The satisfiability threshold for randomly generated binary constraint satisfaction problems

Alan Frieze[1] and Michael Molloy[2]

[1] Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA15213, USA. ***

[2] Department of Computer Science, University of Toronto, Toronto, Ontario M5S 3G4. †

Abstract. We study two natural models of randomly generated constraint satisfaction problems. We determine how quickly the domain size must grow with n to ensure that these models are robust in the sense that they exhibit a non-trivial threshold of satisfiability, and we determine the asymptotic order of that threshold. We also provide resolution complexity lower bounds for these models. One of our results immediately yields a theorem regarding homomorphisms between two random graphs.

1 Introduction

The Constraint Satisfaction Problem (CSP) is a broadly studied generalization of k-SAT. A CSP consists of a set of variables, each of which may receive a value from $\{1, ..., m\}$ and a set of constraints, each of which restricts the values that certain subsets of the variables may receive. In this paper, we focus on the *binary* case, meaning that each constraint is on two variables and lists a set of ordered pairs of values that the two variables may not take.

Over the past several years, much research has gone into the study of random models of CSP's (see eg. [10, 19, 12, 25, 31, 32]). Many different models have been studied. In each model, one first takes a random graph (or in the non-binary case, a random hypergraph) whose vertices are the n variables and then puts a constraint on each pair of variables that is joined by an edge. The random graph is always one of the two standard models: $G_{n,M}$ where we choose a uniformly random set of M edges, and $G_{n,p}$ where each of the $\binom{n}{2}$ possible edges is selected with probability p, independently of the selection of any other edges. In this paper we will use $G_{n,p}$ but, as usual, it is straightforward to show that all of our theorems also hold for $G_{n,M}$ when $M = p\binom{n}{2}$. Where the CSP models differ

*** Supported in part by NSF grant CCR0200945. Research carried out during a visit to the Microsoft Research, Theory Group

† Supported by NSERC and a Sloan Research Fellowship. This research was carried out while the second author was a Visiting Researcher at Microsoft Research.

most notably from each other is in the way that the constraint is chosen for each edge. In this paper, we will focus on two of the most natural ways to do this.

When researchers examine a random model of CSP, they almost always start by looking at the *satisfiability threshold*; i.e. a value p^* and constants $c_1 < c_2$ such that choosing $p = c_1 p^*$ results in a problem that is **whp**[1] satisfiable, while choosing $p = c_2 p^*$ results in a problem that is **whp** unsatisfiable. (We do not address the more specific notion of a *sharp* threshold in this paper.)

Most of the work done thus far both on specific models of random CSP's (eg. [14, 3, 27]) and on families of models (eg. [12, 25, 26]) have focussed on the case where the domain size, m, is constant. Our paper focuses on the case where $m \to \infty$ with n. There has been some previous work on this case (eg. [31, 32, 15, 16, 29, 18]), but not nearly as much as has been done on the constant domain case.

In [2] it was noted that with constant domain sizes, Model A, below, has a fatal flaw which prevents it from exhibiting an interesting satisfiability threshold (described in more detail below). In one of the first studies of a model with non-constant domain size, Xu and Li[31] proved (amongst other things) that if the domain size is $m = n^{1/2+\epsilon}$, then Model A does not have the fatal flaw and does indeed exhibit a non-trivial satisfiability threshold. One of our main contributions is to determine just how high the domain size has to be in order for Model A to exhibit such a threshold. It is easy to see that our second model, Model B, exhibits such a threshold for any domain size $m \to \infty$.

Our second contribution is that we determine, up to a constant multiple, the locations of these thresholds. We were surprised to discover the the threshold for Model B is asymptotically much higher than that of Model A, despite a superficial similarity.

Our final contribution is to prove lower bounds on the resolution complexity for both models.

Next, we will describe our models and results more formally. We will find it convenient to represent our constraints using a $m \times m$ 0-1 *constraint matrix* where the (i, j)th entry is 1 if the constraint permits the pair (i, j) and 0 otherwise.

1.1 Model A

In the first model, for each edge we select a random constraint by forbidding each of the m^2 possible pairs of values with probability p_2.

Model A: The underlying graph G is G_{n,p_1} for some $p_1 = p_1(n) < 1$ where $p_1 \neq o(1/n)$. For each edge e of G there is a random $m \times m$ constraint matrix M_e where $M_e(i, j) = 1$ or 0 independently with probability p_2 or $q_2 = 1 - p_2$ respectively, for some constant $0 < p_2 < 1$.

Ruling out the possibility $p_1 = o(1/n)$ is of technical help. We don't mind ruling out this possibility, since when $p_1 = o(1/n)$, the model creates CSP's that are a.s. trivial in the following sense: the graph G_{n,p_1} is very sparse, and consists

[1] We say that a property holds **whp** (with high probability) if its probability tends to 1 as $n \to \infty$.

of a collection of small vertex-disjoint trees in which all but $o(n)$ of the vertices have degree 0.

We define $d = np_1$ to be (approximately) the average number of constraints that a vertex lies in.

Given m, p_2 we wish to determine the satisfiability threshold, i.e. the range of p_1 over which the random model moves from **whp** satisfiable to **whp** unsatisfiable. It is often easy to get an upper bound on this range using a standard first moment analysis. We can do so for Model A, as follows:

Fact: For $p_1 \geq \frac{2 \ln m}{q_2 n}$, the random CSP is unsatisfiable **whp**.

The proof follows easily by noting that the expected number of satisfying solutions is $m^n (1 - p_1 q_2)^{\binom{n}{2}}$.

Inspired by a familiar pattern of similar random models, it is tempting to assume that $\frac{\ln m}{n}$ is the asymptotic order of a satisfiability threshold and so hypothesize that:

Hypothesis A: There is some constant $c > 0$ so that for $p_1 \leq c\frac{\ln m}{n}$, the random CSP is satisfiable **whp**.

See [19] for a lengthy list of papers in which the authors fell to the temptation of assuming an equivalent hypothesis. In [2], it was observed that for most of those papers, and in fact whenever m, p_2 are both constants, the hypothesis is wrong. In fact, if $p_1 \geq \omega(n)/n^2$ for any $\omega(n)$ that tends to infinity with n, then almost surely the random CSP is trivially unsatisfiable in the sense that it has an edge whose constraint forbids every pair of values; we call such an edge a *blocked edge*. This is the "fatal flaw" alluded to earlier.

In [31] it was shown (amongst other things) that Hypothesis A holds when $m = n^\alpha$ for any constant $\alpha > \frac{1}{2}$. Here, we determine, up to a multiplicative factor of $(1 + o(1))$ exactly how high m must be in order for Hypothesis A to hold:

Theorem 1. *(a) For any constant $\epsilon > 0$, if $m \leq (1 - \epsilon)\sqrt{\ln nd / \ln(1/q_2)}$ then provided $nd \to \infty$, the random CSP has a blocked edge* **whp**.

(b) For any constant $\epsilon > 0$, if $m \geq (1 + \epsilon)\sqrt{\ln nd / \ln(1/q_2)}$ then there is some constant $c > 0$ so that for $p_1 \leq c\frac{\ln m}{n}$, the random CSP is satisfiable **whp**. *Furthermore, an assignment can be found in $O(mn)$ time* **whp**.

