On the hardness and easiness of random 4-SAT formulas

Andreas Goerdt and André Lanka

Technische Universität Chemnitz, Fakultät für Informatik
Straße der Nationen 62, 09107 Chemnitz, Germany
{goerdt, lanka}@informatik.tu-chemnitz.de

Abstract. Assuming 3-SAT formulas are hard to refute with high probability, Feige showed approximation hardness results, among others for the max bipartite clique. We extend this result in that we show that approximating max bipartite clique is hard under the weaker assumption, that random 4-SAT formulas are hard to refute with high probability. On the positive side we present an efficient algorithm which finds a hidden solution in an otherwise random not-all-equal 4-SAT instance. This extends analogous results on not-all-equal 3-SAT and classical 3-SAT. The common principle underlying our results is to obtain efficiently information about discrepancy (expansion) properties of graphs naturally associated to 4-SAT instances. In case of 4-SAT (or k-SAT in general) the relationship between the structure of these graphs and that of the instance itself is weaker than in case of 3-SAT. This causes problems whose solution is the technical core of this paper.

1 Introduction and Results

1.1 Some terminology

Given a standard set of n propositional variables $\text{Var} = \text{Var}_n$ a k-clause is an ordered k-tuple $l_1 \vee \cdots \vee l_k$ where $l_i = x$ or $l_i = \neg x$ for an $x \in \text{Var}_n$. We denote the variable underlying the literal l by $\text{Var}(l)$. Lit_n is the set of literals over Var_n. Altogether we have $2^k n^k$ different k-clauses. A k-SAT formula F simply is a set of k-clauses. A clause C is true in the not-all-equal sense under the truth value assignment a if it contains one literal which evaluates to true and another one which evaluates to false under a. This is naturally extended to formulas. The problem to decide satisfiability in the not-all-equal sense for 3-SAT formulas is $\mathcal{NP}$-complete.

Considering any high probability event, that is the probability goes to 1 when n goes to infinity and we have an underlying family of probability spaces for each n, the following certification problem naturally arises: Given a random instance, how can we be sure that this event really holds for the instance at hand? This question can usually be answered running appropriate (inefficient) algorithms with the given instance. We however are interested in an efficient algorithm satisfying the following requirements: It always stops in polynomial time. It says that the instance belongs to the event considered or it gives an inconclusive answer. If the answer is not the inconclusive one the answer must *always* be correct, that is we have a certificate for the event. Moreover the algorithm must be complete, in that it gives the correct answer with high probability with respect to the random instance. In this case we speak of "efficient certification".

1.2 The hardness result

Given $p = p(n)$ with $0 \le p \le 1$ the random formula $\text{Form}_{n,k,p}$ is obtained as follows: Pick each of the $2^k n^k$ k-clauses independently with probability p. $\text{Form}_{n,k,c/n^{k-1}}$ is unsatisfiable

with high probability when $c > \ln 2$ is a constant. This follows from a simple first moment calculation for the number of satisfying assignments. Thus almost all (that is with high probability) formulas are unsatisfiable, but we have no efficient algorithm to certify this. Feige [6] introduces the random 3-SAT hardness hypothesis: For any constant $c > \ln 2$ there is no efficient certification algorithm of the unsatisfiability of $\text{Form}_{n,3,c/n^2}$. The truth of this hypothesis is supported by the fact that for $p(n) = o(1/n^{3/2})$ no progress concerning the efficient certification of unsatisfiability of $\text{Form}_{n,3,p}$ has been made. The best result known is efficient certification of unsatisfiability of $\text{Form}_{n,3,c/n^{3/2}}$ for some sufficiently large constant c, cf. [7]. Feige shows that the random 3-SAT hardness hypothesis implies several lower bounds on the approximability of combinatorial problems for which such bounds could not be obtained from worst-case assumptions like $\mathcal{P} \neq \mathcal{NP}$. As a random hardness hypothesis is much stronger than a mere worst-case hypothesis like $\mathcal{P} \neq \mathcal{NP}$ it is particularly important to weaken it as much as possible. This motivates to consider random 4-SAT instead of 3-SAT. The random 4-SAT hardness hypothesis reads: For any constant $c > \ln 2$ there is no efficient certification algorithm of the unsatisfiability for $\text{Form}_{n,4,c/n^3}$. The trivial reduction: Given $F = \text{Form}_{n,3,c/n^2}$ place a random literal into each clause of F to obtain a 4-SAT instance G, shows, that the 3-SAT hypothesis is stronger than the 4-SAT hypothesis.

Among the problems considered by Feige is the max clique problem for bipartite graphs. Let $G = (V_1, V_2, E)$ be a bipartite graph. V_1 and V_2 are the sets of vertices (V_1 is the left hand side and V_2 is the right hand side) and $E \subseteq V_1 \times V_2$ is the set of edges. A (bipartite) clique in G is a subgraph $H = (W_1, W_2, F)$ of G with $W_i \subseteq V_i$ such that $F = W_1 \times W_2$. Sometimes we denote such a clique simply by (W_1, W_2). We are interested in the optimization problem maximum clique, that is to determine the maximum size of a clique in G. When the size of H is measured as # of vertices of $H = |W_1| + |W_2|$ the problem is solvable in polynomial time [9], problem GT24. If however the size of H is measured as # of edges of $H = |W_1 \times W_2| = |W_1| \cdot |W_2|$ the problem interestingly becomes $\mathcal{NP}$-hard [15] and approximation algorithms are of interest. This is the version of the problem we consider. The approximation ratio of an algorithm for a maximization problem is the maximum size of a solution divided by the size of the solution found by the algorithm. For the classical clique problem no approximation ratio below $n^{1-\varepsilon}$ for any constant $\varepsilon > 0$ is possible by a polynomial time algorithm (unless $\mathcal{P} = \mathcal{NP}$), cf. [12]. Interesting enough, such results are not known for the bipartite case. Feige shows in [6] that there is a constant $\delta > 0$ such that the bipartite clique problem cannot be approximated with a ratio below n^{δ}, provided the random 3-SAT hardness hypothesis holds. Our hardness result is

Theorem 1. *Under the random 4-SAT hardness hypothesis there is no polynomial time approximation algorithm for the bipartite clique problem with a ratio below n^{δ} for some constant $\delta > 0$, where n is the number of vertices.*

The technical heart of the proof of Theorem 1 is the subsequent Theorem 2 from which Theorem 1 is obtained by means of the derandomized graph product [1].

Theorem 2. *Under the random 4-SAT hardness hypothesis there exist two constants $\varepsilon_1 > \varepsilon_2 > 0$ such that no efficient algorithm is able distinguish between bipartite graphs $G =$*

(V_1, V_2, E) with $|V_1| = |V_2| = n$ which have a clique of size $\geq (n/16)^2(1 + \varepsilon_1)$ and those in which all cliques are of size $\leq (n/16)^2(1 + \varepsilon_2)$.

1.3 The easiness result

Given an assignment ϕ we let $CT_i = CT_{i,\phi}$ be the set of all clauses with exactly i literals true $(= 1)$ under ϕ and $4 - i$ false $(= 0)$. We have that $|CT_i| = \binom{4}{i}n^4$. We let $CT_{nae} = CT_{nae,\phi} = CT_1 \cup CT_2 \cup CT_3$ be the set of all clauses satisfied by ϕ in the not-all-equal sense. We describe the way to generate our random formula I. To this end let $0 < \eta_1, \eta_2, \eta_3 < 1$ be three constants and let $d = d_{\eta_1,\eta_2,\eta_3}$ be a (large) constant. We let $p_i = \eta_i d/n^3$ be three probabilities.

1. We pick any assignment ϕ of Var_n. (This is the hidden solution.) Note that ϕ need not be a random assignment, just any assignment.
2. Let $M = CT_{nae,\phi}$. Pick a uniform random clause $C \in M$ and delete it from M. Include C in the random instance I with probability p_i iff $C \in CT_{i,\phi}$.
3. Repeat 2. until $M = \emptyset$.

All instances generated are satisfiable in the not-all-equal sense and we are left with a classical certification problem. Such certification problems have a long tradition. Seminal work has been done by [2] in that spectral techniques have been introduced to find a hidden 3-coloring in a sparse random graph, that is with a linear number of edges. Note that usually hidden solutions in denser instances are easier to find because the structure gives more information. This approach has been further developed to 2-colorings of random 3-uniform hypergraphs (or not all equal 3-SAT instances) with a linear number of edges (clauses) [3] and – recently – to hidden satisfying assignments in a random 3-SAT formula [8]. By developing this approach further we show

Theorem 3. *Let $0 < \eta_i < 1$ be constants and $d = d_{\eta_1,\eta_2,\eta_3}$ be a (large) constant. There is an efficient certification algorithm which finds an assignment π satisfying a random instance I as above in the not-all-equal sense with high probability. (Note that we fix d and then n gets large.)*

2 Proof of Theorem 1

We prove Theorem 2 in the next but one subsection based on

2.1 Discrepancy certification in random bipartite graphs

Let $B = (V_1, V_2, E)$ be a bipartite graph an $2n$ vertices with $|V_1| = |V_2| = n$. Let

$$E(X, Y) = \{\{x, y\} \in E \mid x \in V_1, y \in V_2\}$$

be the set of edges with one endpoint in $X \subseteq V_1$ and the other in $Y \subseteq V_2$. We abbreviate $|E(X, Y)|$ with $e(X, Y)$.

Definition 4. *We say B as above is of* low discrepancy *with respect to ε iff for all $X \subseteq V_1$, $|X| = \alpha n$ and all subsets $Y \subseteq V_2$, $|Y| = \beta n$ we have that*

$$|e(X, Y) - \alpha\beta \cdot |E|| \leq \varepsilon|E|$$

The random bipartite graph $B_{n,c/n}$ has the set of vertices $V_1 = \{1, \ldots, n\}$ and $V_2 = \{n+1, \ldots, 2n\}$. Each each edge $\{x, y\}$ with $x \in V_1$ and $y \in V_2$ is picked with probability c/n independently. Then $B_{n,c/n}$ enjoys the low discrepancy property for each arbitrarily small constant $\varepsilon > 0$ if only $c = c(\varepsilon)$ is large enough.

At first we show that the number of edges in $B_{n,c/n}$ is with high probability $cn \cdot (1+o(1))$. Note that the number of edges $|E|$ in $B_{n,c/n}$ is binomially distributed with parameters n^2 and c/n. So, the expectation of $|E|$ is $n^2 \cdot c/n = cn$.

The well known tail bound for a random variable Z distributed according to the binomial distribution $\text{Bin}(N, p)$ (Chernoff's bound) reads that for any $1 > \delta > 0$

$$\Pr[Z \geq (1+\delta)\mathbf{E}[Z])] \leq \exp((-1/3)\delta^2 \mathbf{E}[Z])) \tag{1}$$

and

$$\Pr[Z \leq (1-\delta)\mathbf{E}[Z]] \leq \exp((-1/2)\delta^2 \mathbf{E}[Z]))]. \tag{2}$$

Letting $\delta = 1/\log n$ we get

$$\Pr[| \,|E| - cn| \geq cn/\log n] \leq 2\exp((-1/3)cn/\log^2 n) = o(1).$$

So with probability $1 - o(1)$ we have that $|E| = cn \cdot (1 + o(1))$.

To show that $B_{n,c/n}$ has the low discrepancy property take some arbitrary small constant $\varepsilon > 0$. Let $X \subseteq V_1$ and $Y \subseteq V_2$ be two fixed subsets with $|X| = \alpha n$ and $|Y| = \beta n$. Then $e(X, Y)$ is a random variable follows the binomial distribution $\text{Bin}(|X| \cdot |Y|, c/n)$. The expectation of $e(X, Y)$ is

$$\mu = \mathbf{E}[e(X, Y)] = |X| \cdot |Y| \cdot c/n = \alpha n \cdot \beta n \cdot c/n = \alpha\beta \cdot cn$$

Picking c sufficiently large, for example such that $\varepsilon \geq 1/\log c$, we see with (1) and (2) together with $|E| = cn \cdot (1 + o(1)) = \mu/(\alpha\beta) \cdot (1 + o(1))$ that

$$\begin{aligned}
\Pr[|e(X, Y) - \mu| \geq \varepsilon|E|] &= \Pr[|e(X, Y) - \mu| \geq \varepsilon/\alpha\beta \cdot (1 + o(1)) \cdot \mu] \\
&< 2 \cdot \exp(-\varepsilon^2/(\alpha\beta)^2 \cdot (1 + o(1)) \cdot \mu/3) \\
&\leq 2 \cdot \exp(-\varepsilon^2 cn/4) \\
&= o(2^{-2n}).
\end{aligned}$$

As we have at most $2^n \cdot 2^n$ possible sets X and Y, we have the low discrepancy property for all sets X and Y defined as above with high probability. We get

Lemma 5. *Given $\varepsilon > 0$ an arbitrarily small constant and $c = c(\varepsilon)$ sufficiently large but constant $B_{n,c/n}$ has low discrepancy with respect to ε with high probability.*

Moreover, there is a polynomial time algorithm `BipDisc` introduced in [4], which is able to check the property stated in the lemma. This algorithm takes as input a bipartite graph $B = (V_1, V_2, E)$. It tries to certify that for all sets $X \subseteq V_1$ with $|X| = \alpha n$ and $Y \subseteq V_2$ with $|Y| = \beta n$

$$|\alpha \cdot \beta \cdot |E| - e(X, Y)| \leq c_1 \cdot \sqrt{\alpha \cdot \beta \cdot |E|n} + n \cdot e^{-|E|/(c_1 \cdot n)} \tag{3}$$

where c_1 is a constant independent of the rest. If the algorithm gets $B_{n,c/n}$ as input it certifies (3) almost surely.

So for any constant $\varepsilon > 0$ and c large enough, for example so that $\varepsilon \geq 1/\log c$ and $c \geq c_1^3$, we have with $|E| = cn \cdot (1 + o(1))$ that asymptotically

$$
\begin{aligned}
c_1 \cdot \sqrt{\alpha \cdot \beta \cdot |E| n} + n \cdot \mathrm{e}^{-|E|/(c_1 \cdot n)} &\leq c_1 \cdot \sqrt{cn^2} + n \cdot \mathrm{e}^{-c/c_1} \\
&\leq c_1 \cdot |E|/\sqrt{c} + n \\
&\leq c^{-1/6} \cdot |E| + c^{-1}|E| \\
&\leq \varepsilon|E|
\end{aligned}
$$

and Algorithm `BipDisc` certifies low discrepancy for every constant $\varepsilon > 0$ if $c = c(\varepsilon)$ is large enough.

2.2   Proof of Theorem 2

Before we take care on the proof of Theorem 2, we review the following algorithm. It takes as input any 4-SAT formula and bounds the number of variables set to true resp. set to false by a satisfying assignment a. Let T_a (resp. F_a) be the set of variables set to true (resp. false) under a. We denote the set of clauses in F containing only non-negated variables $P = P(F)$. This set is also called positive clauses. The clauses containing only negated variables form the set $N = N(F)$ and are called negative clauses.

Algorithm 6.
    Input: A 4-SAT formula F.

1. Set $S := P(F)$ and $i := 0$.
2. While $(S \neq \emptyset)$ do
3.   Take some clause $C = l_1 \vee l_2 \vee l_3 \vee l_4$ from S.
4.   Delete all clauses containing one of the l_i from S.
5. $i := i + 1$
6. Output i as a lower bound on $|T_a|$
7. Repeat 1-5 for $S := N(F)$.
8. Output i as a lower bound on $|F_a|$.

The idea of the algorithm is the following. In every clause C there must be at least one literal true. If we consider the set $P(F)$, at least one variable per clause must be set to true. As we do not know this variable, we delete all clauses containing a variable from the chosen clause C. If some clauses left, we repeat the procedure, because some more variables must be set to true. Looking on $N(F)$ we get a lower bound on the number of variables set to false by a satisfying assignment. This shows the correctness of the algorithm.

