frac{\alpha-2}{4(\alpha+1)}}\right)$$
$$= o(1), \text{ since } \alpha > 2,$$

when $n \to \infty$. Thus, we've obtained that when $\mathbb{E}\xi^2 > 3\mathbb{E}\xi$, then w.h.p. we expect that some pair of clones will form contradictory paths, witnessing unsatisfiability of the formula ϕ. And so the proposition follows. $\qquad \square$

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 shown in Lemma 9, 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 $n \to \infty$, 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 18. *If $\frac{2T_n}{S_n} = 1 - \mu' < 1$, then paths in ϕ are of length $O(\log n)$, w.h.p.*

Proof. Let $k_0 = \lceil \frac{6}{\mu'} \log n \rceil$, while $\{l_1 \implies l_2\}$ denotes the event that there exists a path (see section 2.3 for the definition) from literal l_1 to l_2. Additionally, let P_k be the number of paths of length k in ϕ. Also recall that by $L(\phi)$ we denote the set of all literals in ϕ, while $\deg(l)$, where l is some literal, denotes the total number of clones associated with the literal l. Then

$$\mathbb{E}P_{k_0} \leq \sum_{l_1,\ldots,l_{k_0} \in L(\phi)} \Pr\left[l_1 \implies l_2 \implies \cdots \implies l_{k_0} \in G_I(\phi)\right]$$

$$\leq \sum_{l_1,\ldots,l_{k_0} \in L(\phi)} \frac{2\deg(\bar{l}_1)\deg(l_2)}{S_n - 1}\frac{2\deg(\bar{l}_2)\deg(l_3)}{S_n - 3}\cdots\frac{2\deg(\bar{l}_{k_0-1})\deg(l_{k_0})}{S_n - 2k_0 + 3}$$

$$\leq \sum_{l_1,l_{k_0}} \frac{2\Delta^2}{S_n - 2k_0} \sum_{l_2,\ldots,l_{k_0-1}} \prod_{i=2}^{k_0-1} \frac{2\deg(\bar{l}_i)\deg(l_i)}{S_n - 2k_0}$$

$$\leq \frac{2n^2\Delta^2}{S_n - 2k_0}\left(\frac{2T_n}{S_n - 2k_0}\right)^{k_0-2}$$

$$\leq (1 + o(1))\frac{2n^2\Delta^2}{S_n}\left((1 + o(1))\frac{2T_n}{S_n}\right)^{k_0-2}.$$

Since w.h.p. $S_n = (1 + o(1)) \, n \mathbb{E}\xi$, and $\frac{2T_n}{S_n} = 1 - \mu'$, we have

$$\mathbb{E}P_{k_0} \leq (1 + o(1))\frac{2n^2\Delta^2}{S_n}\left((1 + o(1))\frac{2T_n}{S_n}\right)^{k_0-2}$$

$$\leq \frac{2n\Delta^2}{\mathbb{E}\xi}\left((1 + o(1))(1 - \mu')\right)^{k_0-2}$$

$$\leq 2n\Delta^2\left((1 + o(1))(1 - \mu')\right)^{k_0-2},$$

where the last inequality follows from the fact that $\mathbb{E}\xi \geq 1$ for any r.v. $\xi \geq 1$. Next, given large enough n, we can assume that $(1 + o(1))(1 - \mu') \leq 1 - \frac{\mu'}{2}$. Hence,

$$\mathbb{E}P_{k_0} \leq 2n\Delta^2\left((1 + o(1))(1 - \mu')\right)^{k_0-2}$$

$$\leq 2n\Delta^2\left(1 - \frac{\mu'}{2}\right)^{k_0-2}$$

$$= O\left(n\Delta^2 e^{-\mu' k_0/2}\right)$$

$$= O\left(\frac{\Delta^2}{n^2}\right), \text{ since } k_0 = \lceil\frac{6}{\mu'}\log n\rceil.$$

Next, since $\alpha > 2$, then $\Delta = o\left(n^{1/2}\right)$ (see Lemma 4). However, then

$$\mathbb{E}P_{k_0} = O\left(\frac{\Delta^2}{n^2}\right) = o(1),$$

and so w.h.p. there are no paths of length greater than k_0 in $G_I(\phi)$. $\qquad\square$

Next we give a straightforward estimation of the number of "short" bicycles.

