# Approximating Highly Satisfiable Random 2-SAT

Andrei A. Bulatov, Cong Wang

Simon Fraser University
{abulatov,cwa9}@sfu.ca

**Abstract.** In this paper we introduce two distributions for the Max-2-SAT problem similar to the uniform distributions of satisfiable CNFs and the planted distribution for the decision version of SAT. In both cases there is a parameter $p$, $0 \leq p \leq \frac{1}{4}d$, such that formulas chosen according to both distributions are $p$-satisfiable, that is, at least $(\frac{3}{4}d + p)n$ clauses can be satisfied. In the planted distribution this happens for a fixed assignment, while for the $p$-satisfiable distribution formulas are chosen uniformly from the set of all $p$-satisfiable formulas. Following Coja-Oghlan, Krivelevich, and Vilenchik (2007) we establish a connection between the probabilities of events under the two distributions. Then we consider the case when $p$ is sufficiently large, $p = \gamma\sqrt{d \log d}$ and $\gamma > 2\sqrt{2}$. We present an algorithm that in the case of the planted distribution for any $\varepsilon$ with high probability finds an assignment satisfying at least $(\frac{3}{4}d + p - \varepsilon)n$ clauses. For the $p$-satisfiable distribution for every $d$ there is $\varepsilon(d)$ (which is a polynomial in $d$ of degree depending on $\gamma$) such that the algorithm with high probability finds an assignment satisfying at least $(\frac{3}{4}d + p - \varepsilon(d))n$ clauses. It does not seem this algorithm can be converted into an expected polynomial time algorithm finding a $p$-satisfying assignment. Also we use the connection between the planted and uniform $p$-satisfiable distributions to evaluate the number of clauses satisfiable in a random (not $p$-satisfiable) 2-CNF. We find the expectation of this number, but do not improve the existing concentration results.

## 1 Introduction

The random SAT and Max-SAT problems have received quite a bit of attention in the last decades. In this paper we work with one of the most popular distributions used for these problems, $\Phi_k(n, dn)$, the uniform one of $k$-CNFs of fixed density $d$: Fix $n$ and $d$ (the number of variables, and the density, it may depend on $n$), and choose $dn$ clauses uniformly at random out of $2^k\binom{n}{k}$ possible ones. It was originally suggested as a source of problems hard for SAT-solvers [18]. On the other hand, instances sampled from this distribution may, with high probability (whp), behave better than in the worst case, which opens up an interesting line of research. For instance, by [19, 14] whp Random Max-3-SAT can be approximated within a factor 1.0957, a marked improvement over Håstad's worst case inapproximability bound ($\frac{22}{21} - \varepsilon$ and $\frac{8}{7} - \varepsilon$ for Max-2-SAT and Max-3-SAT, respectively, see [3, 13]). However, it is not known if this bound can be further improved. Thus, in spite of large amount of research done on this model many questions remain open [9].

The phase transition phenomenon [11] also led to the study of other distributions, biased towards 'more satisfiable' instances. The most interesting of them is $\Phi_k^{sat}(n, dn)$

the uniform distribution of satisfiable formulas from $\Phi_k(n, dn)$. However, this distribution is difficult to sample, and a significant amount of attention has been paid to a related planted distribution $\Phi_k^{pl}(n, dn)$: First, choose a random assignment $f$ of $n$ Boolean variables, and then choose $dn$ $k$-clauses satisfied by $f$ uniformly at random from the $(2^k - 1)\binom{n}{k}$ possible ones. An important property of both satisfiable and planted distributions is that if $d$ is large enough, they demonstrate very high concentration of satisfying assignments. All such assignments have very small distance from each other [17, 6]. This property has been used to solve instances from such distributions efficiently. In [10, 17] different techniques were used to solve whp planted instances provided $d$ is sufficiently large. The technique from [17] was later generalized in [6, 5], where an algorithm was suggested that solves instances from $\Phi_k^{sat}(n, dn)$ in expected polynomial time. One of the improvements made in the latter paper and the one most relevant for this paper is a link between the probability of arbitrary events in the planted and satisfiable models. This allowed to transfer the methods developed in [17] to the more general satisfiable distribution.

While the decision random SAT has been intensively studied, its optimization counterpart, random Max-SAT has received substantially less attention. In this paper we focus on the random Max-2-SAT problem. While 2-SAT is solvable in polynomial time, as we mentioned before, Max-2-SAT is hard to approximate within the factor $\frac{22}{21} - \varepsilon$ [3, 13]. The uniform Max-2-SAT $\Phi_2(n, dn)$ (as from now on we only deal with Max-2-SAT, we will denote this distribution simply by $\Phi(n, dn)$) demonstrates phase transition at density $d = 1$. If density is greater than 1, whp $\varphi \in \Phi(n, dn)$ cannot be satisfied, but a random assignment satisfies the expected 3/4 of all clauses. As it is shown in [8], the number of clauses that can be satisfied whp lies within the interval $[\frac{3}{4}dn + 0.34\sqrt{d}n, \frac{3}{4}dn + 0.51\sqrt{d}n]$, provided $d$ is sufficiently large. For MAX-$k$-SAT, $k > 2$, the similar interval is shown to be much narrower. Achlioptas et al. [1] proved that for any $k \geq 2$ and any $d > 2^k \log 2$, for a random $k$-CNF from $\Phi_k(n, dn)$ the number of clauses that can be satisfied lies in the interval $[1 - 2^{-k} + p(k, d) - \delta_k, 1 - 2^{-k} + p(k, d)]$, where $p(k, d) = 2^k \Psi\left(\frac{2^k \log 2}{d}\right)$ and $\delta_k = O(k2^{-k/2})$. Note that $d = 2^k \log 2$ is the phase transition threshold for random $k$-SAT [2]. Thus the length of the interval decreases exponentially as $k$ grows; however, it gives a weaker bound for $k = 2$ than that from [8].

We first introduce two distributions of Max-2-SAT formulas similar to the satisfiable and planted distributions for the decision problem. An assignment $f$ of variables of a formula $\varphi \in \Phi(n, dn)$ is said to be $p$-satisfying, $0 \leq p \leq \frac{d}{4}$, if it satisfies $\left(\frac{3}{4}d + p\right) n$ clauses. Formula $\varphi$ is called $p$-satisfiable if it has a $p'$-satisfying assignment for some $p' \geq p$. By $\Phi^*(n, dn, p)$ we denote the uniform distribution of $p$-satisfiable formulas. Similar to the satisfiable distribution $\Phi^{sat}$, this distribution is difficult to sample. Therefore we also introduce the planted approximation of $\Phi^*(n, dn, p)$ as follows: Choose a random assignment $f$ to the $n$ variables, and then choose uniformly at random $\left(\frac{3}{4}d + p\right) n$ clauses satisfied by $f$ and $\left(\frac{1}{4}d - p\right) n$ clauses that are not satisfied by $f$.

Our first result is similar to the 'Exchange Rate' lemma from [6]. More precisely, if $\mathcal{P}_d^u(\mathcal{A}), \mathcal{P}_{d,p}^{pl}(\mathcal{A}), \mathcal{P}_{d,p}^{u*}(\mathcal{A})$ denote the probabilities of event $\mathcal{A}$ under the uniform distribution, the planted distribution of $p$-satisfiable formulas, and the uniform distribution of $p$-satisfiable formulas then the following holds.

