

ALMOST EVERY 2-SAT FUNCTION IS UNATE

BY

PETER ALLEN

*Department of Mathematics, London School of Economics
Houghton St., London WC2A 2AE, U.K.
email: p.d.allen@lse.ac.uk*

ABSTRACT

Bollobás, Brightwell and Leader [2] showed that there are at most $2^{\binom{n}{2}+o(n^2)}$ 2-SAT functions on n variables, and conjectured that in fact the number of 2-SAT functions on n variables is $2^{\binom{n}{2}+n(1+o(1))}$. We prove their conjecture.

As a corollary of this, we also find the expected number of satisfying assignments of a random 2-SAT function on n variables. We also find the next largest class of 2-SAT functions and show that if $k = k(n)$ is any function with $k(n) < n^{1/4}$ for all sufficiently large n , then the class of 2-SAT functions on n variables which cannot be made unate by removing $25k$ variables is smaller than $2^{\binom{n}{2}+n-kn}$ for all sufficiently large n .

1. Introduction

Given a collection of n Boolean variables x_1, \dots, x_n , a **satisfying assignment** for a Boolean function S on the n variables is an assignment of True/False to each variable such that $S(x_1, \dots, x_n)$ is True. A Boolean function is defined by its set of satisfying assignments.

Associated with a Boolean variable x is a **positive literal** x and a **negative literal** \bar{x} . The positive literal is True exactly when the variable is True; the negative literal is True exactly when the variable is False.

A **k -clause** is a collection of k literals, no two of which are associated with the same variable. A k -clause is satisfied if and only if at least one of its literals is True. A **k -SAT formula** is a collection of k -clauses, and the formula is satisfied if and only if all of its k -clauses are satisfied. A Boolean function S_F

whose satisfying assignments are exactly those which satisfy a k -SAT formula F is called a **k -SAT function**.

The aim of this paper is to asymptotically enumerate the 2-SAT functions on n variables. This is not equivalent to enumerating the 2-SAT formulae: some 2-SAT functions are given by several different 2-SAT formulae, while others have a unique formula.

We call a 2-SAT formula **monotone** if all of its clauses contain only positive literals, and a 2-SAT function is called monotone if it is given by a monotone formula.

Given a 2-SAT formula F on n variables and a subset R of those variables, we can define a 2-SAT formula by replacing each positive literal associated to a variable in R with the corresponding negative literal, and each negative literal associated to a variable in R with the corresponding positive literal, in every clause of F . We call this process **relabelling the literals**, and say that R is the set of variables that were relabelled. If S is a 2-SAT function, then the same relabelling of the variables R on any formula F which gives rise to S gives rise to the same 2-SAT function, so we can say this is the 2-SAT function obtained by relabelling the variables R . Its satisfying assignments are precisely the satisfying assignments of S with the values of the variables in R reversed.

We say that any 2-SAT function which is the result of relabelling some variables on a monotone function is **unate**.

Let $G(n)$ be the number of 2-SAT functions on n variables. We will see that there are $2^{\binom{n}{2}}$ monotone 2-SAT functions, and asymptotically $2^{\binom{n}{2}+n}$ unate 2-SAT functions, so that certainly $G(n)$ grows at at least this rate.

Bollobás, Brightwell and Leader [2] proved that $G(n) = 2^{\binom{n}{2}+o(n^2)}$, and conjectured that in fact almost every 2-SAT function is unate; in other words, that $G(n) = 2^{\binom{n}{2}+n}(1 + o(1))$. We modify and expand on their argument to obtain a proof of their conjecture.

THEOREM 1.1: *For all sufficiently large n ,*

$$(1 - 2^{-(3n)/4})2^{\binom{n}{2}+n} < G(n) < 2^{\binom{n}{2}+n}(1 + 2^{-n/25}),$$

and almost every 2-SAT function is unate.

The proof of this theorem constitutes the bulk of this paper (Sections 1–7). In Section 8 we prove some refinements as follows.

We will find as a consequence of the proof of Theorem 1.1 the following (logarithms are taken to base 2, as will be the case throughout).

THEOREM 1.2: *The expected number of satisfying assignments of a random 2-SAT function on n variables is*

$$(1 + o(1)) \sum_{k=0}^{5\sqrt{n}} \binom{n}{k} 2^{-\binom{k}{2}} = 2^{(1/2+o(1)) \log^2 n}.$$

We note that this is asymptotically equal to the expected number of independent sets in a random graph on n points.

We are able to prove even more, identifying the next largest class of 2-SAT functions.

THEOREM 1.3: *For all sufficiently large n ,*

$$2^{\binom{n}{2}+n} \left(1 + \binom{n}{2} \left(\frac{3}{4} \right)^{n-2} - 2^{-\frac{n}{2}} \right) < G(n) < 2^{\binom{n}{2}+n} \left(1 + \binom{n}{2} \left(\frac{3}{4} \right)^{n-2} + 2^{-\frac{418n}{1000}} \right).$$

It is well-known that almost every triangle-free graph is bipartite (see Erdős, Kleitman and Rothschild [5]). Prömel, Schickinger and Steger [8] have shown that almost every triangle-free graph which is not bipartite can be made bipartite by removing just one vertex; this theorem gives the equivalent result for our problem. They showed further that almost every triangle-free graph which is not bipartite and cannot be made so by removing one vertex can be made bipartite by removing two vertices, and so on. We will establish an upper bound of $2^{\binom{n}{2}+n-kn}$ on the number of 2-SAT functions which cannot be made unate by removing $25k$ variables, where $k = k(n) < n^{1/4}$, which is a first step towards proving the equivalent result for our problem.

Suppose S_1, S_2 arise from the monotone 2-SAT formulae F_1, F_2 . Suppose that there is a clause (a_1, a_2) in F_1 which is not in F_2 . Then the assignment of True to every variable except for those associated to literals a_1, a_2 , which are assigned False, is not a satisfying assignment for S_1 , but it is for S_2 . Hence, two different monotone 2-SAT formulae always give rise to different monotone 2-SAT functions.

Observe that if F is a 2-SAT formula but not monotone, then there is a clause in it with strictly less than two positive literals. Then we can assign True to every variable not associated to a literal in that clause, and to the variables associated to literals in that clause we assign whichever of True or False makes the clause not satisfied. This is an assignment in which strictly more than $n - 2$ variables are assigned True. Hence, any clause with no negative literals must be satisfied, so it is a satisfying assignment for all monotone 2-SAT functions. It follows that a monotone 2-SAT function arises from exactly one 2-SAT formula, so there are exactly $2^{\binom{n}{2}}$ monotone 2-SAT functions on n variables.

If S is a unate function and F is a formula for S , then relabelling the same variables on F that were relabelled to obtain S from a monotone function must result in a monotone formula, so that S has a unique formula and there are as many unate 2-SAT functions as there are distinct relabellings of monotone 2-SAT formulae. Since two relabellings of a monotone formula F are distinct unless the relabellings differ only on variables whose positive literals do not appear in F , certainly all relabellings of monotone formulae which mention all positive literals are distinct, and a lower bound for the number of unate 2-SAT functions is:

$$2^n \left(2^{\binom{n}{2}} - n 2^{\binom{n-1}{2}} \right) = 2^{\binom{n}{2}+n} \left(1 - 2^{-\binom{n-1}{2}+\log n} \right) = 2^{\binom{n}{2}+n} \left(1 - 2^{-n+\log n+1} \right).$$

We certainly have:

$$G(n) \geq 2^{\binom{n}{2}+n} (1 - 2^{-3n/4}),$$

for sufficiently large n , and we aim to prove an upper bound to match this.

We remark that Bollobás, Brightwell and Leader [2] observe that $G(n)$ can also be interpreted as the number of subsets of the n -cube that are the union of subcubes of codimension 2. However, we will not make use of this viewpoint.

2. Preliminaries

In this section, we reduce the problem to that of counting a class of 2-SAT functions with extra properties. This entire section is essentially identical to the equivalent part of [2], repeated here for convenience.

The 2-SAT function with no satisfying assignment is called the **trivial function**.

Given a 2-SAT function, define its **spine** to be the set of literals which are true in all satisfying assignments; obviously any non-trivial function cannot have both x and \bar{x} in its spine. We will refer to the variable x being in the spine of the function.

Suppose that for some pair of literals u, v , in every satisfying assignment $u \iff v$. Then we say that the literals are **associated**; clearly u, v are associated if and only if \bar{u}, \bar{v} are associated. Then we can say that the corresponding variables are associated, and trivially this is an equivalence relation.

We call a 2-SAT function **elementary** if it has no variables in its spine and no associated pairs of variables. Let there be $H(n)$ elementary 2-SAT functions on n variables.

Given any non-trivial 2-SAT function S on n variables, we can reduce it to an elementary 2-SAT function by ignoring all variables in the spine of S and

all but the lowest numbered in each equivalence class of associated variables, then compressing the labels to obtain an elementary function on $n - l$ variables ($l \geq 0$). This reduction is at worst $\binom{n}{l} (2n - 2l + 2)^l$ -to-1, since for each of the l variables removed ($\binom{n}{l}$ choices of label) we can choose to associate its positive literal to any of the $2n - 2l$ remaining literals, or to put either its positive or negative literal in the spine. Thus

$$H(n) \leq G(n) \leq 1 + \sum_{l=0}^n H(n-l) \binom{n}{l} (2n - 2l + 2)^l,$$

where the 1 is counting the trivial function.

Since every unate function is certainly elementary, we have $2^{\binom{n}{2}+n}(1-2^{-\frac{3n}{4}}) < H(n)$. $H(n)$ is obviously monotone increasing, and we will prove that $H(n) < 2^{\binom{n}{2}+n}(1+2^{-\frac{n}{24}})$ for all sufficiently large n . This will be enough to prove Theorem 1.1.

Proof: Let N be sufficiently large that for all $n \geq N$, $H(n) < 2^{\binom{n}{2}+n}(1+2^{-\frac{n}{24}})$.

Then we have

$$\begin{aligned} G(n) &< 1 + \sum_{l=0}^{n-N} H(n-l) \binom{n}{l} (2n - 2l + 2)^l + \sum_{l=n-N+1}^n H(N) \binom{n}{l} (2n - 2l + 2)^l \\ &< \sum_{l=0}^{n-N} 2^{\binom{n-l}{2}+n-l} (1 + 2^{-\frac{n-l}{24}}) \binom{n}{l} (2n - 2l + 2)^l + NH(N) 2^n (2n)^n \\ &< 2^{\binom{n}{2}+n} (1 + 2^{-\frac{n}{25}}) \end{aligned}$$

for sufficiently large n . ■

We define a bijection between the elementary 2-SAT functions on n variables and a particular class of partial orders on $2n$ points: given any formula F giving rise to an elementary 2-SAT function S_F , let $P_1(F)$ be the relation on the $2n$ points $\{x_1, \dots, x_n, \bar{x}_1, \dots, \bar{x}_n\}$ given by $a < b$ if the clause (\bar{a}, b) appears in F .

Suppose that there were a sequence $a_1 < a_2 < \dots < a_r < a_1$ in $P_1(F)$. Suppose we have a satisfying assignment for S_F with a_1 True. Then as (\bar{a}_1, a_2) must contain a True literal, a_2 is also True in any such assignment. Suppose we have a satisfying assignment for S_F with a_1 False. Then as (\bar{a}_r, a_1) must contain a True literal, a_r must also be False, and by induction a_i must be False for each $1 \leq i \leq r$ in any such assignment. But then a_1 and a_2 are associated, contradicting S_F being elementary. So no such sequence exists. Then let $P(F)$ be the transitive closure of $P_1(F)$; we see that this is a partial order.

The relation $P(F)$ must satisfy $u < v \iff \bar{v} < \bar{u}$, since a sequence of clauses giving the first relation also gives the second. It cannot have $\bar{u} < u$ for any u , since if u is False in a satisfying assignment for S_F , the sequence of clauses certifying $\bar{u} < u$ certify u True, which is a contradiction, so that u is in the spine of S_F .