On $\mathrm{Form}_{n,4,c/n^3}$ the algorithm almost surely certifies that the number of variables set to true is at least $n/16 \cdot (1 + o(1))$. It gives the same lower bound for the variables set to false. To see this, let k be the value of i in Step 6. We have chosen k clauses and have at most $4k$ different variables in these clauses. Let s be the number of clauses containing one of these variables. Then $\mathbf{E}[s]$ is bounded by $4k \cdot 4|P|/n$. Using Chernoff's bound we derive that with high probability $s \leq 16k \cdot |P|/n \cdot (1 + 1/\log k)$. So we deleted at most

$16k|P|/n \cdot (1 + 1/\log k)$ clauses in Step 4. As we reached step 6 S must be empty. This shows, that k is at least $n/16 \cdot (1 + o(1))$. The other bound can be obtained analogously.

We need this algorithm and its answer for the further results. By using it, we can rely on the important property that any satisfying assignment for a given formula F sets a linear number of variables to true and a linear number to false. We need this now and then and state out the importance when we use this fact.

Now we come to the proof of Theorem 2. The proof relies on the certification of low discrepancy of certain bipartite projection graphs of $\text{Form}_{n,4,c/n^3}$. Let F be a 4-SAT formula and $S \subseteq F$ an arbitrary set of clauses from F. Then we define 6 projection graphs $B_{ij} = (V_1, V_2, E_{ij})$, $1 \le i < j \le 4$, of S. The sets V_1 and V_2 are copies of the variables Var of F. So we set $V_1 = \text{Var} \times \{1\}$ and $V_2 = \text{Var} \times \{2\}$. But for clarity of reading we relinquish on $(x,1)$ (resp. $(y,2)$) and use only x (resp. y). So $x \in V_1$ denotes another vertex than $x \in V_2$ even if they mean the same variable in Var.

We have an edge $\{x,y\} \in E_{ij}$ with $x \in V_1$ and $y \in V_2$ if and only if we have a clause $l_1 \vee l_2 \vee l_3 \vee l_4 \in S$ with $\text{Var}(l_i) = x$ and $\text{Var}(l_j) = y$.

Algorithm 7.

Input: A 4-SAT formula F and $\varepsilon > 0$.

1. Apply Algorithm 6 to F. Give an inconclusive answer if one bound is below $n/20$.
2. Check that $|P| = cn \cdot (1 + o(1))$ and $|N| = cn \cdot (1 + o(1))$.
3. Construct the 6 projection graphs of N and the 6 projection graphs of P. Check for every projection that the number of edges is $\ge |N| \cdot (1 - o(1))$ for N and $|P| \cdot (1 - o(1))$ for P. Give an inconclusive answer if this is not the case.
4. Apply the Algorithm `BipDisc` from Section 2 to certify low discrepancy with respect to $\varepsilon > 0$ to all these projection graphs. Give an inconclusive answer if one application gives an inconclusive answer. Give a positive answer otherwise.

Lemma 8. *Algorithm 7 is complete for* $\text{Form}_{n,4,c/n^3}$ *whenever* c *is a sufficiently large constant.*

Proof. Step 1 is complete as Algorithm 6 gives on $\text{Form}_{n,4,c/n^3}$ almost surely two bounds of size $n/16 \cdot (1 + o(1))$. The completeness of Step 2 follows from Chernoff's bound on $|P|$ and $|N|$. Step 4 is passed successfully as follows from the completeness of `BipDisc` for $B_{n,c/n}$ when c is large enough. Note that in our case the projections considered are random bipartite graphs $B_{n,p}$ with $p = 1 - (1 - c/n^3)^{n^2} = c/n \cdot (1 + o(1))$.

Now we calculate the difference between the number of clauses of P and the number of edges in the projection B_{ij}. As every edge is induced by at least one clause, we must have that $|E_{ij}| \le |P|$. But some clauses induce no edge in E_{ij}. We could have pairs of clauses $l_1 \vee l_2 \vee l_3 \vee l_4, g_1 \vee g_2 \vee g_3 \vee g_4 \in P$ inducing the same edge in G_{ij}. This means $l_i = g_i$ and $l_j = g_j$. The expected number of such pairs is $n^2 \cdot n^4 \cdot (c/n^3)^2 = c^2$. By Markov's inequality the number of these pairs exceeds $\log n$ with probability $o(1)$. So we have with high probability more than $|P| - \log n = |P| \cdot (1 - o(1))$ edges in B_{ij}. The same holds for the projections of N. This implies that $\text{Form}_{n,4,c/n^3}$ passes Step 3 successfully with high probability. $\qquad\square$

Let S be a set of clauses. Then we say a clause $C = l_1 \vee l_1 \vee l_3 \vee l_4$ is of *type* $(X_1, X_2, X_3, X_4)_S$ iff $\text{Var}(l_i) \in X_i$ for $i = 1, \dots, 4$ and $C \in S$.

Low discrepancy of the projections implies interesting properties. Let $|F_a| = \alpha n$ and $|T_a| = (1 - \alpha)n$.

Lemma 9. *Let a be a satisfying assignment for F. Then low discrepancy with respect to ε of the projections gives that $1/3 - O(\varepsilon) \le \alpha \le 2/3 + O(\varepsilon)$.*

Note that the above statement is only useful when ε is very small against α. As $\varepsilon > 0$ is a constant α should have a constant lower bound independent of ε. Remember, this feature is assured by the first step of Algorithm 7.

Proof. Consider the projection $B_{1,1} = (V_1, V_2, E_{1,1})$ of P. Low discrepancy of $B_{1,1}$ gives

$$|e(F_a, F_a) - \alpha^2 \cdot |E_{1,1}|| \le \varepsilon \cdot |E_{1,1}|.$$

The edges in $E_{1,1}(F_a, F_a)$ are induced by clauses of type $(F_a, F_a, \mathrm{Var}, \mathrm{Var})_P$. Together with $|E_{1,1}| = |P| \cdot (1 + o(1))$ we have that $|(F_a, F_a, \mathrm{Var}, \mathrm{Var})_P| = (\alpha^2 + O(\varepsilon))|P|$. As a is a satisfying assignment, the third or the fourth variable in these clauses comes from T_a. This means that

$$|(F_a, F_a, T_a, F_a)_P| + |(F_a, F_a, T_a, T_a)_P| \ge (\alpha^2/2 + O(\varepsilon))|P|$$

or

$$|(F_a, F_a, F_a, T_a)_P| + |(F_a, F_a, T_a, T_a)_P| \ge (\alpha^2/2 + O(\varepsilon))|P|.$$

The first possibility gives that $e(F_a, T_a)$ in $B_{1,3}$ is at least $(\alpha^2/2 + O(\varepsilon))|P|$. But by low discrepancy this is at most $(\alpha \cdot (1 - \alpha) + O(\varepsilon))|P|$. From

$$\alpha^2/2 + O(\varepsilon) \le \alpha \cdot (1 - \alpha) + O(\varepsilon)$$

we get $\alpha \le 2/3 + O(\varepsilon)$. Note, the derivation holds only if α is bounded away from 0 by an independent constant as we divide by α. Here again we use the result given by Algorithm 6.

We get the same bound for the second possibility and the graph $B_{1,4}$. The bound $\alpha \ge 1/3 - O(\varepsilon)$ we get by doing the same things for N beginning with $B_{1,1}$ and $E_{1,1}(T_a, T_a)$. $\qquad \square$

We let $\varrho = |P| = |P(F)|$ and $\nu = |N| = |N(F)|$. Then $\varrho_i = \varrho_{i,a}$ is the number of clauses of P which contain exactly i literals true under a. We use the analogous notation $\nu_i = \nu_{i,a}$ for N.

Then low discrepancy of the projections gives some stronger results than Lemma 9.

Theorem 10. *Given $\varepsilon > 0$ a arbitrarily small constant Algorithm 7 certifies that for any assignment a with $|F_a| = \alpha n$ satisfying $\mathrm{Form}_{n,4,c/n^3}$ the following equations, hold:*

(a)
$$\varrho_0 = 0$$
$$\varrho_2 = 6\alpha^2 \varrho - 3\varrho_1 + O(\varepsilon)\varrho$$
$$\varrho_3 = (-12\alpha^2 + 4\alpha)\varrho + 3\varrho_1 + O(\varepsilon)\varrho$$
$$\varrho_4 = (6\alpha^2 - 4\alpha + 1)\varrho - \varrho_1 + O(\varepsilon)\varrho$$

(b) The equations for the ν_i are analogous with $1 - \alpha$ instead of α:

$$\nu_0 = 0$$
$$\nu_2 = 6(1-\alpha)^2\nu - 3\nu_1 + O(\varepsilon)\nu$$
$$\nu_3 = (-12(1-\alpha)^2 + 4(1-\alpha))\nu + 3\nu_1 + O(\varepsilon)\nu$$
$$\nu_4 = (6(1-\alpha)^2 - 4(1-\alpha) + 1)\nu - \nu_1 + O(\varepsilon)\nu$$

Note that (a) and (b) imply that the ϱ_i, ν_i, $i \geq 2$, are determined by α and ϱ_1, ν_1 up to the $O(\varepsilon)$-terms. The claim of Theorem 10 is only useful if α is substantial larger than ε. This again shows the relevance of Algorithm 6. It certifies that α is bounded away from 0 by a fixed constant. This fact allows us to find a sufficiently small constant $\varepsilon > 0$.

Proof. To show that Algorithm 7 correctly certifies the properties of Theorem 10 let $\varepsilon > 0$ be a constant and F be a 4-SAT formula which passes the algorithm successfully. Let a with $|F_a| = \alpha n$ be a satisfying assignment of F. The first equation $\varrho_0 = 0$ trivially holds as a satisfies F.

By low discrepancy we get for any projection B of $P = P(F)$ that in B

$$e(F_a, F_a) = \alpha^2 \cdot \varrho + O(\varepsilon)\varrho$$

No clause from ϱ_3 induces an edge belonging to $E(F_a, F_a)$. Looking at all 6 projections each clause from ϱ_2 induces one edge in one projection and each clause from ϱ_1 induces one edge in three projections. Thus we have

$$6\alpha^2\varrho + O(\varepsilon) \cdot \varrho = \sum_B e_B(F_a, F_a) = 3\varrho_1 + 1\varrho_2 + o(\varrho), \tag{4}$$

where G ranges over all 6 projection of P. The $o(\varrho)$ term accounting for those pairs of clauses inducing the same edge. In each projection B of P we have

$$e_B(T_a, T_a) = (1-\alpha)^2 \cdot \varrho + O(\varepsilon)\varrho$$

and therefore

$$6(1-\alpha)^2 \cdot \varrho = 6\varrho_4 + 3\varrho_3 + \varrho_2 + O(\varepsilon)\varrho. \tag{5}$$

Finally

$$\varrho = \varrho_4 + \varrho_3 + \varrho_2 + \varrho_1. \tag{6}$$

Remember, $\varrho_0 = 0$ as a is a satisfying assignment. The second equation from (a)

$$\varrho_2 = 6\alpha^2\varrho - 3\varrho_1 + O(\varepsilon)\varrho \tag{7}$$

follows directly from (4). Plugging (7) into (5) yields

$$6(1-\alpha)^2\varrho = 6\varrho_4 + 3\varrho_3 + 6\alpha^2\varrho - 3\varrho_1 + O(\varepsilon) \cdot \varrho$$

and simply algebra gives

$$2\varrho - 4\alpha\varrho = 2\varrho_4 + \varrho_3 - \varrho_1 + O(\varepsilon)\varrho. \tag{8}$$

Plugging (7) into (6) gives

$$(1 - 6\alpha^2)\varrho = \varrho_4 + \varrho_3 - 2\varrho_1 + O(\varepsilon)\varrho. \tag{9}$$

Subtracting (9) from (8) we get

$$2\varrho - 4\alpha\varrho - (1 - 6\alpha^2)\varrho = \varrho_4 + \varrho_1 + O(\varepsilon)\varrho.$$

and simple algebra gives the fourth equation of (a)

$$\varrho_4 = (1 + 6\alpha^2 - 4\alpha)\varrho - \varrho_1 + O(\varepsilon)\varrho. \tag{10}$$

Plugging (10) into (8) we get

$$2\varrho - 4\alpha\varrho = 2\varrho + 12\alpha^2\varrho - 8\alpha\varrho - 2\varrho_1 + \varrho_3 - \varrho_1 + O(\varepsilon)\varrho$$

and this gives the third equation from (a)

$$\varrho_3 = -12\alpha^2\varrho + 4\alpha\varrho + 3\varrho_1 + O(\varepsilon)\varrho$$

(b) follows analogously with N and $|T_a| = (1 - \alpha)n$. $\qquad\qquad\square$

We can obtain Lemma 9 from the above equalities, too. For this use $\varrho \geq \varrho_1 + \varrho_4$ to get $\alpha \geq 2/3 + O(\varepsilon)$. The other bound we could get through $\nu \geq \nu_1 + \nu_4$.

To extend the construction from section 4.1 of [6] from 3-SAT to 4-SAT is the purpose of

Definition 11. *Given two sets V_1 and V_2 of 4-clauses, the bipartite graph $BG(V_1, V_2) = (V_1, V_2, E)$ is defined by: For $C \in V_1$, $D \in V_2$ we have an edge $\{C, D\} \in E$ iff $C = u_1 \vee u_2 \vee u_3 \vee u_4$, $D = v_1 \vee v_2 \vee v_3 \vee v_4$ and for all i $\mathrm{Var}(u_i) \neq \mathrm{Var}(v_i)$.*

As we consider clauses as ordered it can be well that

$$\{x_1 \vee x_2 \vee x_3 \vee x_4 ,\ \neg x_2 \vee \neg x_1 \vee x_4 \vee x_3\} \in E$$

provided the x_i are all distinct. However we never have that

$$\{x_1 \vee x_2 \vee x_3 \vee x_4 ,\ \neg x_1 \vee v_1 \vee v_2 \vee v_3\} \in E$$

as $\mathrm{Var}(x_1) = \mathrm{Var}(\neg x_1) = x_1$.

For a set of clauses S and $1 \leq i \leq 4$ the rotations of S are:

$$\mathrm{ROT}_1(S) = \{v_2 \vee v_3 \vee v_4 \vee v_1 \mid v_1 \vee v_2 \vee v_3 \vee v_4 \in S\}$$
$$\mathrm{ROT}_2(S) = \{v_3 \vee v_4 \vee v_1 \vee v_2 \mid v_1 \vee v_2 \vee v_3 \vee v_4 \in S\}$$
$$\mathrm{ROT}_3(S) = \{v_4 \vee v_1 \vee v_2 \vee v_3 \mid v_1 \vee v_2 \vee v_3 \vee v_4 \in S\}$$
$$\mathrm{ROT}_4(S) = S$$

Corollary 12. *There exists a small constant $\delta > 0$ (e.g. $\delta = 1/50$) such that Algorithm 7 certifies the following property for $\mathrm{Form}_{n,4,c/n^3}$ where c is sufficiently large: If $F = \mathrm{Form}_{n,4,c/n^3}$ is satisfiable there must be a bipartite clique of size $\geq (cn/16)^2 \cdot (1 + \delta)$ in one of the following eight graphs:*

$$BG(P, ROT_i(N)) \ \ with \ 1 \leq i \leq 4$$
$$BG(P, ROT_i(P)) \ \ with \ i = 1, 2$$
$$BG(N, ROT_i(N)) \ \ with \ i = 1, 2$$

Proof. We only need to show that Algorithm 7 correctly certifies the property claimed. To this end let F be a 4-SAT formula which passes Algorithm 7 successfully. We distinguish two cases. In the first case is $\varrho_2 \leq 3/8 \cdot \varrho \cdot (1 + \delta)$ and $\nu_2 \leq 3/8 \cdot \nu \cdot (1 + \delta)$. In the second case at least one inequality is violated. We start with the second case.

Assume $\varrho_2 > 3/8 \cdot \varrho(1 + \delta)$. Note that ϱ_2 refers to six subsets of clauses containing two variables true and two variables false under a. So there is at least one subset with cardinality $\geq 1/16 \cdot \varrho(1 + \delta)$. Let for an example $(T_a, F_a, F_a, T_a)_P$ be this subset. Then $BG(P, \mathrm{ROT}_2(P))$ has a large bipartite clique. For the left side of the clique take all clauses of type $(T_a, F_a, F_a, T_a)_P$ in P. The right side is the rotated set of these clauses. Through the rotation the clauses change to $(F_a, T_a, T_a, F_a)_{\mathrm{ROT}_2(P)}$. As $T_a \cap F_a = \emptyset$, $(T_a, F_a, F_a, T_a)_P$ and $(F_a, T_a, T_a, F_a)_{\mathrm{ROT}_2(P)}$ form a bipartite clique. The size of the clique is bounded below by

$$(1/16 \cdot \varrho(1 + \delta))^2 \geq (cn/16)^2 \cdot (1 + \delta)^2 \cdot (1 - o(1)) > (cn/16)^2 \cdot (1 + \delta).$$

For any of the five other types we get the same bound maybe using $BG(P, \mathrm{ROT}_1(P))$. If $\nu_2 > 3/8 \cdot \nu(1 + \delta)$ use $BG(N, \mathrm{ROT}_1(N))$ resp. $BG(N, \mathrm{ROT}_2(N))$ in the same way.

