Electronic Colloquium on Computational Complexity, Report No. 48 (2004)

An approximation hardness result for bipartite Clique

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 on average, Feige showed some approximation hardness results for several problems like min bisection, dense k-subgraph, max bipartite clique and the 2-catalog segmentation problem. We show a similar result for max bipartite clique, but under the assumption, 4-SAT formulas are hard to refute on average. As falsity of the 4-SAT assumption implies falsity of the 3-SAT assumption it seems that our assumption is weaker than that of Feige.

1 Introduction and Results

Given a standard set of n propositional variables $V = V_n$ 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 $x \in V_n$. In the first case we call l_i a non-negated or positive literal in the second case a negated or negative one. We denote the variable underlying the literal l by $V(l)$, thus $V(x) = V(\neg x) = x$. Altogether we have $2^k n^k$ different k-clauses. A k-SAT formula F simply is a set of k-clauses. We write $C_1 \wedge \cdots \wedge C_m$ for the k-SAT formula with clauses $C_1, \ldots, C_m$. Given a truth value assignment a with 0 (standing for false) and 1 (standing for true) of V_n a k-SAT formula F is true under a iff for each clause C of F there is an $x \in V$ such that $x \in C$ and $a(x) = 1$ or $\neg x \in C$ and $a(x) = 0$. In this case we say that a satisfies C or that a makes C true. The set of variables set to true by a is denoted by T_a, the set of variables set to false by F_a.

Given $p = p(n)$ with $0 \leq p \leq 1$ the random formula $\mathrm{Form}_{n,k,p}$ is obtained as follows: Pick each of the $2^k n^k$ k-clauses independently with probability p. An event A of $\mathrm{Form}_{n,k,p}$ (actually a family of events A_n) holds with high probability for $\mathrm{Form}_{n,k,p}$ if and only if $\Pr[\mathrm{Form}_{n,k,p} \in A_n] \to 1$ when n gets large. Its probability is negligible if $\Pr[\mathrm{Form}_{n,k,p} \in A_n] \to 0$. We use the same terminology for random structures different from $\mathrm{Form}_{n,k,p}$, too.

Considering a high probability event, the following certification problem naturally arises: Given a random instance, how can we be sure that this event really holds for the instance. Depending on the event considered 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 be correct, in that the instance really belongs to 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 the case of correctness and completeness we use the term "efficient certification algorithm". $\mathrm{Form}_{n,k,c/n^{k-1}}$ is unsatisfiable with high probability when $c > \ln 2$ is a constant. Conditioning on the high probability event that in this

ISSN 1433-8092

case the number of clauses is approximately $c \cdot 2^k n$ this follows from the fact that when picking $c \cdot 2^k n$ times a random clause from the set of all $2^k \cdot n^k$ k-clauses the expectation of the number of satisfying assignments is $2^n \cdot (1 - 1/2^k)^{c2^k n} \leq 2^n \cdot \mathrm{e}^{-cn}$. This bound approaches 0 when $c > \ln 2$. Concerning the related certification problem Feige [Fe 02] introduces something like a random 3-SAT hardness hypothesis:

> For any constant $c > \ln 2$ there is no efficient certification algorithm of the unsatisfiability of $\mathrm{Form}_{n,3,c/n^2}$.

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 $\mathrm{Form}_{n,3,p}$ has been made. The best known result is efficient certification of unsatisfiability of $\mathrm{Form}_{n,3,c/n^{3/2}}$ for some sufficiently large constant c, cf. [FeOf 03 b]. Feige shows that this hypotheses implies several lower bounds on the approximability of combinatorial problems for which such bounds could not be obtained from worst-case assumptions like $\mathcal{P} \neq \mathcal{NP}$, cf. [Hå99].

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 $\mathrm{Form}_{n,4,c/n^3}$.

The trivial reduction: Given $F = \mathrm{Form}_{n,3,c/n^2}$ 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. This is so because the result of the reduction $G = \mathrm{Form}_{n,4,c/n^3}$ is a truly random formula, and if G is unsatisfiable then F is unsatisfiable.