A satisfying assignment for S_F is an up-set in $P(F)$ containing exactly one of each pair of literals. Furthermore, suppose $u \not< v$ in $P(F)$, then let U be the smallest up-set containing both u and \bar{v} ; there is no x with $u < x$ and $\bar{v} < \bar{x}$, as this implies $x < v$ so $u < v$, so that U contains at most one of each pair of literals. There is no y with $y < v$, $\bar{y} < v$, for this implies $\bar{v} < y < v$. So we can add to U one literal at a time to obtain an up-set containing exactly one of each pair of literals, which is a satisfying assignment for S_F with u and \bar{v} true. Thus $u < v$ in $P(F)$ if and only if $u \Rightarrow v$ is True in every satisfying assignment for S_F . So $P(F)$ depends only on the function S_F and not on the specific formula F giving S_F , and the satisfying assignments of S_F can be found given $P(F)$. Thus there is a 1-1 correspondence between elementary 2-SAT functions on n variables $\{x_1, \dots, x_n\}$ and partial orders on $2n$ points $\{x_1, \dots, x_n, \bar{x}_1, \dots, \bar{x}_n\}$ such that $u < v \iff \bar{v} < \bar{u}$ and there is no u with $\bar{u} < u$. We write $P(S)$ for the partial order corresponding to the elementary function S .

3. Reduction to diagrams

At this point we depart from the method of proof in [2]; we develop a line-and-arrow representation, which turns out to be more amenable to detailed analysis than the coloured graph representation studied in [2].

The trivial function is not elementary, so that every elementary function must have a satisfying assignment. For each elementary 2-SAT function S , pick an elementary 2-SAT function $M(S)$ obtained by relabelling the literals such that x_1, \dots, x_n all True is a satisfying assignment of $M(S)$. Since there are only 2^n ways to relabel the literals, the restriction of M to the elementary 2-SAT functions on n variables is in the worst case 2^n -to-1. Observe that whatever elementary S was chosen, $P(M(S))$ cannot have $x_i < \bar{x}_j$ for any i, j , as this is equivalent to $x_i \Rightarrow \bar{x}_j$ in every satisfying assignment of $M(S)$, contradicting all True being a satisfying assignment. Furthermore, $M(S)$ is unate if and only if S is unate.

Call a 2-SAT function **non-negative** if it is elementary and all True is a satisfying assignment. Then M maps from the class of elementary 2-SAT functions to the non-negative 2-SAT functions.

Given a non-negative 2-SAT function S on n variables $\{x_1, \dots, x_n\}$, we construct a **diagram** $D(S)$, which will be a graph on n points $\{x_1, \dots, x_n\}$ in which some edges are directed (which we call **arrows**) and some are not (which we call **lines**). The directed edges will form a partial order, the **partial order within** D , and we will use equivalently ' $a < b$ ' and 'there is an arrow from a to b '. We do this as follows. First take $P(S)$, the partial order on $2n$ points associated with S . Then whenever $x_i < x_j$ in $P(S)$, we put $x_i < x_j$ in $D(S)$. Whenever $\bar{x}_i < x_j$ is a covering relation in $P(S)$ we put a line $x_i x_j$ in $D(S)$. Observe that if $\bar{x}_i < x_j$ so also $\bar{x}_j < x_i$, and that given $D(S)$ we can certainly recover $P(S)$ and hence S .

Observe that, in $D(S)$, no pair of points a and b are joined by both a line and an arrow: ab and $a < b$. For this would imply that in $P(S)$ we had $\bar{b} < a < b$. We also cannot find each of the following **forbidden structures**, shown in Figure 1:

- (1) a, b, c such that $b < a, c < a$ and there is a line bc .
- (2) a, b, c such that $a < b$ and there are lines ac, bc .
- (3) a, b, c, d such that $a < b, c < d$ and there are lines ac, bd .

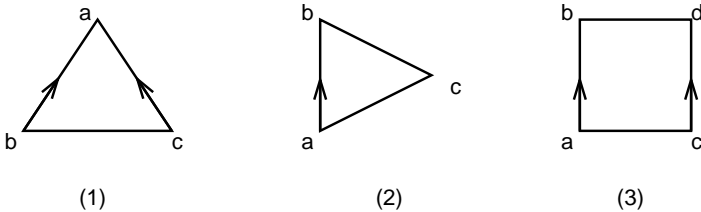


Figure 1. Forbidden structures

For if (1) existed, then in $P(S)$, $\bar{a} < \bar{b} < c < a$. If (2) existed, then in $P(S)$, $\bar{c} < a < b$ so that $\bar{c} < b$ cannot be a covering relation. If (3) existed, then in $P(S)$, $\bar{b} < \bar{a} < c < d$ and $\bar{b} < d$ cannot be a covering relation.

A **valid diagram** is any diagram in which no two points are joined by both an arrow and a line, and which does not contain any of the three forbidden structures. Then certainly the set of valid diagrams on n points contains the set of all diagrams $D(S)$ for S a non-negative 2-SAT function on n variables, and in fact it is not hard to see that every valid diagram D is $D(S)$ for some non-negative 2-SAT function S .

We will bound both the number $F(n)$ of valid diagrams on n points and the number of valid diagrams on n points not corresponding to unate functions. This latter bound will be small enough to give our result.

At this point it is worth observing that a monotone 2-SAT function is always

a non-negative function; it will correspond to a diagram in which there are no arrows, only lines between points. A unate 2-SAT function, by contrast, is not usually non-negative: we get a unate function by relabelling the literals on a monotone function, and the result is only non-negative if there were no clauses consisting of any two of the relabelled literals. So we expect to find that M is in some sense ‘nearly’ 2^n -to-1 on the set of unate 2-SAT functions on n variables, and we expect to find that most unate non-negative 2-SAT functions correspond to diagrams in which there is a large set of points within which no arrows are found, a small set of points within which no lines or arrows are found and between the small and large set no lines are found and no arrows go from the large set to the small set. In fact, we will find that there are a factor of $2^{(\frac{1}{2}+o(1))\log^2 n}$ ‘too many’ valid diagrams, but this is acceptable since the number of valid diagrams corresponding to non-unate 2-SAT functions is exponentially smaller than the number of those that correspond to unate 2-SAT functions, and we already have an accurate count of the unate 2-SAT functions.

In the rest of this paper, all diagrams will consist of a **top set** T , containing all points which are maximal in the partial order within the diagram, and a **bottom set** B , all other points (see Figure 2).

Note that there are no arrows within T , and that every member of B is below at least one point of T .

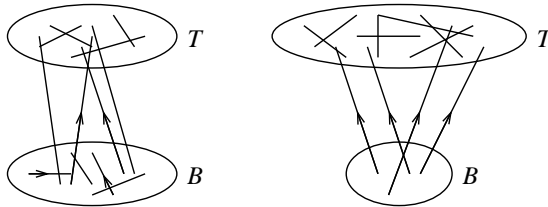


Figure 2. A general valid diagram and a diagram arising from a unate 2-SAT function

We will at times want to say that between a and b in a diagram there is no line or arrow in either direction, then we will simply say that there is **nothing** between a and b ; we will also sometimes want to say that there is either a line or an arrow in one or the other direction between a and b , when we will say there is **something** between a and b .

LEMMA 3.1: *A valid diagram corresponds to a non-negative unate 2-SAT function if and only if there are no lines between its top and bottom sets, and no lines or arrows within its bottom set.*

Proof: If S is a non-negative unate 2-SAT function, then let R be a minimal set of variables such that relabelling the variables R on S gives a monotone function. If a is a point in $D(S)$, and $a \notin R$, then there certainly is no clause (\bar{a}, b) in the 2-SAT formula for S , and so a cannot be below any other point in D , i.e. $a \in T$. If on the other hand $a \in R$, then the clause (\bar{a}, b) must appear in the formula for S (since R is minimal) and $a \in B$. If there is a line tb in D , with $t \in T, b \in B$, then the clause (t, b) is in the formula for S , so that b, \bar{b} are both mentioned in the formula, which is impossible since S is unate. Similarly, if $a < b$ or ab is in D , with $a, b \in B$, then b, \bar{b} are in the formula for S which is impossible. Hence $D(S)$ is of the required form. Now suppose we are given a diagram $D(S)$ of the required form, corresponding to the non-negative function S . Since there is no arrow $a < b$ with a line ac , we see that in $P(S)$ every relation $\bar{x}_i < x_j$ between positive and negative literals is a covering relation, so relabelling the variables B on S gives a unate function. ■

We observe that valid diagrams corresponding to non-negative unate 2-SAT functions are in 1 – 1 correspondence with graphs with a specified independent set B .

Suppose that we attempt to construct a valid diagram. First we choose the set B , then we choose where to put lines within T , then where to put lines between B and T and where to put arrows from B up to T . Then we can choose where to put lines and arrows within B . But this last choice is already restricted. Suppose that we have two points $a, b \in B$. Then either there is, or there is not, a point in T above both a and b . In the former case, we cannot put a line ab in without creating the forbidden structure (1) and hence an invalid diagram. In the latter case, we cannot put an arrow from a to b , for there exists $x \in T$ with $b < x$ and hence if $a < b$ then $a < x$ also; by symmetry, we cannot put an arrow from b to a either. Thus between any two points in b we have either a choice of nothing or a line between a and b , or a choice of nothing or an arrow in one or the other direction.

We can now put a crude upper bound on the number of valid diagrams on n points:

$$\sum_{|B|=0}^{n-1} \binom{n}{|B|} 2^{\binom{|T|}{2}} 3^{|T||B|} 2^{\binom{|B|}{2}} |B|!$$

simply by following the above construction, noting that there are $|B|!$ ways to order the points in B , so that there are at most that many ways to decide the directions of any arrows we might choose in B .

We have now reduced the problem to examining a class of combinatorial structures which are relatively easy to handle; the rest of the paper involves developing and applying tools to handle these structures.

Since we want to find that there are about $2^{\binom{n}{2}}$ valid diagrams, we must find a way of dealing with the possibilities for lines and arrows between B and T , and with the requirement to order B . We will do this by dividing the problem into two cases, when B is small and when it is not. We will need only some simple approximations to deal with the first case, even though it will turn out to be the large case. We will apply the Szemerédi Regularity Lemma followed by an induction argument to dispose of the second case.

4. Some useful facts

The following will be used frequently to restrict choice, relying on the forbidden structure (2) which says that if a point is comparable with one set of points and connected by lines to another set, then there can exist no lines between these two sets.

Let V be a set of points and I an index set. Suppose that there exist for each $i \in I$ sets $S_i, L_i \subset V$, with $S_i \cap L_i = \emptyset$. Call the set

$$\bigcup_{i \in I} \{\{a, b\} : a \in L_i, b \in S_i\}$$

the **forbidden set** (for $(S_i, L_i)_{i \in I}$).

LEMMA 4.1: *Suppose that there is a constant l such that for each i , $|L_i| \geq l$. Let $S = \bigcup_{i \in I} S_i$. Then the forbidden set has size at least $l|S|/2$.*

Proof: Observe that for each $s \in S$, at least l members of the forbidden set contain s . But as a given member of the forbidden set can contain at most two elements of S , this counts each member at most twice, and the forbidden set has size at least $l|S|/2$. ■

We will usually refer to the S_i as the **small sets** and the L_i as the **large sets**.

When we use this, the set I will be some points in the top or bottom of a diagram, the sets L_i and S_i will be points in the other layer connected to i by arrows or lines (in no particular order), and the forbidden structure (2) will dictate that there are no lines between pairs in the forbidden set. This places two restrictions on the available choices: firstly we have to choose the S_i within S , and secondly we will be unable to choose lines between at least $l|S|/2$ pairs.

Some simple bounds will also be useful.

Whenever $k \leq n/3$, we have $\sum_{i=0}^k \binom{n}{i} < 2 \binom{n}{k}$, since

$$\binom{n}{i} = \frac{i+1}{n-i} \cdots \frac{k}{n-k+1} \binom{n}{k} < \left(\frac{1}{2}\right)^{k-i} \binom{n}{k}$$

when $i \leq n/3$.

In [7], Kleitman and Rothschild show that the number of partial orders on n points is asymptotically $2^{n^2/4 + O(n \log n)}$ (this result is sharpened and the proof simplified by Brightwell, Prömel and Steger in [4]). Hence there exists J such that for all $n > J$, there are fewer than $2^{(3n^2 - 5n)/10}$ partial orders on n points. We use J for this number throughout.

5. Counting the case we expect to be large

In this section, we discuss the case $|B| \leq n/100$. Note that we expect to find many valid diagrams in this case, as it covers all the diagrams corresponding to monotone 2-SAT functions and the vast majority of those corresponding to non-negative unate functions.