Note that the "breakpoint" between Cases (a) and (b) occurs when $m = O(\sqrt{\ln n})$. In case (b), Hypothesis A holds, and so $\frac{\ln m}{n}$ is, indeed, the order of the satisfiability threshold. In case (a), **whp** the fact that the random CSP is unsatisfiable can be demonstrated easily by examining a single edge. We show that for $m \geq (\ln n)^{1+\epsilon}$ for any $\epsilon > 0$, this is far from the case. In particular, we show that **whp** there is no short resolution proof of unsatisfiability when p_1 is of the same asymptotic order as the threshold of satisfiability.

Theorem 2. *If $m \geq (\ln n)^{1+\epsilon}$, $d = c \ln m$, for any constants $\epsilon, c > 0$, then* **whp** *the resolution complexity of the random CSP is $2^{\Omega(n/m)}$.*

The resolution complexity of various models of random boolean formula has been well-studied, starting with [11], and continuing through [4],[5],[3] and other papers. This line of inquiry was extended to random models of CSP in [23, 22] and was then continued in [27], where m, the domain-size, was constant. In [32] a similar study was made where $m = n^{\alpha}$.

1.2 Model B

We also consider another model in which every edge receives the same constraint. So that the orientation on the edge does not matter, we insist that the constraint permits (i, j) iff it permits (j, i). Also, we insist that for every i, the constraint forbids (i, i) as otherwise, setting every variable equal to i would yield a trivial satisfying assignment. Let M^* be the set of matrices which correspond to such constraints, i.e. the set of symmetric matrices with all zeroes on the diagonal. As with model A, we restrict our attention to matrices with density some constant p_2 with $0 < p_2 < 1$.

Model B: As with Model A, the underlying graph G is G_{n,p_1} for some $p_1 = p_1(n) < 1$ where $p_1 \neq o(1/n)$. We select a single random $m \times m$ matrix $M \in M^*$ by setting $M(i, j) = M(j, i) = 1$ with probability p_2 independently for each $1 \leq i < j \leq m$, and use $M_e = M$ for every edge.

As with Model A, we set $d = np_1$ to be the average number of constraints that a variable lies in.

Model B has the nice property that, so long as $m \to \infty$, the random matrix is **whp** not all-zeroes and so no edge is blocked. This allows us to prove that we only require $m \to \infty$ in order to get a non-trivial satisfiability threshold. Perhaps surprisingly, that satisfiability threshold is of a higher asymptotic order than in Model A - it is at $d = \Theta(\ln m \ln\ln m)$ rather than $d = \Theta(\ln m)$.

Theorem 3. *Let ϵ be any small positive constant, and consider a random CSP from Model B.*

(a) *If $d \leq (4 - \epsilon)(\ln(1/q_2))^{-1} \ln m \ln\ln m$ then* **whp** *the CSP is satisfiable.*
(b) *If $d \leq (1 - \epsilon)(\ln(1/q_2))^{-1} \ln m \ln\ln m$ then an assignment can be found in polynomial time* **whp**.
(c) *If $0 < q_2 < 1$ is constant and if $d \geq K \ln m \ln\ln m$ for sufficiently large K then* **whp** *the CSP is unsatisfiable.*

It is worth noting that part (c) is analagous to the easy Fact from the previous subsection. However, part (c) is much more difficult to prove.

As with Model A, we can prove high resolution complexity in a restricted range of d, m, p_2.

Theorem 4. *If $m \to \infty$ and $d = c \ln m \ln\ln m$ for some constant $c > 0$, then* **whp** *the resolution complexity of a random CSP from Model B is $2^{\Omega(n/(d^3 m))}$.*

It is interesting to note that studying the satisfiability of a BCSP drawn from Model B is equivalent to the following natural homomorphism problem for random graphs: Consider $n, m \to \infty$, and consider two random graphs $G_1 = G_{n,p_1}$ and $G_2 = G_{m,p_2}$ where $p_1 = d/n$ and $0 < p_2 < 1$ is constant. When is there a homomorphism from G_1 to G_2? Theorem 3 is equivalent to the following:

Theorem 5. *For any $m, n \to \infty$ and any $\epsilon > 0$:*

(a) *If $d \leq (4-\epsilon)(\ln(1/q_2))^{-1} \ln m \ln\ln m$ then* **whp** *there is a homomorphism from G_1 to G_2.*
(b) *If $d \leq (1-\epsilon)(\ln(1/q_2))^{-1} \ln m \ln\ln m$ then such a homomorphism can be found in polytime* **whp**.
(c) *If $0 < q_2 < 1$ is constant and if $d \geq K \ln m \ln\ln m$ for sufficiently large K then* **whp** *there is no homomorphism from G_1 to G_2.*

1.3 Generating Difficult Instances

One of the earliest motivations for the study of random CSP's (see eg. [19]) was the following observation: If one takes a model of random CSP's with a sharp satisfiability threshold and sets the probability parameter to be very close to that threshold, then the resulting CSP will **whp** be very difficult to solve. (See, eg [24] for one of the earliest such studies.) One of the primary traditional motives for proving that models have high resolution complexity is to provide some theoretical support of this observation.

The results in this paper raise the possibility of using Model A as a source for difficult instances. There are a few caveats here: the first is that we have not proven that the satisfiability threshold is sharp. The second is that, as mentioned earlier, Xu and Li[31, 32] have already proven that Model A exhibits a sharp threshold and has high resolution complexity for $m = n^{1/2+\epsilon}$ and for $p_2 \leq \frac{1}{2}$. So even if we had proven Model A to have a sharp threshold for $m = (\ln n)^{1+\epsilon}$, this would only prove that one might generate hard instances using a domain size of roughly $O(\ln n)$ rather than roughly $O(n^{1/2})$. While this improvement is substantial in theory, in practice it is not clear whether it is of much help. Hard instances for complete solvers tend to have very small size (much less than $n = 1000$ variables) and so it is quite possible that this "improvement" would be swallowed up by the other implicit terms in the asymptotics.

2 Some inequalities

We start with the basic Chernoff bounds for the binomial random variable $Bin(N, p)$ viz: Assume that $0 \leq \epsilon \leq 1$.

$$\mathbf{Pr}(Bin(N,p) \leq (1-\epsilon)Np) \leq e^{-\epsilon^2 NP/2}$$
$$\mathbf{Pr}(Bin(N,p) \geq (1+\epsilon)Np) \leq e^{-\epsilon^2 NP/3}$$

In Theorem 6 below we will have a random variable $Z = Z(Y_1, Y_2, \ldots, Y_N)$ where $Y_i \in \Omega_i$ are independent so that Z is defined on $\Omega = \Omega_1 \times \cdots \Omega_N$.

Assumption 1
Suppose that $Y, Y' \in \Omega$ and there exists i such that $Y_j = Y'_j$ for $j \neq i$. Our assumption is that in such a case we have $|Z(Y) - Z(Y')| \leq a$.

Assumption 2
Suppose that, in addition, for any ξ, if $Z(Y) \geq \xi$ then there exist $c(\xi)$ indices $j_1, j_2, \ldots, j_{c(\xi)}$ such that if $Y'_{j_t} = Y_{j_t}$ for $t = 1, 2, \ldots, c(\xi)$ then $Z(Y') \geq \xi$ also.