Lemma 19. *If $\frac{2T_n}{S_n} = 1 - \mu' < 1$, then for any k the expected number of bicycles of length k is at most $(1 + o(1))\frac{2\Delta^2}{S_n}\left((1 + o(1))(1 - \mu')\right)^k$.*

Proof. Let B_k be the number of bicycles of length k. Then simple calculation verifies that

$$\mathbb{E}B_k \leq \sum_{\substack{l_1,\ldots,l_k \in L(\phi) \\ u,v \in \{l_1,\bar{l}_1,\ldots,l_k,\bar{l}_k\}}} \frac{2\deg(\bar{l}_1)\deg(l_2)}{S_n - 1}\cdots\frac{2\deg(\bar{l}_{k-1})\deg(l_k)}{S_n - 2k + 5}\frac{\deg(u)\deg(l_1)}{S_n - 2k + 3}\times$$

$$\times\frac{\deg(v)\deg(\bar{l}_k)}{S_n - 2k + 1}$$

$$\leq (1 + o(1))\frac{2\Delta^2}{S_n}\sum_{\substack{l_1,\ldots,l_k \in L(\phi) \\ u,v \in \{l_1,\bar{l}_1,\ldots,l_k,\bar{l}_k\}}}(1 + o(1))^k\frac{\prod_{i=1}^{k}2\deg(\bar{l}_i)\deg(l_i)}{S_n^k}$$

$$\leq (1 + o(1)) \frac{2\Delta^2}{S_n} (1 + o(1))^k \left(\frac{\sum_{l \in L(\phi)} 2 \deg(\bar{l}) \deg(l)}{S_n} \right)^k$$

$$\leq (1 + o(1)) \frac{2\Delta^2}{S_n} (1 + o(1))^k \left(\frac{2T_n}{S_n} \right)^k$$

$$= (1 + o(1)) \frac{2\Delta^2}{S_n} \left((1 + o(1))(1 - \mu') \right)^k,$$

and the result follows. $\qquad\square$

Hence, the above 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.*

Proof. Let B be the number of all bicycles in ϕ, while B_k is the number of bicycles of length k. Due to Lemma 18, we expect no paths longer than $k_0 = \lceil \frac{6}{\mu'} \log n \rceil$, which means there are no bicycles longer than k_0. Hence, $B = \sum_{k=2}^{k_0} B_k$. Then by Markov's inequality

$$\Pr[B > 0] \leq \mathbb{E}B = \sum_{k=2}^{k_0} \mathbb{E}B_k \leq (1 + o(1)) \frac{2\Delta^2}{S_n} \sum_{k=2}^{k_0} \left((1 + o(1))(1 - \mu') \right)^k.$$

Given large enough n, we can assume that $(1 + o(1))(1 - \mu') \leq 1 - \frac{\mu'}{2}$. Then

$$\Pr[B > 0] \leq (1 + o(1)) \frac{2\Delta^2}{S_n} \sum_{k=2}^{k_0} \left((1 + o(1))(1 - \mu') \right)^k$$

$$\leq (1 + o(1)) \frac{2\Delta^2}{S_n} \sum_{k=1}^{\infty} e^{-\mu' k/2}$$

$$\leq \frac{(1 + o(1))}{e^{\mu'/2} - 1} \frac{2\Delta^2}{S_n}$$

$$= O\left(\frac{\Delta^2}{S_n} \right).$$

Since $\alpha > 2$, we have $S_n = (1 + o(1)) n \mathbb{E}\xi$ and $\Delta = o(n^{1/2})$ w.h.p. Then

$$\Pr[B > 0] = O\left(\frac{\Delta^2}{S_n} \right) = o(1).$$

Hence, w.h.p. we do not expect short bicycles in ϕ, and, since there are no long paths, it follows that w.h.p. ϕ does not have bicycles at all. $\qquad\square$

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.*

Proof. Since $\mathbb{E}\xi^2$ and $\mathbb{E}\xi$ are both finite, we conclude that $\alpha > 2$, and so from Lemma 1 and Lemma 3, it follows that w.h.p.

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

and

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

since $\mathbb{E}\left[d_i^+ d_i^-\right] = \frac{\mathbb{E}\xi^2 - \mathbb{E}\xi}{2}$. Hence, as $\mathbb{E}\xi^2 < 3\mathbb{E}\xi$, it holds

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

Therefore, we can assume that

$$\frac{2T_n}{S_n} = 1 - \mu',$$

where $0 < \mu' < 1$. By Lemma 18 there are no long paths in ϕ, when $\mathbb{E}\xi^2 < 3\mathbb{E}\xi$.

Thus, it follows that there should be no *long* bicycles, and by Corollary 4 there are no bicycles in ϕ at all. Therefore, ϕ is satisfiable, which proves the proposition. $\qquad\square$

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)