THEOREM 5.1: *For all sufficiently large n , there are at most $2^{\binom{n}{2} + \log^2 n}$ valid diagrams on n points with $|B| \leq n/100$, and at most $2^{\binom{n}{2} - n/23}$ of these do not correspond to unate functions. Furthermore, there are at most $2^{\binom{n}{2} - 2n^{5/4}}$ valid diagrams on n points with $3n^{19/20} < |B| \leq n/100$.*

Proof: For any valid diagram and any point $x \in B$, let $\Gamma_{arr}(x) = \{t \in T : t > x\}$ be the **arrow neighbours** of x . Let $\Gamma_{line}(x) = \{t \in T : xt \text{ is a line}\}$ be the **line neighbours** of x . Note that there might be points in B connected to x by arrows or lines, but we do not include them in these sets.

Let $P \subset B$ be the set of points $x \in B$ with both $|\Gamma_{arr}(x)| < n/10$ and $|\Gamma_{line}(x)| < n/10$. Let $I = B - P$.

Now for each point $i \in I$, at least one of $|\Gamma_{arr}(i)|$, $|\Gamma_{line}(i)|$ is at least $n/10$. We apply Lemma 4.1, with this I , with $V = T$, with L_i the larger of $\Gamma_{arr}(i)$ and $\Gamma_{line}(i)$ and with S_i the smaller. Then $l \geq n/10$, and we have $S = \bigcup_{i \in I} S_i$.

Let $D(B, P, S)$ be the number of valid diagrams with sets B, P, S (see Figure 3). Then the number of valid diagrams with $|B| \leq n/100$ is

$$(1) \quad \sum_{|B|, |P|, |S|} \binom{n}{|B|} \binom{|B|}{|P|} \binom{|T|}{|S|} D(B, P, S).$$

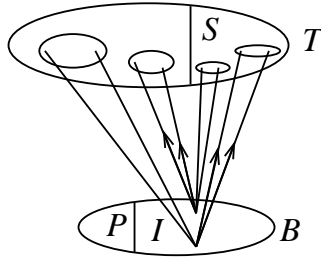


Figure 3. A typical valid diagram with $|B| \leq n/100$

We will attempt to construct valid diagrams, obtaining an upper bound by counting the number of choices at each stage, as follows:

First, we choose the sets B, P, S . Then we choose for each point in P its arrow and line neighbours. Note that $\sum_{j=0}^{n/10} \binom{|T|}{j} < 2 \binom{|T|}{n/10}$, and when n is sufficiently large,

$$\left(2 \binom{|T|}{n/10}\right)^2 < 4(10e)^{n/5} < 2^{|T|} 2^{-n/21}.$$

This approximation will be used to bound the number of choices for lines and arrows connecting to points in P . We choose for each point in I whether its large set is its arrow or line neighbours, then we choose its large and small sets.

We choose where to put lines within T ; from Lemma 4.1 there is a set of size at least $n|S|/20$ in which we cannot choose to put lines.

We choose a total order for B , which will orient any arrows we choose to put in B . For any pair in $B^{(2)}$ we can choose to put either nothing or something, but we have no choice over whether ‘something’ is a line or an arrow.

This gives us the following upper bound on $D(B, P, S)$.

$$\begin{aligned} (2) \quad D(B, P, S) &< \left(2 \binom{|T|}{n/10}\right)^{2|P|} 2^{|I|} 2^{|I||T-S|} 3^{|I||S|} 2^{\binom{|T|}{2} - \frac{n|S|}{20}} |B|! 2^{\binom{|B|}{2}} \\ &< 2^{\binom{n}{2}} 2^{-\frac{|P|n}{21}} 2^{-\frac{n|S|}{20} + |I||S| \log \frac{3}{2}} 2^{|I|} |B|! \end{aligned}$$

Observe that since $|I| \leq |B|$, therefore $|I| \log(3/2) < (7n)/1000 < n/20$. Then $2^{-\frac{n|S|}{20} + |I||S| \log \frac{3}{2}} \leq 1$, so in the sum (1) giving an upper bound on the number of diagrams, the only terms which multiply the $2^{\binom{n}{2}}$ term by an amount greater than 1 are the choices for B , P , S , $2^{|I|}$, and $|B|!$. Together with the sum — over at most n^3 summands — this means that an upper bound for the number of valid diagrams will be $n^3 2^{3n} D(B, P, S)$, taking the worst case for $D(B, P, S)$. We split the proof into cases, of which only the last will be large:

- (i) $|P| \geq 84n^{1/4}$.
 - (ii) $|P| < 84n^{1/4}$ and $|S| \geq 100n^{1/4}$.
 - (iii) $|P| < 84n^{1/4}$, $|S| < 100n^{1/4}$ and $|B| \geq 3n^{19/20}$.
 - (iv) $|B| < 3n^{19/20}$ and at least one of $|S| > 0$, $|P| > 0$ holds.
 - (v) $|S| = |P| = 0$, $|B| < 3n^{19/20}$ and there is a line within B .
 - (vi) $|S| = |P| = 0$, $5\sqrt{n} \leq |B| < 3n^{19/20}$ and there is no line within B .
 - (vii) $|S| = |P| = 0$, $|B| < 5\sqrt{n}$, there is no line within B and there are at least two arrows within B .
 - (viii) $|S| = |P| = 0$, $|B| < 5\sqrt{n}$, there is no line within B and there is exactly one arrow within B .
 - (ix) $|S| = |P| = 0$, $|B| < 5\sqrt{n}$ and there are no lines or arrows within B .
- Following Lemma 3.1, we see that these diagrams correspond to unate 2-SAT functions.

CASE (i): $2^{-(|P|n)/21} \leq 2^{-4n^{5/4}}$, so that using (2) we can see that

$$D(B, P, S) < 2^{\binom{n}{2}} 2^{-\frac{|P|n}{21}} 2^{|I|} |B|! < 2^{\binom{n}{2}-3 \log n - 3n - 3n^{5/4}}$$

for sufficiently large n , so that in this case (1) is bounded from above by $2^{\binom{n}{2}-3n^{5/4}}$.

CASE (ii): $|I| \log(3/2) < (7n)/1000$, so $2^{-\frac{n|S|}{20} + |I||S| \log 3/2} < 2^{-4n^{5/4}}$, so that (2) gives us

$$D(B, P, S) < 2^{\binom{n}{2}-4n^{5/4}} 2^n |B|! < 2^{\binom{n}{2}-3 \log n - 3n - 3n^{5/4}}$$

for sufficiently large n , so that in this case (1) is bounded from above by $2^{\binom{n}{2}-3n^{5/4}}$.

CASE (iii): Since $|P| < 84n^{1/4}$, then $|I| > 2n^{19/20}$ and there must be either $n^{19/20}$ points in I all of whose arrow sets are their large sets, or the same amount of points all of whose arrow sets are their small sets. In the first case, there are at least $\frac{n}{10}n^{19/20}$ arrows going up from I to less than n points in T , so one of these points must be the target of at least $\frac{1}{10}n^{19/20}$ arrows. In the second case, recall that every point in B must be below at least one point in T , so that all of the given $n^{19/20}$ points must be below the points in S , and one point in S must be above at least $\frac{n^{19/20}}{|S|}$ of them. Let C be a maximal set of points in B such that there is one point $t \in T$ with $t > c$ for all $c \in C$.

$$\text{Then } |C| \geq \min\left(\frac{n^{19/20}}{10}, \frac{n^{19/20}}{|S|}\right) \geq \min\left(\frac{n^{19/20}}{10}, \frac{n^{19/20}}{100n^{1/4}}\right) = \frac{n^{14/20}}{100}.$$

Now we see that within $C^{(2)}$, we can find no lines, since structure (1) is forbidden. Therefore, the structure within C is simply a partial order, and we

recall that for $|C| > J$, there are at most $2^{\frac{3|C|^2 - 5|C|}{10}}$ partial orders on $|C|$ points. As a result, in this case, when we choose lines and arrows within B , we find that there are at most

$$2^{\binom{|B|}{2} - \binom{|C|}{2}} |B|! 2^{\frac{3|C|^2 - 5|C|}{10}} = 2^{\binom{|B|}{2} - \frac{|C|^2}{5}} |B|!$$

possible choices.

Observe that $\frac{|C|^2}{5} \geq \frac{n^{7/5}}{50000}$. Consider constructing a diagram in this case: given B, P, S we choose lines and arrows from P to T , from I to $T - S$ and to S , within T and finally within B . We see that we have in this case an upper bound

$$\begin{aligned} D(B, P, S) &< \left(2^{\binom{|T|}{2}}\right)^{2|P|} 2^{|I||T-S|} 3^{|I||S|} 2^{\binom{|T|}{2} - \frac{n|S|}{20}} 2^{\binom{|B|}{2} - \frac{|C|^2}{5}} |B|! \\ &< 2^{\binom{n}{2}} 2^{-\frac{n^{7/5}}{50000}} 2^n n! < 2^{\binom{n}{2} - 3 \log n - 3n - 3n^{5/4}} \end{aligned}$$

for n sufficiently large, so that both $\frac{n^{14/20}}{100} > J$ and the above approximations hold. Then, in this case, (1) is bounded from above by $2^{\binom{n}{2} - 3n^{5/4}}$.

At this point, note that we have the required upper bound $3.2^{\binom{n}{2} - 3n^{5/4}} < 2^{\binom{n}{2} - 2n^{5/4}}$ on the number of valid diagrams with $3n^{19/20} < |B| \leq \frac{n}{100}$.

We also observe that we know that $|B|, |S|, |P|$ and $|I|$ are now all much smaller than n , so that for all sufficiently large n , $n^3 \binom{n}{|B|} \binom{|B|}{|P|} \binom{|T|}{|S|} < 2^{n^{0.99}}$, $2^{|I|}|B|! < 2^{\frac{n}{1000}}$, and $|I| \log \frac{3}{2} < n/1000$. We will use these bounds in the following cases.

CASE (iv): Applying (2) gives us the bound

$$\begin{aligned} D(B, P, S) &< 2^{\binom{n}{2}} 2^{-\frac{|P|n}{21}} 2^{-\frac{|S|n}{20} + |S|\frac{n}{1000}} 2^{|I|} |B|! \\ &< 2^{\binom{n}{2}} 2^{-\frac{|P|n}{21}} 2^{-\frac{|S|n}{21}} 2^{|I|} |B|! < 2^{\binom{n}{2} - \frac{n}{22} - n^{0.99}} \end{aligned}$$

for sufficiently large n , and (1) is bounded from above by $2^{\binom{n}{2} - \frac{n}{22}}$ for all sufficiently large n .

CASE (v): Every point in B is below at least one point in T , by definition. Since $|S| = |P| = 0$, every point in B is in I and has empty small set, hence every point in B has its large set its set of arrow neighbours and has no line neighbours in T . So there are no lines between B and T . This is, of course, also true for the following four cases.

There is a line ab within B ; there are $\binom{|B|}{2}$ ways to choose this line. Then L_a and L_b both have size at least $n/10$, and do not intersect. Hence $L_a \times L_b$ is a

set of size at least $n^2/100$ within T , and any line in it would cause the forbidden structure (3). So we have

$$D(B, P, S) < \binom{|B|}{2} 2^{|B||T|} 2^{\binom{|T|}{2} - \frac{n^2}{100}} 2^{\binom{|B|}{2}} |B|! < 2^{\binom{n}{2} - n^{0.99} - n^{5/4}}$$

for sufficiently large n , so that in this case (1) is bounded from above by $2^{\binom{n}{2} - n^{5/4}}$.

CASE (vi): Since B contains no lines, the diagram structure within B is simply a partial order. Since $|B| \geq 5\sqrt{n}$, whenever $n > J^2$ we can say that there are at most $2^{\frac{3|B|^2 - 5|B|}{10}} = 2^{\binom{|B|}{2} - \frac{|B|^2}{5}}$ ways to choose arrows within B . Then, given B , P and S we can construct diagrams in this case by choosing the arrows from $I = B$ to T , the partial order within B and the lines within T . This gives us the bound

$$D(B, P, S) < 2^{|B||T|} 2^{\binom{|B|}{2} - \frac{|B|^2}{5}} 2^{\binom{|T|}{2}} < 2^{\binom{n}{2} - 5n} < 2^{\binom{n}{2} - n - n^{0.99}}$$

for sufficiently large n , so that in this case (1) is bounded from above by $2^{\binom{n}{2} - n}$.