Moreover, note that $\mathrm{Form}_{n,3,c/n^2}$ can be efficiently certified unsatisfiable in the not-all-equal sense (that is at least one true and one false literal per clause) [GoJu 03]. Such a result is not known for $\mathrm{Form}_{n,4,c/n^3}$, indicating that random 4-SAT is substantially harder to deal with than random 3-SAT. In case of 3 literals per clause, we consider the graph whose vertices are the literals and whose edges are obtained by making a triangle from each clause. Satisfiability in the not-all-equal sense implies the existence of a cut of the graph with at least 2/3 of all edges. The existence of such large cuts can be excluded efficiently with high probability for $\mathrm{Form}_{n,3,c/n^2}$. Similar constructions do not seem to work for 4-SAT.

Among the problems considered by Feige is the 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 by (W_1, W_2) because the set F of edges is well defined. 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

$$\#\text{ of vertices of } H = |W_1| + |W_2|$$

the problem is solvable in polynomial time [GaJo 79], problem GT24. If however the size of H is measured as

$$\# \text{ of edges of } H = |W_1 \times W_2| = |W_1| \cdot |W_2|$$

the problem interestingly becomes $\mathcal{NP}$-hard [Pe 00] 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 possible 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. [Hå99]. Similar results are not known for our bipartite case. Among other results Feige shows in [Fe 02] 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 final main result is

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

We postpone the proof of Theorem 1 from the subsequent Theorem 2 by means of the derandomized graph product to section 4.

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 bipartite cliques are of size $\leq (n/16)^2(1 + \varepsilon_2)$.*

2 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 3. *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 binomial 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 $\mathrm{Bin}(N,p)$ (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 $\mathrm{Bin}(|X|\cdot|Y|, c/n)$. 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 4. *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 [CoGoLa 04], 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 $|Y| = \beta n$

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

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.

3 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 5.
Input: A 4-SAT formula F.

1. Set $S := P(F)$ and $i := 0$.
2. While $(S \neq \emptyset)$ do
3. Take some clause $C = c_1 \vee c_2 \vee c_3 \vee c_4$ from S.
4. Delete all clauses containing one of the c_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 $\text{Form}_{n,4,c/n^3}$ 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 $B_{ij} = (V_1, fV_2, E_{ij})$, $1 \leq i < j \leq 4$, of S. The sets V_1 and V_2 are copies of the variables V of F. So we set $V_1 = V \times \{1\}$ and $V_2 = V \times \{2\}$. 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 V.

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 $V(l_i) = x$ and $V(l_j) = y$.

Algorithm 6.
Input: A 4-SAT formula F and $\varepsilon > 0$.

1. Apply Algorithm 5 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 $\geq |N| \cdot (1 - o(1))$ 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 7. *Algorithm 6 is complete for* $\text{Form}_{n,4,c/n^3}$ *whenever c is a sufficiently large constant.*

Proof. Step 1 is complete as Algorithm 5 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 $|E_{ij}| \leq |P|$. 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. □

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

Lemma 8. *Let a be a satisfying assignment for F. Then low discrepancy with respect to ε of the projections gives that* $1/3 - O(\varepsilon) \leq \alpha \leq 2/3 + O(\varepsilon)$.

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 6.

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

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

The edges in $E_{1,1}(F_a, F_a)$ are induced by clauses in P of type $(F_a, F_a, V, V)_P$, i.e. clauses beginning with two variables from F_a and the third and fourth variable doesn't matters.

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

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

or

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

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

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

we get $\alpha \leq 2/3 + O(\varepsilon)$. 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 5.

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

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 8.

