Technical Report TR04-13, July 2004

Consistency and Random Constraint Satisfaction Models

Yong Gao and Joseph Culberson

Department of Computing Science, University of Alberta
Edmonton, Alberta, Canada, T6G 2E8
{ygao, joe}@cs.ualberta.ca

Abstract. Existing random models for the constraint satisfaction problem (CSP) all require an extremely low constraint tightness in order to have non-trivial threshold behaviors and guaranteed hard instances at the threshold. We study the possibility of designing random CSP models that have interesting threshold and typical-case complexity behaviors while at the same time, allow a much higher constraint tightness. We show that random CSP models that enforce the constraint consistency have guaranteed exponential resolution complexity without putting much restriction on the constraint tightness. A new random CSP model is proposed to generate random CSPs with a high tightness whose instances are always consistent. Initial experimental results are also reported to illustrate the sensitivity of instance hardness to the constraint tightness in classical CSP models and to evaluate the proposed new random CSP model.

1 Introduction

One of the most significant problems with the existing random CSP models is that as a model parameter, the constraint tightness has to be extremely low in order to have non-trivial threshold behaviors and guaranteed hard instances at phase transitions. In [1, 2], it was shown that except for a small range of the constraint tightness, all of the four classical random CSP models are trivially unsatisfiable with high probability due to the existence of the flawed variables. For the case of binary CSPs, the constraint tightness has to be less than the domain size in order to avoid the flawed variables. Recent theoretical results in [3, 4] further indicate that even for a moderate constraint tightness, it is still possible for these classical models to have an asymptotically polynomial complexity due to the appearance of embedded easy subproblems.

Several new models have been proposed to overcome the trivial unsatisfiability. In [2], Gent et al. proposed a CSP model, called the flawless random binary CSP, that is based on the notion of a *flawless conflict matrix*. Instances of the flawless random CSP model are guaranteed to be arc-consistent, and thus do not suffer asymptotically from the problem of flawed variables. In [1], a nogoods-based CSP model was proposed and was shown to have non-trivial asymptotic behaviors. Random CSP models with a (slowly) increasing domain size have also been shown to be free from the problem of flawed variables and have interesting threshold behaviors [5, 6]. However, none of these models have addressed the fundamental requirement of an extremely low constraint tightness in order to have a guaranteed exponential complexity. The flawless random CSP does have a true solubility phase transition at a high constraint tightness, but as we will show later, it still suffers from the embedded easy unsatisfiable subproblems at a moderate constraint tightness. In CSP models with increasing domain size, the (relative) constraint tightness should still remain low. In the nogood-based CSP model, it is impossible to have a high constraint tightness without making the constraint (hyper)graph very dense.

In this paper, we study the possibility of designing non-trivial random CSP models that allow a much higher constraint tightness. For this purpose, we show that consistency, a notion that has been developed to improve the efficiency of CSP algorithms, is in fact the key to the design of random CSP models that have guaranteed exponential resolution complexity without the requirement of an extremely low constraint tightness. We propose a scheme to generate consistent random instances of CSPs that can potentially have a high constraint tightness. Initial experiments show that our model are indeed much harder at phase transition than the classical CSP models and the flawless CSP models.

2 Random Models for CSPs

Throughout this paper, we consider binary CSPs defined on a domain D with $|D| = d$. A binary CSP $\mathcal{C}$ consists of a set of variables $x = \{x_1, \cdots, x_n\}$ and a set of binary *constraints* $(C_1, \cdots, C_m)$. Each constraint C_i is specified by its *constraint scope*, a pair of the variables in x, and a *constraint relation* R_{C_i} that defines a set of incompatible value-tuples in $D \times D$ for the scope variables. An incompatible value tuple is also called a *restriction*, or a *nogood*. Associated with a binary CSP is a *constraint graph* whose vertices correspond to the set of variables and edges correspond to the set of constraint scopes. In the rest of the paper, we will be using the following notation:

1. n, the number of variables;
2. m, the number of constraints;
3. d, the domain size; and
4. t, the constraint tightness, i.e., the size of the restriction set.

Given two variables, their constraint relation can be specified by a 0-1 matrix, called the *conflict matrix*, where an entry 0 at (i, j) indicates that the tuple $(i, j) \in D \times D$ is incompatible. Another way to describe the constraint relation is to use the *compatible graph*, a bipartite graph with the domain of each variable as an independent partite, where an edge signifies the corresponding value-tuple is compatible.

An instance of a CSP is said to be *k-consistent* if and only if for any (k-1) variables, each consistent (k-1)-tuple assignment to the (k-1) variables can be extended to an assignment to any other kth variable such that the k-tuple is also consistent. A CSP instance is called *strongly k-consistent* if and only if it is j-consistent for each $j \leq k$. Of special interest are the strong k-consistency for $k = 1, 2, 3$, also known as node-consistency, arc-consistency, and path-consistency [7].

Definition 1 (Random Binary CSP $\mathcal{B}^{d,t}_{n,m}$). *Let $0 < t < d^2$ be an integer. The $\mathcal{B}^{d,t}_{n,m}$ is a random CSP model such that*

1. *its constraint graph is the standard random graph $G(n, m)$ where m edges of the graph are selected uniformly at random from all the possible $\binom{n}{2}$ edges; and*
2. *for each of the edges of $G(n, m)$, a constraint relation on the corresponding scope variables is specified by choosing t value-tuples from $D \times D$ uniformly at random as its restriction set.*

$\mathcal{B}^{d,t}_{n,m}$ is known in the literature as the Model B. It has been shown in [1, 2] that for $t \geq d$, $\mathcal{B}^{d,t}_{n,m}$ is asymptotically trivial and unsatisfiable, but has a phase transition in satisfiability probability for $t < d$. This motivates the introduction of the flawless conflict matrix to make sure that the random model is arc-consistent [2].

Definition 2 ($\mathcal{B}^{d,t}_{n,m}[1]$, Flawless Random Binary CSP). *In the flawless random binary CSP $\mathcal{B}^{d,t}_{n,m}[1]$, the constraint graph is defined in the same way as that in $\mathcal{B}^{d,t}_{n,m}$. For each constraint edge, the constraint relation is specified in two steps:*

1. *Choosing a random permutation π of $D = \{1, \cdots d\}$; and*
2. *Selecting a set of t value-tuples uniformly at random from $D \times D \setminus \{(i, \pi(i)), 1 \leq i \leq n)\}$ as the restriction set.*

The reason that we use a suffix "[1]" in the symbol $\mathcal{B}^{d,t}_{n,m}[1]$ will become clear after we introduce the generalized flawless random CSPs later in this paper. By specifying a set of tuples $\{(i, \pi(i)), 1 \leq i \leq n)\}$ that will not be considered when choosing incompatible value-tuples, the resulting model is guaranteed to be arc-consistent and consequently will not have flawed variables. However, even though the flawless random binary CSP $\mathcal{B}^{d,t}_{n,m}[1]$ does not suffer the problem of trivial unsatisfiability, it can be shown that $\mathcal{B}^{d,t}_{n,m}[1]$ asymptotically has embedded easy subproblems for $t \geq d - 1$ in the same way as the random binary CSP model $\mathcal{B}^{d,t}_{n,m}$.

Theorem 1. *For $t \geq d - 1$, there is a constant $c^* > 0$ such that for any $\frac{m}{n} > c^*$, with high probability $\mathcal{B}^{d,t}_{n,m}[1]$ is asymptotically unsatisfiable and can be solved in polynomial time.*

A detailed proof of Theorem 1 can be found in the Appendix, Section 6.1. The idea is to show that for $\frac{m}{n} > c^*$, the flawless random CSP $\mathcal{B}^{d,t}_{n,m}[1]$ will with high probability contain an unsatisfiable subproblem called an r-flower. The definition of an r-flower can be found in the Appendix. Furthermore, if a binary CSP instance contains an r-flower, then any path-consistency algorithm (see, e.g., [7]) will produce a new CSP instance in which the center variable of the r-flower has an empty domain. It follows that we can prove that it is unsatisfiable polynomially.