CASE (vii): There are four possibilities for the two arrows that are guaranteed to exist in B : either $a < b < c$, or $a < b, a < c$, or $a > b, a > c$, or $a < b, c < d$. Observe that if $a < b$, then $\Gamma_{arr}(b) \subset \Gamma_{arr}(a)$, so that in each of these four cases we have restrictions on the choices of arrows going from these points upwards; $4^{|T|}$, $5^{|T|}$, $5^{|T|}$ and $9^{|T|}$ choices, respectively. This means that the choice for arrows between B and T is $2^{|B||T| - |T|}$, $2^{|B||T| - |T| \log(8/5)}$, $2^{|B||T| - |T| \log(8/5)}$, $2^{|B||T| - |T| \log(16/9)}$ respectively; so we get

$$D(B, P, S) < 4.2^{\binom{n}{2} - |T| \log \frac{8}{5}} |B|! < 2^{\binom{n}{2} - \frac{3n}{5} - n^{0.99}}$$

for sufficiently large n , so that in this case (1) is bounded from above by $2^{\binom{n}{2} - \frac{3n}{5}}$.

CASE (viii): Suppose the arrow in B is $a < b$. Then $\Gamma_{arr}(b) \subset \Gamma_{arr}(a)$, so that there are only $3^{|T|}$ choices for the arrow neighbours of a, b in T . Then we obtain the bound

$$D(B, P, S) < 2^{\binom{|B|}{2}} 2^{\binom{|T|}{2} + (|B| - 2)|T|} 3^{|T|} < 2^{\binom{n}{2} - \frac{7n}{20} - n^{0.99}}$$

for sufficiently large n , so that in this case (1) is bounded from above by $2^{\binom{n}{2} - \frac{7n}{20}}$.

CASE (ix): We bound directly the number of valid diagrams in this case: we choose B , lines within T and arrows from B to T . This gives us the bound

$$\sum_{|B| < 5\sqrt{n}} \binom{n}{|B|} 2^{\binom{n}{2} - \binom{|B|}{2}} < 2^{\binom{n}{2} + \frac{3 \log^2 n}{4}}$$

for sufficiently large n , since the largest term in the above sum occurs when $\frac{3 \log n}{4} \leq |B| \leq \log n$, so that the largest term is at most $2^{\binom{n}{2} + \log^2 n - \frac{9 \log^2 n}{32}}$.

Adding up the bounds on each case, we see that, for all sufficiently large n , Case (ix) dominates and there are at most $2^{\binom{n}{2} + \log^2 n}$ valid diagrams on n points with $|B| \leq n/100$. Adding up the bounds on Cases (i)-(viii), we see that, for all sufficiently large n , Case (iv) dominates and there are at most $2^{\binom{n}{2} - \frac{n}{23}}$ valid diagrams on n points with $|B| \leq n/100$ which do not correspond to unate 2-SAT functions. ■

6. Counting the case we expect to be small

In this section, we apply (a version of) the Szemerédi Regularity Lemma to restrict the possible partial orders, in a similar way to its application in [2], then use an induction argument to show that there really are very few valid diagrams with $|B| > n/100$. We will need a theorem of Füredi.

THEOREM 6.1: *Let G be a graph on n vertices. Then the proper square of G , the graph G^2 on $V(G)$ with $ab \in E(G^2)$ if and only if $ac, bc \in E(G)$ for some c , has at least $|E(G)| - \lfloor \frac{n}{2} \rfloor$ edges.*

A proof of this is found in Füredi [6]; see also [2].

LEMMA 6.2: *When $|B| > n/100$, for any fixed δ , there exists N such that for all $n > N$, there are at most $2^{\binom{n}{2} - 2n^{5/4}}$ valid diagrams containing a point x such that $|\{y : y < x\}| > \delta n$.*

Proof: Note that if there is a point with δn points below it, then there is a point in T with δn points below it.

There are two cases to consider:

- (1) $|T| < \delta^2 n/5$.
- (2) $|T| \geq \delta^2 n/5$.

CASE (1): We count the number of diagrams with $|T| < \delta^2 n/5$ and a point in T having at least δn points below it as follows. Choose a top set. Choose lines within the top set, and arrows and lines from the top set to the bottom set such that there exists a point x in the top set above at least δn points in the bottom set. Let C be some set of δn points below x . Now choose arrows and lines within $B^{(2)} - C^{(2)}$. Within $C^{(2)}$ we can choose only arrows, so that the structure on C is simply a partial order. This gives an upper bound on the

number of diagrams in this case:

$$\begin{aligned} \sum_{|T|} \binom{n}{|T|} 2^{\binom{|T|}{2}} 3^{|B||T|} 2^{\binom{|B|}{2} - \binom{\delta n}{2}} 2^{\frac{3\delta^2 n^2 - 5\delta n}{10}} |B|! \\ < 2^{\binom{n}{2}} 2^{n^2 \frac{\delta^2}{5} \log \frac{3}{2} - \frac{\delta^2 n^2}{5} + n \log n + n + \log n} \\ < 2^{\binom{n}{2} - 3n^{5/4}} \end{aligned}$$

for all n sufficiently large that both the above approximations hold and $\delta n > J$.

CASE (2): Since both $|T| \geq (\delta^2/5)n$ and $|B| \geq 10^{-2}n$, we will be able to apply a version of the Szemerédi Regularity Lemma to count the number of valid diagrams in this case as follows.

Given a valid diagram, draw a coloured graph G on n vertices corresponding to the points of the diagram as follows. Whenever a point in B is below a point in T , connect them with a red edge in the coloured graph. Whenever two points are connected by lines, or two points in B are connected by arrows, connect them with a blue edge in the coloured graph.

Observe that if two vertices in T are connected by a blue edge, so no vertex in B is connected to one by a red edge and to the other by a blue edge since this would be the forbidden structure, (2) (see Figure 4).

If two vertices a, b in B are connected by a blue edge, so there cannot be three vertices in T , x, y, z , with xa, xb red edges, ya a red edge but yb is not (it does not matter whether yb is a blue edge or not an edge at all), zb a red edge but za is not (again za could be a blue edge or not an edge at all). This is because if the edge ab corresponds to a line, then a, b, x corresponds to the forbidden structure (1), while if it corresponds to an arrow from a to b there would also have to be an arrow from a to z since the partial order is transitive, and vice versa. Since this coloured graph encodes all the information contained in the original diagram except the choice of T and the order on B , there are at most $2^n n!$ times as many valid diagrams in this case as coloured graphs not containing either of these two structures.

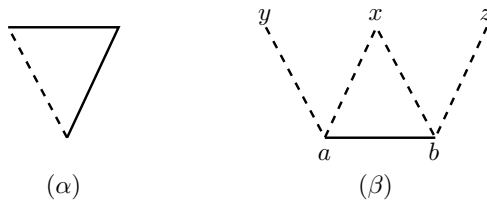


Figure 4. Forbidden coloured subgraphs. Dotted lines are red; solid lines are blue.

Suppose $0 < \epsilon < 1/1000$. If $A, B \subset V(G)$, then let the **red-density** of the pair A, B , $d_r(A, B)$, be the number of red edges between A and B divided by $|A||B|$, and define the **blue-density** $d_b(A, B)$ similarly. Call such a pair ϵ -**uniform** if for every $A' \subset A, B' \subset B$ with $|A'| \geq \epsilon|A|, |B'| \geq \epsilon|B|$, then $|d_b(A, B) - d_b(A', B')| < \epsilon$ and $|d_r(A, B) - d_r(A', B')| < \epsilon$.

Let $V(G)$ be partitioned into m sets $X_1, \dots, X_t, Y_1, \dots, Y_{m-t}$ where all the X_i are subsets of T and all the Y_i are subsets of B . Suppose that each part has size $q > (1 - 2\epsilon)\frac{n}{m}$, except for X_1 and Y_1 which each have size at most ϵn . Call a pair (X_i, Y_j) **rich** if it is ϵ -uniform, $d_r(X_i, Y_j) > 2\epsilon^{1/3}$ and $d_b(X_i, Y_j) > 2\epsilon^{1/3}$. Call any pair which is not ϵ -uniform **bad**, and any ϵ -uniform pair with both $d_r(U, V) \leq 2\epsilon^{1/3}$ and $d_b(U, V) \leq 2\epsilon^{1/3}$ **sparse**. Call the remaining pairs **normal**.

The following two constructions are standard arguments (see for example [2]) whose proof we omit.

Suppose that (X_i, X_j) is a normal pair, and there exists Y_k such that both (X_i, Y_k) and (X_j, Y_k) are rich. Then there is a vertex y in Y_k connected to a vertex a in X_i by a red edge and to a vertex b in X_j by a blue edge, with ab a blue edge. This is the forbidden coloured graph (α) , so that if Y_k is such that $(X_i, Y_k), (X_j, Y_k)$ are rich pairs then (X_i, X_j) must be either a sparse or a bad pair.

Suppose that (Y_i, Y_j) is a normal pair, so $d_b(Y_i, Y_j) \geq 2\epsilon^{1/3}$, and there exists X_k such that both $(X_k, Y_i), (X_k, Y_j)$ are rich. Then we can find vertices $a \in Y_i, b \in Y_j$ and $x, y, z \in X_k$ forming the forbidden coloured graph (β) . Again, if X_k is such that (X_k, Y_i) and (X_k, Y_j) are rich pairs, then (Y_i, Y_j) must be either a sparse or a bad pair.

Now draw a graph H on m vertices corresponding to the parts of $V(G)$. Draw an edge between two vertices in $V(H)$ if and only if they correspond to a rich pair. Then H is bipartite with the X_i making one part and the Y_j the other. Hence the graph H^2 has edges only between vertices X_i, X_j or Y_i, Y_j . No edge of H^2 can correspond to a normal pair in G , so all these edges correspond to bad or sparse pairs in G . We now apply the theorem of Füredi (as in [2]): if there are r edges in H , there are at least $r - \lfloor \frac{m}{2} \rfloor$ edges in H^2 , hence at least as many bad or sparse pairs must be in G .

But it follows from the Szemerédi Regularity Lemma on 2-coloured graphs (proof following the usual method as in Bollobás [1]), that, in fact, every 2-coloured graph has such a partition, for some $1/\epsilon < m < K$, where K depends on ϵ but not on n , with all the parts of size $q \leq n/m$ except for X_1, Y_1 which

have size at most ϵn , and with at most ϵm^2 bad pairs. We will find that choosing ϵ small enough that $62\epsilon^{1/3} \log \frac{e}{2\epsilon^{1/3}} + 116\epsilon < \delta^2/10$ will work.

Consider the number of possibilities for a coloured graph not containing either of the structures (α) or (β) . We must choose T , and the parts X_1, \dots, X_t , Y_1, \dots, Y_{m-t} . We must choose which pairs are to be rich, sparse, bad and normal. We must allow 3^{q^2} possibilities for the edges within every rich or bad pair. We must allow $3^{2n\epsilon n}$ possibilities for the edges with one end in either X_1 or Y_1 . We must allow $3^{mq^2/2}$ possibilities for the edges within parts. Let $\eta = 2\epsilon^{1/3} \log \frac{e}{2\epsilon^{1/3}}$, then within normal pairs there are at most $2^{q^2} 2 \left(\frac{q^2}{2\epsilon^{1/3} q^2} \right) < 2^{q^2+1+\eta q^2}$ possibilities for the edges, and within sparse pairs there are at most $\left(2 \left(\frac{q^2}{2\epsilon^{1/3} q^2} \right) \right)^2 < 2^{2+2\eta q^2}$ possibilities. There are r rich pairs, and at most ϵm^2 bad pairs, hence there are $s \geq r - \epsilon m^2 - \lfloor \frac{m}{2} \rfloor$ sparse pairs. We divide this into two cases and evaluate the number of valid diagrams corresponding to coloured graphs in each case:

First, if $r \geq 5(3\eta + 11\epsilon)m^2$.

We can count the possible graphs in this case simply by enumerating all the possibilities to obtain an upper bound:

$$\begin{aligned}
 & \sum_{|T|, m, r} \binom{n}{|T|} m^n 4^{m^2} 3^{q^2(r + \epsilon m^2 + \frac{m}{2}) + 2\epsilon n^2} 2^{(q^2+1+\eta q^2)(\frac{m^2}{2} - r - s) + (2+2\eta q^2)s} \\
 & < \sum_{|T|, m, r} 2^n m^n 4^{m^2} \left(\frac{3}{4} \right)^{rq^2} 3^{3\epsilon n^2 + \frac{\epsilon n^2}{2}} 2^{(1+\eta q^2)(\frac{m^2}{2} - r - s) + \frac{n^2}{2} + \epsilon n^2 + \lfloor \frac{m}{2} \rfloor q^2 + (2+2\eta q^2)s} \\
 & < \sum_{|T|, m, r} 2^n m^n 4^{m^2} \left(\frac{3}{4} \right)^{rq^2} 2^{\frac{n^2}{2}} 3^{3\epsilon n^2 + \frac{\epsilon n^2}{2}} 2^{m^2 + \eta n^2 + \epsilon n^2 + \frac{\epsilon n^2}{2} + 2m^2 + 2\eta n^2} \\
 & < nK^3 2^n K^n 4^{K^2} 2^{\frac{n^2}{2}} 2^{-(3\eta + 11\epsilon)n^2} 2^{3K^2} 2^{(\frac{7}{2}\epsilon \log 3 + 3\eta + \frac{3}{2}\epsilon)n^2} \\
 & < 2^{\frac{n^2}{2}} 2^{\log n + n + (n+3) \log K + 5K^2 - \epsilon n^2} \\
 & < 2^{\binom{n}{2} - n - n \log n - 3n^{\frac{5}{4}}}
 \end{aligned}$$

for all sufficiently large n .