Theorem 9. *Given $\varepsilon > 0$ a arbitrarily small constant Algorithm 6 certifies that for any assignment a with $|F_a| = \alpha n$ satisfying $Form_{n,4,c/n^3}$ the following equations, hold:*

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

(b) The equations for the ν_i are analogous with $1 - \alpha$ instead of α:
$$\begin{aligned} \nu_0 &= 0 \\ \nu_2 &= 6(1-\alpha)^2\nu - 3\nu_1 + O(\varepsilon)\nu \\ \nu_3 &= (-12(1-\alpha)^2 + 4(1-\alpha))\nu + 3\nu_1 + O(\varepsilon)\nu \\ \nu_4 &= (6(1-\alpha)^2 - 4(1-\alpha) + 1)\nu - \nu_1 + O(\varepsilon)\nu \end{aligned}$$

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 9 is only useful if α is substantial larger than ε. This again shows the relevance of Algorithm 5. 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 6 correctly certifies the properties of Theorem 9 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 clauses inducing no 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 $|T_a| = (1-\alpha)n$. □

We can obtain Lemma 8 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 [Fe 02] from 3-SAT to 4-SAT is the purpose of

Definition 10. *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 $V(u_i) \neq V(v_i)$.*

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 $V(x_1) = V(\neg x_1) = x_1$.

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

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

Corollary 11. *There exists a small constant $\delta > 0$ (e.g. $\delta = 1/50$) such that Algorithm 6 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:*

$$\begin{aligned}
BG(P, ROT_i(N)) &\quad \text{with } 1 \leq i \leq 4 \\
BG(P, ROT_i(P)) &\quad \text{with } i = 1, 2 \\
BG(N, ROT_i(N)) &\quad \text{with } i = 1, 2
\end{aligned}$$

Proof. We only need to show that Algorithm 6 correctly certifies the property claimed. To this end let F be a 4-SAT formula which passed Algorithm 6 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 having 2 true and 2 false variables 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 $BG(N, \mathrm{ROT}_i(N))$ in the same way.

Now we come to the case $\varrho_2 \le 3/8 \cdot \varrho(1+\delta)$ and $\nu_2 \le 3/8 \cdot \nu(1+\delta)$ From the second equalities of (a) and (b) in Theorem 9 we get

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

and

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

As ϱ_1 consists of four subsets of clauses having exactly one variable true under a we have one subset with cardinality $\ge (\alpha^2/2 - (1+\delta)/32 + O(\varepsilon))\varrho$. 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 $1/5 \le \alpha \le 4/5$. Lemma 8 gives us $1/3 - O(\varepsilon) \le \alpha \le 2/3 + O(\varepsilon)$ 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

$$\begin{aligned}
&\left(\frac{(1/3 - O(\varepsilon))^2}{2} - \frac{1+\delta}{32} + O(\varepsilon)\right)\varrho \cdot \left(\frac{(2/3 + O(\varepsilon))^2}{2} - \frac{1+\delta}{32} + O(\varepsilon)\right)\nu \\
&\ge \left(\frac{1}{18} - \frac{1+\delta}{32} + O(\varepsilon)\right)\varrho \cdot \left(\frac{2}{9} - \frac{1+\delta}{32} + O(\varepsilon)\right)\nu \\
&\ge \frac{\varrho \cdot \nu}{250} \ge \frac{(cn \cdot (1+o(1)))^2}{250} = \frac{(cn)^2}{256} \cdot \frac{256}{250} \cdot (1+o(1)) \\
&> \left(\frac{cn}{16}\right)^2 \cdot (1+\delta)
\end{aligned}$$

□

Theorem 12. *Let $\varepsilon > 0$ 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 $1 \le i \le 4$ ($R = T$ 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 $1 \le i \le 4$ let

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

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 $1 \le i \le 4$. The theorem follows when we show that for all sets $K_i \subseteq V$, $L_i = V \setminus K_i$

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

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 binomial distributed with parameters κ and c/n^3, and $\kappa = |K_1| \cdot |K_2| \cdot |K_3| \cdot |K_4|$. Y is also binomial 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

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

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

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

leading to

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

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 11 and Theorem 12 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 11) and the typical formulas whose graphs only have cliques of size e.g. $(cn/16)^2 \cdot (1.001)$ from Theorem 12. This means we could refute 4-SAT on average.