It should be noted that $\mathcal{B}^{d,t}_{n,m}[1]$ does have a non-trivial phase transition since it is satisfiable with high probability if $\frac{m}{n} < \frac{1}{2}$. Theorem 1 does not exclude the possibility that $\mathcal{B}^{d,t}_{n,m}[1]$ will also be able to generate hard instances when $\frac{m}{n}$ is below the upper bound c^*, in particular in the case of a large domain size. Further investigation is required to fully understand the complexity of $\mathcal{B}^{d,t}_{n,m}[1]$ in this regard.

3 Consistency, Resolution Complexity, and Better Random CSP Models

Propositional resolution complexity deals with the minimum length of resolution proofs for an (unsatisfiable) CNF formula. As many backtracking-style complete algorithms can be simulated by a resolution proof, the resolution complexity provides an immediate lower bound on the running time of these algorithms. Since the work of Chvatal and Szemeredi [8], there has been many studies on the resolution complexity of randomly generated CNF formulas [9, 10].

Mitchell [11] developed a framework in which the notion of resolution complexity is generalized to CSPs and the resolution complexity of randomly generated random CSPs can be studied. In this framework, the resolution complexity of a CSP instance is defined to be the resolution complexity of a natural CNF encoding which we give below. Given an instance of a CSP on a set of variables $\{x_1, \cdots, x_n\}$ with a domain $D = \{1, 2, \cdots, d\}$, its CNF encoding is constructed as follows:

1. For each variable x_i, there are d Boolean variables $x_i : 1,\ x_i : 2, \ldots,$ and $x_i : d$, each of which indicates whether or not x_i takes on the corresponding domain value; and there is a clause $x_i : 1 \vee x_i : 2 \vee \cdots \vee x_i : d$ on these d Boolean variables making sure that x_i takes at least one of the domain values;
2. For each restriction $(\delta_1, \cdots, \delta_k) \in D^k$ of each constraint $C(x_{i_1}, \cdots, x_{i_k})$, there is a clause $\overline{x_{i_1} : \delta_1} \vee \cdots \vee \overline{x_{i_k} : \delta_k}$ to respect the restriction.

In [11, 4], upper bounds on the constraint tightness t were established for the random CSPs to have an exponential resolution complexity. For random binary CSP $\mathcal{B}^{d,t}_{n,m}$, the bound is (1) $t < d-1$; or (2) $t < d$ and $\frac{m}{n}$ is sufficient small. For a moderate constraint tightness, recent theoretical results in [3, 4] indicate that it is still possible for these classical models to have an asymptotically polynomial complexity due to the existence of embedded easy subproblems. The primary reason for the existence of embedded easy subproblems is that for a moderate constraint tightness, constraints frequently imply forcers which force a pair of involved variables to take on a single value-tuple.

In the following, we will show that it is not necessary to put restrictions on the constraint tightness in order to have a guaranteed exponential resolution complexity. Based on quite similar arguments as those in [11, 4, 12], it can be shown that if in $\mathcal{B}^{d,t}_{n,m}$, the constraint relation of each constraint is chosen in such a way that the resulting instances are always strongly k-consistent ($k \geq 3$), then $\mathcal{B}^{d,t}_{n,m}$ has an exponential resolution complexity no matter how large the constraint tightness is.

Theorem 2. *Let $\mathcal{B}^{d,t}_{n,m}[SC]$ be a random CSP such that (1) its constraint graph is the standard random graph G(n, m); and (2) for each edge, the constraint relation is such that any instances of $\mathcal{B}^{d,t}_{n,m}[SC]$ is strongly k-consistent. Then, the resolution complexity of $\mathcal{B}^{d,t}_{n,m}[SC]$ is almost surely exponential.*

Proof. See Appendix.

Using the tool developed in [4], the requirement of the strong k-consistency for CSP instances to have an exponential resolution complexity can be further relaxed. We call a CSP instance *weakly path-consistent* if it is arc-consistency and satisfies the conditions of path-consistency for paths of length 3 or more.

Theorem 3. *Let $\mathcal{B}^{d,t}_{n,m}[WC]$ be a random CSP such that (1) its constraint graph is the random graph G(n, m); and (2) for each edge, the constraint relation is such that any instances of $\mathcal{B}^{d,t}_{n,m}[WC]$ is weakly path-consistent and contains no forcer (See Definition 5). Then, The resolution complexity of $\mathcal{B}^{d,t}_{n,m}[WC]$ is almost surely exponential.*

Proof. See Appendix.

The question remaining to be answered is whether or not there are any natural random CSP models that are guaranteed to be strongly k-consistent or weakly path-consistent. In fact, the CSP-encoding of random graph k-coloring problem is strongly k-consistent. Another example is the flawless random binary CSP $\mathcal{B}^{d,t}_{n,m}[1]$ that is guaranteed to be arc-consistent, i.e., strongly 2-consistent. In the rest of this section, we discuss how to generate random CSPs with a high tightness that are strongly 3-consistent or weakly path-consistent.

Definition 3 ($\mathcal{B}^{d,t}_{n,m}[\mathcal{K}]$, Generalized Flawless Random Binary CSP). *In the generalized flawless random binary CSP $\mathcal{B}^{d,t}_{n,m}[\mathcal{K}]$, $\mathcal{K}$ is a random bipartite graph with each partite being the domain D of a variable. The constraint graph is defined in the same way as that in $\mathcal{B}^{d,t}_{n,m}$. For each constraint edge, the constraint relation is specified as follows:*

1. *Generate the bipartite graph $\mathcal{K}$ satisfying certain properties; and*
2. *Select a set of t value-tuples uniformly at random from $(D \times D) \setminus E(\mathcal{K})$ as the restriction set.*

The idea in the generalized flawless random binary CSP is that by enforcing a subset of value-tuples (specified by the edges of the bipartite graph $\mathcal{K}$) to be always compatible, it is possible that the generated CSP instance will always satisfy a certain level of consistency. If we define $\mathcal{K}$ to be a 1-regular bipartite graph, then $\mathcal{B}^{d,t}_{n,m}[\mathcal{K}]$ reduces to the flawless random binary CSP model $\mathcal{B}^{d,t}_{n,m}[1]$.

The following result shows that a connected and l-regular bipartite graph $\mathcal{K}$ with sufficiently large l can be used to generate strongly 3-consistent random CSPs or weakly path-consistent random CSPs.

Theorem 4. *Let $\mathcal{K}$ be an l-regular connected random bipartite graph. Then, $\mathcal{B}^{d,t}_{n,m}[\mathcal{K}]$ is always*

1. *strongly 3-consistent if and only if $l > \frac{d}{2}$; and*
2. *weakly path-consistent if and only if $l > \frac{d-1}{2}$.*

Proof. We only prove the case for the weak path-consistency and the case for the strong 3-consistency is similar.

Consider a path $x_1 - x_2 - x_3 - x_4$ and any assignment $x_1 = i$ and $x_4 = j$. There are l values of x_2 that are compatible to $x_1 = i$ and there are l values of x_3 that are compatible to $x_4 = j$. Since the conflict graph is connected, there are at least $l + 1$ values of x_3 that are compatible to $x_1 = i$. Therefore if $l > (d - 1)/2$, there must be a value of x_3 that is compatible to both $x_1 = i$ and $x_4 = j$.