Let $M = MED(Z)$ denote a *median* of Z i.e. $\mathbf{Pr}(Z \geq M) \geq \frac{1}{2}, \mathbf{Pr}(Z \leq M) \geq \frac{1}{2}$.

Theorem 6. *(***Talagrand's Inequality***) If the random variable Z satisfies Assumptions 1 and 2 then*

$$\mathbf{Pr}(|Z - M| \geq tcM^{1/2}) \leq 2e^{-t^2/(4a^2)}. \tag{1}$$

Proofs of these inequalities can be found, for example, in Janson, Łuczak and Ruciński [20].

3 Model A: Unsatisfiable Region

Let an edge $e = (x, y)$ of G be *blocked* if $M_e = \mathbf{O}$ (the matrix with all zero entries). Of course, any CSP with a blocked edge is unsatisfiable, since there is no possible consistent assignment to x, y. We start with a simple lemma which immediately implies Theorem 1(a):

Lemma 1. *Let $\epsilon > 0$ be a small positive constant and assume that $nd \to \infty$ (so that* **whp** *G has edges). Let $m_0 = \sqrt{(\ln n + \ln d)/\ln(1/q_2)}$. Then*

(a) *$m \geq (1+\epsilon)m_0$ implies that there are no blocked edges,* **whp***.*
(b) *$m \leq (1-\epsilon)m_0$ implies that there are blocked edges,* **whp***.*

Proof Let Z be the number of blocked edges in our instance. Given the graph G, the distribution of Z is $Bin(|E|, q_2^{m^2})$.

$$\mathbf{E}(Z) = \binom{n}{2} p_1 q_2^{m^2} \tag{2}$$

If $m \geq (1+\epsilon)m_0$ then (2) implies that

$$\mathbf{E}(Z) \leq (nd)^{-\epsilon} \to 0$$

and then $Z = 0$ **whp** and (a) follows.

If $m \leq (1-\epsilon)m_0$ then (2) implies that

$$\mathbf{E}(Z) \geq \frac{1}{3}(nd)^{\epsilon} \to \infty.$$

Part (b) now follows from the Chernoff bounds. □

We now consider another simple cause of unsatisfiability that [2] also discovered to be prevalent amongst the models commonly used for experimentation. We say that a vertex (variable) x is *blocked* if for every possible assignment $i \in [m]$ there is some neighbour y which blocks the assignment of i to x, because the ith row of M_e, $e = (x, y)$ is all zero.

Lemma 2. *Let ϵ be a small positive constant, and suppose that $m - \sqrt{\ln n/\ln(1/q_2)} \to \infty$. Then*

(a) $m \geq (1+\epsilon)\sqrt{(\ln n + m\ln d)/\ln(1/q_2)}$ *implies that there are no blocked vertices,* **whp**.

(b) $m \leq (1-\epsilon)\sqrt{(\ln n + m\ln d)/\ln(1/q_2)}$ *implies that there are blocked vertices,* **whp**.

Remark: Note that $m = \sqrt{(\ln n + m\ln d)/\ln(1/q_2)}$, for m slightly smaller than m_0 from Lemma 1.

Proof If the graph G is given and vertex v has degree d_v then

$$\mathbf{Pr}(v \text{ is blocked} \mid G) = (1-(1-q_2^m)^{d_v})^m.$$

This is because for $i \in [m]$, $(1-q_2^m)^{d_v}$ is the probability that no neighbour w of v is such that row i of $M_{(v,w)}$ is all zero. Part (a) now follows from an easy first moment calculation, which we omit.

We turn our attention to proving part (b). Rearranging our assumption yields $\ln d \geq (1-\epsilon)^{-2} m\ln(1/q_2) - \frac{1}{m}\ln n \geq (1-\epsilon)^{-1}(m\ln(1/q_2) - \frac{1}{m}\ln n)$. So we choose d such that $\ln d = (1-\epsilon)^{-1}(m\ln(1/q_2) - \frac{1}{m}\ln n)$, i.e. $d = (q_2^{-m^2}/n)^{1/(m(1-\epsilon))}$ as proving the result for that value of d clearly implies that it holds for all larger values.

Our assumption implies that $d \to \infty$ and so **whp** $n - o(n)$ vertices v have $d_v \in I = [(1-\epsilon)d, (1+\epsilon)d]$. Thus if Z is the number of blocked vertices with $d_v \in I$ then

$$\begin{aligned}
\mathbf{E}(Z) &\geq (n-o(n))(1-(1-q_2^m)^{d(1-\epsilon)})^m \geq (n-o(n))(d(1-\epsilon)q_2^m)^m \\
&\geq (1-o(1))\left(q_2^{-m^2} n\right)^{\epsilon/(1-\epsilon)} (1-\epsilon)^m \\
&\geq (1-o(1))n^{\epsilon/(1-\epsilon)}(1-\epsilon)^{m_0} \qquad \text{(see the Remark preceding this proof)} \\
&\geq n^{\epsilon/2} \to \infty.
\end{aligned}$$

To show that $Z \neq 0$ **whp** we use Talagrand's inequality (1). We condition on G. Then we let each $\Omega_e, e \in E$ be an independent copy of $\{0,1\}^{m^2}$ (the set of $m \times m$ 0-1 matrices). Now changing a single M_e can change z by at most 2 and so Assumption 1 holds with $a = 2$. Then to show that a vertex v is blocked we only have to expose M_e for e incident with v. Thus Assumption 2 holds with $c(\xi) = (1+\epsilon)d\xi$. Thus if $M = Med(Z)$, ((1) gives

$$\mathbf{Pr}(|Z - M| \geq t(1+\epsilon)dM^{1/2}) \leq 2e^{-t^2/16} \tag{3}$$

for any $t > 0$.

Our assumptions imply that $d^2 = o(\mathbf{E}(Z))$ and so (3) implies the result. □

4 Model A: Satisfiable Region

In this section, we prove Theorem 1(b).

So for this section we assume the hypotheses of Theorem 1(b), in particular that:

$$m = (1+\epsilon)\left(\frac{\ln n}{\ln q_2^{-1}}\right)^{1/2}, \; d = c \ln m \text{ and } p_2 \text{ is constant}$$

where c, ϵ are small. (Note that this also implies the result for larger m).

Now let a vertex v be *troublesome* if it has degree $\geq D = 10d$ or there are assignments to its neighbours which leave v without a consistent assignment. Let $\mathcal{T}$ denote the set of troublesome vertices. A subgraph is called troublesome if all of its vertices are troublesome.

Let $\mathcal{A}$ be the event that every set of k_0 vertices contains at most k_0 edges where

$$k_0 = \left\lceil \frac{2\ln n}{d} \right\rceil.$$

Lemma 3.

$$\mathbf{Pr}(\mathcal{A}) = 1 - o(1).$$

Proof

$$\mathbf{Pr}(\overline{\mathcal{A}}) \leq \binom{n}{k_0}\binom{\binom{k_0}{2}}{k_0+1}\left(\frac{d}{n}\right)^{k_0+1} \leq \left(\frac{ne}{k_0}\right)^{k_0} \cdot \left(\frac{d}{n}\right)^{k_0+1} \cdot \left(\frac{k_0 e}{2}\right)^{k_0+1}$$
$$= \frac{k_0 e^{2k_0+1} d^{k_0+1}}{2^{k_0+1}} \cdot \frac{d}{n} = o(1).$$

□

We show next that **whp** the sub-graph induced by $\mathcal{T}$ has no large trees.