Now we come to the case $\varrho_2 \leq 3/8 \cdot \varrho(1 + \delta)$ and $\nu_2 \leq 3/8 \cdot \nu(1 + \delta)$ From the second equalities of (a) and (b) in Theorem 10 we get

$$\varrho_1 = 2\alpha^2 \varrho - \varrho_2/3 + O(\varepsilon)\varrho \geq 2\alpha^2 \varrho - 1/8 \cdot \varrho(1 + \delta) + O(\varepsilon)\varrho$$

and

$$\nu_1 \geq 2(1 - \alpha)^2 \nu - 1/8 \cdot \nu(1 + \delta) + O(\varepsilon)\nu.$$

As ϱ_1 consists of four subsets of clauses having exactly one variable true under a we have one subset with cardinality $\geq (\alpha^2/2 - (1 + \delta)/32 + O(\varepsilon))\varrho$. For example this is $(F_a, T_a, F_a, F_a)_P$. Also we get one subset in N having exactly one variable false under a with at least $((1 - \alpha)^2/2 - (1 + \delta)/32 + O(\varepsilon))\nu$ clauses. Let this subset be $(F_a, T_a, T_a, T_a)_N$. Looking at $BG(P, \mathrm{ROT}_3(N))$ we see that this two subsets form a bipartite clique with at least

$$\left(\frac{\alpha^2}{2} - \frac{1 + \delta}{32} + O(\varepsilon)\right) \varrho \cdot \left(\frac{(1 - \alpha)^2}{2} - \frac{1 + \delta}{32} + O(\varepsilon)\right) \nu \tag{11}$$

edges. Conceive (11) as a function of α. Then it is concave for $1/5 \leq \alpha \leq 4/5$. Lemma 9 gives us $1/3 - O(\varepsilon) \leq \alpha \leq 2/3 + O(\varepsilon)$ as a is a satisfying assignment. Because of the concavity we only have to check these both limits to lower bound (11). For ε and δ sufficiently small we get in both cases a lower bound of

$$\left(\frac{(1/3 - O(\varepsilon))^2}{2} - \frac{1 + \delta}{32} + O(\varepsilon)\right) \varrho \cdot \left(\frac{(2/3 + O(\varepsilon))^2}{2} - \frac{1 + \delta}{32} + O(\varepsilon)\right) \nu$$

$$\geq \left(\frac{1}{18} - \frac{1 + \delta}{32} + O(\varepsilon)\right) \varrho \cdot \left(\frac{2}{9} - \frac{1 + \delta}{32} + O(\varepsilon)\right) \nu$$

$$\geq \frac{\varrho \cdot \nu}{250} \geq \frac{(cn \cdot (1 + o(1)))^2}{250} = \frac{(cn)^2}{256} \cdot \frac{256}{250} \cdot (1 + o(1))$$

$$> \left(\frac{cn}{16}\right)^2 \cdot (1 + \delta)$$

$\square$

Theorem 13. *Let $\varepsilon > 0$ be an arbitrarily small constant and $c = c(\varepsilon)$ large enough. For $F = \mathrm{Form}_{n,4,c/n^3}$ the maximum clique size in the graphs below is with high probability bounded above by $(cn/16)^2 \cdot (1 + \varepsilon)$. This applies to the graphs $BG(R,T)$ where R and T each are one among the sets $ROT_i(N(F))$, $ROT_i(P(F))$ for $1 \leq i \leq 4$ ($R = T$ is also possible).*

Proof. Let $G = BG(R,T) = (R,T,E)$. We show the claim for $R = P(F)$ and $T = \mathrm{ROT}_1(P(F))$. Clearly the remaining cases can be treated similarly. Let $K \subseteq R$ and $L \subseteq T$ such that $K \times L \subseteq E$, meaning that K and L induce a clique in G. For $1 \leq i \leq 4$ let

$$K_i = \{x \mid u_1 \vee u_2 \vee u_3 \vee u_4 \in K, x = \mathrm{Var}(u_i)\}$$

and analogously for L_i. By definition of $BG(R,T)$ and as $K \times L \subseteq E$ we have that $K_i \cap L_i = \emptyset$ for all $1 \leq i \leq 4$. The theorem follows when we show that for all sets $K_i \subseteq \mathrm{Var}$, $L_i = \mathrm{Var} \setminus K_i$

$$|(K_1, K_2, K_3, K_4)_R| \cdot |(L_1, L_2, L_3, L_4)_T| \leq (cn/16)^2 \cdot (1 + \varepsilon)$$

with high probability for $\mathrm{Form}_{n,4,c/n^3}$. Given K_i, L_i let

$$X = |(K_1, K_2, K_3, K_4)_R| \qquad \text{and} \qquad Y = |(L_1, L_2, L_3, L_4)_T|.$$

Then X is binomially distributed with parameters κ and c/n^3, and $\kappa = |K_1| \cdot |K_2| \cdot |K_3| \cdot |K_4|$. Y is also binomially distributed but with the parameters λ and c/n^3, and $\lambda = |L_1| \cdot |L_2| \cdot |L_3| \cdot |L_4|$. Note that X and Y can be dependent because $T = \mathrm{ROT}_1(R)$.

Assume first that $\kappa, \lambda \geq \varepsilon n^4$. In this case we have

$$\mathbf{E}[X] = \kappa \cdot c/n^3 \geq \varepsilon \cdot cn \quad \text{and} \quad \mathbf{E}[Y] = \lambda \cdot c/n^3 \geq \varepsilon \cdot cn.$$

By Chernoff's bound we have

$$\Pr[X \geq \mathbf{E}[X] \cdot (1 + \varepsilon^2)] \leq \exp(-\varepsilon^4/3 \cdot \mathbf{E}[X]) \leq \exp(-\varepsilon^5/3 \cdot cn)$$

and

$$\Pr[Y \geq \mathbf{E}[Y] \cdot (1 + \varepsilon^2)] \leq \exp(-\varepsilon^4/3 \cdot \mathbf{E}[X]) \leq \exp(-\varepsilon^5/3 \cdot cn)$$

Concerning the product we get from these estimates that

$$\Pr[X \cdot Y \geq \mathbf{E}[X] \cdot \mathbf{E}[Y] \cdot (1 + \varepsilon^2)^2] \leq 2 \cdot \exp(-\varepsilon^5/3 \cdot cn)$$

The product $\mathbf{E}[X] \cdot \mathbf{E}[Y]$ is maximized when $|K_i| \cdot |L_i| = n/2$ for $1 \leq i \leq 4$. In this case $\kappa = \lambda = n^4/16$, $\mathbf{E}[Y] = \mathbf{E}[X] = n^4/16 \cdot c/n^3 = cn/16$ and

$$\begin{aligned}
&\Pr[X \cdot Y \geq (cn/16)^2 \cdot (1 + \varepsilon)] \\
&\quad \leq \Pr[X \cdot Y \geq (cn/16)^2 \cdot (1 + \varepsilon^2)^2] \qquad\qquad \text{For } \varepsilon \text{ small enough.} \\
&\quad \leq \Pr[X \cdot Y \geq \mathbf{E}[X] \cdot \mathbf{E}[Y] \cdot (1 + \varepsilon^2)^2] \\
&\quad \leq 2 \cdot \exp(-\varepsilon^5/3 \cdot cn).
\end{aligned}$$

Picking c large enough this probability is $o(2^{-4n})$.

The second case arises for $\kappa < \varepsilon n^4$. As $P(F) = cn \cdot (1 + o(1))$ with high probability, we can condition on the event $Y \leq cn \cdot (1 + o(1))$. Let Z be binomially distributed with

the parameters εn^4 and c/n^3. Then we get

$$\Pr[X \geq c/16^2 \cdot n]$$
$$\leq \Pr[Z \geq c/16^2 \cdot n]$$
$$\leq \Pr[Z \geq 1/(256\varepsilon) \cdot \varepsilon c n]$$
$$\leq \Pr[Z \geq 2 \cdot \varepsilon c n] \qquad\qquad\qquad \text{For } \varepsilon < 1/512.$$
$$= \Pr[Z \geq 2 \cdot \mathbf{E}[Z]]$$
$$\leq \exp(-1/3 \cdot \varepsilon c n)$$

leading to

$$\Pr[X \cdot Y \geq (cn/16)^2 \cdot (1+\varepsilon)] \leq \Pr[X \geq c/16^2 \cdot n] \leq \exp(-1/3 \cdot \varepsilon c n),$$

which is $o(2^{-4n})$ when c is large enough. The third case $\lambda < \varepsilon n^4$ can be handled similarly and is omitted. The claim follows as we have only 2^{4n} possibilities to choose $K_1, \ldots, K_4$.

$\square$

Corollary 12 and Theorem 13 shows the correctness of Theorem 2. If we would have an approximation algorithm with ratio for example 1.01, we could distinguish between the satisfiable formulas inducing graphs with cliques $\geq (cn/16)^2 \cdot (1.02)$ (Corollary 12) and the typical formulas whose graphs only have cliques of size e.g. $(cn/16)^2 \cdot (1.001)$ from Theorem 13. This means we could refute 4-SAT on average.

2.3 The proof of Theorem 1

Let $\varepsilon_1 > \varepsilon_2 > 0$ be constants as in Theorem 2. Let G_l (l for large) be the set of graphs $G = (V_1, V_2, E)$ with $|V_1| = |V_2| = n$ having a bipartite clique of size at least $(n/16)^2 \cdot (1 + \varepsilon_1)$. The set G_s (s for small) contains all the graphs $G = (V_1, V_2, E)$ with $|V_1| = |V_2| = n$ and the maximal clique is at most $(n/16)^2 \cdot (1 + \varepsilon_2)$. The size of the cliques in G_l and G_s differ by a factor $(1 + \varepsilon_1)/(1 + \varepsilon_2)$. This factor we call gap. As the gap of $(1 + \varepsilon_1)/(1 + \varepsilon_2)$ is constant, we have no chance to detect it directly with an approximation algorithm A having ratio n^δ. So we construct from G a graph $\mathcal{G} = (\mathcal{V}_1, \mathcal{V}_2, \mathcal{E})$ with $|\mathcal{V}_1| = |\mathcal{V}_2|$ having significantly more vertices and edges as G. The goal is to enlarge the constant gap to a factor of $|\mathcal{V}_1|^\delta$ for some constant $\delta > 0$. Then we can detect the gap with A. Firstly, we examine the following idea:

1. Choose $k \in \mathbb{N}$.
2. Let $\mathcal{V}_1$ be the set of all k-tuples of vertices in V_1.
3. Let $\mathcal{V}_2$ be the set of all k-tuples of vertices in V_2.
4. Two vertices $x = (x_1, \ldots, x_k) \in \mathcal{V}_1$, $y = (y_1, \ldots, y_k) \in \mathcal{V}_2$ induce an edge $\{x, y\} \in \mathcal{E}$ iff $(\{x_1, \ldots, x_k\}, \{y_1, \ldots, y_k\})$ form a bipartite clique in G.

Let $M \subseteq V_1$, then $\mathcal{V}_1(M)$ denotes the set of all tuples in $\mathcal{V}_1$ consisting only of vertices from M and analogously for $N \subseteq V_2$ and $\mathcal{V}_2(N)$. With the above construction $|\mathcal{V}_1(M)| = |M|^k$ and $|\mathcal{V}_2(N)| = |N|^k$.

From die construction of $\mathcal{E}$ the following two statements hold. Firstly for every bipartite clique (L, R) in G we have that $(\mathcal{V}_1(L), \mathcal{V}_2(R))$ is a bipartite clique in $\mathcal{G}$. Secondly for every

clique $(\mathcal{L}, \mathcal{R})$ in $\mathcal{G}$ let L be the vertices in the tuples of $\mathcal{L}$ and R be the vertices in the tuples of $\mathcal{R}$. Then (L, R) form a bipartite clique in G. Note that $\mathcal{L} \subseteq \mathcal{V}_1(L)$ and $\mathcal{R} \subseteq \mathcal{V}_2(R)$.

For $G \in G_s$ we use this fact to get an upper bound for the clique size in $\mathcal{G}$. Let $(\mathcal{L}, \mathcal{R})$ and (L, R) as above. Then

$$|\mathcal{L}| \cdot |\mathcal{R}| \leq |\mathcal{V}_1(L)| \cdot |\mathcal{V}_2(R)| = |L|^k \cdot |R|^k = (|L| \cdot |R|)^k \leq (n/16)^{2k} \cdot (1 + \varepsilon_2)^k$$

bounds the clique size in $\mathcal{G}$.

From the first statement we get for $G \in G_l$ and (L, R) its maximal clique that $\mathcal{G}$ has a clique of size

$$|\mathcal{V}_1(L)| \cdot |\mathcal{V}_2(R)| = |L|^k \cdot |R|^k = (|L| \cdot |R|)^k \geq (n/16)^{2k} \cdot (1 + \varepsilon_1)^k.$$

Now we have a gap of $((1 + \varepsilon_1)/(1 + \varepsilon_2))^k$. For a bounded k this is still constant. But for unbounded k we cannot construct the sets $\mathcal{V}_1$ and $\mathcal{V}_2$ in polynomial time as they have size n^k. So we have to choose a subset of all tuples.

The next idea is to choose every tuple uniform and independent. Then we have for $M \subseteq \mathcal{V}_1$ that $\mathbf{E}[\mathcal{V}_1(M)] = (|M|/n)^k \cdot |\mathcal{V}_1|$. Together with Chernoff's bounds we have with high probability

$$|\mathcal{V}_1(M)| = (|M|/n)^k \cdot |\mathcal{V}_1| \cdot (1 + o(1))$$

provided $|M|$ and $|\mathcal{V}_1|$ are so that $\mathbf{E}[\mathcal{V}_1(M)]$ is linear in n. We get for $G \in G_l$ and (L, R) its maximal clique that $\mathcal{G}$ has a clique of size

$$|\mathcal{V}_1(L)| \cdot |\mathcal{V}_2(R)| \geq \left(\frac{|L|}{n}\right)^k \cdot |\mathcal{V}_1| \cdot \left(\frac{|R|}{n}\right)^k \cdot |\mathcal{V}_2| \cdot (1 + o(1))$$

$$\geq \left(\frac{1 + \varepsilon_1}{256} + o(1)\right)^k \cdot |\mathcal{V}_1| \cdot |\mathcal{V}_2|.$$

For $G \in G_s$ let $(\mathcal{L}, \mathcal{R})$ and (L, R) as in the second statement above. We get

$$|\mathcal{L}| \cdot |\mathcal{R}| \leq |\mathcal{V}_1(L)| \cdot |\mathcal{V}_2(R)|$$

$$\leq \left(\frac{|L| \cdot |R|}{n^2}\right)^k \cdot |\mathcal{V}_1| \cdot |\mathcal{V}_2| \cdot (1 + o(1))$$

$$\leq \left(\frac{1 + \varepsilon_2}{256} + o(1)\right)^k \cdot |\mathcal{V}_1| \cdot |\mathcal{V}_2|.$$

Both facts together give us a gap of

$$\left(\frac{\frac{1+\varepsilon_1}{256} + o(1)}{\frac{1+\varepsilon_2}{256} + o(1)}\right)^k = \left(\frac{1 + \varepsilon_1}{1 + \varepsilon_2} + o(1)\right)^k \geq (1 + \epsilon)^k$$

for some constant $\epsilon > 0$. Now we choose $k = \lceil \ln n \rceil$, then this gap is at least $(1 + \epsilon)^{\ln n} = n^{\ln(1+\epsilon)}$. Choosing every tuple with probability say n^2/n^k, we get with high probability $|\mathcal{V}_1| = n^2 \cdot (1 + o(1))$ and $|\mathcal{V}_2| = n^2 \cdot (1 + o(1))$. So for $\delta < \ln(1 + \epsilon)/2$ algorithm A with ratio n^δ recognizes this large gap. So through this construction of $\mathcal{G}$ A could decide if a given graph G belongs to G_s or to G_l.