**Proposition 1** *There exists a function* $\xi(d,p,n)$, $\xi(d,p,n) = e^{\Theta(n)}$ *for all* $d,p$, *such that for any* $p_0$, $0 \le p_0 \le \frac{1}{4}d$, *and any event* $\mathcal{A} \subseteq \Phi(n,dn)$ *there is* $p$ *with* $p_0 \le p \le \frac{3}{4}d$ *such that*

$$\mathcal{P}^u_d(\mathcal{A}) \le \mathcal{P}^u_d(C_{p_0}) + \xi(d,p,n)\mathcal{P}^{pl}_{d,p}(\mathcal{A}),$$

*where* $C_{p_0}$ *is the set of all formulas from* $\mathcal{A}$ *that are not* $p_0$*-satisfiable. In particular,*

$$\mathcal{P}^{u*}_{d,p_0}(\mathcal{A}) \le \xi(d,p,n)\mathcal{P}^{pl}_{d,p}(\mathcal{A}).$$

As is shown in [6], satisfiable random 3-SAT formulas possess a very nice structure that whp allows one to find a satisfying assignment for formulas, provided the density is sufficiently large. We show that it suffices to assume sufficiently high level of satisfiability to infer a similar result. The level of satisfiability that allows our proofs to go through is $2\sqrt{2}\sqrt{d\log d}$. Throughout the paper $\log$ denotes the natural logarithm. More precisely for the planted distribution we prove the following

**Theorem 2.** *There is a polynomial time algorithm that for any* $\varepsilon > 0$ *and any* $p > \gamma\sqrt{d\log d}$, $\gamma > 2\sqrt{2}$ *(*$\gamma$ *does not have to be a constant), where* $d$ *is sufficiently large, given* $\varphi \in \Phi^{pl}(n,dn,p)$ *whp finds an at least* $(p-\varepsilon)$*-satisfying assignment.*

Observe that the probability that the algorithm succeeds does not necessarily go to 1 uniformly over $\varepsilon$. This theorem can be restated in terms of approximability: For any $\beta = \frac{3}{4}d + \gamma\sqrt{d\log d}$, $\gamma > 2\sqrt{2}$, and any $\varepsilon > 0$ there is a $(\beta-\varepsilon,\beta)$-approximation algorithm for $\Phi^{pl}(n,dn,p)$ that succeeds whp. That is, an algorithm that given an instance in which at least $\beta$-fraction of clauses are satisfiable whp returns an assignment that satisfies at least $\beta-\varepsilon$-fraction of clauses.

Using the 'Exchange Rate' proposition we then can transfer this result to the uniform distribution

**Theorem 3.** *There is an algorithm that for every* $p > \gamma\sqrt{d\log d}$, $\gamma > 2\sqrt{2}$, *given* $\varphi \in \Phi^*(n,dn,p)$ *finds a* $p'$*-satisfying assignment,* $p' \ge p - \varepsilon + o(\varepsilon)$, *where* $\varepsilon = \frac{1}{2}d^{1-\frac{4\gamma^2}{9}} + \sqrt{6}d^{\frac{1}{2}-\frac{2\gamma^2}{9}}$.

In both cases we use the same algorithm, very similar to those from [17, 6, 5, 7], although the analysis is somewhat different, in particular, it has to be tighter.

Finally, we make an attempt to estimate the range $[p_1,p_2]$ such that whp a formula from $\Phi(n,dn)$ is $p_1$-satisfiable, but not $p_2$-satisfiable. As was mentioned before, the best result known to date is $[0.34\sqrt{d}, 0.51\sqrt{d}]$ [8]. Although we were unable to improve upon this result, we use a careful analysis of the majority vote algorithm and the 'Exchange Rate' proposition to find the likely location of the threshold for the level of satisfiability. It turns out to be very close to $0.5\sqrt{d}$.

## 2 Preliminaries

### 2.1 Random Max-2-SAT

We denote the four possible types of 2-clauses (depending on the polarity of literals) by $(+,+), (+,-), (-,+), (-,-)$. The distributions of 2-CNFs we consider are the following:

– The uniform distribution $\Phi(n, dn)$ with $n$ variables and $dn$ 2-clauses, $d$ constant. Whether or not repetitions of variables in a clause or repetitions of clauses are allowed is immaterial. A slightly different distribution is $\Phi(n, \varrho)$, in which each of the possible $\binom{n}{2}$ pairs is included with probability $\varrho$. It is well known that properties of $\Phi(n, dn)$ and $\Phi(n, \varrho)$ for $\varrho = \frac{2d}{n}$ are nearly identical.
– The uniform distribution of $p$-satisfiable 2-CNFs $\Phi^*(n, dn, p)$, $0 \le p \le \frac{d}{4}$, with $n$ variables, $dn$ clauses. Every $\varphi \in \Phi^*(n, dn, p)$ has a $p'$-satisfying assignment for $p' \ge p$.
– The planted distribution $\Phi^{pl}(n, dn, p)$ of level $p$. To generate a formula from $\Phi^{pl}(n, dn, p)$, we first choose a random assignment $f$ of the $n$ variables, and then select $\left(\frac{3}{4}d + p\right) n$ random 2-clauses that are satisfied by $f$ and $\left(\frac{1}{4}d - p\right) n$ random 2-clauses that are not satisfied by $f$. Since the statistical properties of a planted formula do not depend on the planted assignment, we will always assume that the planted assignment $f$ assigns 1 to all variables. Under this assumption planted formulas look particularly simple: they contain $\left(\frac{3}{4}d + p\right) n$ clauses of types $(+,+)$, $(+,-)$, and $(-,+)$, and $\left(\frac{1}{4}d - p\right) n$ clauses of type $(-,-)$.

The degrees of variables of formulas in $\Phi(n, dn)$ or $\Phi(n, \varrho)$ are not completely independent that makes the analysis of algorithms more difficult. To overcome this difficulty [15] suggested the *Poisson cloning model*. Under this model random formulas are generated as follows: First, take Poisson random variables $\deg(x)$ with the mean $2d$ for each of the $n$ variables $x$. Then take $\deg(x)$ copies, or clones, of each variable $x$. If the sum of $d(x)$'s is even, then generate a uniform random matching on the set of all clones; for each edge in the matching select its type $(+,+), (+,-), (-,+), (-,-)$ at random. The clause $x^\alpha \vee y^\beta$ is in the formula $\varphi \in \Phi^{poi}(n, dn)$ if a clone of $x$ is matched with a clone of $y$ with an edge of type $(\alpha, \beta)$. If the sum of $d(x)$'s is odd, one of the edges must be a loop.