Lemma 4. Whp *there are no troublesome trees with* $\geq k_0$ *vertices.*

Proof If $\mathcal{T}$ contains a tree of size greater than k_0 then it contains one of size k_0. Let Z be the number of troublesome trees with k_0 vertices. Let Ω be the set of trees/unicyclic graphs spanning $[k_0]$. For each $T \in \mathcal{T}$ we define $\mathcal{G}_T$ to be the event that the subgraph of G induced by$[k_0]$ is T. Then for any subset J of $[k_0]$ we may write

$$\mathbf{E}(Z \cdot 1_{\mathcal{A}}) \leq \binom{n}{k_0} \sum_{T\in\Omega} \left(\frac{d}{n}\right)^{k_0-1} \prod_{i\in J} \mathbf{Pr}(x_i \in \mathcal{T} \mid \mathcal{G}_T \wedge (x_j \in \mathcal{T},\, \forall j \in J, j < i)). \tag{4}$$

Fix $T \in \Omega$ and let I_1 be the set of vertices of T with degree at most 4 in T. Then $|I_1| \geq k_0/2$. Note next that I_1 contains an independent set I of size at least $k_0/10$.

Now if $i \in I$ then

$$\mathbf{Pr}(x_i \in \mathcal{T} \mid \mathcal{G}_T \wedge (x_j \in \mathcal{T},\, \forall j \in I, j < i)) \leq \binom{n}{D-4}\left(\frac{d}{n}\right)^{D-4} + \sum_{t=0}^{D} m^t (1-p_2^t)^m.$$

The first term bounds the probability that x_i has at least $D-4$ neighbours outside the tree and assuming the degree of x_i is at most D, the second term bounds the probability that the $\le D$ neighbours have an assignment which can not be extended to x_i. We use the fact that I is an independent set to gain the stochastic independence we need.

Thus, applying (4) with $J = I$ we obtain

$$\begin{aligned}
&\mathbf{E}(Z \cdot 1_{\mathcal{A}}) \\
&\le \binom{n}{k_0} k_0^{k_0-2} k_0^2 \left(\frac{d}{n}\right)^{k_0-1} \left(\binom{n}{D-4}\left(\frac{d}{n}\right)^{D-4} + \sum_{t=0}^{D} m^t (1-p_2^t)^m\right)^{k_0/10} \qquad (5)\\
&\le n(de)^{k_0} \left(\left(\frac{de}{D-4}\right)^{D-4} + Dm^D e^{-mp_2^D}\right)^{k_0/10} \qquad = o(1).
\end{aligned}$$

□

Now we deal with troublesome cycles in a similar manner.

Lemma 5. Whp *there are no troublesome cycles.*

Proof It follows from Lemma 4 that we need only consider cycles of length less than k_0, since a cycle on at least k_0 vertices contains a tree on at least k_0 vertices. If Z now denotes the number of troublesome cycles of length less than k_0 then arguing as in (4), (5) we see that

$$\begin{aligned}
&\mathbf{E}(Z) \le \\
&\sum_{k=3}^{k_0-1} \binom{n}{k} \frac{(k-1)!}{2} \left(\frac{d}{n}\right)^k \left(\binom{n}{D-2}\left(\frac{d}{n}\right)^{D-2} + \sum_{t=0}^{D} m^t (1-p_2^t)^m\right)^{\lfloor k/2 \rfloor} \\
&\qquad = o(1).
\end{aligned}$$

□

Let a tree be *small* if it contains less than k_0 vertices. We have therefore shown that **whp** the troublesome vertices $\mathcal{T}$ induce a forest of small trees. We show next that **whp** there at most $n^{1+o(1)}$ small trees.

Lemma 6. Whp *there are at most $n^{1+o(1)}$ small trees.*

Proof Let σ_T denote the number of small trees. Then

$$\mathbf{E}(\sigma_T) = \sum_{k=1}^{k_0-1} \binom{n}{k} k^{k-2} \left(\frac{d}{n}\right)^{k-1} \le \sum_{k=1}^{k_0-1} n(de)^k = n^{1+o(1)}.$$

The result now follows from the Markov inequality. □

Our method of finding an assignment to our CSP is to (i) make a consistent assignment to the vertices of $\mathcal{T}$ first and then (ii) extend this assignment "greedily" to the non-troublesome vertices.

It is clear from the definition of troublesome that it is possible to carry out Step (ii). We wish to show that (i) can be carried out successfully **whp**. For this purpose we show that **whp** G does not contain a small tree which cannot be given a consistent assignment.

So we fix a small tree T and a vertex $v \in T$ and root T at v. Let $L < k_0$ denote the depth of T and let $X_i, 0 \leq i \leq L$ denote the vertices at level i, where level L is the root and level 0 is the lowest level. Let d_ℓ be the maximum number of descendants of a vertex in X_ℓ.

For $u \in X_\ell$ let $S(u)$ be the the set of values δ such that there is a consistent assignment to the sub-tree of T rooted at u in which u receives δ. We let $t = \lceil 10/\epsilon \rceil$ and define the events

$$\mathcal{B}_{u,i} = \left\{ \frac{(i-1)m}{t} \leq |S(u)| \leq \frac{im}{t} \right\}.$$

Then for $1 \leq i \leq t$ and $0 \leq \ell \leq L$, let

$$\pi_{i,\ell} = \max_{u \in X_\ell} \mathbf{Pr}\left(\bigcup_{j=1}^{i} \mathcal{B}_{u,j} | \overline{\mathcal{B}_{w,1}} \text{ for every descendent } w \text{ of } u \right).$$

In other words, $\pi_{i,\ell}$ is the maximum over all $u \in X_\ell$ of the probability that $|S(u)| \leq \frac{im}{t}$ conditional on the event that $\frac{(i-1)m}{t} \leq |S(w)| > \frac{m}{t}$ for every descendent w of u. Note that $\pi_{t,\ell} = 1$ and that $\pi_{i,0} = 0$ for all $i < \ell$.

We will prove by induction on ℓ that for $\eta = \epsilon/3$ and for $1 \leq i \leq t$ we have

$$\pi_{i,\ell} \leq t^\ell n^{-(1+\eta)\frac{t-i}{t}}. \tag{6}$$

In particular, this implies that $\pi_{1,\ell} \leq t^\ell n^{-(1+\eta)(t-1)/t} < \frac{1}{2}$. The probability that there is no consistent assignment for T is clearly at most the probability that $\mathcal{B}_{u,1}$ holds for at least one $u \in T$ which is at most

$$|T| \times t^L n^{-(1+\eta)(t-1)/t} < k_0 t^{k_0} n^{-(1+\eta)(t-1)/t}.$$

Therefore

$$\begin{aligned}&\mathbf{Pr}(\exists \text{a troublesome tree which cannot be consistently assigned})\\ &\qquad\qquad\leq o(1) + n^{1+o(1)} k_0 t^{k_0} n^{-(1+\eta)(t-1)/t} = o(1)\end{aligned}$$

which implies that Step (i) can be completed **whp**.