But as we are interested in deterministic algorithms we do not want to choose the tuples randomized. We use the so called *derandomized graph product* as introduced in [1]. This makes use of *Ramanujan graphs*, cf. [14]. These regular graphs have good expansion properties. The above construction of $\mathcal{V}_1$ and $\mathcal{V}_2$ is substituted by the following procedure:

1. Choose $k \in \mathbb{N}$ odd and a large constant $d \in \mathbb{N}$.
2. Construct a Ramanujan graph H with n vertices and degree d.
3. Identify every vertex from V_1 with exactly one vertex in H.
4. Enumerate all walks in H of length $k-1$. Each of this walks can be seen as a tuple of k vertices from V_1 (in order of appearance on the walk). Put each such tuple into $\mathcal{V}_1$.
5. Identify every vertex from V_2 with exactly one vertex in H.
6. Enumerate all walks in H of length $k-1$. Put for each walk the relating tuple into $\mathcal{V}_2$.

Note that $|\mathcal{V}_1| = |\mathcal{V}_2| = n \cdot d^{k-1}$ as we have n vertices in H each has d neighbors.

Fact 14. *From [1] section 2 we have for every set $M \subseteq V_1$*

$$|\mathcal{V}_1(M)| \leq |M| \cdot d^{k-1} \cdot \left(\frac{|M|}{n} + \frac{2}{\sqrt{d}} \cdot \left(1 - \frac{|M|}{n} \right) \right)^{k-1}$$

$$|\mathcal{V}_1(M)| \geq |M| \cdot d^{k-1} \cdot \left(\frac{|M|}{n} - \frac{2}{\sqrt{d}} \cdot \left(1 - \frac{|M|}{n} \right) \right)^{k-1}$$

and the same for $N \subseteq V_2$ and $\mathcal{V}_2(N)$.

The first inequality evaluates to

$$|\mathcal{V}_1(M)| \leq |M| \cdot d^{k-1} \cdot \left(\frac{|M|}{n} + \frac{2}{\sqrt{d}} \cdot \left(1 - \frac{|M|}{n} \right) \right)^{k-1}$$

$$\leq \frac{|M|}{n} \cdot n d^{k-1} \cdot \left(\frac{|M|}{n} + O\left(\frac{1}{\sqrt{d}} \right) \right)^{k-1}$$

$$\leq |\mathcal{V}_1| \cdot \left(\frac{|M|}{n} + O\left(\frac{1}{\sqrt{d}} \right) \right)^{k}$$

and the second to $|\mathcal{V}_1(M)| \geq |\mathcal{V}_1| \cdot \left(\frac{|M|}{n} + O\left(\frac{1}{\sqrt{d}} \right) \right)^{k}$ where the constant behind the O is < 0. We get for every $N \subseteq V_2$ on the same way $|\mathcal{V}_2(N)| = |\mathcal{V}_2| \cdot \left(\frac{|N|}{n} + O\left(\frac{1}{\sqrt{d}} \right) \right)^{k}$

With the same calculations as in the randomized case above we get for $G \in G_l$ in $\mathcal{G}$ a clique of size at least $|\mathcal{V}_1| \cdot |\mathcal{V}_2| \cdot ((1+\varepsilon_1)/256 + O(1/\sqrt{d}))^k$. In the case $G \in G_s$ we have in $\mathcal{G}$ only cliques of size at most $|\mathcal{V}_1| \cdot |\mathcal{V}_2| \cdot ((1+\varepsilon_2)/256 + O(1/\sqrt{d}))^k$. So we have a gap of

$$\left(\frac{\frac{1+\varepsilon_1}{256} + O(1/\sqrt{d})}{\frac{1+\varepsilon_2}{256} + O(1/\sqrt{d})} \right)^{k} = \left(\frac{1 + \varepsilon_1 + O(1/\sqrt{d})}{1 + \varepsilon_2 + O(1/\sqrt{d})} \right)^{k} = (1+\epsilon)^{k}$$

for some constant $\epsilon > 0$ provided d large enough. If we set k to the smallest odd integer $\geq \ln n$, we have a gap of at least $(1+\epsilon)^{\ln n} = n^{\ln(1+\epsilon)}$. The number of vertices in $\mathcal{G}$ is bounded above by $2n \cdot d^{k-1} = O(n^{1+\ln d})$, remember d is a constant. So an approximation ratio for A of n^δ with $\delta < \frac{\ln(1+\epsilon)}{1+\ln d}$ and constant suffices to detect the gap. As this contradicts the random 4-SAT hardness hypothesis, we found a δ for Theorem 1. The certification algorithm for unsatisfiability of $Form_{n,4,c/n^3}$ could be the following:

Algorithm 15. Input a 4-SAT formula F.
Step 1. Apply Algorithm 7 to F.
Step 2. Construct $\mathcal{G} = (\mathcal{V}_1, \mathcal{V}_2, \mathcal{E})$ as described for every BG from Corollary 12.
Step 3. Apply A to every $\mathcal{G}$.
Step 4. If A detects a clique of size $\geq |\mathcal{V}_1| \cdot |\mathcal{V}_2| \cdot ((1 + \varepsilon_1)/256 + O(1/\sqrt{d}))^k/|\mathcal{V}_1|^\delta$ give an inconclusive answer, otherwise give a positive answer.

The correctness of the algorithm follows from Corollary 12. Its completeness from Theorem 13 and the completeness of Algorithm 7.

3 Proof of Theorem 3

3.1 The algorithm

For $U_i \subseteq \mathrm{Var}$ we say that a clause $l_1 \vee l_2 \vee l_3 \vee l_4$ is of *type* $\{U_1, U_2, U_3, U_4\}$ if there is a permutation $g_1 \vee g_2 \vee g_3 \vee g_4$ of the l_i such that $\mathrm{Var}(g_i) \in U_i$. Remember $\mathrm{Var}(g_i)$ denotes the variable underlying the literal g_i. Note, the definition of the type is slightly different to that in Section 1, p. 6. Given a (4-SAT) formula F, $\{U_1, U_2, U_3, U_4\}_F$ is the set of clauses of type $\{U_1, U_2, U_3, U_4\}$ in F. We write $\{U_1, U_2, -, -\} = \{U_1, U_2, V, V\}$ and $\{U_1, U_2, -, -\}_F$ then stands for the subset of clauses C of F in which we have two positions one of which is filled with a literal over U_1 and the other one with a literal over U_2. (Note that the literals in the two positions can be equal.)

Given a formula F and an assignment ϕ we let $CT_i(F) = CT_{i,\phi}(F) = F \cap CT_{i,\phi}$ be the set of those clauses of F with exactly i literals true under ϕ. The support in CT_1 of the variable x with respect to F and ϕ is

$$\mathrm{Supp}_{1,F,\phi}(x) = |\{C \in CT_{1,\phi}(F) \,|\, x \in C \text{ and } \phi(x) = 1 \text{ or } \neg x \in C, \ \phi(x) = 0\}|.$$

Similarly for CT_3 we have

$$\mathrm{Supp}_{3,F,\phi}(x) = |\{C \in CT_{3,\phi}(F) \,|\, x \in C \text{ and } \phi(x) = 0 \text{ or } \neg x \in C, \ \phi(x) = 1\}|.$$

Thus $\mathrm{Supp}_{F,\phi}(x) = \mathrm{Supp}_{1,F,\phi}(x) + \mathrm{Supp}_{3,F,\phi}(x)$ is the number of clauses of F which have exactly one literal true or exactly one literal false under ϕ and this literal is either x or $\neg x$. $\mathrm{Occ}_F(x) = \mathrm{Occ}_F(\neg x) = |\{C \in F \,|\, x \in C \text{ or } \neg x \in C\}|$ is the number of clauses of F which contain x or $\neg x$. Note that $\mathrm{Occ}(x)$ need not be equal to the number of actual occurrences of $x, \neg x$ as $x, \neg x$ can occur several times inside a clause.

Recall the generation procedure of our formulas from Subsection 1.3. Let ϕ be the assignment picked in Step 1, then we have that $|CT_{i,\phi}(I)|$ follows the binomial distribution with parameters $|CT_{i,\phi}|$ and p_i. For the expectation we have $\mathbf{E}[|CT_i(I)|] = p_i \cdot 4n^4$ for $i = 1, 3$ and $\mathbf{E}[|CT_2(I)|] = p_2 \cdot 6n^4$. For $x \in \mathrm{Var}$ we can decompose $\mathrm{Occ}_I(x) = X_1 + X_2 + X_3$ where X_i follows the binomial distribution with parameters $4n^4 - 4(n-1)^4 = 16n^3 + O(n^2)$ and p_i for $i = 1, 3$. X_2 follows the binomial distribution with $6n^4 - 6(n-1)^4 = 24n^3 + O(n^2)$. We have that

$$\mathbf{E}[\mathrm{Occ}_I(x)] = (16\eta_1 + 24\eta_2 + 16\eta_3)d + O(1/n).$$

We let $\mu = 16\eta_1 + 24\eta_2 + 16\eta_3$ and $d_i = \eta_i d$ throughout. $\mathrm{Supp}_{F,\phi}(x) = Y_1 + Y_3$ where Y_i follows the binomial distribution with parameter $4n^3 + O(n^2)$ and p_i. Then $\mathbf{E}[\mathrm{Supp}(x)] =$

$4\eta d + O(1/n)$ where $\eta = \eta_1 + \eta_3$ throughout. $R(I)$ is the subset of those variables for which $\mathrm{Occ}_I(x)$ and $\mathrm{Supp}_{I,\phi}(x)$ are approximately right (like the expectation). Given an assignment ϕ and $\varepsilon > 0$ we define

$$R(I) \;=\; R_{\phi,\varepsilon}(I) \;=\; \{x \in V \mid |\mathrm{Occ}_I(x) - \mu d| \le \varepsilon d \;,\; |\mathrm{Supp}_{I,\phi}(x) - 4\eta d| \;\le\; \varepsilon d\}.$$

Concerning $R'(I) \supseteq R(I)$ we are slightly more generous concerning the support:

$$R'(I) \;=\; R'_{\phi,\varepsilon}(I) \;=\; \{x \in V \mid |\mathrm{Occ}_I(x) - \mu d| \le \varepsilon d, |\mathrm{Supp}_{I,\phi}(x) - 4\eta d| \;\le\; 4\varepsilon d\}.$$

Given a set of variables $W \subseteq V$, the boundary of W with respect to a (random) instance I and $\varepsilon > 0$ is $\partial(W) = \partial_{I,\varepsilon}(W) = \{x \in W \mid |\{x, W, W, W\}_I| \le (\mu - 2\varepsilon)d\}$. (Recall that $\mathrm{E}[\mathrm{Occ}(x)] \sim \mu d$.) The core of W with respect to I and $\varepsilon > 0$, $\mathcal{C}(W) = \mathcal{C}_{I,\varepsilon}(W)$, is the largest subset $W' \subseteq W$ with $\partial(W') = \emptyset$. It can be obtained by the following algorithm which iteratively deletes a variable from the current boundary:

$W' := W$
while $\partial(W') \ne \emptyset$ do
Pick any $x \in \partial(W')$; $W' := W' \setminus \{x\}$.
The correctness follows with the invariant $\mathcal{C}(W) \subseteq W'$.

The following algorithm to find a satisfying assignment in the not-all-equal-sense yields the main result. It is inspired by the algorithm in [8] for 3-SAT.

Algorithm 16. Input: Constants d, η_i, ε and a 4-SAT formula I over Var_n (generated as above).

1. Construct the graph $G = G_I = (V, E)$ with $V = \mathrm{Lit}_n$. For $l \ne k \in \mathrm{Lit}_n$ we have $\{l, k\} \in E$ iff we have a clause $C \in I$ with $l, k \in C$. (Note that we have no loops or multiple edges.)
2. Construct $G' = (V, E')$ by deleting all edges incident with vertices $l \in V$ with $d_l \ge 180d$. Here d_l is the degree of l. Compute the eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_{2n}$ of the adjacency matrix A of G'. We have $\sum \lambda_i = 0$ and $\lambda_1 \ge a$ where a is the average degree of G', cf. [16]. Let $e_{2n} = (a_1, \ldots, a_{2n})^t$ (t for transpose) be the eigenvector of the most negative eigenvalue λ_{2n}. Let a_i be the entry corresponding to the variable x_i and a_{n+i} be the entry corresponding to $\neg x_i$.
3. We construct the assignment π: For the variable x_i we let $\pi(x_i) := 1$ if the entry $a_i \ge 0$ and $a_{n+i} < 0$. If $a_i < 0$ and $a_{n+i} \ge 0$ we let $\pi(x_i) := 0$. The variables which have no truth value by now can be given any truth value (for example all 1).
4. For $i = 1, 2, \ldots, \log n$ do
    $W := \{x \in V \mid |\{C \in F \mid (x \in C \text{ or } \neg x \in C) \text{ and } C \text{ false under } \pi\}| \ge 5\varepsilon d\}$.
    For all $x \in W$ do $\pi(x) := 1 - \pi(x)$.
5. We consider the core $\mathcal{C}_{I,\varepsilon}(R'_{\pi,\varepsilon}(I))$. (For $R'_{\pi,\varepsilon}(I)$ recall the definition above.) Modify π to a partial assignment by unassigning all variables not belonging to the core $\mathcal{C}' = \mathcal{C}_{I,\varepsilon}(R'_{\pi,\varepsilon}(I))$.
6. Construct the graph $\Gamma = (\mathrm{Var}, E)$ where

$$\{x, y\} \in E \text{ iff } x, y \in V \setminus \mathcal{C}' \text{ and } \{x, y, -, -\}_I \ne \emptyset.$$

   (Note that $\pi(x), \pi(y)$ is undefined at present.) Determine the connected components of Γ. If any of these has more that $\log n$ vertices the algorithm fails. Otherwise it searches

for a satisfying assignment by trying out all possibilities for each connected component by itself and assigning the unassigned variables of π such that a not-all-equal solution of I is obtained, if possible. If no such assignment is found the inconclusive one is the answer.

For the subsequent analysis of this algorithm we let ϕ be the assignment fixed in Step 1 of the generation algorithm. Usually we denote by I a random instance.

Theorem 17. *For every constant $\delta > 0$ and for all sufficiently large constants d we have for the assignment π after Step 3 of the algorithm that $|\{x \in \text{Var} \,|\, \pi(x) \neq \phi(x)\}| \leq \delta n\}$, or the symmetric statement $|\{x \in \text{Var} \,|\, \pi(x) \neq 1 - \phi(x)\}| \leq \delta n\}$.*

3.2 Proof of Theorem 17

The proof of Theorem 17 follows [8], but due to the application of some recent lemmas from [5] is somewhat simpler.

Let G and G' be the graphs constructed from the random instance I in Step 1 and Step 2. Let I be a random NAE-4-Sat formula generated as described. Assume every literal in I occurs in every position in every clause from $CT_{i,\phi}$ (notation see Subsection 1.3) exactly how it is expected. Then every literal true under ϕ occurs in $4d_1$ clauses from $CT_{1,\phi}$. Note, the 4 comes from the 4 positions the true literal can occupy. Additionally every literal true under ϕ occurs in $12d_2$ clauses from $CT_{2,\phi}$ and in $12d_3$ clauses in $CT_{3,\phi}$. In the same way we get that every false literal in ϕ occurs in $CT_{1,\phi}$ in $12d_1$ clauses, in $12d_2$ clauses from $CT_{2,\phi}$ an in $4d_3$ clauses from $CT_{3,\phi}$.

Now consider the matrix A. To simplify the following things, we assume that every clause in I induces exactly 6 edges to G. Let v_T (resp. v_F) be the characteristic 0/1-vector for the literals true (resp. false) under ϕ. Note that $v_T^t \cdot v_T = n$, $v_F^t \cdot v_F = n$ and $v_F^t \cdot v_T = 0$.