Let $\mathcal{P}^u_\varrho(\mathcal{A})$ and $\mathcal{P}^{poi}_\varrho(\mathcal{A})$ denote the probability of event $\mathcal{A}$ under distributions $\Phi(n, \varrho)$ and $\Phi^{poi}(n, \varrho)$. In the case of 2-SAT the result from [15] is the following

**Theorem 4 ([15]).** *Suppose $\varrho = \Theta(n^{-1})$. Then for any event $\mathcal{A}$*

$$c_1 \mathcal{P}^{poi}_\varrho(\mathcal{A}) \le \mathcal{P}^u_\varrho(\mathcal{A}) \le c_2 \left( \mathcal{P}^{poi}_\varrho(\mathcal{A})^{\frac{1}{2}} + e^{-n} \right),$$

*where*

$$c_1 = \sqrt{2} e^{\frac{\varrho(n-1)}{2} + \frac{\varrho^2 n(n-1)}{4}} + O(n^{-\frac{1}{2}}), \qquad c_2 = \sqrt{c_1} + o(1).$$

### 2.2 Random graph

We will need two standard results about random graphs. It is convenient for us to state them as follows. They can be proved in the standard way (see, e.g. Theorems 2.8, 2.9 of [4].

**Lemma 1.** *Let $G = G(n, dn)$ (or $G = G(n, \varrho)$, $\varrho = \frac{2d}{n}$) be a random graph with $n$ vertices and $dn$ edges (or a random graph with $n$ vertices and density $\varrho$), and let*

$0 < \alpha < 1$ *and* $\beta > \alpha^2 d + \sqrt{3d}\alpha\sqrt{\alpha(1-\log\alpha)}$. *Then whp there is no set of vertices* $S \subseteq V$ *such that* $|S| = \alpha n$ *and the subgraph* $G_{|S}$ *of* $G$ *induced by* $S$ *contains* $\beta n$ *edges.*

*More precisely, if* $\beta = \alpha^2 d + \tau\sqrt{3d}\alpha\sqrt{\alpha(1-\log\alpha)}$, $\tau > 1$, *then the probability such a set exists is at most* $\exp\left[-\alpha(1-\log\alpha)(\tau^2-1)n\right]$.

**Lemma 2.** *Let* $G = G(n, dn)$ *(or* $G = G(n, \varrho)$, $\varrho = \frac{2d}{n}$*) be a random graph with* $n$ *vertices and* $dn$ *edges (or a random graph with* $n$ *vertices and density* $\varrho$*), and let* $0 < \alpha < 1$ *and* $\beta > 2\alpha d + \sqrt{6d}\alpha\sqrt{1-\log\alpha}$. *Then whp there is no set of variables* $S \subseteq V$ *such that* $|S| = \alpha n$ *and* $\deg S = \sum_{x\in S} \deg x \geq \beta n$.

*More precisely, if* $\beta = 2\alpha d + \tau\sqrt{6d}\alpha\sqrt{1-\log\alpha}$, $\tau > 1$, *then the probability such a set exists is at most* $\exp\left[-\alpha(1-\log\alpha)(\tau^2-1)n\right]$.

## 3 From planted to uniform

First, we establish a connection between the probabilities of events under the uniform and planted distributions. Note that the density $d$ is NOT assumed to be large in this section. The probability of event $\mathcal{A}$ w.r.t. the uniform distribution over this set is denoted by $\mathcal{P}^u_d(\mathcal{A})$. Observe that, $\Phi^{pl}(n, dn, p)$ is uniform on the set of 2-CNFs with $(\frac{3}{4}d + p)n$ clauses of types $(+,+), (+,-), (-,+)$ and $(\frac{1}{4}d - p)n$ clauses type $(-,-)$. By $\mathcal{P}^{u*}_{d,p}(\mathcal{A})$ and $\mathcal{P}^{pl}_{d,p}(\mathcal{A})$ we denote the probability of event $\mathcal{A}$ under distributions $\Phi^*(n, dn, p)$ and $\Phi^{pl}(n, dn, p)$, respectively.

**Proposition 5** *(1) Let* $C_{p_0}$ *denote the event that a formula from* $\Phi(n, dn)$ *does not have a* $p$*-satisfying assignment for any* $p \geq p_0$. *For any* $p_0 \leq \frac{d}{4}$ *and event* $\mathcal{A}$ *there is* $p \in [p_0, \frac{d}{4}]$ *such that*

$$\mathcal{P}^u_d(\mathcal{A}) \leq \mathsf{Prob}[C_{p_0}] + \left(\frac{d}{4} - p_0\right) n \cdot \exp\left[n \log\left(1 + \exp\left[-\frac{\tau p^2}{d}\right]\right)\right] \mathcal{P}^{pl}_{d,p}(\mathcal{A}),$$

*where* $\tau = \frac{16}{9}$ *if* $p \leq \frac{d}{8}$ *and* $\tau = \frac{2}{3}$ *if* $p > \frac{d}{8}$.

*(2) For any* $p_0 \in [0, \frac{d}{4}]$ *there is* $p \in [p_0, \frac{d}{4}]$ *such that*

$$\mathcal{P}^{u*}_{d,p_0}(\mathcal{A}) \leq \left(\frac{d}{4} - p_0\right) n \cdot \exp\left[n \log\left(1 + \exp\left[-\frac{\tau p^2}{d}\right]\right)\right] \mathcal{P}^{pl}_{d,p}(\mathcal{A}),$$

*where* $\tau = \frac{16}{9}$ *if* $p \leq \frac{d}{8}$ *and* $\tau = \frac{2}{3}$ *if* $p > \frac{d}{8}$.

The proof follows the lines of Lemma 1 from [5] with certain modifications.

## 4 Approximating Max-2-SAT

In this section we present an algorithm to find an approximate solution of a Max-2-SAT instance from $\Phi^{pl}(n, dn, p)$ or $\Phi^*(n, dn, p)$ and prove Theorems 2 and 3. Some parts of the proof follow [17, 6, 7]. We try to highlight as many of the new proofs as the page limit allows.

### 4.1 Core and Majority Vote

Let $\varphi$ be a 2-CNF and $f$ its (partial) assignment. Variable $x$ *supports* a clause $C$ if the value of $x$ under $f$ satisfies $C$, while the value of the other variable from $C$ exists, but does not satisfy $C$.

A set of variables $W \subseteq V$ is called a *core* of $\varphi$ with respect to an assignment $f$, if every variable $x \in W$ supports at least $\frac{d}{2} - \frac{2}{3}\gamma\sqrt{d \log d}$ clauses in $\varphi(W)$. A core that contains at least $(1 - d^{-\nu})n$ variables, $\nu \geq 2$, will be called a *large core*.

We show that similar to [17] every highly satisfiable instance has a large core. Majority Vote is a one of the key ingredients in this proof. Majority Vote is a simple heuristic to solve Max-2-SAT instances. Given a 2-CNF $\varphi$ it sets a variable $x$ to 0 if $\neg x$ occurs in more clauses than $x$; it sets $x$ to 1 if $x$ occurs in more clauses than $\neg x$; and breaks the tie at random with probability $1/2$. In this section we use Majority Vote to construct an algorithm to solve highly satisfiable Random Max-2-SAT instances.

Let $\varphi$ be sampled from $\Phi^{pl}(n, dn, p)$ or $\Phi^*(n, dn, p)$ and $f$ some assignment. The following procedure defining set $W$ will be called $\mathsf{Core}(f)$:

- let $U \subseteq V$ be the set of variables on which the majority vote assignment disagrees with $f$;
- let $B$ be the set of variables that support less than $\frac{d}{2}$ clauses with respect to $f$;
- now $W$ is given inductively
  - $W_0 = V - (U \cup B)$;
  - while there is $x_i \in W_i$ supporting less than $\frac{d}{2} - \frac{2}{3}\gamma\sqrt{d \log d}$ clauses in $\varphi(W_i)$ with respect to $f$, set $W_{i+1} = W_i - \{x_i\}$;
- set $W = W_i$.