To see the "only if" part, we will show that there is a connected bipartite graph $K(V, U)$ on two sets of vertices $V = \{v_1, v_2, \cdots, v_d\}$ and $U = \{v_1, u_2, \cdots, u_d\}$ such that the neighbors of the first l vertices in V are the first $l+1$ vertices in U. First, we construct a complete bipartite graph on the vertex sets $\{v_1, v_2, \cdots, v_l\}$ and $\{u_1, u_2, \cdots, u_l\}$; Second, we construct an l-regular connected bipartite graph on the vertex sets $\{v_{l+1}, \cdots, v_d\}$ and $\{u_{l+1}, \cdots, u_d\}$ such that (v_{l+1}, u_{l+1}) is an edge. We then replace the two edges (v_l, u_l) and (v_{l+1}, u_{l+1}) with two new edges (v_l, u_{l+1}) and (v_{l+1}, u_l). This gives the bipartite graph $K(V, U)$. The existence of such a bipartite graph $K(V, U)$ shows that when $l \leq \frac{d-1}{2}$, it is possible to have a constraint relation such that a constraint path of length 3 is not consistent. □

The generalized random CSP model $\mathcal{B}^{d,t}_{n,m}[\mathcal{K}]$ with a connected regular bipartite $\mathcal{K}$ allows a constraint tightness up to $\frac{(d+1)d}{2}$. The above theorem also indicates that this is the best possible constraint tightness when using an arbitrary connected bipartite graph $\mathcal{K}$. To achieve higher constraint tightness, we propose a recursive scheme to generate a bipartite graph $\mathcal{K}$ that is more efficient in its use of edges.

Definition 4 (Consistency Core). *Let $D_1 = D_2$ be the domains of two variables with $|D_1| = |D_2| = d$. The consistency core for the domains D_1 and D_2 is a bipartite graph $\mathcal{G}_{core}(D_1, D_2)$ on D_1 and D_2, and is defined recursively as follows.*

1. *Let $\{D_{ij}, 1 \leq j \leq s\}$ be a partition of D_i such that $|D_{ij}| \geq 3$.*

2. *If $s < 3$, then $\mathcal{G}_{core}(D_1, D_2)$ is equal to an l_0-regular connected bipartite graph on $D_1(\pi_1) = \{\pi_1(1), \cdots, \pi_1(d)\}$ and $D_2(\pi_2) = \{\pi_2(1), \cdots, \pi_2(d)\}$ where π_1, π_2 are two permutations of $\{1, 2, \cdots, d\}$ and $l_0 > \frac{d}{2}$.*
3. *For $s \geq 3$, let π_1, π_2 be two permutations of $S = \{1, 2, \cdots, s\}$ and*

$$G(S(\pi_1), S(\pi_2), l)$$

be an l-regular connected bipartite graph on $S(\pi_1) = \{\pi_1(1), \cdots, \pi_1(s)\}$ and $S(\pi_2) = \{\pi_2(1), \cdots, \pi_2(s)\}$. The edge set of $\mathcal{G}_{core}(D_1, D_2)$ is defined to be the union of the edge sets of all consistency cores $\mathcal{G}_{core}(D_{1i}, D_{1j})$ where i and j are integers such that $(i, j) \in G(S(\pi_1), S(\pi_2), l)$.

Theorem 5. *If a consistency core is used for $\mathcal{K}$, then $\mathcal{B}^{d,t}_{n,m}[\mathcal{K}]$ is*

1. *strongly 3-consistent if and only if $l > \frac{s}{2}$; and*
2. *weakly path-consistent if and only if $l > \frac{s-1}{2}$.*

Proof. By induction on the domain size and using the previous theorem.

Using the consistency core, we can define random CSP models with constraint tightness well above $\frac{(d+1)d}{2}$. For example, if the domain size d is 12, the random generalized random CSP model $\mathcal{B}^{d,t}_{n,m}[\mathcal{K}]$ with a consistency core $\mathcal{K}$ allow a constraint tightness up to $144 - 6 * 8 = 96$.

Example 1. Consider the consistency core $\mathcal{K}$ where the domain size is $|D| = 9$ and assume that all the permutations used in Definition 4 are identity permutations and $l = s = 3$. Figure 1 shows the consistency core where the edges connected to two vertices in the lower partite was depicted. Using such a consistency core, a constraint on two variables x_i, x_j in $\mathcal{B}^{d,t}_{n,m}[\mathcal{K}]$ with $t = 45$, has a set of restrictions

$$\{(i, j);\ i = 3a_1 + a_2 \text{ and } j = 3b_1 + b_2 \text{ are integers such that } a_1 \neq b_1 \text{ and } a_2 \neq b_2\}.$$

An instance of this CSP model can be viewed as a generalized 3-colorability problem.

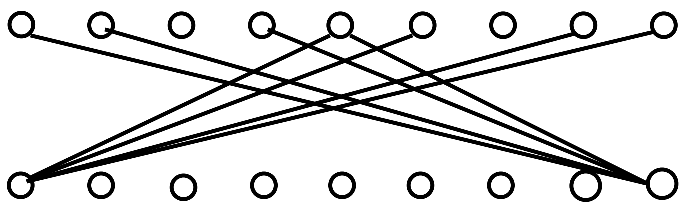

Fig. 1. A special type of consistency core with the domain size 9

4 Experiments

In this section, we report results of two sets of preliminary experiments designed to (1) study the effect of an increase in the constraint tightness on the typical-case complexity; and (2) compare typical-case instance hardness between the classical random CSPs, flawless random CSPs, and the generalized flawless random CSPs.

4.1 Effect of an increase in the constraint tightness

In [3, 4], upper bounds on the constraint tightness have been established for random CSPs to have an exponential resolution complexity for all the constant constraint-to-variable ratio $\frac{m}{n}$. Molloy [4] further showed that for the constraint tightness above the upper bound, the existence of forcers can be compensated by sufficiently low constraint-to-variable ratio so that one can still have typical instances with exponential resolution complexity.

We have conducted the following experiments to gain further understanding on the effect of an increase in the constraint tightness (and hence an increase in the likelihood of the existence of a forcer in a constraint) on the typical-case hardness of random CSPs. The experiments also help understand the behavior of CSP models, such as the flawless CSP model, that only enforce the arc-consistency (strong 2-consistency).

In the experiments, we start with a random 3-CNF formula whose clauses are treated as constraints. We then incrementally increase the tightness of each constraint by adding more clauses defined over the same set of variables. There are two reasons why we have based our experiments on random SAT models. First, the typical-case complexity of the random SAT model is well-understood and therefore, experiments based on the random SAT model will enable us to have an objective comparison on the impact of an increase in the constraint tightness. Secondly, the complexity of Boolean-valued random CSPs obtained by increasing the tightness of the random 3-CNF formula has been characterized in great detail. We have a clear picture regarding the appearance of embedded easy subproblems in these Boolean-valued random CSPs [3].

Let $\mathcal{F}(n, m)$ be a random 3-CNF formula with n variables and m clauses. We construct a new random 3-CNF formula $\mathcal{F}(n, m, a)$ as follows:

1. $\mathcal{F}(n, m, a)$ contains all the clauses in $\mathcal{F}(n, m)$;
2. For each clause C in $\mathcal{F}(n, m)$, we generate a random clause on the same set of variables of C, and add this new clause to $\mathcal{F}(n, m, a)$ with probability a.

In fact, $\mathcal{F}(n, m, a)$ is the random boolean CSP model with a real-valued constraint tightness $1 + a$ and has been discussed in [3]. For $a > 0$, it is easy to see that $\mathcal{F}(n, m, a)$ is always strongly 2-consistent, but is not 3-consistent asymptotically with probability 1.

Figure 2 shows the median of the number of branches used by the SAT solver zChaff on 100 instances of $\mathcal{F}(n, m, a), a = 0.0, 0.1, 0.2$.