It follows that for sufficiently large n , at most $2^{\binom{n}{2} - 3n^{5/4}}$ valid diagrams give rise to these coloured graphs.

Second, suppose $r < 5(3\eta + 11\epsilon)m^2$.

Here we do not count graphs directly: instead we count diagrams, using the information we now have about the arrows and lines between B and T to obtain an upper bound. We still need to allow for choice of T, r, m , the

partition, which pairs are to be rich, sparse, bad and normal, and choices of arrows and lines between B and T corresponding to rich, bad and normal/sparse pairs. But now we can count the choices within T as $2^{\binom{|T|}{2}}$, and within B as $2^{\binom{|B|}{2} - \binom{\delta n}{2}} 2^{\frac{3\delta^2 n^2 - 5\delta n}{10}} |B|!$ (since by assumption there is a point in T above $\delta n > J$ points in B). Let the number of valid diagrams on n points which correspond to coloured graphs with less than $5(3\eta + 11\epsilon)m^2$ rich pairs in a Szemerédi partition be $D_g(n)$, then

$$D_g(n) < \sum_{|T|, m, r} \binom{n}{|T|} m^n 4^{m^2} 2^{\binom{|T|}{2}} 3^{rq^2 + 3\epsilon n^2} 2^{(q^2 + 1 + \eta q^2)(t-1)(m-t-1)} \\ \times 2^{\binom{|B|}{2} - \binom{\delta n}{2}} 2^{\frac{3\delta^2 n^2 - 5\delta n}{10}} |B|!$$

since there are at most $(t-1)(m-t-1)$ pairs which are normal or sparse between B and T , and the given upper bound for the number of ways to choose a normal pair is also an upper bound for the number of ways to choose a normal or sparse pair. Simplifying:

$$D_g(n) < \sum_{|T|, m, r} 2^n m^n 4^{m^2} 2^{\binom{|T|}{2}} 3^{rq^2 + 3\epsilon n^2} 2^{(q^2 + 1 + \eta q^2)(t-1)(m-t-1)} 2^{\binom{|B|}{2} - \frac{\delta^2 n^2}{5}} |B|! \\ < nK^3 2^n K^n 4^{K^2} 2^{\binom{n}{2}} 2^{5(3\eta + 11\epsilon)n^2 \log 3 + 3\epsilon n^2 \log 3 + K^2 + \eta n^2 - \frac{\delta^2 n^2}{5}} n!$$

since $q^2(t-1)(m-t-1) \leq |T||B|$, so

$$D_g(n) < nK^3 2^n K^n 2^{3K^2} 2^{\binom{n}{2}} 2^{(31\eta + 116\epsilon)n^2 - \frac{\delta^2 n^2}{5}} n!. \\ < nK^3 2^n K^n 2^{3K^2} 2^{\binom{n}{2}} 2^{-\frac{\delta^2 n^2}{10}} n!.$$

Again this is less than $2^{\binom{n}{2} - 3n^{5/4}}$ for all n sufficiently large that both the above approximations hold and $\delta n > J$.

It follows that for all sufficiently large n , there are at most $3 \cdot 2^{\binom{n}{2} - 3n^{5/4}} < 2^{\binom{n}{2} - 2n^{5/4}}$ valid diagrams with $|B| > n/100$ and with a point above more than δn others. ■

Suppose $\delta = 10^{-9}$. Then observe that, if $n > 10^8$, the number of partial orders where no point has more than δn points below it is at most

$$\left(2 \binom{n}{\delta n}\right)^n < 2^n \left(\frac{10^9 \epsilon n}{n}\right)^{10^{-9} n^2} < 2^{\frac{40}{10^9} n^2 + n} < 2^{\frac{n^2}{2 \cdot 10^7}}.$$

This will enable us to prove the following theorem.

THEOREM 6.3: *For all sufficiently large n , there are at most $2^{\binom{n}{2}-n^{5/4}}$ valid diagrams on n points with $|B| > n/100$.*

Proof: Let $\delta = 10^{-9}$. Let $N > 10^{60}$ be large enough so that both the conclusion of the previous lemma holds for $\delta = 10^{-9}$ and for all $k > N/2$, we have at most $2^{\binom{k}{2}-2k^{5/4}}$ valid diagrams on k points with $3k^{19/20} < |B| \leq k/100$, as provided by Theorem 5.1. Let $F_s(n)$ be the number of valid diagrams on n points with $|B| > n/100$.

Let $A = F_s(N)$. We will prove by induction that $F_s(n) < A2^{\binom{n}{2}-n^{5/4}}$, for every n . When $n \leq N$, this is trivially true by the choice of A . Now suppose that $n > N$, and the induction hypothesis holds for every $k < n$.

Given a valid diagram D on n points with $|B| > n/100$, we can let $P \subset T$ be the set of points each of which is connected to at most $|B|/200$ points in B by lines. Then we can apply Lemma 4.1, with $I = T - P$ and $V = B$. For each $i \in I$, we let S_i be the set of points below i , and L_i be the points in B connected to i by lines (see Figure 5). Then $l \geq |B|/200$, and by Lemma 4.1 we have a forbidden set of size at least $(|B||S|)/400$.

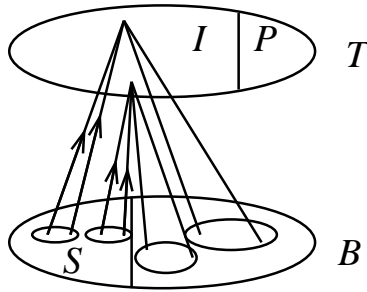


Figure 5. A typical valid diagram with $|B| > n/100$

We consider three cases:

- (a) diagrams such that there is a point above at least δn others,
- (b) diagrams such that there is no point above δn others, and $|S| > |B|/2$,
- (c) diagrams such that there is no point above δn others, and $|S| \leq |B|/2$.

CASE (a): By the previous lemma, there are at most $2^{\binom{n}{2}-2n^{5/4}}$ such valid diagrams.

CASE (b): We count the number of valid diagrams by choosing the partial order within the diagram, then choosing the lines within the top set (now fixed

by choice of the partial order) and between the top and bottom sets (fixing P and S); then we choose the lines within B , observing that there are at most $\binom{|B|}{2} - \frac{|B||S|}{400} \leq \binom{|B|}{2} - \frac{|B|^2}{800} \leq \binom{|B|}{2} - \frac{n^2}{8 \cdot 10^6}$ places where we can choose to put lines within B , to obtain an upper bound:

$$2^{\frac{n^2}{2 \cdot 10^7}} 2^{\binom{|T|}{2} + |B||T| + \binom{|B|}{2} - \frac{|B||S|}{400}} < 2^{\binom{n}{2} - \frac{n^2}{2 \cdot 10^7}} < 2^{\binom{n}{2} - 2n^{5/4}}$$

since $n > N > 10^{40}$.

CASE (c): Since $|S| \leq |B|/2$, $|B - S| \geq |B|/2$. Every point in $B - S$ is under some point in T , and no point in $B - S$ is under any point in I , hence every point in $B - S$ is under some point in P . But no point in P can be above more than δn points, so that $|P| \geq \frac{|B|}{2\delta n} > \frac{1}{200\delta}$. Let $P' \subset P$ be a set with $|P'| = \frac{1}{200\delta}$. Now observe that the diagram D' on the $k = n - 1/(200\delta) < n$ points $T \cup B - P'$, with $a < b$ in D' if and only if $a < b$ in D , and ab a line in D' if and only if ab is a line in D , must be a valid diagram.

Since $|B| > n/100$, and the points P' can be above at most $|P'|\delta n$ points of the top set of D' , D' must have bottom set of size at least $n/100 - |P'|\delta n = n/200 > k/200$. Since $k/200 > 3k^{19/20}$, either D' is a valid diagram on $k > n/2 > N/2$ points with the size of its bottom set in $(3k^{19/20}, k/100]$, or D' is a valid diagram on $k < n$ points with bottom set larger than $k/100$. There are at most $2^{\binom{k}{2} - 2k^{5/4}}$ possible diagrams in the first case by Theorem 5.1, and by the induction hypothesis there are at most $A2^{\binom{k}{2} - k - k^{5/4}}$ possible diagrams in the second case. So there are at most $2^{\binom{k}{2} - 2k^{5/4}} + A2^{\binom{k}{2} - k - k^{5/4}} < (A+1)2^{\binom{k}{2} - k^{5/4}}$ possibilities for D' . Now the following construction includes every diagram in this case.

Choose a top set T , with $\frac{1}{200\delta} \leq |T| \leq \frac{99n}{100}$. Choose a set P' of $\frac{1}{200\delta}$ points in T . Choose lines within P' and from P' to $T - P'$. Choose at most δn points in B to be below each point in P' . Choose at most $\frac{|B|}{200}$ lines going from each point in P' to B . Choose any valid diagram with sufficiently large bottom set on $B \cup T - P'$. Hence an upper bound for the number of valid diagrams in this case is:

$$\begin{aligned} & \sum_{|T|} \binom{n}{|T|} \left(\frac{|T|}{\frac{1}{200\delta}} \right) 2^{\binom{\frac{1}{200\delta}}{2} + \frac{1}{200\delta}(|T| - \frac{1}{200\delta})} \left(2^{\binom{|B|}{2}} 2^{\binom{|B|}{\frac{|B|}{200}}} \right)^{\frac{1}{200\delta}} (A+1) 2^{\binom{k}{2} - k^{\frac{5}{4}}} \\ & < n 2^{2n} 2^{\binom{n}{2}} 2^{-\frac{|B|}{200\delta}} \left(\frac{200e|B|}{|B|} \right)^{\frac{|B|}{20000\delta}} (A+1) 2^{-k^{\frac{5}{4}}} \\ & < A 2^{\binom{n}{2}} 2^{1+\log n + 2n + \frac{|B| \log(200e)}{20000\delta} - \frac{|B|}{200\delta} - k^{\frac{5}{4}}} \\ & < A 2^{\binom{n}{2}} 2^{1+\log n + 2n - \frac{|B|}{400\delta} - k^{\frac{5}{4}}} \end{aligned}$$

$$\begin{aligned}
&< A2^{\binom{n}{2}} 2^{1+\log n+2n-\frac{n}{4000\delta}-n^{\frac{5}{4}}+(n^{\frac{5}{4}}-k^{\frac{5}{4}})} \\
&< A2^{\binom{n}{2}-2n-n^{\frac{5}{4}}} 2^{1+\log n+4n+\frac{5}{800\delta}n^{\frac{1}{4}}-\frac{n}{4000\delta}} \\
&< A2^{\binom{n}{2}-2n-n^{\frac{5}{4}}}
\end{aligned}$$

since $n > 10^{60}$ and $\delta = 10^{-9}$.

Then there are at most $A2^{\binom{n}{2}-2n-n^{5/4}} + 2.2^{\binom{n}{2}-2n^{5/4}} < A2^{\binom{n}{2}-n-n^{5/4}}$ valid diagrams with $|B| > n/100$. Therefore the induction hypothesis holds for n .

By induction, $F_s(n) < A2^{\binom{n}{2}-n-n^{5/4}}$ for all n . If n is sufficiently large that $2^n > A$, then we have $F_s(n) < 2^{\binom{n}{2}-n^{5/4}}$, which completes the proof. ■

7. Upper bound for $H(n)$

Collecting results, from Theorem 5.1 we have that there are at most $2^{\binom{n}{2}-\frac{n}{23}}$ valid diagrams on n points with $|B| \leq n/100$ that do not correspond to unate 2-SAT functions, for all sufficiently large n . From Theorem 6.3, we have that there are at most $2^{\binom{n}{2}-n^{5/4}}$ valid diagrams on n points with $|B| > n/100$, for all sufficiently large n .