In the rest of this subsection $f$ is the planted assignment (recall it is assumed to be the all-ones assignment).

**Lemma 3.** *Let $\varphi \in \Phi(n, dn, p)$, $p = \gamma\sqrt{d \log d}$, $\gamma > 2\sqrt{2}$, where $d$ is sufficiently large and let $W$ be the result of $\mathsf{Core}(f)$ where $f$ is the planted assignment and $W^* = V - W$. Then for any $\alpha \in [d^{-\frac{\gamma^2}{4}}, d^{-1/2}]$,*

$$\mathsf{Prob}[|W^*| \geq \alpha n] \leq \exp\left[-\frac{\gamma^2 \log d}{48} d^{-\frac{\gamma^2}{4}} n\right].$$

**Proposition 6** *Let $\varphi$ be sampled according to $\Phi^*(n, dn, p)$ or $\Phi^{pl}(n, dn, p)$, $p = \gamma\sqrt{d \log d}$, $\gamma > 2\sqrt{2}$, where $d$ is sufficiently large. Then whp there exists a large core with respect to the assignment given by Majority Vote.*

*Proof.* Consider first the planted distribution. Let $f, f_m$ be the planted assignment and the assignment given by Majority Vote. By Lemma 3 with probability at least $1 - \exp\left[-\frac{\gamma^2 \log d}{48} d^{-\frac{\gamma^2}{4}} n\right]$ there is a core of size at least $\left(1 - d^{-\frac{\gamma^2}{4}}\right) n$ with respect to the planted assignment, which is $p$-satisfying. However, since $f$ coincides with $f_m$ on $W$, it is also a large core with respect to $f_m$. Thus, observing that if $\gamma > 2\sqrt{2}$ then $\frac{\gamma^2}{4} > 2$, the claim of the proposition is true whp for the planted distribution.

By Proposition 5 there is a core of size $\left(1 - d^{-\frac{\gamma^2}{4}}\right) n$ with respect to the assignment given by the Majority Vote with probability at least

$$1 - \exp\left[n \log\left(1 + e^{-\frac{\frac{16}{9}p^2}{d}}\right)\right] \cdot \exp\left[-\frac{\gamma^2 \log d}{48} d^{-\frac{\gamma^2}{4}} n\right] \geq 1 - \exp\left[-d^{-\frac{\gamma^2}{4}}\right].$$

### 4.2 Cores and connected components

Next we show that a large core and its complement satisfy certain strong conditions.

**Proposition 7** *Let $\varphi$ be sampled according to $\Phi^*(n, dn, p)$, $p = \gamma\sqrt{d \log d}$, $\gamma > 2\sqrt{2}$, where $d$ is sufficiently large. Let $W$ be a core of $\varphi$ with respect to an at least $p$-satisfiable assignment $f$ and such that $|W| \geq (1 - d^{-\nu})n$, $\nu > 2$. Then whp the largest connected component in subformula of $\varphi$ induced by $V - W$ is of size at most $\log n$.*

Observe that if we choose a random set $X$ of variables of size $\alpha n$ then, as the probability of each pair of variables to form a clause is $\frac{d}{n}$, by the results of [4] if $\alpha < \frac{1}{2d}$, the connected components of the formula induced by $X$ whp are of size at most $\log n$. However, high probability in this case means $1 - n^{\log(\alpha d)}$, and we cannot claim that each of the $\binom{n}{\alpha n}$ sets of $\alpha n$ variables satisfies this condition. Therefore, we follow [17] and provide a fairly sophisticated proof that the complement of a core has connected components of size at most $\log n$.

Fix a set of variables $T \subseteq V$, $t = |T| = \log n$, and a collection $\tau$ of $t - 1$ clauses that induce a tree on $T$. We aim to bound the probability that $\tau \subseteq \varphi$. Let $J \subseteq T$ be a set of variables which appear in at most 2 clauses of $\tau$. By a simple counting argument we have $|J| \geq \frac{1}{2}t$.

Lemma 3 amounts to saying that for $\varphi \in \Phi^{pl}(n, dn, p)$ whp there exists a $p$-satisfying assignment $f$ such that $\mathsf{Core}(f)$ produces sets $U, B, W$ satisfying the conditions proved there. Now by Proposition 5 the same is true whp for $\varphi \in \Phi^*(n, dn, p)$. More precisely, we have the following

**Lemma 4.** *Let $\varphi$ be sampled according to $\Phi^*(n, dn, p)$ or $\Phi^{pl}(n, dn, p)$, $p = \gamma\sqrt{d \log d}$, $\gamma > 2\sqrt{2}$, where $d$ is sufficiently large. Then for any $\alpha \in [d^{-\frac{\gamma^2}{4}}, d^{-1/2}]$ whp there exists a $p$-satisfying assignment $f_\varphi$ such that if $U$ is the set of variables on which Majority Vote disagrees with $f_\varphi$ and $B$ is the set of variables that support at most $\frac{d}{2}$ clauses with respect to $f_\varphi$, then $|U \cup B| \leq \frac{\alpha}{2}n$.*

Let $\varphi$ be sampled from $\Phi^{pl}(n, dn, p)$ or $\Phi^*(n, dn, p)$ and $f$ some assignment. The following procedure, denoted $\mathsf{Core}'(f)$, is a modification of $\mathsf{Core}$ from Section 4.1.

- let $U' \subseteq V$ be the set of variables on which the majority vote assignment disagrees with $f$ or is right with advantage smaller than 4;
- let $B'$ be the set of variables that support less than $\frac{d}{2}$ clauses with respect to the planted assignment;
- now $W$ is given inductively
  - $W_0' = V - (U \cup B \cup (T - J))$, where $J$ is defined as above;

- while there is $x_i \in W_i$ supporting less than $\frac{d}{2} - \frac{2}{3}\gamma\sqrt{d \log d}$ clauses in $\varphi(W_i')$ , set $W_{i+1}' = W_i' - \{x_i\}$;

– set $W' = W_i'$.

The following lemma is proved in the same way as Lemma 3.