As expected, an increase in the tightness results in a shift of the location of the hardness peak toward smaller m/n. More significant is the magnitude of

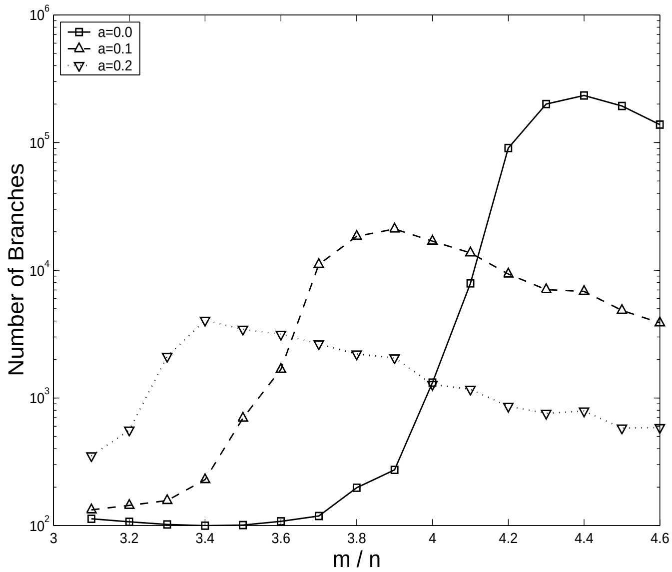

Fig. 2. Effects of an increase in the constraint tightness on the instance hardness for $\mathcal{F}(n, m, 0.0), \mathcal{F}(n, m, 0.1),$ and $\mathcal{F}(n, m, 0.2)$. $n = 250$.

the decrease of the hardness as a result of a small increase in the constraint tightness.

From [3], the upper bounds on m/n for $\mathcal{F}(n, m, a)$ to have an exponential resolution complexity are respectively 23.3 if $a = 0.1$ and 11.7 if $a = 0.2$. Since the constraint-to-variable rations m/n considered in the experiment are well below these bounds above which embedded 2SAT subproblems appear with high probability, it seems that the impact of forcers on the instance hardness goes beyond simply producing embedded easy subproblems. As forcers can appear at a relatively low constraint tightness even in CSP models such as the flawless model, approaches that are solely based on restricting constraint tightness to generate interesting and typically hard instances cannot be as effective as has been previously believed.

4.2 Comparisons between three random CSP Models

This set of experiments is designed to investigate the effectiveness of the generalized flawless random CSP model. We generate random instances of the classical random models $\mathcal{B}^{d,t}_{n,m}$, flawless random model $\mathcal{B}^{d,t}_{n,m}[1]$, and the generalized random model $\mathcal{B}^{d,t}_{n,m}[\mathcal{K}]$ with the domain size $d = 4$. For $\mathcal{B}^{d,t}_{n,m}[\mathcal{K}]$, we have used a 2-regular connected bipartite graph as $\mathcal{K}$. These instances are then encoded as CNF formulas and solved by the SAT solver zChaff [13]. It looks unnatural that we have tested random CSP instances by converting them to SAT instances and using a SAT solver. This is justified by the following considerations. First, all of the existing research on the resolution complexity of random CSPs have been carried out by studying the resolution complexity of a SAT encoding of CSPs as described in Section 3. We have used the same encoding in the experiments. Secondly, it has been shown that as far as the complexity of solving unsatisfiable CSP instances is concerned, many of the existing CSP algorithms can be efficiently simulated by the resolution system of the corresponding SAT encodings of the CSPs [14].

The experiments show that the threshold of the solution probability of the generalized random CSP model $\mathcal{B}_{n,m}^{d,t}[\mathcal{K}]$ is much sharper than those of $\mathcal{B}_{n,m}^{d,t}$ and $\mathcal{B}_{n,m}^{d,t}[1]$. More importantly, instances of $\mathcal{B}_{n,m}^{d,t}[\mathcal{K}]$ at the phase transition are much harder than those of $\mathcal{B}_{n,m}^{d,t}$ and $\mathcal{B}_{n,m}^{d,t}[1]$, as shown in Tables 1-3 where the median of the number of branches of zChaff for 100 instances of each of the three random CSP models is listed at different stages of the solubility phase transition: Table 1 is for the constraint density $\frac{m}{n}$ where the maximum median of the number of branches is observed; Table 2 is for the constraint density $\frac{m}{n}$ where the solubility probability is less than 0.1; and Table 3 is for the constraint density $\frac{m}{n}$ where the solubility probability is greater than 0.9.

It can be seen that while the classical random CSP model and flawless matrix CSP model have little difference, the proposed strong flawless random CSP model $\mathcal{B}_{n,m}^{d,t}[\mathcal{K}]$ with $\mathcal{K}$ being a connected 2-regular bipartite graph is significantly harder in all of the cases except row 1 in Table 3. It is also interesting to notice that the most significant difference in the hardness among the three models is at the phase where instances of the random CSP models are almost always unsatisfiable. A plausible explanation for this phenomenon is that consistency is a property that may also help improve the efficiency of search algorithms in solving satisfiable instances.

Table 1. Maximum Median Number of Branches of zChaff on random instances of three random CSP models , over all $\frac{m}{n}$. Domain size $d = 4$ and $\mathcal{K}$ is 2-regular.

	Number of Branches		
(n, t)	$\mathcal{B}_{n,m}^{d,t}$	$\mathcal{B}_{n,m}^{d,t}[1]$	$\mathcal{B}_{n,m}^{d,t}[\mathcal{K}]$
$(100, 6)$	230	224	399
$(300, 6)$	1830	1622	4768
$(500, 6)$	7152	6480	45315
$(300, 8)$	843	1010	2785

Table 2. Median Number of Branches of zChaff on random instances of three random CSP models at the smallest $\frac{m}{n}$ where the solution probability is less than 0.1. Domain size $d = 4$ and $\mathcal{K}$ is 2-regular.

	Number of Branches		
(n,t)	$\mathcal{B}^{d,t}_{n,m}$	$\mathcal{B}^{d,t}_{n,m}[1]$	$\mathcal{B}^{d,t}_{n,m}[\mathcal{K}]$
$(100, 6)$	116	154	241
$(300, 6)$	819	700	4768
$(500, 6)$	1398	1649	45315
$(300, 8)$	204	269	1118

Table 3. Median Number of Branches of zChaff on random instances of three random CSP models at the largest $\frac{m}{n}$ where the solution probability is greater than 0.9. Domain size $d = 4$ and $\mathcal{K}$ is 2-regular.

	Number of Branches		
(n,t)	$\mathcal{B}^{d,t}_{n,m}$	$\mathcal{B}^{d,t}_{n,m}[1]$	$\mathcal{B}^{d,t}_{n,m}[\mathcal{K}]$
$(100, 6)$	211	212	199
$(300, 6)$	1327	1595	2809
$(500, 6)$	7152	6450	8787
$(300, 8)$	843(0.67)	709	2785

5 Conclusions

In this paper, we have shown that consistency, a notion that has been introduced in an effort to improve the efficiency of CSP algorithms, also plays an important role in the design of random CSP models that have interesting threshold behavior and guaranteed exponential complexity at phase transitions, while at the same time allow a much higher constraint tightness. We have also proposed a scheme to generate random consistent random CSPs by generalizing the idea of flawless random CSPs. Initial experiments show that the proposed model is indeed significantly harder than existing random CSP models.

6 Appendix

In this section, we present more concepts related to the resolution complexity results stated in this paper and prove Theorems 1, 2, and 3.

6.1 Theorem 1

This subsection is devoted to Theorem 1. First, let us formalize some definitions such as a forcer, a forbidding cycle, and an r-flower. Following [11], we call an expression of the form $x : \alpha$ a literal for a CSP. A literal $x : \alpha$ evaluates to TRUE at an assignment if the variable x is assigned the value α. A nogood for a CSP, denoted by $\eta(x_1 : \alpha_1, \cdots, x_l : \alpha_l)$, is a disjunction of the negations of the literals $x_i : \alpha_i, 1 \leq i \leq l$. A nogood is equivalent to a restriction $\{\alpha_1, \cdots, \alpha_l\}$ on the set of variables $\{x_1, \cdots, x_l\}$, and the restrictions of a constraint correspond to a set of nogoods defined over the same set of variables.

Definition 5 (Forcers [4]). *A constraint C_f with $\mathrm{var}(C) = \{x_1, x_2\}$ is called an (α, β)-forcer if its restriction set corresponds to the set of nogoods*

$$\mathrm{NG}(C_f) = \{\eta(x_1 : \alpha, x_2 : \gamma); \gamma \neq \beta\}.$$

We say that a constraint C contains an (α, β)-forcer C_f defined on the same set of variables as C if $\mathrm{NG}(C_f) \subseteq \mathrm{NG}(C)$.