We calculate $A \cdot v_T$. In a row in A corresponding to a literal true under ϕ the number of 1s we count is $12d_2$ (every clause in $CT_{2,\phi}$ contributes one 1 for the sum) $+ \, 2 \cdot 12d_3$ (as each clauses in $CT_{3,\phi}$ contributes two 1s for the sum). In a row corresponding to a literal false under ϕ we get a value of $12d_1 + 2 \cdot 12d_2 + 3 \cdot 4d_3$. So we can say

$$Av_T = (12d_2 + 24d_3) \cdot v_T + (12d_1 + 24d_2 + 12d_3) \cdot v_F \qquad \text{and}$$
$$Av_F = (12d_1 + 24d_2 + 12d_3) \cdot v_T + (24d_1 + 12d_3) \cdot v_F,$$

where the calculation for Av_F is omitted.

Note that for every linear combination v of v_T and v_F Av gives again a linear combination of these both vectors. So, v_T and v_F span an eigenspace of A. The eigenvalues and the eigenvectors of this space can be found by solving the following eigenvalue problem

$$12 \begin{pmatrix} d_2 + 2d_3 & d_1 + 2d_2 + d_3 \\ d_1 + 2d_2 + d_3 & 2d_1 + d_3 \end{pmatrix} \cdot \begin{pmatrix} \beta \\ \gamma \end{pmatrix} = \alpha \cdot \begin{pmatrix} \beta \\ \gamma \end{pmatrix} \tag{12}$$

By calculating and solving the characteristic polynomial of the above matrix we get

$$\alpha_\pm = 12(d_1 + d_2 + d_3 \pm \sqrt{(d_1 + d_2 + d_3)^2 + (d_1 + 2d_2 + d_3)^2 - (d_2 + 2d_3)(2d_1 + d_2)})$$
$$= 12(d_1 + d_2 + d_3 \pm \sqrt{(d_1 + 2d_2 + d_3)^2 + (d_1 - d_3)^2})$$

And we have $\alpha_+ \geq 12(2d_1 + 3d_2 + 2d_3)$ and $\alpha_- \leq -12d_2$. So $\alpha_+ = \Theta(d)$ and $\alpha_- = -\Theta(d)$ where the constants behind the Θ are positive and depend only of η_1, η_2, η_3. The above linear system of equations yields two eigenvectors $f = \beta_- v_T + \beta_- v_F$ and $g = \beta_+ v_F + \gamma_+ v_F$ of A. We normalize f and g so that $\beta_-^2 + \gamma_-^2 = \beta_+^2 + \gamma_+^2 = 1$. This insures that $g^t g = f^t f = n$.

Lemma 18. *Each of $\beta_-, \beta_+, \gamma_-, \gamma_+$ is in absolut value $\geq 1/4$ and $\beta_- \beta_+ + \gamma_- \gamma_+ = 0$. Moreover, β_- and γ_- have different signs.*

Proof. We can follow from (12) that

$$12((d_2 + 2d_3)\beta_+ + (d_1 + 2d_2 + d_3)\gamma_+) = \alpha_+ \beta_+ \qquad \text{and}$$
$$12((d_1 + 2d_2 + d_3)\beta_+ + (2d_1 + 2d_2)\gamma_+) = \alpha_+ \gamma_+.$$

By substituting α we get

$$2d_3 - 2d_1 + (d_1 + 2d_2 + d_3) \cdot \left(\frac{\gamma_+}{\beta_+} - \frac{\beta_+}{\gamma_+} \right) = 0. \tag{13}$$

Assume $|\beta_+| < 1/4$. Then $|\gamma_+| > \sqrt{15}/4$ because $\beta_+^2 + \gamma_+^2 = 1$. So the term $|\frac{\gamma_+}{\beta_+} - \frac{\beta_+}{\gamma_+}| > \sqrt{15} - 1/\sqrt{15}$ exceeds 2 in absolut value. This shows that the left side of (13) cannot reach 0 as either all coefficients of the d_i are positive or all coefficients are negative. The remaining cases can be treated similarly.

The second fact follows directly from $f^t g = 0$ (f and g are orthogonal):

$$0 = f^t g = (\beta_- v_T + \gamma_- v_F)^t (\beta_+ v_T + \gamma_+ v_F) = \beta_- \beta_+ v_T^t v_T + \gamma_- \gamma_+ v_F^t v_F = n(\beta_- \beta_+ + \gamma_- \gamma_+).$$

The signs of β_- and γ_- must be different because $\alpha_- < 0$ and all entries of A are ≤ 0. $\square$

Unfortunately, our random instances I the matrix A has rarely that regular structure. But we will show that the smallest eigenvalue is near α_- and the corresponding eigenvector is similar to f.

Let $E_{T,T}$ be the set of edges in G' with both endpoints are set to true under ϕ. $E_{T,F}$ and $E_{F,F}$ are defined analogously. The following lemma results from standard calculations which are omitted, cf. [8, Appendix A].

Lemma 19. *Let F be a random NAE-4-Sat formula generated as described. For the graph $G' = G'(F)$ the following holds with high probability:*

1. $|E_{T,T}| = (6d_2 + 12d_3 + o(1))n$
2. $|E_{T,F}| = (12d_1 + 24d_2 + 12d_3 + o(1))n$
3. $|E_{F,F}| = (12d_1 + 6d_2 + o(1))n$
4. $|E \setminus E'| \leq 2^{-2d/C}n$, where C is a constant independent from d.

Lemma 20. *With the above notation the following holds with high probability for some constant C independent of d.*

1. $\alpha_- - 2^{-d/C} \leq f^t A f \leq \alpha_- + 2^{-d/C}$
2. $\alpha_+ - 2^{-d/C} \leq g^t A g \leq \alpha_+ + 2^{-d/C}$

Proof. We show the proof for the upper bound in 1. explicit. The other statements can be shown analogously.

$$
\begin{aligned}
f^t A f &= (\beta_- v_T + \gamma_- v_F)^t A (\beta_- v_T + \gamma_- v_F) \\
&= \beta_-^2 v_T^t A v_T + \gamma_- \beta_- v_F^t A v_T + \beta_- \gamma_- v_F^t A v_T + \gamma_-^2 v_F^t A v_F \\
&= \beta_-^2 \cdot 2|E'_{T,T}| + 2\gamma_- \beta_- |E'_{T,F}| + \gamma_-^2 \cdot 2|E'_{F,F}| \\
&\leq \beta_-^2 \cdot 2|E_{T,T}| + 2\gamma_- \beta_- |E_{T,F}| + \gamma_-^2 \cdot 2|E_{T,T}| + 2 \cdot 2^{-2d/C} n \\
&= \beta_-^2 \cdot 2(6d_2 + 12d_3)n + 2\gamma_- \beta_-(12d_1 + 24d_2 + 12d_3)n + \gamma_-^2 \cdot 2(12d_1 + 6d_2)n + \\
&\quad o(n) + 2 \cdot 2^{-2d/C} n \\
&= \begin{pmatrix} \beta_- & \gamma_- \end{pmatrix} \cdot \begin{pmatrix} 12d_2 + 24d_3 & 12d_1 + 24d_2 + 12d_3 \\ 12d_1 + 24d_2 + 12d_3 & 24d_1 + 12d_3 \end{pmatrix} \cdot \begin{pmatrix} \beta_- \\ \gamma_- \end{pmatrix} + 3 \cdot 2^{-2d/C} n \\
&\leq \alpha_- n + 2^{-d/C} n
\end{aligned}
$$

$\square$

As $A = A(G'(F))$ is real valued and symmetric, A has $2n$ (not necessary different) eigenvalues $\lambda_1 \geq \ldots \geq \lambda_{2n}$. Let $e_1, \ldots, e_{2n}$ be a set of corresponding eigenvectors with $\|e_i\| = 1$ and $Ae_i = \lambda_i e_i$, where $\|\cdot\|$ denotes the standard Euclidean norm. As $\lambda_{2n} = \min_{v \neq 0} v^t A v / (v^t v)$ we get that $\lambda_{2n} \leq f^t A f / (f^t f) \leq \alpha_- + 2^{-d/C}$.

Lemma 21. *For any unit vector v perpendicular on v_T and v_F we have with high probability that $\|Av\| = O(\sqrt{d})$.*

Proof. Note, that we have some dependencies between the entries of A. The edges induced by different clauses are independent, but the edges that come from the same clause are not. To avoid problems, we color every 4-clique that comes from a clause in F. We use six colors and every edge from a 4-clique gets a different color uniformly. This gives six partitions of edges. For every partition c we get a matrix A^c containing exactly the entries belonging to the edges in c. So we have six matrices A^c and now there are no dependencies between the entries in any A^c. Surely, the matrices are not independent of each other, but this does not disturb.

Next, we divide every A^c into four blocks $A^c_{i,j}$ with $i, j \in \{F, T\}$. For example, the block $A^c_{T,F}$ contains all entries a_{ij} from A with i true under ϕ and j false under ϕ and having color c.

The reason for these blocks is, that now $A^c_{i,j}$ and the vector v have the following behavior. Every $A^c_{i,j}$ is like a truly random $n \times n$-matrix, where each entry is included with the same probability d'/n.

Although d' is different for every $A^c_{i,j}$, it is constant and linear in d and so can be sufficiently large. The vector v behaves to $A^c_{i,j}$ like a vector perpendicular to the n-dimensional all-1-vector $\mathbf{1}$.

For such random matrices B it is known, for example from [5], that whp. for all unit vectors v perpendicular to $\mathbf{1}$ $\|Bv\| = O(\sqrt{d'})$ holds. Note, in case of $A_{F,T}$ and $A_{T,F}$ we have matrices that are not symmetric and the diagonal elements can be different from 0. For these both use Lemma 45 of [5] and for $A_{T,T}$ and $A_{F,F}$ use Lemma 39 of the cited paper.

With

$$\|Av\| = \left\| \sum_{c=1}^{6} \sum_{i,j \in \{F,T\}} A_{i,j}^c v \right\| \leq \sum_{c=1}^{6} \sum_{i,j \in \{F,T\}} \|A_{i,j}^c v\| \leq 24 \cdot O(\sqrt{d})$$

follows the lemma. □

Fact 22. *Let A, f, g and v as above, then the following facts hold with high probability*

1. $|f^t Av| = O(\sqrt{dn})$
2. $|g^t Av| = O(\sqrt{dn})$
3. $|f^t Ag| = O(2^{-d/C}n)$

Proof. The first fact can be seen easily as $|f^t Av| = |(f, Av)| \leq \|f\| \cdot \|Av\| = \sqrt{n} \cdot O(\sqrt{d})$. In the same way one can conclude 2. A similar calculation as in the proof of Lemma 20 in conjunction with the fact $\beta_- \beta_+ + \gamma_- \gamma_+ = 0$ yields $|f^t Ag| = O(2^{-d/C}n)$.

Lemma 23. *For every constant $\delta > 0$ there exists a constant d' so that for all $d > d'$ the following holds with high probability. There are at most δn coordinates where f and e_{2n} have different signs or at most δn coordinates where $-f$ and e_{2n} have different signs.*

Proof. Note that the vector e_{2n} can be expressed as linear combination of f, g, and v with v perpendicular to f and g and $\|v\| = 1$. As f and g both are linear combinations of v_T and v_F, v fulfills the conditions of Lemma 21. Let $e_{2n} = c_1 f/\|f\| + c_2 g/\|g\| + c_3 v$. As e_{2n}, $f/\|f\|$, $g/\|g\|$, and v are unit vectors, we have $c_1^2 + c_2^2 + c_3^2 = 1$.

With Lemma 21 and Fact 22 in mind we calculate

$$\begin{aligned}
e_{2n}^t A e_{2n} &= \left(c_1 \frac{f}{\|f\|} + c_2 \frac{g}{\|g\|} + c_3 v \right)^t A \left(c_1 \frac{f}{\|f\|} + c_2 \frac{g}{\|g\|} + c_3 v \right) \\
&= c_1^2 \frac{f^t Af}{\|f\|^2} + c_2^2 \frac{g^t Ag}{\|g\|^2} + c_3^2 v^t Av + 2 \left(c_1 c_2 \frac{f^t Ag}{\|f\| \cdot \|g\|} + c_1 c_3 \frac{f^t Av}{\|f\|} + c_2 c_3 \frac{g^t Av}{\|g\|} \right) \\
&\geq c_1^2 \frac{f^t Af}{\|f\|^2} + c_2^2 \frac{g^t Ag}{\|g\|^2} - c_3^2 \|v\| \cdot \|Av\| + O(\sqrt{d}) \\
&\geq c_1^2 \alpha_- + c_2^2 \alpha_+ + O(\sqrt{d}),
\end{aligned}$$

where the constant behind the O is negative. As $e_{2n}^t A e_{2n} = \lambda_{2n} \leq \alpha_- + 2^{-d/C}$ we get

$$\begin{aligned}
\alpha_- + 2^{-d/C} &\geq c_1^2 \alpha_- + c_2^2 \alpha_+ + O(\sqrt{d}) \\
\alpha_- &\geq c_1^2 \alpha_- + c_2^2 \alpha_+ + O(\sqrt{d}) \\
1 &\leq c_1^2 + c_2^2 \frac{\alpha_+}{\alpha_-} + O(\sqrt{d}/\alpha_-).
\end{aligned}$$

As shown earlier both α_- and α_+ are linear in d and $\alpha_- < 0$. This gives for some constant $c' > 0$ independent from d

$$1 \leq c_1^2 - c' \cdot c_2^2 + O(1/\sqrt{d})$$

where the constant behind the O now is positive. Now set $c_1 = 1 - \delta'$ with $0 \leq \delta' \leq 1$

$$\begin{aligned}
1 &\leq 1 - 2\delta' + \delta'^2 - c' \cdot c_2^2 + O(1/\sqrt{d}) \\
0 &\leq -2\delta' + \delta'^2 - c' \cdot c_2^2 + O(1/\sqrt{d}) \\
0 &\leq -\delta' - c' \cdot c_2^2 + O(1/\sqrt{d})
\end{aligned}$$

As the both first terms have negative signs, they must be in absolut value $O(1/\sqrt{d})$. And so by setting d sufficiently large δ' becomes arbitrarily small. From $e_{2n} = c_1 f/\|f\| + c_2 g\|f\| + c_3 v$ we get $\|e_{2n} - f/\|f\| \| = c_2^2 + c_3^2 \leq 2\delta'$. Let x be the number of entries where f and e_{2n} have different signs. Remember, each entry of f is at least $1/4$ in absolut value. So in every of the x entries e_{2n} must be at least $1/(4\sqrt{n})$ in absolut value. This gives $2\delta' \geq \|e_{2n} - f/\|f\| \| \geq x/16n$ respective $x \leq 32\delta' n$. As δ' can be made sufficiently small by increasing d, the claim holds. For the case $2 \geq \delta' > 1$ the above argumentation works with $-f$ instead of f. $\qquad\square$

Now Theorem 17 follows as β_-, γ_- have different signs and therefore we know that the positive entries of f either correspond to the literals set to 1 by ϕ and the negative ones to those set to 0 or the other way round.

3.3 Proof of Theorem 3

The remaining part of the algorithm is also analyzed based on Flaxman's work, but some subtle details have to be taken care of. With Theorem 17 we assume that for the assignment π after Step 3 $|\{\pi(x) = \phi(x)\}| \geq (1 - \delta)n$. If $|\{\pi(x) = 1 - \phi(x)\}| \geq (1 - \delta)n$ we proceed analogously. We pick an ε sufficiently small (for this d must be sufficiently large.)

Lemma 24.