Since each non-negative 2-SAT function on n variables can be obtained by applying M to at most 2^n elementary functions on n variables, we have that there are at most $2^n(2^{\binom{n}{2}-n/23} + 2^{\binom{n}{2}-n^{5/4}})$ elementary 2-SAT functions which are not unate, for all sufficiently large n .

There are at most $2^{\binom{n}{2}+n}$ unate 2-SAT functions. So we obtain an upper bound for $H(n)$ valid for all sufficiently large n :

$$H(n) < 2^{\binom{n}{2}+n}(1 + 2^{-n/23} + 2^{-n^{5/4}}) < 2^{\binom{n}{2}+n}(1 + 2^{-n/24})$$

for all sufficiently large n .

8. Improvements

Now we can easily prove Theorem 1.2, restated here for convenience.

THEOREM 1.2: *For sufficiently large n , the expected number of satisfying assignments of a random 2-SAT function on n variables is*

$$(1 + o(1)) \sum_{k=0}^{5\sqrt{n}} \binom{n}{k} 2^{-\binom{k}{2}} = 2^{(1/2+o(1)) \log^2 n}.$$

Proof: The average number of satisfying assignments of a 2-SAT function on n variables is equal to the number of pairs (F, A) , where F is a 2-SAT function

on n variables and A is a satisfying assignment of F , divided by the number of 2-SAT functions on n variables. Since we know the latter, we only need to compute the former. First we find the number of such pairs where F is an elementary 2-SAT function on n variables:

Observe that if (F, A) is a pair as above, with F elementary, then we can let the set of variables which are assigned False in A be X , and relabelling the variables X on F must give a non-negative 2-SAT function, which corresponds to a unique valid diagram D . Thus there are as many pairs with F elementary as there are pairs (D, X) where D is a valid diagram and X a subset of the n variables. There are 2^n possible X , so there are $2^n F(n)$ such pairs. Following the logic in Section 2, the total number of pairs (F, A) where F is any 2-SAT function on n variables is $(1 + o(1))2^n F(n)$. Now recall that when the valid diagrams were split into several types, the largest was the class of diagrams in Case (ix) of Theorem 5.1. There are

$$2^{\binom{n}{2}} \sum_{k=0}^{5\sqrt{n}} \binom{n}{k} 2^{-\binom{k}{2}}$$

diagrams in that case, so that there are

$$(1 + o(1))2^{\binom{n}{2}+n} \sum_{k=0}^{5\sqrt{n}} \binom{n}{k} 2^{-\binom{k}{2}}$$

pairs (F, A) where F is a 2-SAT function on n variables and A is a satisfying assignment for F . Hence the average number of satisfying assignments of a 2-SAT function on n variables is

$$(1 + o(1)) \sum_{k=0}^{5\sqrt{n}} \binom{n}{k} 2^{-\binom{k}{2}}$$

as required. It is trivial to see that the sum is dominated by the terms with k approximately $\log n$, so that the approximation $2^{(1/2+o(1))\log^2 n}$ is valid. ■

We now make some improvements to Theorem 1.1. First, observe that in the proof of this theorem, we could only bound above the class of non-unate 2-SAT functions by $2^{\binom{n}{2}+\frac{24n}{25}}$ because of the bound in Case (iv) of Theorem 5.1.

Given a 2-SAT function S , let F be the unique maximal 2-SAT formula for S . If V is a subset of the domain of S , we can define the formula F' which contains exactly the clauses of F that do not contain literals associated to the variables in V . We say we can **remove the variables** V from S to get S' , a 2-SAT

function on $n - |V|$ variables given by the formula F' . Observe that if R is any subset of the domain of S , then relabelling the variables R on S and removing the variables V gives the same result as removing the variables V from S then relabelling the variables R .

If we can remove a set of k variables from a 2-SAT function S to obtain a unate 2-SAT function, then we say that S is **k -nearly-unate**. If S is k -nearly-unate but not $(k - 1)$ -nearly-unate, we say that S is **exactly k -nearly-unate**.

Observe that if S is an elementary function, and the set V of variables is removed from S to give S' , then $P(S')$ is precisely the partial order on the $2n - 2|V|$ literals induced by $P(S)$. If also S is non-negative, however, $D(S')$ is not in general the diagram given by simply removing the points V and lines and arrows meeting them from $D(S)$. The arrows in $D(S')$ correspond to the arrows in $D(S)$ not meeting V . But if ab is a line in $D(S)$, and $a \in V$, and $c > a$, then cb may be a line in $D(S')$ (if $\bar{c} < b$ is a covering relation in $P(S')$). If ab is a line in $D(S)$, and $a, b \in V$, and $a < c, b < d$, then cd may be a line in $D(S')$. However, if e, f are points in $D(S')$, then ef can only be a line if either it is a line in $D(S)$ or one of the two above situations occurs.

We will require the following simple lemma.

LEMMA 8.1: *Let G be any graph, and k any integer. Then we can find either a set E consisting of at least $k/2$ independent edges of G , or a set Z consisting of at most k vertices of G which meets every edge of G .*

Proof: Given a graph G , let E be a maximal set of independent edges of G . Then either E has size at least $k/2$, or the set $Z = \bigcup E$ has size at most k . Z must meet every edge of G since E is maximal. ■

THEOREM 8.2: *For any $k = k(n)$ such that $k(n) < n^{1/4}$ for all sufficiently large n , the set of 2-SAT functions on n variables which are not k -nearly-unate has size at most $2^{\binom{n}{2} + n - kn/25}$ for all sufficiently large n .*

Proof: We follow essentially the same logic as was used to prove Theorem 1.1.

We observe that for sufficiently large n , if $k(n) = 0$ then the conclusion certainly holds for that n by Theorem 1.1, so in the remainder of the proof we shall assume $k > 0$.

First we show that for any such k , for sufficiently large n , the set $H_k(n)$ of elementary 2-SAT functions on n variables which are not k -nearly-unate has size at most $2^{\binom{n}{2} + n - kn/24}$.

We let $D_k(n)$ be the set of valid diagrams corresponding to the non-negative 2-SAT functions on n variables which are not k -nearly-unate.

As before, we divide $D_k(n)$ into two parts: the diagrams with $|B| > n/100$, and the diagrams with $|B| \leq n/100$. Theorem 6.3 tells us that the first part has size at most $2^{\binom{n}{2} - n^{5/4}}$ for sufficiently large n , so we only need to bound the second part.

We bound above the set of valid diagrams with $|B| \leq n/100$ corresponding to 2-SAT functions on n points which are not k -nearly-unate. For such a diagram, we define the sets $\Gamma_{arr}(b)$, $\Gamma_{line}(b)$ for $b \in B, I, P, S_a, L_a$ for $a \in I, S$ as in Theorem 5.1. We also define sets S'_p, L'_p , for some $p \in P$, and the set S' as follows. If $p \in P$, and there is $a \in I$ with ap a line in D and L_a is the set of arrow neighbours of a , then let $S'_p = \Gamma_{arr}(p) - S$, $L'_p = \Gamma_{arr}(a) - S$. Let S' be the union of the defined S'_p . We now split the set of these valid diagrams into five parts:

CASE (i): $|P| \geq 84n^{1/4}$.

CASE (ii): $|P| < 84n^{1/4}$ and $|S| \geq 100n^{1/4}$.

CASE (iii): $|P| < 84n^{1/4}$, $|S| < 100n^{1/4}$ and $|B| \geq 3n^{19/20}$.

CASE (iv): $|P| < 84n^{1/4}$, $|S| < 100n^{1/4}$, $|B| < 3n^{19/20}$ and $|P| + |S| + |S'| > k$.

CASE (v): $|P| < 84n^{1/4}$, $|S| < 100n^{1/4}$, $|B| < 3n^{19/20}$ and $|P| + |S| + |S'| \leq k$.

By identical logic to that in Theorem 5.1, each of Cases (i), (ii), (iii) contains at most $2^{\binom{n}{2} - 3n^{5/4}}$ valid diagrams, for sufficiently large n .

We now provide a bound for Case (iv).

Observe that, if $p \in P$ is connected by a line to $a \in I$, with L_a the set of arrow neighbours of a , then $L'_p = \Gamma_{arr}(a) - S$ cannot intersect S'_p since the structure (1) is forbidden. Furthermore, there can be no lines between S'_p and L'_p since structure (3) is forbidden. Now $|L'_p| \geq n/10 - |S| > n/11$ for sufficiently large n . This means that by Lemma 4.1 there is a forbidden set in T of size $\frac{|S'|n}{22}$, which does not intersect the forbidden set between the sets L_i and S_i ($i \in I$) since no member of this new forbidden set has an end in S . So we can construct any diagram in this case in the usual way: given B, P, S , choose the lines and arrows from P to T , whether the members of I have as their large set their set of line or arrow neighbours, the lines and arrows from I to $T - S$, the lines and arrows from I to S , the lines and arrows within B , which fixes S' , and finally the lines within T , taking account of both the forbidden sets. This allows us to

use in this case the bound

$$\begin{aligned} D(B, P, S) &< 2^{|P||T|} 2^{-\frac{|P||n|}{21}} 2^{|I||T-S|} 3^{|I||S|} 2^{\binom{|B|}{2}} |B|! 2^{\binom{|T|}{2} - \frac{|S||n|}{20} - \frac{|S'||n|}{22}} \\ &< 2^{\binom{n}{2}} 2^{|I|} |B|! 2^{-\frac{(|P|+|S|+|S'|)n}{22}} \\ &< 2^{\binom{n}{2} - n^{0.99} - \frac{kn}{23}} \end{aligned}$$

for sufficiently large n , so that in this case the number of valid diagrams is bounded from above by $2^{\binom{n}{2} - \frac{kn}{23}}$.

Finally we bound Case (v):

Given a diagram D in Case (v), we draw a graph G with $V(G) = I$. If $a, b \in I$, $L_a = \Gamma_{line}(a)$, $L_b = \Gamma_{arr}(b)$, and either ab is a line or ab being a line would create one of the forbidden structures (2), (3) (so that ab could potentially be a line in some diagram D' obtained by removing variables), then we put $ab \in E(G)$. If $a < b$ in D , $L_a = \Gamma_{line}(a)$ and $L_b = \Gamma_{line}(b)$, then we put $ab \in E(G)$. If $a < b$ in D , $L_a = \Gamma_{arr}(a)$ and $L_b = \Gamma_{arr}(b)$, then we put $ab \in E(G)$. Otherwise we do not put $ab \in E(G)$.

We split this case into three sub-cases:

Subcase (a): There is a line ab in I , with both L_a the set of arrow neighbours of a and L_b the set of arrow neighbours of b .

In this case, there is a forbidden set of size at least $n^2/100$ in T , and we can bound above the number of valid diagrams in this case as in Case (v) of Theorem 5.1, obtaining a bound $2^{\binom{n}{2} - n^{5/4}}$ for sufficiently large n .

Subcase (b): There is no line between any two points $a, b \in I$ with both L_a the set of arrow neighbours of a and L_b the set of arrow neighbours of b . We cannot find any set of $q = k - |P| - |S| - |S'|$ points in I such that every edge in $E(G)$ touches the set.

In this case, by Lemma 8.1 there must be a set E of $\lceil \frac{q}{2} \rceil$ edges in $E(G)$, no two of which meet at any point.

Suppose we know for each point $i \in I$ whether its large set is its set of arrow or line neighbours. We have previously used the bound $2^{|I|}$ on the choices of lines and arrows between I and any given $t \in T - S$. But now observe that if $ab \in E$, we have only 3 choices for the lines and arrows from a, b to t ; if ab is in E since $a < b$ in D with both $L_a = \Gamma_{line}(a)$ and $L_b = \Gamma_{line}(b)$, then we cannot choose to put lines from both a and b to t since structure (2) is forbidden, and so on. Since the edges in E are independent, we obtain a bound $2^{|I| - q3^{q/2}}$ on the number of choices of lines and arrows between I and t .