**Lemma 5.** *Let $\varphi \in \Phi(n, dn, p)$, $p = \gamma\sqrt{d \log d}$, $\gamma > 2\sqrt{2}$, where $d$ is sufficiently large, $W'$ be the result of $\mathsf{Core}'(f)$ where $f$ is the planted assignment and $W^* = V - W'$. Then for any $\alpha \in [d^{-\frac{\gamma^2}{4}}, \sqrt{d}]$,*

$$\mathsf{Prob}[|W^*| \geq \alpha n] \leq \exp\left[-\frac{\gamma^2 \log d}{48} d^{-\frac{\gamma^2}{4}} n\right].$$

Observe also that Lemmas 3 and 5 remain true if instead of a formula $\varphi \in \Phi^*(n, dn, p)$ we apply the procedures to $\varphi$ along with a small number of extra clauses. More precisely,

**Lemma 6.** *Let $\varphi \in \Phi^*(n, dn, p)$, $p = \gamma\sqrt{d \log d}$, $\gamma > 2\sqrt{2}$, where $d$ is sufficiently large, and $\psi$ a set of clauses of size $\log n$. Also let $W^* = V - W$ or $W^* = V - W'$ where $W, W'$ are the sets obtained by applying $\mathsf{Core}(f_\varphi)$ and $\mathsf{Core}'(f_\varphi)$ to $\varphi \cup \psi$; here $f_\varphi$ is as found in Lemma 4. Then for any $\alpha \in [d^{-\frac{\gamma^2}{4}}, \sqrt{d}]$,*

$$\mathsf{Prob}[|W^*| \geq \alpha n] \leq \exp\left[-\frac{\gamma^2 \log d}{48} d^{-\frac{\gamma^2}{4}} n\right].$$

**Lemma 7.** *Let $\varphi \in \Phi^*(n, dn, p)$ and let $W(\varphi \cup \tau)$ be the value of $W$ if $\mathsf{Core}(f_\varphi)$ is applied to $\varphi \cup \tau$. Let $W'(\varphi)$ be the value of $W'$ obtained by $\mathsf{Core}'(f_\varphi)$ applied to $\varphi$. Then $W'(\varphi) \subseteq W(\varphi \cup \tau)$.*

*Proof.* We proceed by induction. First, show that $W_0'(\varphi) \subseteq W_0(\varphi \cup \tau)$. If $x \in W_0'(\varphi)$ then Majority Vote is right with advantage at least 3, and $x$ supports at least $\frac{d}{2}$ clauses. Since the degree of $x$ in $\tau$ is at most 2, Majority Vote cannot be flipped by adding $\tau$ to $\varphi$. Also, support can only increase after adding $\tau$. Therefore $s \in W_0(\varphi \cup \tau)$.

Now, suppose $W_i'(\varphi) \subseteq W_i(\varphi \cup \tau)$. We show that the statement holds for $i + 1$. If $x$ supports the required number of clauses with only variables from $W_i'(\varphi)$ then it also supports sufficiently many variables with respect to $W_i(\varphi \cup \tau)$. Therefore if $x \in W_{i+1}'(\varphi) \subseteq W_{i+1}(\varphi \cup \tau)$. The lemma is proved.

Proposition 7 now follows from a sequence of lemmas we borrow from [17]. In all of them we assume $p = \gamma\sqrt{d \log d}$, $\gamma > 2\sqrt{2}$, $d$ is large enough.

**Lemma 8.** $\mathsf{Prob}[\tau \subseteq \varphi$ *and* $T \cap W = \varnothing] \leq \mathsf{Prob}[\tau \subseteq \varphi] \cdot \mathsf{Prob}[J \cap W' = \varnothing]$.

**Lemma 9.** *Let $\varrho \in [\frac{t}{n}, 1]$ be such that $|W'| = (1 - \varrho)n$. Then* $\mathsf{Prob}[J \cap W' = \varnothing] \leq (6\varrho)^{t/2}$.

**Corollary 1.** *Let $\varrho \in [\frac{t}{n}, 1]$ be such that $|W'| = (1 - \varrho)n$. Then*

$$\mathsf{Prob}[\tau \subseteq \varphi \textit{ and } T \cap W = \varnothing] \leq (6\varrho)^{t/2} \left(\frac{d}{n}\right)^t.$$

**Lemma 10.** *The probability of a tree of size* $t = \log n$ *in the formula induced by* $W^*$ *is at most* $\exp[-(\delta \log d - 1 - \log(\sqrt{6})) \log n]$ *for some* $\delta > 0$.

*Proof.* We assume that $\varrho$ in the previous lemmas is $o(d^{-(2+2\delta)})$. By Cayley theorem there are $t^{t-2}$ trees on a given set of $t$ vertices. Then

$$\begin{aligned}
&\mathsf{Prob}[\text{there is a tree of size} \geq \beta n] \leq \mathsf{Prob}[\exists \tau \subseteq \varphi, |\tau| = t \text{ and } T = V(\tau) \subseteq W = \varnothing] \\
&\quad \leq \binom{n}{t} t^{t-2} \cdot \left(\frac{d}{n}\right)^t \cdot (6\varrho)^t \leq \left(\frac{en}{t}\right)^t t^t \cdot \left(\frac{d}{n}\right)^t \cdot (6\varrho)^{t/2} \\
&\quad \leq (\sqrt{6\varrho} ed)^t \leq (\sqrt{6} e d^{-\delta})^{\log n} = \exp[-(\delta \log d - 1 - \log(\sqrt{6})) \log n],
\end{aligned}$$

which completes the proof.

Proposition 7 follows.

### 4.3 The algorithm

The solution algorithm follows the approach from [17]. However, the brute force search part of that algorithm is not applicable in our case, as we cannot guarantee that sufficiently few wrongly assigned variables remain assigned after the unassigning step.

Here $\varphi^h(A)$ denotes the formula $\varphi$ after substituting the partial assignment $h$ and restricted to $V - A$.

---

**Algorithm 1** The algorithm

---

**Require:** $\varphi \in \Phi^*(n, dn, p)$
1: $f_1 \leftarrow \mathsf{MajorityVote}(\varphi)$
2: $i \leftarrow 1$
3: **while** there is $x$ supporting less than $\frac{d}{2} - \frac{2}{3}\gamma\sqrt{d \log d}$ clauses with respect to $f_i$ **do**
4: $\quad f_{i+1} \leftarrow f_i$ with $x$ unassigned
5: $\quad i \leftarrow i + 1$
6: **end while**
7: let $h$ be the final partial assignment
8: let $A$ be the set of assigned variables in $h$
9: exhaustively search the formula $\varphi^h(A)$, component by component looking for the assignment satisfying the maximum number of clauses

---

**Lemma 11.** *Let* $f$ *be the assignment given in line 1 of the algorithm; suppose* $f$ *has a large core* $W$. *Then no variable from* $W$ *is unassigned in lines 3–6 of the algorithm.*

*Proof.* Every $x \in W$ supports at least $\frac{d}{2} - \frac{2}{3}\gamma\sqrt{d \log d}$ clauses from $\varphi(W)$. Then the statement follows by induction.

**Lemma 12.** *Let* $\varphi \in \Phi^{pl}(n, dn, p)$, $p = \gamma\sqrt{d \log d}$, $\gamma > 2\sqrt{2}$, *where* $d$ *is sufficiently large, and let* $f$ *be the planted assignment. Let* $U$ *denote the set of assigned variables after line 7 of the algorithm, whose value differs from that under* $f$. *Then* $|U| \geq \chi n$ *with probability at most* $\exp[-2\chi n]$ *for any* $\chi < d^{-1/2}$.

*Proof.* (Sketch.) Let $\varphi \in \Phi^{pl}(n, dn, p)$, and $p = \gamma\sqrt{d \log d}$. Let $Z$ denote the set of unassigned variables, and $U$ the set of variables $x$ with $h(x) \neq f(x)$. In other words, $U = \{x | h(x) = 0\}$. Let $|U| = \chi n$. By Lemmas 3 and 11 and $|Z \cup U| \leq d^{-1/2}n$; in particular, $\chi < d^{-1/2}$.