1. *With probability $1 - \mathrm{e}^{-\Omega(n)}$ we have $|R_{\phi,\varepsilon}(I)| \geq (1 - e^{-d/C})n$ for a constant C independent of d.*
2. *If $\delta = \delta(\varepsilon)$ is a sufficiently small constant, then we have with probability $1 - O(n^{-\sqrt{d}})$ for all $U \subset \mathrm{Var}, |U| \leq 2\delta n$ that $|\{U, U, -, -\}_I| \leq 1/9 \cdot \varepsilon d|U|$.*

Proof. 1. Choose some arbitrarily small constant ε' with $0 < \varepsilon' < \varepsilon/5$. Let $\mathrm{Occ}^i(x)$ be the number of clauses in I having a literal over x at the i'th position. Let Z_x^i be an indicator variable with $Z_x^i = 1$ if $|\mathrm{Occ}^i(x) - (4d_1 + 6d_2 + 4d_3)| > \varepsilon'd$ and $Z_x^i = 0$ otherwise. Further below we show that $\Pr[Z_x^i = 1] \leq \mathrm{e}^{-d/C'}$. Let $Z^i = \sum_x Z_x^i$, then Z^i is binomially distributed with parameters n and $\Pr[Z_x^i = 1]$. This gives $\mathbf{E}[Z^i] \leq \mathrm{e}^{-d/C'}n$. We use the general Chernoff bound which holds for any $\delta \geq 0$

$$\Pr[Z^i \geq (1 + \delta) \cdot \mathbf{E}[Z^i]] \leq \left(\frac{\mathrm{e}^\delta}{(1 + \delta)^{1+\delta}}\right)^{\mathbf{E}[Z^i]}.$$

So we get for any $m \geq \mathbf{E}[Z^i]$

$$\Pr[Z^i \geq m] = \Pr\left[Z^i \geq \frac{m}{\mathbf{E}[Z^i]} \cdot \mathbf{E}[Z^i]\right] \leq \left(\frac{\mathrm{e}^{m/\mathbf{E}[Z^i]-1}}{(m/\mathbf{E}[Z^i])^{m/\mathbf{E}[Z^i]}}\right)^{\mathbf{E}[Z^i]}$$

$$= \frac{\mathrm{e}^{m-\mathbf{E}[Z^i]}}{(m/\mathbf{E}[Z^i])^m} = \mathrm{e}^{-\mathbf{E}[Z^i]} \cdot \left(\frac{\mathrm{e} \cdot \mathbf{E}[Z^i]}{m}\right)^m$$

$$\leq \left(\frac{\mathrm{e} \cdot \mathbf{E}[Z^i]}{m}\right)^m \tag{14}$$

Now we choose $m = 3e^{-d/C'} \geq 3 \cdot \mathbf{E}[Z^i]$ and obtain

$$\Pr\left[Z^i \geq 3e^{-d/C'}n\right] \leq \left(\frac{e \cdot \mathbf{E}[Z^i]}{3\mathbf{E}[Z^i]}\right)^{3e^{-d/C'}n} = (e/3)^{\Omega(n)} = e^{-\Omega(n)}$$

So we have for each fixed i that with probability $e^{-\Omega(n)}$ more than $3e^{-d/C'}n$ variables x have $|\text{Occ}^i(x) - (4d_1 + 6d_2 + 4d_3)| > \varepsilon'd$. Thus with probability $\leq 1 - 4e^{-\Omega(n)}$ we have at each position at most $3e^{-d/C'}n$ variables with this property. Summing over the 4 positions we get with probability $1 - 4e^{-\Omega(n)}$ that at most $12e^{-d/C'}n$ variables x have

$$\left|\sum_{i=1}^{4} \text{Occ}^i(x) - (16d_1 + 24d_2 + 16d_3)\right| > \varepsilon'd. \tag{15}$$

Since $\sum_{i=1}^{4}\text{Occ}^i(x) \geq \text{Occ}(x)$ and $\varepsilon > \varepsilon'$ we have with probability $1 - 4e^{-\Omega(n)}$ that there are at most $12e^{-d/C'}$ variables x with $\text{Occ}(x) - (16d_1 + 24d_2 + 16d_3) > \varepsilon$.
The difference between $\sum_{i=1}^{4}\text{Occ}^i(x)$ and $\text{Occ}(x)$ is bounded by 4-times the number of clauses containing x twice or more. Similarly to the above calculation one can show that with probability $1 - e^{-\Omega(n)}$ there are at most $e^{-d/C'}n$ variables x with $|\{x, x, -, -\}_I| > \varepsilon'd$. So for at least $(1 - e^{-d/C'})n$ variables x we have $\text{Occ}(x) \geq \sum_{i=1}^{4}\text{Occ}^i(x) - 4\varepsilon'd$ yielding together with (15)

$$(16d_1 + 24d_2 + 16d_3) - \text{Occ}(x) \leq (16d_1 + 24d_2 + 16d_3) - \left(\sum_{i=1}^{4}\text{Occ}^i(x) - 4\varepsilon'd\right) < 5\varepsilon'd$$

for at least $(1 - 13e^{-d/C'})n$ variables x with probability $1 - 5e^{-\Omega(n)}$. All these variables fulfil the requirements of $R_{\phi,\varepsilon}(I)$ with respect to Occ as $\varepsilon' < \varepsilon/5$
Remark, the proof of $\Pr[Z_x^i = 1] \leq e^{-d/C'}$ is still missing. We do this know. Fix x and i. Remember, $Z_x^i = 1$ iff $|\text{Occ}^i(x) - (4d_1 + 6d_2 + 4d_3)| \geq \varepsilon'd$. We partition the clauses containing x at the i'th position into three groups depending on the number of literals true under ϕ. We denote by $\text{Occ}_j^i(x)$ the number of clauses in I containing exactly j literals true under ϕ and having x at its i'th position. Clearly we have $\text{Occ}^i(x) = \sum_{j=1}^{3}\text{Occ}_j^i(x)$. Each $\text{Occ}_j^i(x)$ is binomially distributed for $j = 1, 2, 3$ with parameters $4n^3$ and d_1/n^3, $6n^3$ and d_2/n^3, $4n^3$ and d_3/n^3. Using Chernoff's bound it is easy to show that for each j

$$\Pr[|\text{Occ}_j^i(x) - \mathbf{E}[\text{Occ}_j^i(x)]| \geq \varepsilon'd/3] \leq e^{-d/(2C')}$$

for some constant C' independent of d. This leads to
(a) $\Pr[|\text{Occ}_1^i(x) - 4d_1| \geq \varepsilon'd/3] \leq e^{-d/(2C')}$
(b) $\Pr[|\text{Occ}_2^i(x) - 6d_2| \geq \varepsilon'd/3] \leq e^{-d/(2C')}$
(c) $\Pr[|\text{Occ}_3^i(x) - 4d_3| \geq \varepsilon'd/3] \leq e^{-d/(2C')}$.
As

$$|\text{Occ}^i(x) - (4d_1 + 6d_2 + 4d_3)| \leq |\text{Occ}_1^i(x) - 4d_1| + |\text{Occ}_2^i(x) - 6d_2| + |\text{Occ}_3^i(x) - 4d_3|$$

we get finally

$$
\begin{aligned}
\Pr[Z_x^i = 1] &= \Pr[|\mathrm{Occ}^i(x) - (4d_1 + 6d_2 + 4d_3)| \geq \varepsilon' d] \\
&\leq \Pr[|\mathrm{Occ}_1^i(x) - 4d_1| + |\mathrm{Occ}_2^i(x) - 6d_2| + |\mathrm{Occ}_3^i(x) - 4d_3| \geq \varepsilon' d] \\
&\leq \Pr[|\mathrm{Occ}_1^i(x) - 4d_1| \geq \varepsilon' d/3] + \Pr[|\mathrm{Occ}_2^i(x) - 6d_2| \geq \varepsilon' d/3] \\
&\quad + \Pr[|\mathrm{Occ}_3^i(x) - 4d_3| \geq \varepsilon' d/3] \\
&\leq 3e^{-d/(2C')} \\
&\leq e^{-d/C'}.
\end{aligned}
$$

It remains to show that with probability $(1 - e^{-\Omega(n)})$ at most $\mathrm{e}^{-d/C''}n$ variables have the wrong support. This gives that at most $e^{-d/C''}n + 13e^{-d/C'}n < e^{-d/C}n$ variables are not in $R_{\phi,\varepsilon}$ with probability $1 - e^{-\Omega(n)}$. The proof for Supp is omitted because it is analogously to the above proof for Occ.

2. Fix any set $U \subset \mathrm{Var}$ with $|U| = \alpha n \leq 2\delta n$. Fix i and j with $1 \leq i < j \leq 4$ and let $X_{i,j}$ denote the number of clauses in I having variables from U at it's i'th and j'th position. Note that $\sum_{i,j=1;\ i<j}^4 X_{i,j} \geq |\{U, U, -, -\}_I|$. We show that with probability $\leq (c \cdot \alpha^2)^{\alpha n \cdot \sqrt{d}}$ the value of $X_{i,j}$ exceeds $\varepsilon' \alpha \cdot nd$ for $\varepsilon' < \varepsilon/54$. Then a simple union bound shows that with probability $\leq 6 \cdot (c \cdot \alpha^2)^{\alpha n \cdot \sqrt{d}}$

$$
|\{U, U, -, -\}_I| \geq 6\varepsilon' \alpha \cdot nd > 1/9 \cdot \varepsilon \alpha \cdot nd.
$$

As we have at most $\binom{n}{\alpha n} \leq (\mathrm{e}/\alpha)^{\alpha n}$ such sets U we get finally

$$
\begin{aligned}
\Pr[\text{There is a set } U \text{ with} &|\{U, U, -, -\}_I| \geq 1/9 \cdot \alpha\varepsilon \cdot nd] \\
&\leq 6 \left(c \cdot \alpha^2\right)^{\alpha n \cdot \sqrt{d}} \cdot \left(\frac{\mathrm{e}}{\alpha}\right)^{\alpha n} \\
&\leq \left(c \cdot \alpha^2\right)^{\alpha n \cdot \sqrt{d}} \cdot \left(\frac{\mathrm{e}}{\alpha}\right)^{\alpha n \cdot \sqrt{d}} \\
&\leq (c \cdot \mathrm{e} \cdot \alpha)^{\alpha n \cdot \sqrt{d}}
\end{aligned}
$$

Seeing $(c \cdot \mathrm{e} \cdot \alpha)^{\alpha n \cdot \sqrt{d}}$ as a function of α, it is convex for $\alpha > 0$. As $1 \leq \alpha n \leq 2\delta n$ it suffices to check the bounds $\alpha = 1/n$ and $\alpha = 2\delta$. For $\alpha = 1/n$ we get a value of $(c \cdot \mathrm{e}/n)^{\sqrt{d}} = O(n^{-\sqrt{d}})$. To bound the other case we choose δ sufficiently small, i.e. so that $c \cdot \mathrm{e} \cdot \alpha \leq c \cdot \mathrm{e} \cdot 2\delta < 1$. Then we get a value of $(c \cdot \mathrm{e} \cdot 2\delta)^{2\delta n \cdot \sqrt{d}} = \mathrm{e}^{-\Omega(n)}$ since the basis is < 1. The claim follows.

We are left to bound $\Pr[X_{i,j} > \varepsilon' \alpha \cdot nd]$ for a fixed set U. The random variable $X_{i,j}$ follows the binomial distribution. Its expectation is bounded by $16(\alpha n)^2 \cdot n^2 \cdot d/n^3 = 16\alpha^2 dn$ as we have $(\alpha n)^2 \cdot n^2$ possibilities to choose the variables, 16 ways to set the negation signs and each clause is chosen with probability at most d/n^3.

We make use of inequality (14) and get

$$\Pr[x_{i,j} \geq \varepsilon'\alpha \cdot nd] \leq \left(\frac{\mathrm{e} \cdot \mathbf{E}[X_{i,j}]}{\varepsilon'\alpha \cdot nd}\right)^{\varepsilon'\alpha \cdot nd}$$

$$\leq \left(\frac{\mathrm{e} \cdot 16\alpha^2 nd}{\varepsilon'\alpha \cdot nd}\right)^{\varepsilon'\alpha \cdot nd}$$

$$\leq \left(\frac{16\mathrm{e} \cdot \alpha}{\varepsilon'}\right)^{\varepsilon'\alpha \cdot nd}$$

$$< \left(\frac{256\mathrm{e}^2 \cdot \alpha^2}{\varepsilon'^2}\right)^{\alpha n \cdot \sqrt{d}} \qquad \text{We can assume } \varepsilon' \cdot \sqrt{d} > 2.$$

$$< (C \cdot \alpha^2)^{\alpha n \cdot \sqrt{d}}$$

for some constant C independent of d and n.

$\square$

The core $\mathcal{C}_{I,\varepsilon}(R_{\phi,\varepsilon}(I))$ is only slightly smaller than $R_{\phi,\varepsilon}(I)$ itself.

Lemma 25. *If I fulfils the properties of Lemma 24, then we have $|\mathcal{C}_{I,\varepsilon}(R_{\phi,\varepsilon}(I))| \geq (1 - 2^{-d/C})n$.*

Note, that the lemma means that we have $|\mathcal{C}_{I,\varepsilon}(R_{\phi,\varepsilon}(I))| \geq (1 - 2^{-d/C})n$ with probability at least $1 - O(n^{-\sqrt{d}})$.

Proof. Let $R = R_{\phi,\varepsilon}(I)$ and $\mathcal{C} = \mathcal{C}_{I,\varepsilon}(R)$ and recall the algorithm from Section 3.1 to generate $\mathcal{C}$. We show that the while loop of this algorithm is executed $m \leq e^{-d/C}n$-times. Then the result follows, for $|\mathrm{Var} \setminus \mathcal{C}| = |\mathrm{Var} \setminus R| + m \leq 2e^{-d/C}n \leq 2^{-d/C}n$.

Assume that the loop of the algorithm is executed at least m-times and consider the first m executions of the loop. (We specify m further below.) Let $x_i = x$ after the i'th execution of the loop and let $\mathcal{C}_0 = R$ and $\mathcal{C}_i = \mathcal{C}_{i-1} \setminus \{x_i\}$. Then $\mathcal{C}_i$ is the value of W' of the algorithm after the i'th execution of the while loop. Let $U_i = \mathrm{Var} \setminus \mathcal{C}_i$. As $x_i \in \partial(\mathcal{C}_{i-1}) \subseteq R$ we have that $|\{x_i, -, -, -\}_I| \geq (\mu - \varepsilon)d$. As $x_i \in \partial(\mathcal{C}_{i-1})$ we have that $|\{x_i, \mathcal{C}_{i-1}, \mathcal{C}_{i-1}, \mathcal{C}_{i-1}\}_I| \leq (\mu - 2\varepsilon)d$. Therefore $|\{x_i, U_{i-1}, -, -\}_I| \geq \varepsilon d$ and thus $\sum_{i-1}^{m} |\{x_i, U_{i-1}, -, -\}_I| \geq m\varepsilon d$. Clauses from $\{x_i, x_{i+1}, x_{i+2}, U_{i-1}\}_I$ are counted 3-times in the sum. No clause is counted 4 or more times. Thus the number of different clauses contributing to the sum is $\geq 1/3m\varepsilon d$. As for all i $\{x_i, U_{i-1}, -, -\}_I \subseteq \{U_m, U_m, -, -\}_I$ we get that $|\{U_m, U_m, -, -\}_I| \geq 1/3m\varepsilon d$. Now assuming $m = e^{-d/C}n$ we have that $|U_m| = 2m \leq 2\delta n$ and $|\{U_m, U_m, -, -\}_I| \geq 1/6|U_m|\varepsilon d$ contradicting item 2 of Lemma 24. $\square$

Lemma 26. *Let $\mathcal{C} = \mathcal{C}_{I,\varepsilon}(R_{\phi,\varepsilon}(I))$, $\pi_i =$ the assignment π after the i'th execution of the loop in step 4 of the algorithm, and let $B_i = \{x \in \mathcal{C} \mid \pi_i(x) \neq \phi(x)\}$. If the properties of Lemma 24 hold for I then we have for all $i \leq \log n$ that $|B_i| \leq |B_{i-1}|/2$.*

This lemma directly implies that after Step 4 all variables from $\mathcal{C}$ have the right truth value.

Corollary 27. *With high probability we have after step 4 that for the core $\mathcal{C}$ as above* $\mathcal{C} \subseteq \{x \in \mathrm{Var}_n \,|\, \pi(x) = \phi(x)\}$.

Proof of Lemma 26. From Theorem 17 we know that $|B_0| \leq \delta n$. Further below we show that for all $x \in B_i$ we have $|\{x, B_{i-1}, -, -\}_I| \geq 2\varepsilon d$. This implies the claim as follows. By induction we can assume that $|B_{i-1}| \leq \delta n$. Assuming that $|B_i| > |B_{i-1}|/2$ we let $B' \subseteq B_i$ with $|B'| = \lfloor |B_{i-1}|/2 \rfloor + 1$. From the statement above we get that $\sum_{x \in B'} |\{x, B_{i-1}, -, -\}_I| \geq |B'|2\varepsilon d$. For $x_1, \ldots, x_4 \in B' \cap B_{i-1}$ all distinct a clause like $x_1 \vee x_2 \vee x_3 \vee x_4$ is counted 4-times. No clause is counted more than 4-times. This implies that $|\{B', B_{i-1}, -, -\}_I| \geq |B'|2\varepsilon d/4$. Now consider $B = B' \cup B_{i-1}$, then we have that $|B| \leq 2\delta n$, but $|\{B, B, -, -\}_I| \geq |\{B', B_{i-1}, -, -\}_I| \geq |B|2\varepsilon d/16$ as $|B'| \geq |B|/4$ in contradiction to item 2 of Lemma 24.