Definition 6 (Forbidding cycles and r-flowers [4]). *An α-forbidding cycle for a variable x_0 is a set of constraints $C_1(x_0, x_1), C_2(x_1, x_2), \ldots, C_{r-1}(x_{r-2}, x_{r-1})$, and $C_r(x_{r-1}, x_0)$ such that $C_1(x_0, x_1)$ is an (α, α_1)-forcer, $C_r(x_{r-1}, x_0)$ is an (α_{r-1}, α_r)-forcer ($\alpha_r \neq \alpha$), and $C_i(x_{i-1}, x_i)$ is an (α_{i-1}, α_i)-forcer ($2 \leq i \leq r-1$). We call x_0 the center variable of the α-forbidding cycle.*

An r-flower $R = \{\mathcal{C}_1, \cdots, \mathcal{C}_d\}$ consists of d (the domain size) forbidding cycles each of them has the length r such that

1. *$\mathcal{C}_i, 1 \leq i \leq d$, have the same center variable x;*
2. *each $\mathcal{C}_i$ is an α_i forbidding cycle of the common center variable x; and*
3. *these forbidding cycles do not share any other variables.*

The following facts are straightforward to establish:

1. An r-flower consists of $s = d(r-1) + 1 = dr - d + 1$ variables and dr constraints;
2. The total number of r-flowers is
$$\binom{n}{s} s!(d-1)^d d^{d(r-1)}.$$
3. A constraint in the flawless CSP model contains an (α, β)-forcer only if the pair (α, β) is one of the pre-selected tuples in the flawless constraint matrix.

In the following, we assume that $r = o(\sqrt{n})$. The probability for a constraint to contain a forcer and the probability for the flawless random CSP to contain an r-flower are given in the following lemma.

Lemma 1. *Consider the flawless random CSP $\mathcal{B}_{n,m}^{d,t}[1]$ and define $f_e = \frac{\binom{d^2-d-d+1}{t-d+1}}{\binom{d^2-d}{t}}$.*

1. *The probability that a given constraint $C(x_1, x_2)$ contains an (α, β)-forcer is*
$$\frac{1}{d} f_e. \tag{1}$$

2. Let R be an r-flower and let $c = m/n$,

$$P\{R \text{ appears in } \mathcal{B}_{n,m}^{d,t}[1]\} = \Theta(1)(2cf_e)^{dr}\frac{1}{n^{dr}}\frac{1}{d^{dr}}. \quad (2)$$

Proof. Equation (1) follows from the following two observations: (a) $\frac{1}{d}$ is the probability for (α, β) to be one of the pre-selected tuples in the flawless conflict matrix; and (b) f_e is the probability for the $d-1$ tuples, $(\alpha,\gamma), \gamma \neq \beta$, to be in the set of t restrictions selected uniformly at random from $d^2 - d$ tuples.

To calculate the probability that a given r-flower R appears in $\mathcal{B}_{n,m}^{d,t}[1]$, notice that the probability of selecting all the constraint edges in R is

$$\frac{\binom{N-dr}{cn-dr}}{\binom{N}{cn}} = \frac{cn(cn-1)\cdots(cn-dr+1)}{N(N-1)\cdots(N-dr+1)} = \Theta(1)\left(\frac{2c}{n}\right)^{dr}$$

where $N = \binom{n}{2}$. Since for each fixed choice of dr constraint edges in the r-flower, the probability for these constraints to contain the r-flower $(\frac{1}{d}f_e)^{dr}$, Equation (2) follows. □

*Proof (***Proof of Theorem 1***).* Let $c^* = \frac{d}{2f_e}$. We will show that if $c = \frac{m}{n} > c*$, then

$$\lim_n P\{\mathcal{B}_{n,m}^{d,t}[1]\} \text{ contains an r-flower } \} = 1. \quad (3)$$

Let I_R be the indicator function of the event that the r-flower R appears in $\mathcal{B}_{n,m}^{d,t}[1]$ and let $X = \sum_R I_R$ where the sum is over all the possible r-flowers. Then, $\mathcal{B}_{n,m}^{d,t}[1]$ contains an r-flower if and only if $X > 0$.

By Lemma 1 and the fact that $s = dr - d + 1$, we have

$$\begin{aligned} E[X] &= \sum_R E[I_R] \\ &= \Theta(1)\binom{n}{s}s!(d-1)^d d^{d(r-1)}(2cf_e)^{dr}\frac{1}{n^{dr}}\frac{1}{d^{dr}} \\ &= \Theta(1)n(n-1)\cdots(n-s+1)d^{dr}(2cf_e)^{dr}\frac{1}{n^{dr}}\frac{1}{d^{dr}} \\ &= \Theta(1)n^{1-d}(2cf_e)^{dr}. \end{aligned}$$

Therefore, if $c > c^*$ and $r = \lambda \log n$ with λ sufficiently large, we have $\lim_n E[X] = \infty$.

If we can show that $E[X^2] \leq E^2[X](1 + o(1))$, then an application of the Chebyshev inequality will establish that $\lim_n P\{X = 0\} = 0$. To get an upper bound on $E[X^2]$, we need a counting argument to upper bound the number of r-flowers sharing a given number of edges. This is done by considering how the shared edges form connected components [4, 3, 15]. Here, we follow the way that is used by Molloy [4], from which we have

$$
\begin{aligned}
E[X^2] &= \sum_A E[I_A^2] + \sum_A \sum_{B:B\cap A=\emptyset} E[I_A I_B] + \sum_A I_A \left(\sum_{i=1}^{s} \sum_{j=1}^{i} N_{ij} (P_{ij})^{dr-i} \right) \\
&= \sum_A E[I_A^2] + \sum_A \sum_{B:B\cap A=\emptyset} E[I_A] E[I_B] + \sum_A I_A \left(\sum_{i=1}^{s} \sum_{j=1}^{i} N_{ij} (P_{ij})^{dr-i} \right) \\
&\leq E^2[X] + \sum_A I_A \left(\sum_{i=1}^{s} \sum_{j=1}^{i} N_{ij} (P_{ij})^{dr-i} \right) \qquad (4)
\end{aligned}
$$

where (1) N_{ij} is the number of the r-flowers that share exactly i constraint edges with A and these i constraints forms j connected components in the constraint graph of A; and (2) $(P_{ij})^{dr-i}$ is the probability that conditional on I_A, the random CSP contains the $dr - i$ constraints of a specific r-flower as described in Lemma (1). In [4], N_{ij} is upper bounded by

$$
\left((2+r^2)^d (dr^2)^{j-1}\right)^2 j! n^{s-i-j} d^{s-i-j+d-1},
$$

where $((2+r^2)^d(dr^2)^{j-1})^2 j!$ upper bounds the number of ways to choose and arrange the j shared connected components for two r-flowers; n^{s-i-j} upper bounds the number of ways of choosing the remaining non-shared variables——the number of variables in each of the j shared connected component is at least one plus the number of edges in that shared component; and $d^{s-i-j+d-1}$ upper bounds the number of ways of choosing the forcing values in these non-sharing variables. The shared variables have to take the same forcing values as those in A due to the assumption that $t < d$ made in [4].