We can construct every diagram in this case as follows: Given B , P , S , we choose the $\lceil \frac{q}{2} \rceil$ edges in E , and whether they are to correspond to lines or arrows in one or the other direction in D . We choose for each $i \in I$ whether L_i is to be the set of line or arrow neighbours of i . We choose the lines and arrows from I to $T - S$, taking account of the above restriction, and from I to S . We choose the lines and arrows from P to T , the lines and arrows within B , and the lines within T , taking account of both the forbidden sets (note S' and the L'_p are already chosen). This gives us the bound

$$\begin{aligned} D(B, P, S) &< \left(3 \binom{|I|}{2} \right)^{\lceil \frac{q}{2} \rceil} 2^{|I|} 2^{(|I|-q)|T-S|} 3^{\frac{q}{2}|T-S|} 3^{|I||S|} 2^{|P||T|} 2^{-\frac{|P||n|}{21}} 2^{\binom{|B|}{2}} |B|! \\ &\quad \times 2^{\binom{|T|}{2} - \frac{n|S|}{20} - \frac{n|S'|}{22}} \\ &< (3|B|^2)^k 2^{|I|} |B|! 2^{\binom{n}{2}} 2^{-\frac{(|P|+|S|+|S'|)n}{22} - \frac{k-|P|-|S|-|S'|}{2} \frac{98n}{100} \log \frac{4}{3}} \\ &< 2^{\binom{n}{2} - \frac{kn}{23} - n^{0.99}} \end{aligned}$$

for sufficiently large n , so that the number of diagrams in this subcase is bounded from above by $2^{\binom{n}{2} - \frac{kn}{23}}$.

Subcase (c): There is no line between points $a, b \in I$ with both L_a the set of arrow neighbours of a and L_b the set of arrow neighbours of b . We can find a set Z of $k - |P| - |S| - |S'|$ points in I which meets every edge in $E(G)$.

Let D' be the diagram corresponding to the non-negative 2-SAT function obtained by removing the k variables $V = P \cup S \cup S' \cup Z$ from the 2-SAT function corresponding to D . Recall that the arrows in D' correspond to the arrows in D not meeting V , but there may be some lines in D' which do not correspond to lines in D , but exist because the covering relations have changed. We will continue to use T , B to refer to the top and bottom sets of D , and will use T' , B' for the top and bottom sets of D' .

Suppose $a \in I$ has $L_a = \Gamma_{line}(a)$, and $a < b$ for some $b \in I$. Then b has $L_b = \Gamma_{line}(b)$, otherwise $S_a \supset L_b$, but $|S_a| < 100n^{1/4} < |L_b|$. So either a or b must be in Z . It follows that any point $a \in I - Z$ with $L_a = \Gamma_{line}(a)$ is maximal in D' .

Suppose that a point $b \in I - Z$ has $L_b = \Gamma_{arr}(b)$, and in D' there is a line db for some d . Certainly db cannot have existed in D , as that would imply either a line between two points whose large sets are their sets of arrow neighbours, $d \in S$, $d \in Z$ or $d \in P$. Furthermore, $d \notin B$, as that would imply either $d \in P$ or $d \in Z$, so $d \in T$. Recall that $d \notin V$. There are three possibilities.

First, there could be $e \in V$, with de a line in D and $e < b$. But then $e \in I$, since e has at least $|L_b|$ arrow neighbours in T , and this would imply $d \in S$.

Second, there could be $f \in V$ with $f < d$ and fb a line in D . But then $f \notin I$, since this would either imply $d \in S$ or fb would be a line between two points in I with large sets their sets of arrow neighbours. So $f \in P$, and then we have $d \in S'$.

Third, there could be $a, c \in V$ with $a < b$, $c < d$, and ac a line in D . Since $a < b$, so a has at least $|L_b|$ arrow neighbours, and $a \in I$. Since $|S| < 100n^{1/4} < |L_b|$, so L_a would be the set of arrow neighbours of a . Then if $c \in I$, we would have either $d \in S$ or ac would be a line between two points in I with large sets their sets of arrow neighbours. So $c \in P$, and we have $d \in S'$.

Thus we see that we cannot have any point $b \in I - Z$ with $L_b = \Gamma_{arr}(b)$ which meets a line in D' .

Now the top set T' of D' consists of $T - S - S'$ together with those points in $I - Z$ whose large set was their line set, while the bottom set B' of D' consists of those points in $I - Z$ whose large set was their arrow set. Then there are no lines within B' . There are no arrows within B' , since Z was removed. There are no lines between B' and T' as above. Now, by Lemma 3.1, D' corresponds to a unate 2-SAT function on $n - k$ points, contradicting the original diagram on n points being in $D_k(n)$. Hence there are no diagrams in this subcase.

Adding up the bounds from all the cases, we see that for any $k = k(n)$ with $k(n) < n^{1/4}$ for all sufficiently large n , we have

$$|D_k(n)| < 3.2 \binom{n}{2}^{-3n^{\frac{5}{4}}} + 2.2 \binom{n}{2}^{-n^{\frac{5}{4}}} + 2.2 \binom{n}{2}^{-\frac{kn}{23}} < 2 \binom{n}{2}^{-\frac{kn}{24}}$$

for all sufficiently large n , and $|H_k(n)| < 2^{\binom{n}{2} + n - \frac{kn}{24}}$.

Since every 2-SAT function which is not k -nearly-unate can be reduced to an elementary 2-SAT function on $n - l$ variables which is not $(k - l)$ -nearly-unate if $k \geq l$, or to an elementary 2-SAT function on $n - l$ variables if $k < l$, we can follow the logic in Section 2 and bound above the size of the set of 2-SAT functions on n variables which are not k -nearly-unate by

$$\begin{aligned} 1 + \sum_{l=0}^k \binom{n}{l} (2n + 2 - 2l)^l H_{k-l}(n - l) + \sum_{l=k+1}^n \binom{n}{l} (2n + 2 - 2l)^l H(n - l) \\ < 1 + \sum_{l=0}^k \binom{n}{l} (2n + 2 - 2l)^l 2^{\binom{n-l}{2} + n - l - \frac{(k-l)(n-l)}{24}} \\ + \sum_{l=k+1}^n \binom{n}{l} (2n + 2 - 2l)^l H(n - l), \end{aligned}$$

which sum is, by the same logic as used to prove Theorem 1.1, dominated by the term $2^{\binom{n}{2} + n - \frac{kn}{24}}$, hence, for sufficiently large n , bounded from above by $2^{\binom{n}{2} + n - \frac{kn}{25}}$. ■

Finally, we prove Theorem 1.3. We find the next largest class of 2-SAT functions after the unate functions: we will see that this is a subset of the set of 1-nearly-unate 2-SAT functions. Let $W(n)$ be the set of 2-SAT functions which are not unate and which are given by taking a monotone function and changing exactly one positive literal x in one clause of its formula to the corresponding negative literal \bar{x} . Let $V(n)$ be the set of 2-SAT functions given by relabelling variables on the elements of $W(n)$.

Note that applying the above process to a monotone function results in a unate function if and only if x is mentioned just once in the formula for the monotone function.

We can easily find the size of $W(n)$. Observe that all elements of $W(n)$ are non-negative; then they are in 1-1 correspondence with the diagrams on n points which have $|T| = n - 1$, and $x \in B$ is linked to exactly one element $y \in T$ by an arrow, and to at least one other element of T by a line. As structure (2) is forbidden, for every other $z \in T$ there cannot be lines from z to both y and x . Therefore there are at most $n(n-1)2^{\binom{n-2}{2}}3^{n-2}$ such valid diagrams. There are at least

$$n(n-1)(2^{\binom{n-2}{2}} - n2^{\binom{n-3}{2}})(3^{n-2} - 2\left(\binom{n-2}{0} + \binom{n-2}{1}\right)2^{n-2})$$

such valid diagrams in which every point has at least two lines connected to it.

Therefore

$$\begin{aligned} \binom{n}{2} 2^{\binom{n}{2}} 2^{-2(n-2)} (1 - o(2^{-\frac{n}{2}})) 3^{n-2} (1 - o(2^{-\frac{n}{2}})) &< |W(n)| \\ &< \binom{n}{2} 2^{\binom{n}{2}} 2^{-2(n-2)} 3^{n-2} \end{aligned}$$

so

$$\binom{n}{2} 2^{\binom{n}{2}} \left(\frac{3}{4}\right)^{n-2} (1 - o(2^{-\frac{n}{2}})) < |W(n)| < \binom{n}{2} 2^{\binom{n}{2}} \left(\frac{3}{4}\right)^{n-2}$$

for sufficiently large n .

Now $|V(n)| \leq 2^n |W(n)|$, but also if R, R' are subsets of the n variables and w, w' are elements of $W(n)$ corresponding to diagrams in which every point has at least two lines connected to it, then the functions given by relabelling the variables R on w and R' on w' are distinct unless $R = R'$, $w = w'$. Thus

$$\binom{n}{2} 2^{\binom{n}{2} + n} \left(\frac{3}{4}\right)^{n-2} (1 - o(2^{-\frac{n}{2}})) < |V(n)| < \binom{n}{2} 2^{\binom{n}{2} + n} \left(\frac{3}{4}\right)^{n-2}.$$

Now we improve the bounds from Theorem 5.1 to show that $V(n)$ really is the next largest class of 2-SAT functions after the unate functions. The following proof is essentially a more precise, but much longer, replacement for the argument in Case (iv) of Theorem 5.1; it was left to this point to make that theorem more easily understood.

We observe that three types of diagram correspond to functions in $V(n)$:

(p) the diagrams in which B contains no lines, there is a point x in B which is connected to exactly one point in T by an arrow and to at least one other point in T by a line, any arrows in B go to x , and all other points in B are connected to points in T by arrows only;

(q) the diagrams in which B contains no arrows or lines, there is a point x in B which is connected to exactly one point in T by a line and to at least one other point in T by an arrow, and all other points in B are connected to points in T by arrows only;

(r) the diagrams in which B contains exactly one arrow and no lines, and there are no lines between B and T .

We restate Theorem 1.3 for convenience.

THEOREM 1.3: *For sufficiently large n ,*

$$2^{\binom{n}{2}+n} \left(1 + \binom{n}{2} \left(\frac{3}{4} \right)^{n-2} - 2^{-\frac{n}{2}} \right) < G(n) < 2^{\binom{n}{2}+n} \left(1 + \binom{n}{2} \left(\frac{3}{4} \right)^{n-2} + 2^{-\frac{418n}{1000}} \right)$$

and the next largest class of 2-SAT functions after the unate 2-SAT functions is $V(n)$.

Proof: We find that the bound in Theorem 6.3 is already good enough, as are the bounds in Theorem 5.1 Cases (i), (ii), (iii), (v), (vi), (vii). We observe that the diagrams in Case (viii) already correspond to 2-SAT functions in $V(n)$ (in form (r)), and that those in Case (ix) correspond to the unate functions, which we have already enumerated with sufficient accuracy. Hence we only need to improve the bounds given in Case (iv) of Theorem 5.1.

Case (iv) of Theorem 5.1 covered diagrams such that $|P| < 84n^{\frac{1}{4}}$, $|S| < 100n^{\frac{1}{4}}$, $|B| < 3n^{19/20}$ and at least one of $|P| > 0$, $|S| > 0$ holds.

We now split this into several cases and analyse each.

- (a) $|P| < 84n^{\frac{1}{4}}$, $1 < |S| < 100n^{\frac{1}{4}}$, $|B| < 3n^{\frac{19}{20}}$.
- (b) $|S| \leq 1$, $0 < |P| < 84n^{\frac{1}{4}}$ and $|B| < 3n^{\frac{19}{20}}$.
- (c) $|S| = 1$, $|P| = 0$, $|B| < 3n^{\frac{19}{20}}$ and there are at least two points in B with non-empty small set.

- (d) $|S| = 1$, $|P| = 0$, $|B| < 3n^{\frac{19}{20}}$, only one point $b \in B$ has $|S_b| = 1$ and there is a line within B .
- (e) $|S| = 1$, $|P| = 0$, $5\sqrt{n} \leq |B| < 3n^{\frac{19}{20}}$, only one point $b \in B$ has $|S_b| = 1$ and there is no line within B .
- (f) $|S| = 1$, $|P| = 0$, $|B| < 5\sqrt{n}$, only one point $b \in B$ has $|S_b| = 1$, there is no line within B but there is an arrow within B .
- (g) $|S| = 1$, $|P| = 0$, $|B| < 5\sqrt{n}$, only one point $b \in B$ has $|S_b| = 1$ and there are no lines or arrows within B .

CASE (a): As in Theorem 5.1, we will use the bounds $|I| \log(3/2) < n/1000$, $n^3 \binom{n}{|B|} \binom{|B|}{|P|} \binom{|T|}{|S|} < 2^{n^{0.99}}$ and $2^{|I|}|B|! < 2^{\frac{n}{1000}}$ in the following cases.

We divide this into two sub-cases.

Subcase (1): $|S| > 40$.

We use (2) from Theorem 5.1 which gives us

$$D(B, P, S) < 2^{\binom{n}{2}} 2^{-\frac{|P||n|}{21}} 2^{-\frac{40n}{21}} < 2^{\binom{n}{2} - n - n^{0.99}}$$

for all sufficiently large n , so that the sum (1) from Theorem 5.1 is bounded from above by $2^{\binom{n}{2} - n}$.