We now show the statement above by the case distinction that for $x \in B_i$ either $x \in B_{i-1}$ or $x \notin B_{i-1}$. Let

$$a = |\{C \in F \,|\, (x \in C \text{ or } \neg x \in C) \text{ and } C \text{ false under } \pi_i\}|.$$

If $x \in B_i$ and $x \in B_{i-1}$ we have that $\pi_{i-1}(x) = \pi_i(x) = \neg\phi(x)$. Moreover, as the value of x has not been changed by the loop we know that $a \leq 5\varepsilon d$. As $x \in \mathcal{C} \subseteq R(I)$ we have that $\mathrm{Supp}_{I,\phi}(x) \geq (4\eta - \varepsilon)d$. Therefore we have at least $(4\eta - \varepsilon)d - 5\varepsilon d = (4\eta - 6\varepsilon)d$ clauses $C \in F$ with the property: There is a literal $l \in C$ which makes C true under π_{i-1}, but for the underlying variable y we have that $\pi_{i-1}(y) \neq \phi(y)$. As $x \in \mathcal{C}$ we have that $|\{x, -, -, -\}_I| \leq (\mu + \varepsilon)d$ and $|\{x, \mathcal{C}, \mathcal{C}, \mathcal{C}\}_I| \geq (\mu - 2\varepsilon)d$. Therefore $|\{x, \mathrm{Var} \setminus \mathcal{C}, -, -\}_I| \leq 3\varepsilon d$. Hence, among the $(4\eta - 6\varepsilon)d$ clauses containing x above, we have at least $(4\eta - 6\varepsilon)d - 3\varepsilon d = (4\eta - 9\varepsilon)d \geq 2\varepsilon d$ (ε sufficiently small) clauses which contain a literal over a variable y from C which is false under π_{i-1}. For this y we clearly have $y \in B_{i-1}$.

If $x \in B_i$ and $x \notin B_{i-1}$ the value of x has been changed in the loop of the algorithm and we know that $a \geq 5\varepsilon d$. Each of these a clauses obviously contains a literal over a variable y such that $\pi_{i-1}(y) = \neg\phi(y)$. We show that for at least $2\varepsilon d$ of these a clauses we have that $y \in B_{i-1}$. This follows as $|\{x, -, -, -\}_I| \leq (\mu + \varepsilon)d$ and $|\{x, \mathcal{C}, \mathcal{C}, \mathcal{C}\}_I| \geq (\mu - 2\varepsilon)d$. Therefore $|\{x, \mathrm{Var} \setminus \mathcal{C}, -, -, \}_I| \leq 3\varepsilon d$ and we get $|\{x, B_{i-1}, -, -, \}_I| \geq 2\varepsilon d$. $\qquad\square$

After Step 5 of the algorithm the core $\mathcal{C}$ remains correctly assigned.

Lemma 28. *Let π be the partial assignment obtained after executing Step 5. If I complies with the items of Lemma 24 then we have:*

1. $\mathcal{C} = \mathcal{C}_{I,\varepsilon}(R_{\phi,\varepsilon}(I)) \subseteq \{x \,|\, \pi(x) \text{ defined }\}$.

2. For all x with $\pi(x)$ defined we have $\pi(x) = \phi(x)$.

Proof. Let π' be the value of the assignment π before Step 5.

1. By Corollary 27 we have that for $x \in \mathcal{C}$ $\pi'(x) = \phi(x)$. We show below that $\mathcal{C} \subseteq R'_{\pi',\varepsilon}$ which clearly implies that $\mathcal{C} \subseteq \mathcal{C}_{I,\varepsilon}(R'_{\pi,\varepsilon})$ and the variables from $\mathcal{C}$ are still correctly assigned in π after step 5. Let $x \in \mathcal{C}$. Then $x \in R_{\phi,\varepsilon}(I)$ and we have $|\{x, -, -, -\}_I| \leq (\mu + \varepsilon)d$, and $|\{x, \mathcal{C}, \mathcal{C}, \mathcal{C}\}_I| \geq (\mu - 2\varepsilon)d$. From this we directly get that $|\{x, \mathrm{Var} \setminus \mathcal{C}, -, -\}_I| \leq 3\varepsilon d$. Again as $x \in R_{\phi,\varepsilon}(I)$ we have that $\mathrm{Supp}_{I,\phi,\varepsilon}(x) \geq (4\eta - \varepsilon)d$. As π' is equal to ϕ when restricted to $\mathcal{C}$ we have that $\mathrm{Supp}_{I,\pi',\varepsilon}(x) \geq (4\eta - 4\varepsilon)d$. As $\mathrm{Supp}_{I,\phi,\varepsilon}(x) \leq (4\eta + \varepsilon)d$ we get similarly that $\mathrm{Supp}_{I,\pi',\varepsilon}(x) \leq (4\eta + 4\varepsilon)d$. Which shows that $\mathcal{C} \subseteq R'_{\pi,\varepsilon}$ and the proof is finished.

2. Let $\mathcal{C}' = \mathcal{C}_{I,\varepsilon}(R'_{\pi',\varepsilon}(I))$ and let $\pi(x)$ be defined after step 5 that is $x \in \mathcal{C}' \subseteq R'_{\pi',\varepsilon}$. Thus $\mathrm{Occ}_I(x) \leq (\mu + \varepsilon)d$ and $|\{x, \mathcal{C}', \mathcal{C}', \mathcal{C}'\}_I| \geq (\mu - 2\varepsilon)d$. This directly implies that $|\{x, \mathrm{Var} \setminus \mathcal{C}', -, -\}_I| \leq 3\varepsilon d$. Moreover, we have $\mathrm{Supp}_{I,\pi'}(x) \geq (4\eta - 4\varepsilon)d$ and we have $\geq (4\eta - 7\varepsilon)d$ clauses in $\{x, \mathcal{C}', \mathcal{C}', \mathcal{C}'\}_I$ which are also counted in $\mathrm{Supp}_{I,\pi'}(x)$. If x is not correctly assigned we would have that $\phi(x) = \neg\pi'(x)$ and all these $\geq (4\eta - 7\varepsilon)d$ clauses have a literal over a variable y which is also incorrectly assigned under π' that is $\phi(y) = \neg\pi'(y)$. With $U = \{y \mid \pi'(y) = \neg\phi(y)\}$ we have for that $|\{x, U, -, -\}_I| \geq (4\eta - 7\varepsilon)d$. Summing over all such x we get that $\sum_{x \in U} |\{x, U, -, -\}_I| \geq |U|(4\eta - 7\varepsilon)d$. In this sum each clause can be counted at most 4-times and we have that $|\{U, U, -, -\}| \geq |U|(4\eta - 7\varepsilon)d/4$. As $|U| \leq 2^{-d/C}$ by Corollary 27 this contradicts item 2 of Lemma 24. □

Each connected component of Γ of size $\geq \log n$ contains a connected component of size $\log n$. We show that the expected value of such components goes to 0. To this end let $T' = (V(T'), E(T'))$ be a fixed tree (connected graph without cycles) with $V(T') \subseteq \mathrm{Var}$ and $|V(T')| = \log n$. Let $T \subseteq CT_{nae,\phi}$ be a fixed set of 4-clauses such that for each $\{x, y\} \in E(T')$ we have a clause $C \in \{x, y, -, -\}_T$. Let T be a minimal set with this property. The tree T' induced by T occurs only then in Γ if firstly $V(T') \cap \mathcal{C}_I(R(I)) = \emptyset$ and secondly $T \subseteq I$. Thus we have to bound

$$\mathrm{Pr}[T \subseteq I \text{ and } V(T') \cap \mathcal{C}_I(R(I))] = \mathrm{Pr}[T \subseteq I] \cdot \mathrm{Pr}[V(T') \cap \mathcal{C}_I(R(I)) \mid T \subseteq I].$$

Lemma 29. *Let T and T' fixed as above then*

1. $\mathrm{Pr}[T \subseteq I] \leq (d/n^3)^{|T|}$
2. $\mathrm{Pr}[V(T') \cap \mathcal{C}_I(R(I)) \mid T \subseteq I] = O\left(n^{-\sqrt{d}}\right)$

The first item is easy to see as each of the $|T|$ clauses is chosen with probability at most d/n^3. The second one is more difficult to show – not only because of the condition $T \subseteq I$.

We need to disregard those vertices from $V(T')$ which occur to often in T. To this end let

$$J = \{x \in V(T') \mid |\{x, -, -, -\}_T| \leq 8\}.$$

Note that $J \subseteq V(T')$.

Given any 4-SAT formula F we define the set of variables

$$\hat{R}(F) = \{x \notin V(T) \text{ or } x \in J \mid (\mu - \varepsilon)d \leq \mathrm{Occ}_F(x) \leq (\mu + \varepsilon)d - 8 \text{ and}$$
$$(4\eta - \varepsilon)d \leq \mathrm{Supp}_F(x) \leq (4\eta + \varepsilon)d - 8\}.$$

Similarly we abbreviate $\hat{R}(F) = \hat{R}_{\phi,\varepsilon}(F)$. Clearly we have $\hat{R}(F) \subseteq R(F)$, but more holds:

Lemma 30. *We have for all F that $\mathcal{C}_{F,\varepsilon}(\hat{R}(F)) \subseteq \mathcal{C}_{F \cup T,\varepsilon}(R(F \cup T))$.*

Proof. Let $x \in \hat{R}(F)$, then from the definition of $\hat{R}(F)$ we have that $(\mu + \varepsilon)d - 8 \geq \mathrm{Occ}_F(x) \geq (\mu - \varepsilon)d$ and additionally $x \notin V(T)$ or $x \in J$. If $x \notin V(T)$, then $\mathrm{Occ}_F(x) = \mathrm{Occ}_{F \cup T}(x)$. So x complies the requirements of $R(F \cup T)$ with respect to Occ. If $x \in J$ then x occurs in at most 8 clauses from T, so we have $\mathrm{Occ}_{F \cup T}(x) \leq \mathrm{Occ}_F(x) + 8$. Again

x fulfills the conditions of $R(F \cup T)$ with respect to Occ. In the same way we get that any $x \in \hat{R}(F)$ complies the conditions of $R(F \cup T)$ with respect to Supp. We see that $\hat{R}(F) \subseteq R(F \cup T)$.

Let $\hat{\mathcal{C}}_0 = \hat{R}(F)$ and let $\mathcal{C}_0 = R(F \cup T)$. We define $\mathcal{C}_{i+1} = \mathcal{C}_i \setminus \partial(\mathcal{C}_i)$ and $\hat{\mathcal{C}}_{i+1} = \hat{\mathcal{C}}_i \setminus \partial(\hat{\mathcal{C}}_i)$. Note there exist i_0 and j_0 so that $\mathcal{C}_F(\hat{R}(F)) = \hat{\mathcal{C}}_i$ for any $i > i_0$ and $\mathcal{C}_{F \cup T}(R(F \cup T)) = \mathcal{C}_j$ for any $j > j_0$. If we can show $\hat{\mathcal{C}}_i \subseteq \mathcal{C}_i$ for any $i \geq 0$ the claim follows.

We use induction over i. For $i = 0$ the property holds as shown above. If $x \in \hat{\mathcal{C}}_{i+1}$, then $x \in \hat{\mathcal{C}}_i$ and $x \notin \partial(\hat{\mathcal{C}}_i)$. So the number of clauses of type $\{x, \hat{\mathcal{C}}_i, \hat{\mathcal{C}}_i, \hat{\mathcal{C}}_i\}$ exceeds $(\mu - 2\varepsilon)d$.

As $\hat{\mathcal{C}}_i \subseteq \mathcal{C}_i$ by induction we have that $x \in \mathcal{C}_i$ and additional $|\{x, \hat{\mathcal{C}}_i, \hat{\mathcal{C}}_i, \hat{\mathcal{C}}_i\}_F| \leq |\{x, \mathcal{C}_i, \mathcal{C}_i, \mathcal{C}_i\}_{F \cup T}|$. The second statement gives $x \notin \partial(\mathcal{C}_i)$. Together with the first statement this implies $x \in (\mathcal{C}_i \setminus \partial(\mathcal{C}_i)) = \mathcal{C}_{i+1}$ and shows $\hat{\mathcal{C}}_{i+1} \subseteq \mathcal{C}_{i+1}$. $\qquad\square$

Lemma 31. $\Pr_I[J \cap \mathcal{C} = \emptyset \,|\, T \subseteq I] \leq \Pr_I[J \cap \hat{\mathcal{C}} = \emptyset]$ where $\hat{\mathcal{C}} = \mathcal{C}_{I,\varepsilon}(\hat{R}(I))$

Proof. Let $F_1 \supseteq T$ be a formula with $J \cap \mathcal{C}_{F_1}(R(F_1)) = \emptyset$ and let $F' = F \setminus T$. Then for any $M \subseteq T$ we have $J \cap \mathcal{C}_{F' \cup M}(\hat{R}(F' \cup M)) = \emptyset$ by Lemma 30 ($F' \cup M$ is in the lemma F).

The probability that F_1 is chosen conditioned on the event our random I contains T equals the probability that $F' \cup M$ is chosen conditioned on $T \cap I = M$: Since each clause is chosen independently and $F' \cap T = \emptyset$ we have for our random I that

$$
\begin{aligned}
\Pr[I = F_1 \,|\, T \subseteq I] &= \Pr[I = (F' \cup T) \,|\, T \subseteq I] \\
&= \Pr[I \setminus T = F' \,|\, T \subseteq I] \\
&= \Pr[I \setminus T = F' \,|\, T \cap I = M] \\
&= \Pr[I = F' \cup M \,|\, T \cap I = M].
\end{aligned}
$$

This gives for all $M \subseteq T$

$$
\Pr_I[J \cap \mathcal{C}_I(R(I)) = \emptyset \,|\, T \subseteq I] \leq \Pr_I[J \cap \mathcal{C}_I(\hat{R}(I)) = \emptyset \,|\, M \subseteq I].
$$

We conclude

$$
\begin{aligned}
\Pr[J \cap \mathcal{C}_I(\hat{R}(I)) = \emptyset] &= \sum_{M \subseteq T} \Pr[T \cap I = M] \cdot \Pr[J \cap \mathcal{C}_I(\hat{R}(I)) = \emptyset \,|\, T \cap I = M] \\
&\geq \sum_{M \subseteq T} \Pr[T \cap I = M] \cdot \Pr[J \cap \mathcal{C}_I(R(I)) = \emptyset \,|\, T \subseteq I] \\
&= \Pr[J \cap \mathcal{C}_I(R(I)) = \emptyset \,|\, T \subseteq I]
\end{aligned}
$$

$\qquad\square$

We are left to show

Lemma 32. $\Pr[J \cap \hat{\mathcal{C}} = \emptyset] = O(n^{-\sqrt{d}})$.

Proof. We denote by ξ the event that the following two properties hold for our random instance I.

1. The number of variables x in Var with $(\mu - \varepsilon)d \leq \text{Occ}_I(x) \leq (\mu + \varepsilon)d - 8$ and $(4\eta - \varepsilon)d \leq \text{Supp}_I(x) \leq (4\eta + \varepsilon)d - 8$ is at least $(1 - e^{-d/C'})n$ for some constant C' independent of d.

2. Property 2. from Lemma 24 holds: For all $U \subset \text{Var}, |U| \leq 2\delta n$ we have $|\{U, U, -, -\}_I| \leq 1/9 \cdot \varepsilon d|U|$.