Since in our case we $d - 1 \leq t \leq d^2 - d$, it is possible for shared variables to take different forcing values in different r-flowers. Thus, an upper bound for N_{ij} is

$$
\left((2+r^2)^d (dr^2)^{j-1}\right)^2 j! n^{s-i-j} d^{s}.
$$

But in our case, the probability corresponding to $(P_{ij})^{dr-i}$ is

$$
\begin{aligned}
&\frac{\binom{N-dr-(dr-i)}{cn-i-(dr-i)}}{\binom{N-dr}{cn-i}} \left(\frac{1}{d} f_e\right)^{dr-i} = \Theta(1) \left(\frac{cn-i}{N-dr}\right)^{dr-i} \left(\frac{1}{d} f_e\right)^{dr-i} \\
&= \Theta(1)(2cf_e)^{dr-i} \frac{1}{n^{dr-i}} \frac{1}{d^{dr-i}}.
\end{aligned}
$$

Therefore, with $c^* = \frac{d}{2f_e}$, we have

$$\begin{aligned}
&\sum_{i=1}^{s}\sum_{j=1}^{i} N_{ij}(2cf_e)^{dr-i}\frac{1}{n^{dr-i}}\frac{1}{d^{dr-i}} \\
&\leq \sum_{i=1}^{s}\left[(2+r^2)^{2d}r^{-4}n^{s-i}d^s(2cf_e)^{dr-i}\frac{1}{n^{dr-i}}\frac{1}{d^{dr-i}}\sum_{j=1}^{i}(\frac{d^2r^4j}{n})^j\right] \\
&\leq \sum_{i=1}^{s} O(r^{4d-4})n^{1-d}(2cf_e)^{dr}\frac{(2cf_e)^{-i}}{d^{-i}}O(\frac{r^4}{n}) \\
&\leq E[X]O(r^{4d-4})O(\frac{r^4}{n})\sum_{i=1}^{s}(\frac{d}{2cf_e})^i \\
&\leq E[X]O(\frac{r^{4d}}{n}),
\end{aligned}$$

where the last inequality is because $c > \frac{d}{2f_e}$. From this and formula (4), the proof is completed. □

Remark 1. The relatively loose upper bound $c^* = \frac{d}{2f_e}$ in the above proof may be improved by a factor of d by making a further distinction among the r-flowers that share forcing values at different number of shared variables. But for the purpose of showing that the flawless random CSP also has potential embedded easy sub-problems, our upper bound for the constraint-variable ratio c is sufficient since the domain size d is a constant.

6.2 Proof of Theorems 2 and 3

Given a CNF formula $\mathcal{F}$, we use $\mathrm{Res}(\mathcal{F})$ to denote the resolution complexity of $\mathcal{F}$, i.e., the length of the shortest resolution refutation of $\mathcal{F}$. The width of deriving a clause A from $\mathcal{F}$, denoted by $w(\mathcal{F} \vdash A)$, is defined to be the minimum over all the resolution refutations of the maximum clause size in a resolution refutation. The width $w(\mathcal{F})$ of a formula $\mathcal{F}$ is the largest size of the clauses in it. Ben-Sasson and Wigderson [16] established a relationship between $\mathrm{Res}(\mathcal{F})$ and $w(\mathcal{F} \vdash \emptyset)$:

$$\mathrm{Res}(\mathcal{F}) = e^{\Omega(\frac{(w(\mathcal{F}\vdash\emptyset)-w(\mathcal{F}))^2}{n})}.$$

This relationship indicates that to give an exponential lower bound on the resolution complexity, it is sufficient to show that every resolution refutation of $\mathcal{F}$ contains a clause whose size is linear in n, the number of variables.

Let $\mathcal{T}$ be an instance of the CSP and let $\mathrm{CNF}(\mathcal{T})$ be the CNF encoding of $\mathcal{T}$. Mitchell [11] provided a framework within which one can investigate the resolution complexity of $\mathcal{T}$, i.e., the resolution complexity of the CNF formula $\mathrm{CNF}(\mathcal{T})$ that encodes $\mathcal{T}$, by working directly on the structural properties of $\mathcal{T}$. A sub-instance $\mathcal{J}$ of $\mathcal{T}$ is a CSP instance such that $\mathrm{var}(\mathcal{J}) \subset \mathrm{var}(\mathcal{T})$ and

$\mathcal{J}$ contains all the constraints of $\mathcal{T}$ whose scope variables are in $\text{var}(\mathcal{J})$. The following crucial concepts make it possible to work directly on the structural properties of the CSP instance when investigating the resolution complexity of the encoding CNF formula.

Definition 7 (Implies. Defined in [11]). *For any assignment α to the variables in the CSP instance $\mathcal{T}$, we write $\hat{\alpha}$ for the truth assignment to the variables in* $\text{CNF}(\mathcal{T})$ *that assigns a variable $x : a$ the value TRUE if and only if $\alpha(x) = a$.*

Let C be a clause over the variables of $\text{CNF}(\mathcal{T})$*. We say that a sub-instance $\mathcal{J}$ of $\mathcal{T}$ implies C, denoted as $\mathcal{J} \models C$, if and only if for each assignment α satisfying $\mathcal{J}$, $\hat{\alpha}$ satisfies C.*

Definition 8 (Clause Complexity [11]). *Let $\mathcal{T}$ be a CSP instance. For each clause C defined over the Boolean variables in* $\text{var}(\text{CNF}(\mathcal{T}))$*, define*

$$\mu(C, \mathcal{T}) = \min\{|\text{var}(\mathcal{J})|; \mathcal{J} \textit{ is a sub-instance and implies } C\}.$$

The following two concepts slightly generalize those used in [11, 4] and enable us to have a uniform treatment when establishing resolution complexity lower bounds.

Definition 9 (Boundary). *The boundary $\mathcal{B}(\mathcal{J})$ of a sub-instance $\mathcal{J}$ is defined to be the set of CSP variables such that for each $x \in \mathcal{B}(\mathcal{J})$ if and only if the following is true: If $\mathcal{J}$ minimally implies a clause C defined on some boolean variables in* $\text{var}(\text{CNF}(\mathcal{T}))$*, then C contains at least one of the boolean variables, $x : a, a \in D$, that encode the CSP variable x.*

Definition 10 (Sub-critical Expansion [11]). *Let $\mathcal{T}$ be a CSP instance. The sub-critical expansion of $\mathcal{T}$ is defined as*

$$e(\mathcal{T}) = \max_{0 \le s \le \mu(\emptyset, \mathcal{T})} \min_{s/2 \le |\text{var}(\mathcal{J})| \le s} |\mathcal{B}(\mathcal{J})| \tag{5}$$

where the minimum is taken over all the sub-instances of $\mathcal{T}$ such that $s/2 \le |\text{var}(\mathcal{J})| \le s$.

The following theorem relates the resolution complexity of the CNF encoding of a CSP instance to the sub-critical expansion of the CSP instance.

Theorem 6. *[11] For any CSP instance $\mathcal{T}$, we have*

$$w(\text{CNF}(\mathcal{T}) \vdash \emptyset) \ge e(\mathcal{T}) \tag{6}$$

Proof. For any resolution refutation π of $\text{CNF}(\mathcal{T})$ and $s \le \mu(\emptyset, \mathcal{T})$, Lemma 1 of [11] shows that π must contain a clause C with

$$s/2 \le \mu(C, \mathcal{T}) \le s.$$

Let $\mathcal{J}$ be a sub-instance such that $|\text{var}(\mathcal{J})| = \mu(C, \mathcal{T})$ and $\mathcal{J}$ minimally implies C. Since $\mathcal{J}$ minimally implies C, according to the definition of the boundary, $w(C) \ge |\mathcal{B}(\mathcal{J})|$. (6) follows. □

To establish an asymptotically exponential lower bound on $\text{Res}(\mathcal{C})$ of a random CSP $\mathcal{C}$, it is enough to show that there is a constant $\beta^* > 0$ that does not depend on n such that

$$\lim_n P\{e(\mathcal{C}) \geq \beta^* n\} = 1. \tag{7}$$

For any $\alpha > 0$, let $\mathcal{A}_m(\alpha)$ be the event $\{\mu(\emptyset, \mathcal{C}) > \alpha n\}$ and $\mathcal{A}_s(\alpha, \beta^*)$ be the event

$$\left\{ \min_{\frac{\alpha n}{2} \leq |\text{var}(\mathcal{J})| \leq \alpha n} \mathcal{B}(\mathcal{J}) \geq \beta^* n \right\}.$$