(6) is clearly true for the base case of $\ell = 0$ since $\pi_{j,0} = 0$ for $j < t$ and $\pi_{t,0} = 1$. For $\ell > 0$, note that for each child w of u, the conditional probability

that $|S(w)| \leq \frac{im}{t}$ is at most $\pi_{i,\ell-1}/(1-\pi_{1,\ell-1}) < 2\pi_{i,\ell-1}$. Thus, we have:

$$\pi_{i,\ell} \leq \sum_{k_2+\cdots+k_t=d_\ell} \binom{d_\ell}{k_2,...,k_t} \prod_{j=2}^{t} \left(\frac{\pi_{j,\ell-1}}{1-\pi_{1,\ell-1}}\right)^{k_j} \binom{m}{\frac{t-i}{t}m} \left(1-\prod_{j=2}^{t}(1-q_2^{\frac{m(j-1)}{t}})^{k_j}\right)^{\frac{t-i}{t}m} \tag{7}$$

$$\leq 2^m \sum_{k_2+\cdots+k_t=d_\ell} \binom{d_\ell}{k_2,...,k_t} \prod_{j=2}^{t} \left(\frac{\pi_{j,\ell-1}}{1-\pi_{1,\ell-1}}\right)^{k_j} \left(\sum_{j=2}^{t} k_j q_2^{\frac{m(j-1)}{t}}\right)^{\frac{t-i}{t}m}$$

$$\leq 2^m \sum_{j=2}^{t} \frac{\pi_{j,\ell-1}}{1-\pi_{1,\ell-1}} (d_\ell q_2^{\frac{j-1}{t}m})^{\frac{t-i}{t}m} \sum_{k_2+\cdots+k_t=d_\ell} \binom{d_\ell}{k_2,...,k_t}$$

$$\leq t^{d_\ell} 2^{m+1} \sum_{j=2}^{t} (d_\ell q_2^{\frac{j-1}{t}m})^{\frac{t-i}{t}m} \pi_{j,\ell-1}. \tag{8}$$

Explanation of (7)**:** Suppose that there are k_j descendants w of u for which $\mathcal{B}_{w,j}$ occurs. If $u \in \mathcal{B}_{u,i}$ then r assignment values will be forbidden to it, $\frac{t-i}{t}m \leq r \leq \frac{t-i+1}{t}m$. The product bounds the probability that these values are forbidden and that $\mathcal{B}_{w,j}$ occurs for the corresponding descendants.

Then applying (6) inductively to (8) and recalling that $m^2 = (1+\epsilon)^2 \ln n / \ln(q_2^{-1})$ we obtain

$$\pi_{i,\ell} \leq \sum_{j=2}^{t} t^{d_\ell} 2^{m+1} d_\ell^{\frac{t-i}{t}m} q_2^{\frac{(j-1)(t-i)}{t^2}m^2} t^{\ell-1} n^{-(1+\eta)\frac{t-j}{t}}$$

$$\leq t^{\ell-1} \sum_{j=2}^{t} n^{-\frac{(j-1)(t-i)}{t^2}(1+\epsilon)-\frac{t-j}{t}(1+\eta)}.$$

In going from the first to second inequality we use the fact that since $\ell, d_\ell \leq k_0$ we find that $2^{m+1} t^{d_\ell} d_\ell^{\frac{t-i}{t}m} = n^{o(1)}$. This term is then absorbed by using $1+\epsilon$ in place of $(1+\epsilon)^2$.

Now consider the expression

$$\varDelta = \frac{(j-1)(t-i)}{t^2}(1+\epsilon) + \frac{t-j}{t}(1+\eta) - \frac{t-i}{t}(1+\eta)$$

$$= \frac{(j-1)(t-i)}{t^2}(1+\epsilon) + \frac{i-j}{t}(1+\eta).$$

To complete the inductive proof of (6) we have only to show that $\varDelta$ is non-negative.

Now $\varDelta$ is clearly non-negative if $i \geq j$ and so assume that $j > i$. Now for a fixed j, $\varDelta$ can be thought of as a linear function of i and so we need only check non-negativity for $i = 1$ or $i = j-1$.

For $i = 1$ we need

$$(j-1)(t-1)(1+\epsilon) \geq (j-1)t(1+\eta) \tag{9}$$

and this holds for $\epsilon \leq 1$.

For $i = j - 1$ we need

$$(j-1)(t-j+1)(1+\epsilon) \geq t(1+\eta).$$

But here $j \geq 2$ and the LHS is at least $(t-1)(1+\epsilon)$ and the inequality reduces to (9) (after dividing through by $j-1$). This competes the proof of (6), and thus proves that satisfiability claim in Theorem 1(b).

It only remains to discuss the time to find an assignment. Once we have assigned values to $\mathcal{T}$ then we can fill in an assignment in $O(mn)$ time. So let us now fix a small tree T of troublesome vertices. Choose a root $v \in T$ arbitrarily. Starting at the lowest levels we compute the set of values $S_\ell(u)$ available to a vertex $u \in X_\ell$. For each descendant w of u we compute $T_\ell(w) = \{a \in S_{\ell+1}(w) : M_{(u,w)}(a) = 1\}$ and then we have $S_\ell(u) = \bigcap_w T_\ell(w)$. At the leaves, $S_L = [m]$ and so in this way we can assign a value to the root and then work back down the tree to the leaves giving an assignment to the whole of T. Thus the whole algorithm takes $O(mn)$ time as claimed.

This concludes the proof of Theorem 1(b). $\square$

5 Model A: Resolution complexity

In this section, we prove Theorem 2.

For a boolean CNF-formula F, a *resolution refutation* of F with length r is a sequence of clauses $C_1, ..., C_r = \emptyset$ such that each C_i is either a clause of F, or is derived from two earlier clauses $C_j, C_{j'}$ for $j, j' < i$ by the following rule: $C_j = (A \vee x), C_{j'} = (B \vee \overline{x})$ and $C_i = (A \vee B)$, for some variable x. The *resolution complexity* of F, denoted $\mathbf{RES}(F)$, is the length of the shortest resolution refutation of F. (If F is satisfiable then $\mathbf{RES}(F) = \infty$.)

Mitchell[23] discusses two natural ways to extend the notion of resolution complexity to the setting of a CSP. These two measures of resolution complexity are denoted $\mathbf{C} - \mathbf{RES}$ and $\mathbf{NG} - \mathbf{RES}$. Here, our focus will be on the $\mathbf{C} - \mathbf{RES}$ measure, as it was in [22] and in [27].