4 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

$$\begin{aligned}|\mathcal{V}_1(L)| \cdot |\mathcal{V}_2(R)| &\geq \left(\frac{|L|}{n}\right)^k \cdot |\mathcal{V}_1| \cdot \left(\frac{|R|}{n}\right)^k \cdot |\mathcal{V}_2| \cdot (1 + o(1)) \\ &\geq \left(\frac{1+\varepsilon_1}{256} + o(1)\right)^k \cdot |\mathcal{V}_1| \cdot |\mathcal{V}_2|.\end{aligned}$$

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

$$\begin{aligned}|\mathcal{L}| \cdot |\mathcal{R}| &\leq |\mathcal{V}_1(L)| \cdot |\mathcal{V}_2(R)| \\ &\leq \left(\frac{|L| \cdot |R|}{n^2}\right)^k \cdot |\mathcal{V}_1| \cdot |\mathcal{V}_2| \cdot (1 + o(1)) \\ &\leq \left(\frac{1+\varepsilon_2}{256} + o(1)\right)^k \cdot |\mathcal{V}_1| \cdot |\mathcal{V}_2|.\end{aligned}$$

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 [AlFeWiZu 95]. This makes use of *Ramanujan graphs*, cf. [LuPhSa 88]. 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 13. *From [AlFeWiZu 95] 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

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

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 14. Input a 4-SAT formula F.
Step 1. Apply Algorithm 6 to F.
Step 2. Construct $\mathcal{G} = (\mathcal{V}_1, \mathcal{V}_2, \mathcal{E})$ as described for every BG from Corollary 11.
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 11. Its completeness from Theorem 12 and the completeness of Algorithm 6.

Acknowledgment. We are grateful to Amin Coja-Oghlan for the algorithm `BipDisc`.

References

[AlFeWiZu 95] Alon, N., Feige, U., Widgerson, A., Zuckerman, D.: Derandomized Graph Products. Computational Complexity 5 (1995), 60–75.

[AlKa 94] Alon, N., Kahale, N.: A spectral technique for colouring random 3-colourable graphs. Proc. STOC (1994) 346–355.

[CoGoLaSch 03] Coja-Oghlan, A., Goerdt, A., Lanka, A., and Schädlich, F.: Certifying unsatisfiability of random $2k$-Sat instances using approximation techniques. Proc. 14th FCT (2003) 15-26

[CoGoLa 04] Coja-Oghlan, A., Goerdt, A., and Lanka, A.: Strong Refutation Heuristics for Random k-SAT. (http://www.tu-chemnitz.de/informatik/HomePages/TI/publikationen.php)

[Fe 02] Feige, U.: Relations between average case complexity and approximation complexity. Proc. 24th STOC (2002) 534–543

[FeOf 03 a] Feige, U., Ofek, E.: Spectral techniques applied to sparse random graphs. Report MCS03-01, Weizmann Institute of Science (2003) (http://www.wisdom.weizmann.ac.il/math/research.shtml)

[FeOf 03 b] Feige, U., Ofek, E.,: Easily refutable subformulas of large random 3CNF formulas. (http://www.wisdom.weizmann.ac.il/~erano/)

[FrGo 01] Friedman, J., Goerdt, A.: Recognizing more unsatisfiable random 3-Sat instances efficiently. Proc. ICALP 2001 310–321.

[FüKo 81] Füredi, Z., Komloś, J.: The eigenvalues of random symmetric matrices. Combinatorica **1**, 1981 233–241

[GaJo 79] Garey, M.R., Johnson, D.S.: Computers and Intractability. 1979

[GoJu 03] 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

[GoLa 03] 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).

[Hå99] J. Håstad, Clique is Hard to Approximate within $n^{1-\varepsilon}$, Acta Mathematica vol. 182, 1999, 105–142

[KrSu 03] Krivelevich, M., Sudakov, B.: The largest eigenvalue of sparse random graphs. Combinatorics, Probability and Computing 12 (2003) 61–72

[LuPhSa 88] Lubotsky, A., Phillips, R., Sarnak, P.: Ramanujan Graphs. Combinatorica **8**(3), 1988 261–277

[Pe 00] 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)

ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'