Subcase (2): $2 \leq |S| \leq 40$.

We observe that as there are at least two points in S , there are at most $\binom{40}{2}$ pairs in $T^{(2)}$ with both points in S , and the forbidden set has size at least $2l - 800$, where l is the minimum of the $|L_i|$ for $i \in I$. We can construct diagrams in this subcase, given B, P, S , by choosing the lines and arrows from the points in P to T , choosing for each point in I lines and arrows to S and lines to T , using the fact that there are at most $\binom{|T|}{l}$ ways to choose lines from the point i with $|L_i| = l$ ($|I|$ ways to choose i), choosing lines and arrows within B and finally lines within T , taking account of the forbidden set. This gives us

$$\begin{aligned} D(B, P, S) &< 2^{|P||T|} 2^{-\frac{|P||n|}{21}} 3^{|S||I|} |I| 2^{(|I|-1)|T|} \binom{|T|}{l} 2^{\binom{|B|}{2}} |B|! 2^{\binom{|T|}{2} - 2l + 800} \\ &< 2^{\binom{n}{2}} 2^{-\frac{|P||n|}{21}} 2^{40|I| \log \frac{3}{2}} 2^{-|T|} \binom{|T|}{l} 2^{-2l + 800} |B|!. \end{aligned}$$

Now $2^{40|I| \log(3/2)} 2^{800} |B|! < 2^{n^{0.99}}$ for all sufficiently large n , so

$$D(B, P, S) < 2^{\binom{n}{2}} 2^{-2l - \frac{99n}{100}} \binom{n}{l} 2^{n^{0.99}} < 2^{\binom{n}{2} - \frac{n}{2} - n^{0.99}}$$

for all sufficiently large n , and the sum (1) is bounded from above by $2^{\binom{n}{2} - n/2}$.

CASE (b): We divide this into two subcases.

Subcase (1): There is a point $p \in P$ with $|S_p| > \sqrt{n}$.

Given B, P, S , we can construct the valid diagrams in this case by choosing the lines and arrows within B and between B and T , and then choosing the lines within T . But since $|L_p| > |S_p|$ there are at least n pairs in $T^{(2)}$ which cannot be chosen as lines, and we can bound above the number of possibilities by

$$\begin{aligned} D(B, P, S) &< 2^{\binom{|B|}{2}} |B|! 2^{|P||T|} 2^{-\frac{|P||n|}{21}} 2^{|I|} 2^{|I||T-S|} 3^{|I||S|} 2^{\binom{|T|}{2} - n} \\ &< 2^{\binom{n}{2} - n + (1 + \log \frac{3}{2})|I|} |B|! \\ &< 2^{\binom{n}{2} - \frac{n}{2} - n^{0.99}} \end{aligned}$$

for all sufficiently large n , so that in this subcase (1) is bounded from above by $2^{\binom{n}{2} - n/2}$.

Subcase (2): Every point $p \in P$ has $|S_p| \leq \sqrt{n}$.

Now instead of there being $(2^{\binom{|T|}{n/10}})^2$ ways to choose the arrows and lines from each point in P , there are only $4 \binom{|T|}{n/10} \binom{|T|}{\sqrt{n}} < 2^{n/2}$ ways, for sufficiently large n . Hence we can bound above the number of possibilities by

$$\begin{aligned} D(B, P, S) &< 2^{\binom{|B|}{2}} |B|! 2^{|P||T|} 2^{-\frac{49|P||n|}{100}} 2^{|I|} 2^{|I||T-S|} 3^{|I||S|} 2^{\binom{|T|}{2}} \\ &< 2^{\binom{n}{2} - \frac{97|P||n|}{200}} \\ &< 2^{\binom{n}{2} - \frac{48n}{100} - n^{0.99}} \end{aligned}$$

for sufficiently large n , so that in this subcase (1) is bounded from above by $2^{\binom{n}{2} - \frac{48n}{100}}$.

CASE (c): Let the two points in B with non-empty small set be a, b . Then the forbidden set has size at least $l = \max(|L_a|, |L_b|)$. Observe that there are at most $l \binom{|T|}{l}$ ways to choose the large sets of each of a, b , and that

$$\left(l \binom{n}{l} \right)^2 2^{-l}$$

is maximised at $l = (\sqrt{2} - 1)n$. We construct the diagrams in this case by choosing B, l, S , the two points with non-empty small set, whether the large set of each point in B will be its arrow neighbours or line neighbours, the small sets of all the points in B , then we choose the large sets of both the points with non-empty small set, the large sets of the rest of the points in B , the lines and

arrows within B and finally the lines within T . This allows us to bound (1) in this case by

$$\begin{aligned} \sum_{|B|, l} \binom{n}{|B|} \binom{|T|}{1} \binom{|B|}{2} 2^{|B|} 2^{|B||S|} \left(l \binom{|T|}{l} \right)^2 2^{(|B|-2)|T|} 2^{\binom{|B|}{2}} |B|! 2^{\binom{|T|}{2}-l} \\ < 2^{2n^{0.99}} 2^{\binom{n}{2} - \frac{198n}{100}} \left(\frac{n}{n(\sqrt{2}-1)} \right)^2 2^{-(\sqrt{2}-1)n} \\ < 2^{\binom{n}{2} - \frac{42n}{100}} \end{aligned}$$

for all sufficiently large n .

CASE (d): We know that every point in B except b has only arrow neighbours in T . If the line in B does not touch b , or if L_b is the set of arrow neighbours of b , then we will be able to write down exactly the same bound as in Case (v) of Theorem 5.1. If b has its large set its line neighbours, and the line in B is ab for some $a \in B$, then we find that there can be no arrows from a to L_b or S , since structures (2) and (1) are forbidden. If $|L_a| = l$, then we find a forbidden set of size l in T . We can construct these diagrams by choosing a, b , arrows from $B - \{a, b\}$ to T , lines from b to T and arrows from a to $T - S - L_b$, the arrow and line connections within B , and finally lines in T . We can bound this above by:

$$\begin{aligned} D(B, P, S) &< \sum_l 2^{\binom{|B|}{2}} 2^{(|B|-2)|T|} \binom{|T|}{l} 2^{|T|-l} 2^{\binom{|B|}{2}} |B|! 2^{\binom{|T|}{2}-l} \\ &< 2^{\binom{n}{2} - n/2 - n^{0.99}} \end{aligned}$$

for all sufficiently large n , so that in this case (1) is bounded by $2^{\binom{n}{2} - \frac{n}{2}}$.

CASE (e): We use exactly the same argument as in Case (vi) of Theorem 5.1 and get the same bound.

CASE (f): Subcase (1): If b has L_b , the set of its arrow neighbours in T , or there is an arrow in B that does not go to or from b , then let the arrow be $c < d$. We note that if $d < t$, then $c < t$ and so there are at most $3^{|T|}$ choices for arrows between $\{c, d\}$ and T . We note that there is also a forbidden set in T of size at least $n/10$ between L_b and S_b , so that following the usual logic the number of diagrams in this subcase is bounded from above by:

$$\begin{aligned} \sum_{|B|} \binom{n}{|B|} \binom{|T|}{1} 2^{\binom{|B|}{2}} 2^{\binom{|B|}{2}} |B|! 2^{(|B|-2)|T|} 3^{|T|} 2^{\binom{|T|}{2} - \frac{n}{10}} < 2^{\binom{n}{2} - |T| \log \frac{4}{3} - \frac{n}{10}} \\ < 2^{\binom{n}{2} - \frac{n}{2}}. \end{aligned}$$

Subcase (2): If b has L_b , the set of its line neighbours, and the only arrows in B go to b , then for every point a with $a < b$, L_a is the set of arrow neighbours of a , since S_a is empty. These diagrams correspond to 2-SAT functions in $V(n)$, in the form (p).

Subcase (3): If b has L_b , the set of its line neighbours, and there is an arrow in B from b to a , then every point in L_a is above b , which contradicts $|S_b| = 1$. Thus there are no valid diagrams in this subcase.

CASE (g): These diagrams correspond to 2-SAT functions in $V(n)$, in forms (p) and (q).

We have now improved the bounds on Case (iv) of Theorem 5.1, so that we may say that for sufficiently large n there are at most $2^{\binom{n}{2} - 419n/1000}$ valid diagrams which do not correspond either to unate 2-SAT functions or to 2-SAT functions in $V(n)$.

Following the same logic as was used to prove Theorem 1.1, we now obtain, for sufficiently large n , an upper bound $2^{\binom{n}{2} + n - 418n/1000}$ on the number of 2-SAT functions which are neither unate nor in $V(n)$. ■

9. Further thoughts

We could certainly continue to extract further classes of 2-SAT functions in decreasing size order, either in the manner of Theorem 1.3 or by enumerating the large classes of k -nearly-unate functions for suitable k and appealing to Theorem 8.2 to show that there are no larger classes left uncounted.

It is obvious that Theorem 8.2 is not best possible: the constant $1/25$ is certainly too small. Also, we conjecture that in fact the result holds for any function $k(n)$. However, observe that the class of 2-SAT functions on n variables which consist of the first k variables all having their positive literals in the spine, and the remaining $n - k$ variables forming a monotone function, has size $2^{\binom{n-k}{2}} = 2^{\binom{n}{2} - kn + \frac{k^2+k}{2}}$. All functions in this class are exactly k -nearly-unate, so that we do not expect a sharp upper bound on the number of non- k -nearly-unate functions to be of a substantially different form to that given.

Prömel, Schickinger and Steger [8] found that almost every triangle-free graph which is not bipartite, and cannot be made bipartite by removing one vertex, can be made bipartite by removing two vertices, and so on; Theorem 8.2 suggests that the corresponding result holds for our problem. We conjecture that the class of 2-SAT functions on n variables which are exactly k -nearly-unate is larger than the class of those which are exactly $(k+1)$ -nearly-unate by a factor

of at least $2^{\epsilon n}$ for all sufficiently large n , all k , and some constant $\epsilon > 1/25$.

Our results only hold for a number n of variables greater than some (very large) number N . It is worth considering whether we really needed to apply the Szemerédi Regularity Lemma, which was responsible for causing N to be so large; to obtain our results we did not make as much use of the lemma as in [2], where, (implicitly) the Regularity Lemma is applied for an infinite sequence of ϵ 's tending to zero, and we did not need the lemma at all to deal with the large case. However, even if a method of avoiding it could be found, other parts of the proof require N to be so large as to be useless for practical application.

This paper does not attempt to attack the problem posed in [2] of determining the asymptotic behaviour of the number of k -SAT functions for $k > 2$; in [2], an upper bound $2\sqrt{\pi(k+1)}\binom{n}{k}$ is given, and the conjecture offered that the exponent should in fact be $\binom{n}{k}(1+o(1))$. It does not seem likely that the methods applied here will be of any use for this more general problem; the first step of moving from 2-SAT functions to partial orders has no obvious analogue in the case of 3-SAT or higher. The problem of determining the asymptotic behaviour of the number of k -SAT functions on n variables in the case $k > n/2$ is also studied in a paper of Bollobás and Brightwell [3].

References

- [1] B. Bollobás, *Modern Graph Theory*, Springer-Verlag, Berlin, 1998, xiii+394pp.
- [2] B. Bollobás, G. Brightwell and I. Leader, *The number of 2-SAT functions*, Israel Journal of Mathematics **133** (2003), 45–60.
- [3] B. Bollobás and G. Brightwell, *The number of k -SAT functions*, Random Structures and Algorithms **22** (2003), 227–247.
- [4] G. Brightwell, H. J. Prömel and A. Steger, *The average number of linear extensions of a partial order*, Journal of Combinatorial Theory, Series A **73** (1996), 193–206.
- [5] P. Erdős, D. Kleitman and B. Rothschild, *Asymptotic enumeration of K_n -free graphs*, in *International Colloquium on Combinatorial Theory*, volume 2 of *Atti dei Convegni Lincei*, 1976, pp. 19–27.
- [6] Z. Füredi, *The maximum number of edges in a minimal graph of diameter 2*, Journal of Graph Theory **16** (1992), 81–98.
- [7] D. Kleitman and B. Rothschild, *Asymptotic enumeration of partial orders on a finite set*, Transactions of the American Mathematical Society **205** (1975), 205–220.
- [8] H. J. Prömel, T. Schickinger and A. Steger, *On the structure of clique-free graphs*, Random Structures and Algorithms **19** (2001), 37–53.