Notice that

$$\begin{aligned} P\{e(\mathcal{C}) \geq \beta^* n\} &\geq P\{\mathcal{A}_m(\alpha) \cap \mathcal{A}_s(\alpha, \beta^*)\} \\ &\geq 1 - P\{\overline{\mathcal{A}_m(\alpha)}\} - P\{\overline{\mathcal{A}_s(\alpha, \beta^*)}\}. \end{aligned} \tag{8}$$

We only need to find appropriate α^* and β^* such that

$$\lim_n P\{\overline{\mathcal{A}_m(\alpha^*)}\} = 0 \tag{9}$$

and

$$\lim_n P\{\overline{\mathcal{A}_s(\alpha^*, \beta^*)}\} = 0. \tag{10}$$

Event $\mathcal{A}_m(\alpha^*)$ is about the size of minimally unsatisfiable sub-instances. For the event $\mathcal{A}_s(\alpha^*, \beta^*)$, a common practice is to identify a special subset of boundaries and show that this subset is large. For different random CSP models and under different assumptions on the model parameters, there are different ways to achieve this. Following [12], we say a graph G is (r, q)-dense if there is a subset of r vertices that induces at least r edges of G.

*Proof (***Proof of Theorem 2).** Recall that the constraint graph of $\mathcal{B}^{d,t}_{n,m}[SC]$ is the standard random graph $G(n, m)$. Since each instance of $\mathcal{B}^{d,t}_{n,m}[SC]$ is strongly k-consistent, variables in a minimal unsatisfiable sub-instance $\mathcal{J}$ with $|\text{var}(\mathcal{J})| = r$ must have a vertex degree greater than or equal to k, and consequently, the constraint sub-graph $H(\mathcal{J})$ must contains at least $\frac{rk}{2}$ edges. Thus,

$$\begin{aligned} P\{\overline{\mathcal{A}_m(\alpha^*)}\} &= P\{\mu(\emptyset, \mathcal{B}^{d,t}_{n,m}[SC]) \leq \alpha^* n\} \\ &\leq P\left\{ \bigcup_{r=k+1}^{\alpha^* n} \{G(n, m) \text{ is } (r, rk/2)\text{-dense }\} \right\}. \end{aligned}$$

Let $\mathcal{B}^k(\mathcal{J})$ be the set of variables in $\text{var}(\mathcal{J})$ whose vertex degrees are less than k. Again, since instances of $\mathcal{B}^{d,t}_{n,m}[SC]$ are always strongly k-consistent, we have $\mathcal{B}^k(\mathcal{J}) \subset \mathcal{B}(\mathcal{J})$ and thus, $|\mathcal{B}(\mathcal{J})| \geq |\mathcal{B}^k(\mathcal{J})|$. Therefore, the probability $P\{\overline{\mathcal{A}_s(\alpha^*, \beta^*)}\}$ can be bounded as

$$P\{\overline{\mathcal{A}_s(\alpha^*, \beta^*)}\} \leq P\{\overline{\mathcal{A}^k_s(\alpha^*, \beta^*)}\}$$

where $\mathcal{A}_s^k(\alpha^*, \beta^*)$ is the event

$$\left\{ \min_{\alpha^* n/2 \le |\text{var}(\mathcal{J})| \le \alpha^* n} \mathcal{B}^k(\mathcal{J}) \ge \beta^* n \right\}.$$

Random graph arguments (see, e.g. [12]) show that there exist constants α^* and β^* such that $P\{\overline{\mathcal{A}_m(\alpha^*})\}$ and $P\{\overline{\mathcal{A}_s^k(\alpha^*, \beta^*)}\}$ both tend to 0. Indeed, let β^* be such that $\frac{(1-\beta^*)k}{2} > 1$, $c = \frac{m}{n}$, and $N = \frac{n(n-1)}{2}$. We have

$$\begin{aligned}
P\{\overline{\mathcal{A}_m(\alpha^*)}\} &\le P\left\{ \bigcup_{r=k+1}^{\alpha^* n} \{G(n,m) \text{ is } (r, rk/2)\text{-dense }\} \right\} \\
&\le \sum_{r=k+1}^{\alpha^* n} P\{G(n,m) \text{ is } (r, \frac{rk}{2})\text{-dense}\} \\
&\le \sum_{r=k+1}^{\alpha^* n} \binom{n}{r} \binom{\frac{r(r-1)}{2}}{\frac{rk}{2}} \binom{N-\frac{rk}{2}}{m-\frac{rk}{2}} \binom{N}{m}^{-1} \\
&\le \sum_{r=k+1}^{\alpha^* n} (\frac{en}{r})^r (\frac{e(r-1)}{k})^{\frac{rk}{2}} (\frac{2c}{n})^{\frac{rk}{2}} \\
&= \sum_{r=k+1}^{\alpha^* n} \left[\frac{en}{r} (\frac{2ec(r-1)}{kn})^{\frac{k}{2}} \right]^r \\
&= \sum_{r=k+1}^{\alpha^* n} \left[(\frac{k}{2})^{\frac{k}{2}} e^{\frac{k+2}{2}} c^{\frac{k}{2}} (\frac{r}{n})^{\frac{k-2}{2}} \right]^r \\
&\le \sum_{r=k+1}^{\lfloor \log n \rfloor} \left[(\frac{k}{2})^{\frac{k}{2}} e^{\frac{k+2}{2}} c^{\frac{k}{2}} (\frac{\log n}{n})^{\frac{k-2}{2}} \right] \\
&\quad + \sum_{r=\lfloor \log n \rfloor}^{\alpha^* n} \left[(\frac{k}{2})^{\frac{k}{2}} e^{\frac{k+2}{2}} c^{\frac{k}{2}} (\alpha^*)^{\frac{k-2}{2}} \right]^{\log n} \qquad (11)
\end{aligned}$$

Similarly, we have for $\overline{\beta} = \frac{2\beta^*}{\alpha^*}$,

$$\begin{aligned}
P\{\overline{\mathcal{A}_s^k(\alpha^*, \beta^*)}\} &= P\left\{ \bigcup_{r=\frac{\alpha^* n}{2}}^{\alpha^* n} \{\exists \text{ a size-r sub-instance } \mathcal{J} \text{ s.t. } |\mathcal{B}^k(\mathcal{J})| \le \beta^* n\} \right\} \\
&\le P\left\{ \bigcup_{r=\frac{\alpha^* n}{2}}^{\alpha^* n} \{G(n,m) \text{ is } (r, \frac{r(1-\overline{\beta})k}{2})\text{-dense}\} \right\} \\
&\le \sum_{r=\frac{\alpha^* n}{2}}^{\alpha^* n} \left[\left(\frac{2c}{(1-\overline{\beta})k} \right)^{\frac{(1-\overline{\beta})k}{2}} e^{\frac{(1-\overline{\beta})k+2}{2}} (\alpha^*)^{\frac{(1-\overline{\beta})k-2}{2}} \right]^r \qquad (12)
\end{aligned}$$

where the second inequality is because of the fact that for a sub-instance $\mathcal{J}$ with size r and $|\mathcal{B}^k(\mathcal{J})| \leq \beta^* n$, its constraint graph contains at least $r - \beta^* n = r - \frac{\alpha^*}{2}\overline{\beta} n \geq r - \overline{\beta} r$ vertices whose degree is at least k.