We evaluate the number of clauses supported by variables from $U$. There are two kinds of such clauses . In clauses of the first kind both variables are from $U$, we call such clauses *internal*. In clauses of the second kind the second variable belongs to $V - (Z \cup U)$, we call such clauses *external*. Every internal clause has the type $(-, +)$ or $(+, -)$, every external clause has the type $(-, -)$. Using Lemma 1 we bound the number of internal clauses by $\beta_1 n < (\frac{1}{2} + o(1))(\chi^2 d + 2\chi\sqrt{3d})n$. Since every variable from $U$ supports at least $\frac{d}{2} - \frac{2}{3}\gamma\sqrt{d \log d}$ clauses, the total number of internal and external clauses is at least $\left(\frac{d}{2} - \frac{2}{3}\gamma\sqrt{d \log d}\right)\chi n$. By Lemma 2 and the bound on the number of internal variables we show that the probability that there exist a sufficient number of external clauses is bounded as in the statement of the lemma; that completes the proof.

**Proposition 8** *Let $\varphi \in \Phi^{pl}(n, dn, p)$, $p = \gamma\sqrt{d \log d}$, $\gamma > 2\sqrt{2}$, where $d$ is sufficiently large, let $U$ be the set of assigned variables after line 7 of the algorithm, whose value differs from that under $f$, and let $|U| = \chi n$. The exhaustive search of the algorithm whp completes in polynomial time and returns an extension of $h$ which is at least $p'$-satisfiable for $p' = p - \chi(\frac{1}{2}d - \sqrt{6d} + o(d^{-1}))$ with probability at least $1 - e^{-2\chi n}$ for any $\chi < d^{-1/2}$.*

*Proof.* (Sketch.) This proof is similar to the previous one. As $\varphi$ has a large core with respect to $f$, by Lemma 11 there are at most $d^{-2}n$ unassigned variables. Therefore it follows from Proposition 7 that whp the algorithm completes in polynomial time .

To prove the second claim of the lemma observe that if all unassigned variables are assigned according to $f$ then the only clauses that are satisfied under $f$ but not satisfied under the assignment produced by the algorithm are the clauses of the type $(+, +)$, both of whose variables belong to $U$, and the clauses of the type $(+, -)$, whose first variable is from $U$, and the other is from $V - U$.

By Lemma 1 we bound the number of clauses of the first kind by $(1 + o(1))\left(\frac{1}{4}\chi^2 d + \chi\sqrt{3d}\right)n$. Furthermore, by Lemma 2 the number of clauses of the second kind can be bounded by $(1 + o(1))\chi\left(\frac{1}{2}d + \sqrt{6d}\right)n$. Since $\chi < 1$, the number of clauses of the second kind dominates. Thus, in the planted case the assignment produced by the algorithm has level at least $p - \chi(\frac{1}{2}d - \sqrt{6d} + o(d^{-1}))$.

*Proof.* (of Theorem 2). If $\gamma > 2\sqrt{2}$ then $\frac{\gamma^2}{4} > 2$. By Lemma 3 in this case the conditions of Lemma 12 hold. Therefore whp $|U| < \chi n$, where $\chi$ is such that $\chi(\frac{1}{2}d + \sqrt{6d} + o(d^{-1})) < \varepsilon$, and the result follows by Proposition 8.

*Proof.* (of Theorem 3). First, we compute the 'exchange rate' from Proposition 5 for $p = \gamma\sqrt{d \log d}$:

$$\exp\left[n \log\left(1 + \exp\left[-\frac{16p^2}{9d}\right]\right)\right] < \exp\left[nd^{-\frac{16\gamma^2}{9}}\right].$$

By Lemma 11 and Proposition 5, if $2\chi > d^{-\frac{16\gamma^2}{9}}$ whp the set $U$ produced in line 7 of the algorithm has size $|U| < \chi n$. We now conclude by Proposition 8.

## 5 Majority vote

In this section we use Majority Vote to estimate the likely level of satisfiability of Random Max-2-SAT instances. The results of this section are true for ANY value of $d$.

### 5.1 Majority Vote and its performance

First, we use the simple approach from [12] to estimate the number of clauses unsatisfied by Majority Vote algorithm under the uniform distribution.

**Proposition 9** *Given a random formula of density $d$ Majority Vote whp does not satisfy $\beta^2(d)dn + o(n)$ clauses, where $\beta(d) = \mathbb{E}[\min(X, Y)]$, and $X, Y$ are independent identical Poisson variables with the mean $d$.*

*Remark 1.* Analytically, the expected ratio of clauses unsatisfied by Majority Vote can be expressed as follows

$$\beta(d) = 1 - \frac{1}{d}\int_0^\infty t \cdot e^{-d} \frac{d^t}{\Gamma(t)} \frac{\Gamma(t+1,d)}{\Gamma(t)} dt.$$

Since there is no convenient approximation or asymptotic for function $\beta(d)$, we can only estimate it numerically. Table 1 contains some values of it.

| $d$ | $\alpha$ | $\beta$ | Unsat | $d$ | $\alpha$ | $\beta$ | Unsat | $d$ | $\alpha$ | $\beta$ | Unsat |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 25 | 0.440 | 0.560 | 4.838 | 64 | 0.463 | 0.537 | 13.740 | 121 | 0.474 | 0.526 | 27.143 |
| 36 | 0.450 | 0.550 | 7.305 | 81 | 0.468 | 0.532 | 17.964 | 144 | 0.476 | 0.524 | 32.610 |
| 49 | 0.458 | 0.542 | 10.273 | 100 | 0.471 | 0.529 | 22.175 | | | | |

**Table 1.** Minimum of two Poisson distributions. $\alpha$ and $\beta$ denote the values $\alpha(d) = 1 - \beta(d)$ and $\beta(d)$; Unsat stands for the number of unsatisfied clauses.

### 5.2 Majority Vote on planted instances

Here we study how Majority Vote performs on planted instances of different level. As before, we assume that the planted assignment $f$ is the all-ones one, and is $p$-satisfying, $0 \le p \le \frac{d}{4}$.

**Proposition 10** *Let $\varphi \in \Phi^{pl}(n, dn, p)$, let $f$ be the planted assignment and $h$ the assignment produced by Majority Vote.*
*(1) Whp the assignments $f$ and $h$ disagree on $\beta(d,p)n+o(n)$ variables, where $\beta(d,p) = \mathsf{Prob}[X \geq Y]$ and $X, Y$ are independent Poisson variables with means $\mu_1 = d - \frac{4}{3}p$ and $\mu_2 = d + \frac{4}{3}p$.*
*(2) Let $x(d,p,n)$ be a random variable equal to the number of clauses unsatisfied by $h$. Then*

$$\mathrm{num}(d,p) = \frac{1}{n}\mathbb{E}[x(d,p,n)] = \frac{1}{4}d - p + \frac{1}{3}p\beta(d,p)(2-\beta(d,p)).$$

*(3)* $\mathsf{Prob}[|x(d,p,n) - \mathrm{num}(d,p)n| > \delta\sqrt{d}n] < e^{-16\delta^2 n}$.

*Proof.* Using the Poisson cloning model proofs of items (1) and (2) are straightforward.