Given an instance $\mathcal{I}$ of a CSP in which every variable has domain $\{1, ..., m\}$, we construct a boolean CNF-formula CNF($\mathcal{I}$) as follows. For each variable x of $\mathcal{I}$, there are m variables in CNF($\mathcal{I}$), denoted $x:1, x:2, ..., x:m$, and there is a *domain clause* $(x : 1 \vee ... \vee x : m)$. For each pair of variables x, y and each *restriction* (i, j) such that $M_{(x,y)}(i, j) = 0$, CNF($\mathcal{I}$) has a *conflict clause* $(\overline{x : i} \vee \overline{y : j})$. We also add $\binom{m}{2}$ 2-clauses for each x which specify that $x : i$ can be true for at most one value of i. It is easy to see that CNF($\mathcal{I}$) has a satisfying assignment iff $\mathcal{I}$ does. We define the resolution complexity of $\mathcal{I}$, denoted $\mathbf{C} - \mathbf{RES}(\mathcal{I})$ to be equal to $\mathbf{RES}(\text{CNF}(\mathcal{I}))$.

A variable x is *free* if any assignment which satisfies $\mathcal{I} - x$ can be extended to a satisfying assignment of $\mathcal{I}$. The *boundary* $\mathcal{B}(\mathcal{I})$ is the set of *free* variables. We extend a key result from [23] to the case where m grows with n:

Lemma 7. *Suppose that there exist $s, \zeta > 0$ such that*

(a) Every subproblem on at most s variables is satisfiable, and
(b) Every subproblem $\mathcal{I}'$ on v variables where $\frac{1}{2}s \leq v \leq s$ has $|\mathcal{B}(\mathcal{I}')| \geq \zeta n$.

then $\mathbf{C}-\mathbf{RES}(\mathcal{I}) \geq 2^{\Omega(\zeta^2 n/m)}$.

The proof is a straightforward adaptation of the proof of the corresponding work in [23] and so we omit it.

We assume now the hypotheses of Theorem 2, in particular that ϵ is a small positive constant and

$$m \geq (\ln n)^{1+\epsilon},\ d = c \ln m \text{ and } p_2 \text{ is constant.} \tag{10}$$

Let γ be a sufficiently small constant. Let $\mathcal{T}_1$ denote the set of vertices v for which there are γd neighbours W and a set of assignments of values to W for which v has no consistent assignment.

Lemma 8.

$$\mathbf{Pr}(\mathcal{T}_1 \neq \emptyset) = o(1).$$

Proof

$$\begin{aligned}
\mathbf{E}(|\mathcal{T}_1|) &\leq n \sum_{t=\gamma d}^{n-1} \binom{n}{t} \left(\frac{d}{n}\right)^t \binom{t}{\gamma d} m^{\gamma d} (1-p_2^{\gamma d})^m \\
&\leq n \sum_{t=\gamma d}^{n-1} \left(\frac{de}{t}\right)^t \left(\frac{tem}{\gamma d}\right)^{\gamma d} e^{-mp_2^{\gamma d}} \\
&\leq ne^{-m^{1-\epsilon/2}} \left(\sum_{t=\gamma d}^{10d} (de)^{10d} (10e\gamma^{-1}m)^{\gamma d} + \sum_{10d}^{n-1} (mn)^{\gamma d} \right) = o(1).
\end{aligned}$$

□

Now we show that **whp** every set of $s \leq s_0 = \alpha n$ vertices, $\alpha = \gamma/3$ has less than $\gamma ds/2$ edges. Let $\mathcal{B}$ denote this event.

Lemma 9.

$$\mathbf{Pr}(\mathcal{B}) = 1 - o(1).$$

Proof

$$\mathbf{Pr}(\overline{\mathcal{B}}) \leq \sum_{s=\gamma d}^{\alpha n} \binom{n}{s} \binom{\binom{s_0}{2}}{\gamma ds/2} \left(\frac{d}{n}\right)^{\gamma ds/2} \leq \sum_{s=\gamma d}^{\alpha n} \left(\left(\frac{se}{\gamma n}\right)^{-1+\gamma d/2} \cdot \frac{e^2}{\gamma} \right)^s = o(1).$$

□

Let us now check the conditions of Lemma 7. Condition (a) holds because Lemma 9 implies that if $s = |S| \leq \alpha n$ then we can order S as $v_1, v_2, \ldots, v_s$ so that v_j has less than αd neighbours among $v_1, v_2, \ldots, v_{j-1}$ for $1 \leq j \leq s$. Because we can assume that $\mathcal{T}_1 = \emptyset$ (Lemma 8) we see that it will be possible to sequentially assign values to $v_1, v_2, \ldots, v_s$ in order. Lemma 9 implies that at least $\frac{1}{2}$ the vertices of S have degree $\leq \alpha d$ in S and now $\mathcal{T}_1 = \emptyset$ implies that (b) holds with $\zeta = 1/2$.

We conclude that with the parameters as stated in (10), $\mathbf{C}-\mathbf{RES}(\mathcal{I})$ is **whp** as large as is claimed by Theorem 2.

6 Model B: Satisfiable Region

We have a blocked edge iff $M = \mathbf{O}$ and this happens with probability $q_2^{m(m-1)}$ and so there is not much more to say on this point.

Secondly, if $M \neq \mathbf{O}$ then there are two values x, y which can be assigned to adjacent vertices. This implies that for any bipartite subgraph H of G there is a satisfying assignment for H just using x, y. So, in particular there will be no blocked vertices.

Proof of Theorem 3(a,b) Let H be the graph defined by treating M as its adjacency matrix. Thus $H = G_{m,p_2}$. As such it has a clique I of size $(2 - o(1)) \ln m/(\ln 1/q_2)$ **whp**.

If we can properly colour G with I (i.e. give adjacent vertices different values in I) then we will have a satisfying assignment for our CSP. Now the chromatic number of G is $(1 + o(1))d/(2 \ln d)$ **whp**. So the CSP is satisfiable **whp** if

$$(2 - o(1)) \ln m/(\ln 1/q_2) \geq (1 + o(1))d/(2 \ln d)$$

and this holds under assumption (a).

For (b) we observe that we can find a clique of size $(1-o(1)) \ln m/(\ln 1/q_2)$ in polynomial time and we can colour G with $(1+o(1))d/\ln d$ colours in polynomial time. □

7 Model B: Unsatisfiable Region

In this section, we prove Theorem 3(c). We first observe

Lemma 10. *There exists a constant ϵ_0 such that for $\epsilon \leq \epsilon_0$ there exist $R_0 = R_0(\epsilon), Q_0 = Q_0(\epsilon)$ such that if $Q \geq Q_0, R \geq R_0$ and $s_0 = R \ln m$ then*

(a) **whp** *every pair of disjoint sets $S_1, S_2 \subseteq [m]$, $|S_1| = s_1 \geq s_0, |S_2| = s_2 \geq s_0$ contains at most $(1 - \epsilon)s_1 s_2$ edges of H between S_1 and S_2;*
(b) **whp** *every $S \subseteq [m]$, $|S| = s \geq s_0$ contains at most $Q \ln m$ members with degree greater than $(1 - \epsilon)s$ in the subgraph of H induced by S.*