Similar to the proof of Lemma 24 one can show that ξ holds with probability $1 - O(n^{-\sqrt{d}})$. We get

$$\Pr[J \cap \hat{\mathcal{C}} = \emptyset] = \Pr[\xi] \cdot \Pr[J \cap \hat{\mathcal{C}} = \emptyset \,|\, \xi] + \Pr[\bar{\xi}] \cdot \Pr[J \cap \hat{\mathcal{C}} = \emptyset \,|\, \bar{\xi}]$$
$$= \Pr[\xi] \cdot \Pr[J \cap \hat{\mathcal{C}} = \emptyset \,|\, \xi] + O(n^{-\sqrt{d}})$$
$$\leq \Pr[J \cap \hat{\mathcal{C}} = \emptyset \,|\, \xi] + O(n^{-\sqrt{d}})$$

Similar to Lemma 25 one can conclude that if ξ holds $|\hat{\mathcal{C}}|$ is bounded below by $(1 - 2^{-d/C'})n$ for some constant C' independent of d.

Conditioned on the event that $|\hat{\mathcal{C}}| = m$ all possible $\hat{\mathcal{C}}$ have the same probability to appear. To show this let $\hat{\mathcal{C}}$ and $\hat{\mathcal{C}}'$ be two cores of cardinality m. Let F be an arbitrary formula F with k clauses inducing $\hat{\mathcal{C}}$. By a simple bijective renaming of the variables F can be translated to a formula F' with k clauses inducing $\hat{\mathcal{C}}'$. So for any k the number of formulas with k clauses inducing $\hat{\mathcal{C}}$ is equal to the number of formulas with k clauses inducing $\hat{\mathcal{C}}'$. So $\hat{\mathcal{C}}$ and $\hat{\mathcal{C}}'$ have the same probability to appear.

As J is fixed, we see abbreviating $|J| = j$ that

$$\Pr[J \cap \hat{\mathcal{C}} = \emptyset \,|\, |\hat{\mathcal{C}}| = m] = \frac{\binom{n-j}{m}}{\binom{n}{m}} = \frac{\binom{n-m}{j}}{\binom{n}{j}} = \frac{(n - m) \cdot (n - m - 1) \cdot \ldots \cdot (n - m - j + 1)}{n \cdot (n - 1) \cdot \ldots \cdot (n - j + 1)}$$

One can see that the last fraction is maximized when j and m are minimized. We show below that $j \geq 1/2 \cdot \log n$. Since we condition on ξ we have $m \geq (1 - 2^{-d/C'})n$ and get

$$\Pr[J \cap \hat{\mathcal{C}} = \emptyset \,|\, \xi] \leq \frac{(2^{-d/C}n) \cdot (2^{-d/C}n - 1) \cdot \ldots \cdot (2^{-d/C}n - j + 1)}{n \cdot (n - 1) \cdot \ldots \cdot (n - j + 1)}$$
$$\leq \frac{(2^{-d/C}n) \cdot (2^{-d/C}(n - 1)) \cdot \ldots \cdot (2^{-d/C}(n - j + 1))}{n \cdot (n - 1) \cdot \ldots \cdot (n - j + 1)}$$
$$= \left(2^{-d/C}\right)^j$$
$$\leq 2^{-d/C \cdot 1/2 \cdot \log n}$$
$$= n^{-d/(2C)}.$$

This leads to
$$\Pr[J \cap \hat{\mathcal{C}} = \emptyset] \leq n^{-d/2C} + O(n^{-\sqrt{d}}) = O(n^{-\sqrt{d}}).$$

It remains to show that $|J| \geq 1/2 \cdot \log n$: As T' is a tree, there are $V(T') - 1$ edges inside T'. So we need at most $V(T') - 1$ clauses in T to "cover" each edge of T'. Now assume $|J| < |V(T')|/2$. Then we have at least $|V(T')|/2$ variables in $V(T')$ occurring more than 8 times in clauses from T. Each clause from T can cover at most 4 of these variables. So we must have more than $|V(T')|$ clauses in T. This is a contradiction to the minimality of T.

$\square$

We come to the calculation of the expected number of connected components of size $\log n$ in Γ.

Lemma 33. *With high probability we have that Γ contains no connected component of size larger than $\log n$.*

Proof.

$$\Pr[\,\Gamma \text{ has a connected component larger than } \log n]$$

$$\leq \mathbf{E}[\#\text{connected components larger than } \log n] \qquad \text{(Markov-Inequality)}$$

$$\leq \sum_{T,T'} \Pr[T \subseteq I \text{ and } V(T') \cap \mathcal{C} = \emptyset]$$

$$\leq \sum_{T,T'} \left(\frac{d}{n^3}\right)^{|T|} \cdot O(n^{-\sqrt{d}}) \qquad\qquad (16)$$

where the sums goes over all trees T' with $V(T') \subseteq \text{Var}$ and all minimal sets of clauses T so that for any edge $\{x,y\} \in E(T')$ we have a clause in $\{x,y,-,-\}_T$.

At next we bound the number of possibilities of such pairs (T,T'). For this look at the following generation process for T and T'.

1. Choose $\log n$ variables for T'.
2. Choose e_3, the number of clauses covering three edges in T'.
3. Choose e_2, the number of clauses covering exactly two edges in T' incident with 4 variables.
4. Choose e_2', the number of clauses covering exactly two edges in T' incident with exactly 3 variables. Let $e_1 = \log n - 1 - 3e_3 - 2e_2 - 2e_2'$ be the number of clauses covering exactly one edge in T'.
5. For each of the e_3 clauses, choose a different 4-tuple of variables from T' and three edges to connect these variables.
6. For each of the e_2 clauses, choose a different 4-tuple of variables from T' and two edges lying between these variables.
7. For each of the e_2' clauses, choose 3 variables from T', 1 arbitrarily variable and two edges lying between. For each of these clauses we must have a different 4-tuple of variables.
8. For each of the e_1 clauses, choose a different pair of variables from T', connect them and choose 2 arbitrarily variables.
9. For every generated tuple of 4 variables choose a clause comprising these variables.

Clearly, the generated T' is not necessarily a tree and maybe T is not minimal. But we can build any pair (T,T') satisfying our requirements. We get an upper bound for the number of possible pairs (T,T') generated by the above process:

1. $\binom{n}{\log n}$ possibilities.
2. A number between 0 and $1/3 \cdot \log n$.
3. A number between 0 and $1/2 \cdot \log n$.
4. A number between 0 and $1/2 \cdot \log n$.
5. $\binom{\log^4 n}{e_3}$ possibilities to choose the 4-tuples and for each 16 ways to connect the four vertices.

6. $\binom{\log^4 n}{e_2}$ possibilities to choose the 4-tuples and for each 3 ways to connect the four vertices.

7. $\binom{\log^3 n}{e_2'}$ possibilities to choose the 3-tuples, for each 3 ways to connect it, and for each n possibilities to choose the last variable.

8. $\binom{\log^2 n}{e_1}$ possibilities to choose the pairs and for each pair n^2 possibilities to choose the remaining two variables.

9. For each of the $e_1 + e_2 + e_2' + e_3$ 4-tuples we have 24 possibilities to permutate them and 16 ways to set the negation signs. This leads to $384^{|T|}$.

So the number of possibilities for (T, T') is bounded above by

$$\sum_{e_3=0}^{1/3 \cdot \log n} \sum_{e_2+e_2'=0}^{(\log n - 3e_3)/2} \binom{n}{\log n} \cdot 16^{e_3} \binom{\log^4 n}{e_3} \cdot 3^{e_2} \binom{\log^4 n}{e_2} \cdot (3n)^{e_2'} \binom{\log^3 n}{e_2'} \cdot n^{2e_1} \binom{\log^2 n}{e_1} \cdot 384^{|T|}$$

Note that the value of e_1 is fixed by the other three values as $3e_3 + 2(e_2 + e_2') + e_1 = |E(T')| = \log n - 1$. We have that each e_i and e_2' is ≥ 0 and bounded above by $|T| \leq \log n$. We simplify the above expression by the inequality $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$ to a weaker upper bound:

$$\sum_{e_3=0}^{1/3 \cdot \log n} \sum_{e_2+e_2'=0}^{(\log n - 3e_3)/2} C^{\log n} \cdot \left(\frac{en}{\log n}\right)^{\log n} \cdot \left(\frac{e \log^4 n}{e_3}\right)^{e_3} \cdot \left(\frac{e \log^4 n}{e_2}\right)^{e_2} \cdot \left(\frac{e \log^3 n}{e_2'}\right)^{e_2'} \cdot n^{e_2'+2e_1} \left(\frac{e \log^2 n}{e_1}\right)^{e_1}$$

for some constant C. Since any e_i and e_2' are bounded by $\log n$, we have at most

$$\sum_{e_3=0}^{1/3 \cdot \log n} \sum_{e_2+e_2'=0}^{(\log n - 3e_3)/2} C'^{\log n} \cdot \frac{n^{\log n} \cdot (\log n)^{4e_3 + 4e_2 + 3e_2' + 2e_1} \cdot n^{e_2'+2e_1}}{(\log n)^{\log n} \cdot e_3^{e_3} \cdot e_2^{e_2} \cdot e_2'^{e_2'} \cdot e_1^{e_1}} \tag{17}$$

valid pairs (T, T'). With $|T| = e_3 + e_2 + e_2' + e_1$ and $3e_3 + 2(e_2 + e_2') + e_1 = \log n - 1$ in mind we plug (17) into (16):

$\Pr[\Gamma$ has a connected component larger than $\log n]$

$$\leq \sum_{e_3=0}^{1/3 \cdot \log n} \sum_{e_2+e_2'=0}^{(\log n - 3e_3)/2} C'^{\log n} \cdot \frac{n^{\log n} \cdot (\log n)^{4e_3 + 4e_2 + 3e_2' + 2e_1} \cdot n^{e_2'+2e_1}}{(\log n)^{\log n} \cdot e_3^{e_3} \cdot e_2^{e_2} \cdot e_2'^{e_2'} \cdot e_1^{e_1}} \left(\frac{d}{n^3}\right)^{|T|} \cdot O(n^{-\sqrt{d}})$$

$$\leq \sum_{e_3=0}^{1/3 \cdot \log n} \sum_{e_2+e_2'=0}^{(\log n - 3e_3)/2} C'^{\log n} \cdot \frac{n^{\log n} \cdot (\log n)^{e_3 + 2e_2 + e_2' + e_1} \cdot n^{e_2'+2e_1}}{e_3^{e_3} \cdot e_2^{e_2} \cdot e_2'^{e_2'} \cdot e_1^{e_1}} \left(\frac{d}{n^3}\right)^{|T|} \cdot O(n^{-\sqrt{d}})$$

$$\leq \sum_{e_3=0}^{1/3 \cdot \log n} \sum_{e_2+e_2'=0}^{(\log n - 3e_3)/2} C'^{\log n} \cdot \frac{n^{\log n} \cdot (\log n)^{e_3 + 2e_2 + e_2' + e_1} \cdot n^{e_2'+2e_1} \cdot d^{|T|}}{e_3^{e_3} \cdot e_2^{e_2} \cdot e_2'^{e_2'} \cdot e_1^{e_1} \cdot n^{3e_2 + 3e_2 + 3e_2' + 3e_1}} \cdot O(n^{-\sqrt{d}})$$

$$\leq \sum_{e_3=0}^{1/3 \cdot \log n} \sum_{e_2+e_2'=0}^{(\log n - 3e_3)/2} C'^{\log n} \cdot \frac{(\log n)^{e_3 + 2e_2 + e_2' + e_1} \cdot d^{|T|}}{e_3^{e_3} \cdot e_2^{e_2} \cdot e_2'^{e_2'} \cdot e_1^{e_1} \cdot n^{e_2}} \cdot O(n^{-\sqrt{d}})$$

$$\leq \sum_{e_3=0}^{1/3 \cdot \log n} \sum_{e_2+e_2'=0}^{(\log n - 3e_3)/2} C'^{\log n} \cdot \frac{(\log n)^{e_3 + e_2 + e_2' + e_1} \cdot d^{|T|}}{e_3^{e_3} \cdot e_2^{e_2} \cdot e_2'^{e_2'} \cdot e_1^{e_1}} \cdot O(n^{-\sqrt{d}}) \tag{18}$$

Now we lower bound the denominator. The product is minimized, when the logarithm of the product is minimized. So we deal with $e_3 \log e_3 + e_2 \log e_2 + e_2' \log e_2' + e_1 \log e_1$. Since $x \log x$ is a convex function for $x > 0$ and $e_3 + e_2 + e_2' + e_1$ is fixed to $|T|$, we are able to use Jensen's inequality. This gives that the sum is minimized when each of the e's is at $|T|/4$. So the above product is bounded below by $(|T|/4)^{e_3 + e_2 + e_2' + e_1}$. As $1/3 \cdot \log n \le |T| \le \log n$ holds, we have

$$\frac{(\log n)^{e_3 + e_2 + e_2' + e_1}}{e_3^{e_3} \cdot e_2^{e_2} \cdot e_2'^{e_2'} \cdot e_1^{e_1}} \le \frac{(\log n)^{e_3 + e_2 + e_2' + e_1}}{(|T|/4)^{e_3 + e_2 + e_2' + e_1}} \le \frac{(\log n)^{e_3 + e_2 + e_2' + e_1}}{(1/12 \cdot \log n)^{e_3 + e_2 + e_2' + e_1}} \le 12^{\log n}.$$

Then (18) simplifies to

$$\sum_{e_3=0}^{1/3 \cdot \log n} \sum_{e_2+e_2'=0}^{(\log n - 3e_3)/2} C'^{\log n} \cdot 12^{\log n} \cdot d^{|T|} \cdot O(n^{-\sqrt{d}}) \le \log^2 n \cdot n^{\log C''} \cdot n^{\log d} \cdot O(n^{-\sqrt{d}}) = o(1)$$

for d large enough but still constant. We are done. $\qquad\square$

References

1. Alon, N., Feige, U., Widgerson, A., Zuckerman, D.: Derandomized Graph Products. Computational Complexity 5 (1995), 60–75.
2. Alon, N., Kahale N. : A spectral technique for coloring random 3-colourable graphs. DIMACS TR 94-35, 1994.
3. Chen H., Frieze, A.: Coloring bipartite hypergraphs. Proc. 5th IPCO 1996, LNCS, 345 - 358.
4. Coja-Oghlan, A., Goerdt, A., and Lanka, A.: Strong Refutation Heuristics for Random k-SAT. RANDOM 2004. To appear. (http://www.tu-chemnitz.de/informatik/HomePages/TI/publikationen.php.)
5. Coja-Oghlan, A., Goerdt, A., Lanka, A., and Schädlich, F.: Techniques from combinatorial approximation algorithms yield efficient algorithms for random 2k-SAT. Theoretical Computer Science. To appear. (http://www.tu-chemnitz.de/informatik/HomePages/TI/publikationen.php)
6. Feige, U.: Relations between average case complexity and approximation complexity. Proc. 24th STOC (2002) 534–543
7. Feige, U., Ofek, E.,: Easily refutable subformulas of large random 3CNF formulas. (http://www.wisdom.weizmann.ac.il/~erano/)
8. Flaxman A.: A spectral techn ique for random satisfiable 3CNF formulas. Proc. SoDA 2002, SIAM.
9. Garey, M.R., Johnson, D.S.: Computers and Intractability. 1979
10. Goerdt, A., Lanka, A.: Recognizing more random unsatisfiable 3-SAT instances efficiently. Proc. Typical Case Complexity and Phase Transitions, Satellite Workshop of Logic in Computer Science 2003 (Ottawa). To appear
11. Goerdt, A., Jurdzinski, T.: Some results on random unsatisfiable k-SAT instances and approximation algorithms applied to random structures. Combinatorics, Probability and Computing **12** (2003) 245 – 267
12. J. Håstad, Clique is Hard to Approximate within $n^{1-\varepsilon}$, Acta Mathematica vol. 182, 1999, 105–142
13. Janson, S., Łuczak, T., Ruciński, A.: Random graphs. John Wiley and Sons 2000
14. Lubotsky, A., Phillips, R., Sarnak, P.: Ramanujan Graphs. Combinatorica **8**(3), 1988 261–277
15. Peeters R.: The maximum edge biclique problem ist $\mathcal{NP}$-complete. Research Memorandum 789, Faculty of Economics and Business Administration, Tilberg University, 2000 (http://econpapers.hhs.se/paper/dgrkubrem/2000789.htm)
16. Strang, G.: Linear Algebra and its Applications. Harcourt Brace Jovanovich 1988