(3) To evaluate the concentration of the number of unsatisfied clauses we again use the Poisson cloning approach as follows. For each variable $v$ we create two sets of clones $v^+$ and $v^-$ such that $|v^+|$ is a Poisson variable with the mean $d + \frac{4}{3}p$ and $|v^-|$ is a Poisson variable with the mean $d - \frac{4}{3}p$, which are the expected numbers of positive and negative occurrences of $v$, respectively. We now select a random matching among the generated clones.

We consider 2 cases depending on $p$. First, when $p = C\sqrt{d}$, $C$ constant; we will use this case to compare against the uniform distribution. The second case includes all remaining values of $p$.

Case 1. $p = C\sqrt{d}$, $C$ constant.

Let the number of clones satisfied by Majority Vote is $sn$, for a random variable $s$, and the number of unsatisfied clauses is $xn$ for a random variable $x$. Clearly the expectation of $x$ is $\frac{s^2}{4d'}n$, where $2d'n$ is the total number of clones. We evaluate the probability that $x$ deviates by at least $\delta\sqrt{d}$ from its expectation conditioned on the assumption that $2d' = 2d$. Since $p = C\sqrt{d}$, whp $s = d - \alpha\sqrt{d}$. In fact, $\mathbb{E}[s] = \mathbb{E}[\min(X,Y)]$, where $X, Y$ are defined as in part (1) of the proposition. We need a tighter estimation, therefore, rather than using Chernoff bound we resort to a more straightforward method.

We count the total number of formulas with $2dn$ clones, and the number of those of them having $sn$ unsatisfied literals and $xn$ unsatisfied clauses. The first number, $N$ equals the number of matchings of a complete graph, see, e.g., [16]

$$N = \left(\frac{2dn}{e}\right)^{dn} \frac{e^{\sqrt{2dn}}}{(4e)^{1/4}}(1+o(1)).$$

The second number $M$ is the number of ways to complete the following process: choose a matching of $xn$ edges in the complete graph with $sn$ vertices (i.e., choose $xn$ unsatisfied clauses among pairs of unsatisfied literals), then choose matches for the remaining $(s-2x)n$ unsatisfied literals among $(2d-s)n$ satisfied ones, and, finally, choose a matching among the $(2d-2s+2x)n$ remaining satisfied literals. Thus,

$M = M_1 \cdot M_2 \cdot M_3$, where

$$M_1 = \frac{(sn)!}{2^{xn}((s-2x)n)!(xn)!}, \qquad M_2 = \frac{((2d-s)n)!}{((2d-2s+2x)n)!},$$
$$M_3 = \left(\frac{(2d-2s+2x)n}{e}\right)^{(d-s+x)n} \frac{e^{\sqrt{(2d-2s+2x)n}}}{(4e)^{1/4}}(1+o(1)).$$

Using Stirling's approximation up to a factor of $1+o(1)$ we can represent the probability we are looking for, that is, the fraction $\frac{M}{N}$, as $\frac{M}{N} = n^A \cdot e^{Bn} \cdot e^{o(n)}$, where $B$ is a constant, while $A$ may depend on $n$. As is easily seen, $A = 0$. A straightforward (although somewhat tedious) calculation shows that $B < -16\delta^2(1+o(1))$.

Thus,

$$\mathsf{Prob}[|x - \mathbb{E}[x]| \geq \delta\sqrt{d}] \leq e^{-16\delta^2 n}.$$

Case 2. $p = \Omega(\sqrt{d})$.

In this case the result follows from Chernoff bound.

*Remark 2.* Analytically, $\beta(d,p)$ can be expressed as follows

$$\beta(d,p) = e^{-\mu_2}\int_0^{\mu_1} e^{-t} I_0(2\sqrt{\mu_2 t})dt.$$

Table 2 below contains values of $\beta(d,p)$, that is, the probability that majority vote assigns values corresponding to the all-ones assignment, and the number of unsatisfied clauses left by majority vote. These numbers depend on two parameters, $d$ the density, and $p$ the parameter regulating the number of clauses satisfied by the all-ones assignment. Recall that for the uniform distribution whp $p = \Theta(\sqrt{d})$, more precise estimations give $0.34\sqrt{d} \leq p \leq 0.51\sqrt{d}$ [8].

| $d$ / $p$ | $0.34\sqrt{d}$ | $0.42\sqrt{d}$ | $0.51\sqrt{d}$ | $\frac{1}{16}d$ | $\frac{1}{8}d$ | $\frac{1}{4}d$ |
|---|---|---|---|---|---|---|
| 25 | 0.284/5.661 | 0.235/5.321 | 0.191/4.916 | 0.302/5.761 | 0.133/4.182 | 0.010/0.230 |
| 49 | 0.277/11.395 | 0.229/10.914 | 0.186/10.343 | 0.219/10.796 | 0.054/7.040 | 0.0005/0.138 |
| 64 | 0.275/15.012 | 0.227/14.460 | 0.184/13.806 | 0.184/13.806 | 0.032/8.752 | |
| 100 | 0.272/23.745 | 0.225/23.052 | 0.182/22.233 | 0.126/20.765 | 0.010/12.950 | |
| 144 | 0.270/34.479 | 0.223/33.645 | 0.180/32.660 | 0.083/28.978 | 0.0025/18.298 | |
| 256 | 0.268/61.947 | 0.221/60.830 | 0.179/59.514 | 0.031/49.453 | | |

**Table 2.** Values of $\beta(d,p)$ and $\mathrm{num}(d,p)$. Each entry has the form $\beta(d,p)/\mathrm{num}(d,p)$. The values in the empty cells are too small and cannot be reliably approximated.

### 5.3 Satisfiability of Random Max-2-SAT

We now combine Propositions 9, 10, and Proposition 5 to estimate the maximal expected number of clauses that can be simultaneously satisfied in $\varphi \in \Phi(n, dn)$.

**Proposition 11** *Let $r(d) = \beta^2(d)d$ be the fraction of clauses unsatisfied by Majority Vote on $\varphi \in \Phi(n, dn)$, and let $p = p(d)$ be the value of $p$ such that*

$$r(d) = \frac{1}{4}d - p + \frac{1}{3}p\beta(d,p)(2 - \beta(d,p)).$$

*Then whp the maximum number of simultaneously satisfiable clauses of $\varphi \in \Phi(n, dn)$ belongs to the interval $[p(d) - \delta, p(d) + \delta]$, $\delta(d) = 0.2\sqrt{d}$.*