Proof
(a) We can bound the probability that there are sets S_1, S_2 with more than the stated number of edges between S_1 and S_2 by

$$\sum_{s_1=s_0}^{m} \sum_{s_2=s_0}^{m} \binom{m}{s_1} \binom{m}{s_2} \binom{s_1 s_2}{\epsilon s_1 s_2} p_2^{(1-\epsilon)s_1 s_2}$$
$$\leq \sum_{s_1=s_0}^{m} \sum_{s_2=s_0}^{m} \left(\frac{me}{s_1}\right)^{s_1} \left(\frac{me}{s_2}\right)^{s_2} \left(\left(\frac{e}{\epsilon}\right)^{\epsilon} p_2^{1-\epsilon}\right)^{s_1 s_2} = o(1).$$

(b) We choose $\epsilon > 0$ so that $p_2 < 1 - 3\epsilon$. Given S, we consider a set $L \subset S$ of size $Q \ln m$. For $R > Q\epsilon^{-1}$ we have $|L| < \epsilon|S|$ and so if each $i \in L$ has at least $(1-\epsilon)s$ neighbours in S then it has at least $(1-2\epsilon)s$ neighbours in $S-L$. By the Chernoff bound, this occurs with probability at most $\left(e^{-\zeta s}\right)^{|L|}$, for some $\zeta > 0$ and this is less than m^{-2s} for Q sufficiently high. Therefore, the expected number of S, L violating part (b) is at most

$$\sum_{s=s_0}^{m} \binom{m}{s} \binom{s}{Q \ln m} m^{-2s} < \sum_{s=s_0}^{m} \left(\frac{em}{s}\right)^s 2^s m^{-2s} < \sum_{s \geq s_0} m^{-s} = o(1).$$

□

Proof of Theorem 3(c) Consider an assignment σ for our CSP and let N_i be the set of variables that are assigned the value i by σ. We observe that if σ is consistent then each N_i is an independent set in G and so **whp** G is such that we must have

$$|N_i| \leq \frac{3n \ln d}{d} < \frac{4n}{K \ln m} \qquad \text{for } i = 1, 2, \ldots, m. \tag{11}$$

Thus, we will restrict our attention to assignments which satisfy (11). We will prove that the expected number of such assignments that are consistent is $o(1)$, thus proving part (c) of Theorem 3.

We say that a pair of vertices is *forbidden* by σ if that pair cannot form an edge of G without violating σ. Note that every pair in the same set N_i is forbidden, and a pair in $N_i \times N_j$ is forbidden iff ij is not an edge of H. We will show that the number of forbidden pairs is at least $n^2/\ln\ln m$. It follows that

$$\mathbf{Pr}(\sigma \text{ is consistent}) \leq (1-p_1)^{n^2/\ln\ln m} \leq e^{-nd/\ln\ln m} = o(m^{-n}),$$

assuming that $d \geq K \ln m \ln\ln m$ for sufficiently large K. Since this probability is $o(m^{-n})$ we can multiply by m^n, which is an overcount of the number of assignments satisfying (11), and so obtain the desired first moment bound.

Let $n_i = |N_i|$ and let $I = \{i : n_i \geq n/(2m)\}$. Now

$$\sum_{i \in I} n_i = n - \sum_{i \notin I} n_i \geq n - m \cdot \frac{n}{2m} = \frac{n}{2}. \tag{12}$$

For the following analysis we choose constants:

$$\epsilon, \qquad Q = \max\{Q_0, 100\epsilon^{-1}\}, \qquad K_1 = 100R_0, \qquad K = 100K_1Q$$

where $\epsilon \leq \epsilon_0, Q_0, R_0$ are from Lemma 10.

We partition I into 3 parts:

- $I_1 = \{i : n/(K_1 \ln m \ln\ln m) \leq n_i < 4n/K \ln m\}$
- $I_2 = \{i : n/(K_1 \ln m)^2 \leq n_i < n/(K_1 \ln m \ln\ln m)\}$
- $I_3 = \{i : n/(2m) \leq n_i < n/(K_1 \ln m)^2\}$

Case 1: $\sum_{i \in I_1} n_i \geq \frac{n}{6}$ Let H_1 be the subgraph of H induced by I_1, and for each $i \in I_1$, we let $\overline{d}(i)$ be the degree of i in $\overline{H_1}$. Note that the total number of forbidden pairs of vertices for G is at least

$$\frac{1}{2} \sum_{i \in I_1} \overline{d}(i) n_i \times \frac{n}{K_1 \ln m \ln\ln m}, \tag{13}$$

since for all $i' \in I_1, n_{i'} \geq n/(K_1 \ln m \ln\ln m)$.

By (11), we have $|I_1| \geq (K \ln m)/24$, so $(K \ln m)/Q < \epsilon |I_1|$. Thus, by Lemma 10(b) then there are at most $Q \ln m$ members $i \in I_1$ with $\underline{d}(i) < (K \ln m)/Q$. Again using (11), these members contribute at most $4Qn/K < n/12$ to $\sum_{i \in I_1} n_i$. Therefore, the sum in (13) is at least

$$\frac{1}{2} \times \frac{K \ln m}{Q} \times \frac{n}{12} \times \frac{n}{K_1 \ln m \ln\ln m} \geq \frac{n^2}{\ln\ln m}.$$

Case 2: $\sum_{i \in I_2} n_i \geq \frac{n}{6}$ We let $I(j) = \{i \in I_2 :\ n/2^j \leq n_i \leq n/2^{j-1}\}$, for

$\log_2(K_1 \ln m \ln\ln m) \leq j \leq 2\log_2(K_1 \ln m)$. We set $t_j = \sum_{i \in I(j)} n_i$ and $s_j = |I(j)| \geq t_j \times (K_1 \ln m \ln\ln m/n)$. We set $J = \{j : t_j \geq n/(100 \ln\ln m)\}$ and note that $s_j \geq s_0$ (from Lemma 10) for each $j \in J$. Note also that

$$\sum_{j \in J} t_j \geq \frac{n}{6} - 2\log_2(K_1 \ln m) \times \frac{n}{100 \ln\ln m} \geq \frac{n}{8}.$$

Consider $I(j)$ for any $j \in J$. By Lemma 10, there are at least $\epsilon\binom{s_j}{2}$ pairs $i, i' \in I(j)$ such that every pair of vertices in $N_i \times N_{i'}$ is forbidden. Also, for any i, every pair in $N_i \times N_i$ is forbidden. Since the sizes of the sets $N_i, i \in I(j)$ differ by at most a factor of 2, this implies that the number of forbidden pairs in $\cup_{i \in I(j)} N_i$ is at least $\frac{\epsilon}{8} t_j^2$. Now consider any pair $I(j), I(j')$ with $j, j' \in J$. By Lemma 10(a), there are at least $\epsilon s_j s_{j'}$ pairs $i \in I(j), i' \in I(j')$ such that every pair of vertices in $N_i \times N_{i'}$ is forbidden, and this implies that the number of forbidden pairs in $\cup_{i \in I(j)} N_i \times \cup_{i \in I(j')} N_i$ is at least $\frac{\epsilon}{4} t_j t_{j'}$. Thus, the total number of forbidden pairs is at least

$$\frac{\epsilon}{8}\left(\sum_{j \in J} t_j^2 + \sum_{j, j' \in J; j < j'} 2 t_j t_{j'}\right) = \frac{\epsilon}{8}\left(\sum_{j \in J} t_j\right)^2 \geq \frac{\epsilon n^2}{8^3} > \frac{n^2}{\ln\ln m}.$$

Case 3: $\sum_{i \in I_3} n_i \geq \frac{n}{6}$**.** Here we follow essentially the same argument as in Case 2. Again, let $I(j) = \{i \in I :\ n/2^j \leq n_i \leq n/2^{j-1}\}$, but this time we consider $2\log_2(K_1 \ln m) < j \leq \log_2(2m)$. Again, $t_j = \sum_{i \in I(j)} n_i$ and $s_j = |I(j)|$, but note that this time we have

$$s_j \geq \frac{t_j}{n/(K_1 \ln m)^2}.$$

Here, we set $J = \{j : t_j \geq n/K_1 \ln m\}$ and so again we have $s_j \geq s_0$ for every $j \in J$.

$$\sum_{j \in J} t_j \geq \frac{n}{4} - \log_2(2m) \times \frac{n}{K_1 \ln m} \geq \frac{n}{8}.$$

The same argument as in Case 2 now goes through to imply that the total number of forbidden pairs is at least

$$\frac{\epsilon}{8}\left(\sum_{j \in J} t_j\right)^2 > \frac{n^2}{\ln\ln m}.$$

$\square$

8 Model B: Resolution complexity

Proof of Theorem 5 First note that **whp** every set of 10 vertices in H has a common neighbour, since the probability of at least one such set not having a common neighbour is less than $\binom{m}{10} q_2^{m-10} = o(1)$. Assuming that H has this property, every vertex of degree at most 10 in G will be in the boundary.

A straightforward first moment argument shows that a.s. every subgraph G' of G with at most $n/d^{3/2}$ vertices has at most $5|G'|$ edges. (We omit the standard calculation.) Therefore, every such G' has at least $|G'|/11$ vertices of degree at most 10. This implies both conditions of Lemma 7 with $s = n/d^{3/2}$ and $\zeta = 1/(22d^{3/2})$ and thus implies Theorem 4. $\square$

We remark that the exponent "3" of d in the statement of Theorem 4 can be replaced by values arbitrarily close to 2 by replacing "10" with a larger value in this proof.

References

1. D. Achlioptas, P. Beame and M. Molloy. *A sharp threshold in proof complexity.* Proceedings of STOC 2001, 337 - 346.
2. D. Achlioptas, L. Kirousis, E. Kranakis, D. Krizanc, M. Molloy, and Y. Stamatiou. *Random constraint satisfaction: a more accurate picture.* Constraints **6**, 329 - 324 (2001). Conference version in Proceedings of CP 97, 107 - 120.
3. P. Beame, J. Culberson and D. Mitchell. *The resolution complexity of random graph k-colourability.* In preparation.
4. P. Beame and T. Pitassi. *Simplified and improved resolution lower bounds.* Proceedings of FOCS 1996, 274 - 282.
5. P. Beame, R. Karp, T. Pitassi and M. Saks. *The efficiency of resolution and Davis-Putnam procedures.* Proceedings of STOC 1998 and SIAM Journal on Computing, **31**, 1048 - 1075 (2002).
6. E. Ben-Sasson and A. Wigderson. *Short proofs are narrow - resolution made simple.* Proceedings of STOC 1999 and Journal of the ACM **48** (2001)
7. B. Bollobás, Random graphs, Second Edition, Cabridge University Press, 2001.
8. B. Bollobás, *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs*, European Journal on Combinatorics **1** (1980) 311–316.
9. E. A. Bender and E. R. Canfield, *The asymptotic number of labelled graphs with given degree sequence*, Journal of Combinatorial Theory (A) **24** (1978) 296–307.
10. D. G. Bobrow and M. Brady, eds., Special Volume on Frontiers in Problem Solving: Phase Transitions and Complexity, Guest Editors: T. Hogg, B. A. Hubermann, and C. P. Williams, *Artificial Intelligence* **81** (1996), nos. 1 and 2.
11. V. Chvatal and E. Szemeredi. *Many hard examples for resolution.* Journal of the ACM **35** (1988) 759 - 768.
12. N. Creignou and H. Daude. *Generalized satisfiability problems: minimal elements and phase transitions.* Theoretical Computer Science **302** (2003), 417 - 430. Preliminary version in proceedings of SAT 2002.

13. R. Dechter, *Constraint networks*, in Encyclopedia of Artificial Intelligence, S. Shapiro (ed.), Wiley, New York, 2nd ed. (1992) 276–285.
14. O. Dubois and J. Mandler. *The 3-XORSAT Threshold.* Proceedings of FOCS 2002, 769 - 778.
15. M. Dyer, A. Frieze and M. Molloy, *A probabilistic analysis of randomly generated binary constraint satisfaction problems.* Theoretical Computer Scince **290**, 1815 - 1828 (2003).
16. A. Flaxman. *A sharp threshold for a random constraint satisfaction problem.* preprint (2003).
17. E. C. Freuder, *A sufficient condition for backtrack-free search*, Journal of the ACM **29** (1982) 24–32.
18. A. Frieze and N. Wormald, *??????*
19. I. Gent, E. MacIntyre, P. Prosser, B. Smith and T. Walsh. *Random constraint satisfaction: flaws and structure.* Constraints **6**, 345 - 372 (2001).
20. S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley, 2000.
21. A. K. Mackworth, *Constraint satisfaction*, in Encyclopedia of Artificial Intelligence, S. Shapiro (ed.), Wiley, New York, 2nd ed. (1992) 285-293.
22. D. Mitchell, *The Resolution complexity of random constraints.* Proceedings of Principles and Practices of Constraint Programming - CP 2002.
23. D. Mitchell, *The Resolution Complexity of Constraint Satisfaction.* Ph.D. Thesis, University of Toronto, 2002.
24. D. Mitchell, B. Selman and H. Levesque. "Hard and Easy Distributions of SAT Problems." Proceedings of AAAI 1992, 459 - 465.
25. M. Molloy, *Models for Random Constraint Satisfaction Problems.* Proceedings of STOC 2002, 209 - 217. Longer version to appear in SIAM J. Computing.
26. M. Molloy, *When does the giant component bring unsatisfiability?* Combinatorica (to appear).
27. M. Molloy and M. Salavatipour, *The resolution complexity of random constraint satisfaction problems.* Submitted.
28. B. Pittel, J. Spencer and N. Wormald, *Sudden emergence of a giant k-core in a random graph*, Journal of Combinatorial Theory (B) **67** (1996) 111–151.
29. Barbara M Smith. *Constructing an Asymptotic Phase Transition in Random Binary Constraint Satisfaction Problems.* Journal of Theoretical Computer Science **265**, 265 - 283 (2001).
30. D. Waltz, *Understanding line drawings of scenes with shadows*, The Psychology of Computer Vision, McGraw-Hill, New York, (1975) 19-91.
31. Ke Xu and Wei Li. *Exact Phase Transitions in Random Constraint Satisfaction Problems.* Journal of Artificial Intelligence Research **12**, 93 - 103 (2000).
32. K. Xu and W. Li. *Many hard examples in exact phase transitions with application to generating hard satisfiable instances.* Technical Report cs.CC/0302001, Computing Research Repository (CoRR), 2003.