*Proof.* Let $\overline{C}_p$ be the event that $\varphi \in \Phi(n, dn)$ has a $p'$-satisfying assignment, for some $p' \geq p$. Let also $M_p$ denote the event that Majority Vote finds a $p'$-satisfying assignment, $p' \geq p$, and $\overline{M}_p$ the event that it does not happen. By Proposition 9 $\mathcal{P}^u_d(M_{d-(r(d)+\varepsilon)})$ is exponentially small for any $\varepsilon > 0$. On the other hand, by Proposition 5 for $p \geq p(d) + \delta$

$$\begin{aligned}
&\mathcal{P}^u_d(M_{d-(r(d)+\varepsilon)} \wedge \overline{C}_p) \\
&= \mathcal{P}^u_d(\overline{C}_p)\mathcal{P}^{u*}_{d,p}(M_{d-(r(d)+\varepsilon)}) = \mathcal{P}^u_d(\overline{C}_p)(1 - \mathcal{P}^{u*}_{d,p}(\overline{M}_{d-(r(d)+\varepsilon)})) \\
&\geq \mathcal{P}^u_d(\overline{C}_p)\left(1 - \left(\frac{d}{4} - p\right) n \cdot \exp\left[n \log\left(1 + e^{-\frac{\frac{16}{9}p^2}{d}}\right)\right] \mathcal{P}^{pl}_{d,p'}(\overline{M}_{d-(r(d)+\varepsilon)})\right)
\end{aligned}$$

for some $p' \geq p$. Since $\mathcal{P}^{pl}_{d,p'}(\overline{M}_{d-(r(d)+\varepsilon)}) \leq \mathcal{P}^{pl}_{d,p}(\overline{M}_{d-(r(d)+\varepsilon)})$ we assume $p' = p$. By [8] whp $p(d) > \frac{1}{3}\sqrt{d}$. Therefore,

$$\exp\left[n \log\left(1 + e^{-\frac{\frac{16}{9}p^2}{d}}\right)\right] \leq e^{0.599n};$$

and if $n$ is sufficiently large

$$\left(\frac{d}{4} - p\right) n \cdot \exp\left[n \log\left(1 + e^{-\frac{\frac{16}{9}p^2}{d}}\right)\right] \leq e^{0.6n}.$$

On the other hand, since $r(d) = \mathrm{num}(d, p(d))$, if we prove that $\mathrm{num}(d, p(d)) > \mathrm{num}(d, p(d) + \delta) + \delta$, we get

$$\begin{aligned}
\mathcal{P}^{pl}_{d,p}(\overline{M}_{d-(r(d)+\varepsilon)}) &= \mathsf{Prob}[x(d, p(d) + \delta, n) > (r(d) + \varepsilon)n] \\
&\leq \mathsf{Prob}[|x(d, p(d) + \delta, n) - \mathrm{num}(d, p(d) + \delta)n| > \delta n] < e^{-16\frac{\delta^2}{d}n}.
\end{aligned}$$

Thus,

$$\mathcal{P}^u_d(M_{d-(r(d)+\varepsilon)} \wedge \overline{C}_p) \geq \mathcal{P}^u_d(\overline{C}_p)(1 - e^{0.6n} \cdot e^{-0.64n}),$$

and as it is exponentially small, $\mathcal{P}^u_d(\overline{C}_p)$ is exponentially small.

The lower bound is similar.

It remains to show that $\mathrm{num}(d, p(d)) > \mathrm{num}(d, p(d)+\delta)+\delta$. Let $g(x) = x\beta(d, x)(2-\beta(d, x)$. It is not hard to see that $g(x)$ decreases on $[0, \frac{d}{4}]$. Thus,

$$\begin{aligned}
\mathrm{num}(d, p(d) + \delta) + \delta &= \frac{d}{4} - p(d) - \delta + \frac{1}{3}g(p(d) + \delta) + \delta \\
&< \frac{d}{4} - p(d) + \frac{1}{3}g(p(d)) = \mathrm{num}(d, p(d)).
\end{aligned}$$

This completes the proof.

The values of $p(d)$ for some values of $d$ are given in Table 3. It seems that the fraction of satisfiable clauses should be very close to $\frac{3}{4}d + \frac{1}{2}\sqrt{d}$.

| $d$ | $5^2$ | $6^2$ | $7^2$ | $8^2$ | $9^2$ | $10^2$ | $11^2$ | $12^2$ | $13^2$ | $14^2$ | $15^2$ | $16^2$ | $17^2$ | $18^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $p(d)$ | 2.572 | 3.069 | 3.565 | 4.061 | 4.558 | 5.054 | 5.550 | 6.047 | 6.543 | 7.039 | 7.535 | 8.032 | 8.528 | 9.025 |
| $p(d)/\sqrt{d}$ | 0.514 | 0.512 | 0.509 | 0.507 | 0.506 | 0.505 | 0.504 | 0.504 | 0.503 | 0.503 | 0.502 | 0.502 | 0.502 | 0.501 |

**Table 3.** Values of $p(d)$ depending on $d$

## References


1. Achlioptas, D., Naor, A., Peres, Y.: On the maximum satisfiability of random formulas. J. ACM 54(2) (2007)
2. Achlioptas, D., Peres, Y.: The threshold for random $k$-SAT is $2^k \ln 2 - o(k)$. J. AMS 17, 947–973 (2004)
3. Bellare, M., Goldreich, O., Sudan, M.: Free bits, PCPs, and nonapproximability-towards tight results. SIAM J. Comput. 27(3), 804–915 (1998)
4. Bollobás, B.: Random Graphs, Cambridge Studies in Advanced Mathematics, vol. 73. Cambridge University Press (2001)
5. Coja-Oghlan, A., Krivelevich, M., Vilenchik, D.: Why almost all k-CNF formulas are easy, http://www.wisdom.weizmann.ac.il/ vilenchi/papers/2007/UniformSAT.pdf
6. Coja-Oghlan, A., Krivelevich, M., Vilenchik, D.: Why almost all satisfiable k-CNF formulas are easy. In: DTMC. pp. 89–102 (2007)
7. Coja-Oghlan, A., Krivelevich, M., Vilenchik, D.: Why almost all $k$-colorable graphs are easy to color. Theory Comput. Syst. 46(3), 523–565 (2010)
8. Coppersmith, D., Gamarnik, D., Hajiaghayi, M.T., Sorkin, G.B.: Random MAX SAT, random MAX CUT, and their phase transitions. In: SODA. pp. 364–373 (2003)
9. Feige, U.: Relations between average case complexity and approximation complexity. In: IEEE Conference on Computational Complexity. p. 5 (2002)
10. Flaxman, A.: A spectral technique for random satisfiable 3CNF formulas. In: SODA. pp. 357–363 (2003)
11. Friedgut, E.: Sharp thresholds of graph properties, and the $k$-SAT problem. J. Amer. Math. Soc. 12(4), 1017–1054 (1999)
12. Hajiaghayi, M., Kim, J.H.: Tight bounds for random MAX 2-SAT, http://research.microsoft.com/en-us/um/redmond/groups/theory/jehkim/a-yspapers/max2sat.pdf
13. Håstad, J.: Some optimal inapproximability results. J. ACM 48(4), 798–859 (2001)
14. Interian, Y.: Approximation algorithm for random MAX-$k$SAT. In: SAT (Selected Papers. pp. 173–182 (2004)
15. Kim, J.H.: The poisson cloning model for random graphs, random directed graphs and random $k$-SAT problems. In: COCOON. p. 2 (2004)
16. Knuth, D.E.: The Art of Computer Programming, Volume III: Sorting and Searching. Addison-Wesley (1973)
17. Krivelevich, M., Vilenchik, D.: Solving random satisfiable 3CNF formulas in expected polynomial time. In: SODA. pp. 454–463 (2006)
18. Mitchell, D.G., Selman, B., Levesque, H.J.: Hard and easy distributions of sat problems. In: AAAI. pp. 459–465 (1992)
19. de la Vega, W.F., Karpinski, M.: 1.0957-approximation algorithm for random MAX-3SAT. RAIRO - Operations Research 41(1), 95–103 (2007)