There exist α^* and β^* be such that (1) $\frac{2\beta^*}{\alpha^*} < 1$; (2) $\frac{(1-\beta^*)k}{2} > 1$; and (3) the right hand side of formula (11) and the right hand side of formula (12) both tend to zero. This completes the proof of Theorem 2. □

We now prove Theorem 3. First from the definition of $\mathcal{B}^{d,t}_{n,m}[WC]$, we have the following

Lemma 2. *For the random CSP $\mathcal{B}^{d,t}_{n,m}[WC]$, we have*

1. *Every sub-instance whose constraint graph is a cycle is satisfiable;*
2. *For any path of length ≥ 3, any compatible assignments to the two variables at the ends of the path can be extended to assignments that satisfy the whole path.*

In an effort to establish exponential lower bounds on the resolution complexity for a classical random CSP models with a tightness higher than those in [11], Molloy and Salavatipour [4] introduced a collection of sub-instances, denoted here as $\mathcal{B}_M(\mathcal{J})$, and used its size to give a lower bound on the size of the boundary. For binary CSPs whose constraints are arc-consistent and contain no forcer, $\mathcal{B}_M(\mathcal{J})$ consists of two parts: $\mathcal{B}^1_M(\mathcal{J})$ and $\mathcal{B}^2_M(\mathcal{J})$, defined respectively as follows:

1. $\mathcal{B}^1_M(\mathcal{J})$ contains the set of single-edge sub-instances $\mathcal{X}$, i.e., var$(\mathcal{X}) = 2$, such that at least one of the variables has a vertex degree 1 in the original constraint graph;
2. $\mathcal{B}^2_M(\mathcal{J})$ contains the set of sub-instances $\mathcal{X}$ whose induced constraint graph is a pendant path of length 4, i.e., a path of length 4 such that no vertex other than the endpoints has a vertex degree greater than 2 in the original constraint graph.

It can be shown that

Lemma 3 ([4]). *For any weakly path-consistent CSP sub-instance $\mathcal{J}$, we have*

$$|\mathcal{B}(\mathcal{J})| \geq |\mathcal{B}^1_M(\mathcal{J})| + \frac{|\mathcal{B}^2_M(\mathcal{J})|}{4}.$$

Proof. The variable with degree one in any sub-instance in $\mathcal{B}^1_M(\mathcal{J})$ has to be in $\mathcal{B}(\mathcal{J})$; At least one internal variable in any pendant path $\mathcal{B}^2_M(\mathcal{J})$ has to be in $\mathcal{B}(\mathcal{J})$. It is possible that several pendant paths of length 4 share a common internal variable that is in $\mathcal{B}(\mathcal{J})$, e.g., in a very long pendant path. But a variable can only appear in at most three pendant paths of length 4.

With the above preparations, the proof provided for Theorem 1 of [4] readily applies to our case. To make this report self-contained, we give the proof below.

*Proof (***Proof of Theorem 3).** By Lemma 2, any minimally unsatisfiable sub-instance $\mathcal{J}$ is such that (1) its constraint graph cannot be a single cycle; and (2) $\mathcal{B}_M(\mathcal{J})$ is empty since $|\mathcal{B}_M^1(\mathcal{J})| = 0$ and $|\mathcal{B}_M^2(\mathcal{J})| = 0$ for a minimally unsatisfiable sub-instance. According to Lemma 11 of [4], the constraint graph of $\mathcal{J}$ has at least $(1 + \frac{1}{12})\text{var}(\mathcal{J})$ edges. Therefore, due to the locally sparse property of random graphs (e.g., Lemma 10 in [4]), there is a constant $\alpha^* > 0$ such that formula (9) holds, i.e.,

$$\lim_n P\{\overline{\mathcal{A}_m(\alpha^*)}\} = 0.$$

To establish formula (10), due to Lemma 3 we have

$$P\{\mathcal{A}_s(\alpha^*, \beta^*)\} \geq P\{\mathcal{A}_{s,M}(\alpha^*, \beta^*)\}$$

where $\mathcal{A}_{s,M}(\alpha^*, \beta^*)$ is the event

$$\left\{ \min_{\alpha^* n/2 \leq |\text{var}(\mathcal{J})| \leq \alpha^* n} |\mathcal{B}_M(\mathcal{J})| \geq \beta^* n \right\}.$$

Now suppose on the contrary that there is a sub-instance $\mathcal{J}$ with $\alpha^* n/2 \leq |\text{var}(\mathcal{J})| \leq \alpha^*$ such that $|\mathcal{B}_M^1(\mathcal{J})| + |\mathcal{B}_M^2(\mathcal{J})| \leq \zeta n$. Then, from Lemmas 10 and 11 of [4], the constraint graph of $\mathcal{J}$ contains only cycle components——Lemma 11 of [4] asserts that the edges-to-vertices ratio of the constraint graph of $\mathcal{J}$ has to be bigger than one. If we remove all the cycle components from the constraint graph of $\mathcal{J}$, the edges-to-vertices ratio of the remaining graph becomes even bigger. But this is impossible from Lemma 10 of [4] because the constraint graph of $\mathcal{J}$, and hence the remaining graph, has less than $\alpha^* n$ vertices.

It is well-known that w.h.p. a random graph has fewer than $\log n$ cycle components of length at most 4——for the random graph $G(m, n)$ with $m/n = c$ being constant, the number of cycle components with a fixed length has asymptotically Poisson distribution [17]. Thus, the number of variables that are in cycle components of length 4 is at most $4 \log n$. Since any cycle component of length $l > 4$ contain l pendant paths of length 4, the total number of variables in cycle components of length greater than 4 is at most $|\mathcal{B}_M^2(\mathcal{J})| < \zeta n$. Therefore, we have $\text{var}(\mathcal{J}) < \zeta n + 4 \log n < \alpha^* n/2 \leq \text{var}(\mathcal{J})$ for sufficiently small ζ, a contradictory.

We, therefore, conclude that there is a β^* such that w.h.p, for any sub-instance $\mathcal{J}$ with $\alpha^* n/2 \leq |\text{var}(\mathcal{J})| \leq \alpha^*$, $|\mathcal{B}_M(\mathcal{J})| \geq \beta^* n$, i.e., formula (10) holds. □

References

1. Achlioptas, D., L.M.Kirousis, E.Kranakis, D.Krizanc, M.Molloy, Y.C.Stamation: Random constraint satisfaction: A more accurate picture. In: Proceedings of CP97, Springer (1997) 107–120
2. Gent, I., MacIntyre, E., Prosser, P., Smith, B., Walsh, T.: Random constraint satisfaction: Flaws and structure. Constraints **6** (2001) 345–372

3. Gao, Y., Culberson, J.: Resolution complexity of random constraint satisfaction problems: Another half of the story. In: LICS'03 Workshop on Typical Case Complexity and Phase Transitions. (2003)
4. Molloy, M., Salavatipour, M.: The resolution complexity of random constraint satisfaction problems. In: Proceedings of FOCS 2003. (2003)
5. Xu, K., Li, W.: Exact phase transitions in random constraint satisfaction problems. Journal of Artificial Intelligence Research **12** (2000) 93–103
6. Smith, B.M.: Constructing an asymptotic phase transition in random binary constraint satisfaction problems. Theoretical Computer Science **265** (2001) 265–283
7. Mackworth, A.K.: Consistency in netowrks of relations. Artificial Intelligence **8** (1977) 99–118
8. Chvatal, V., Szemeredi, E.: Many hard examples for resolution. Journal of the Association for Computing Machinery **35** (1988) 759–768
9. Beame, P., Karp, R.M., Pitassi, T., Saks, M.E.: On the complexity of unsatisfiability proofs for random k -CNF formulas. In: ACM Symposium on Theory of Computing. (1998) 561–571
10. Achlioptas, D., Beame, P., Molloy, M.: A sharp threshold in proof complexity. In: ACM Symposium on Theory of Computing. (2001) 337–346
11. Mitchell, D.: Resolution complexity of random constraints. In: Proceedings Principles and Practices of Constraint Programming (CP-2002), Springer (2002) 295–309
12. Beame, P., Culberosn, J., Mitchell, D., Moore, C.: The resolution complexity of random graph k-colorability. Electronic Colloquium on Computational Complexity, TR04-012 (2004)
13. Zhang, L., Madigan, C., Moskewicz, M., Malik, S.: Efficient conflict driven learning in a boolean satisfiability solver. In: Proceedings of International Conference on Computer Aided Design (ICCAD2001). (2001)
14. Mitchell, D.G.: The Resolution Complexity of Constraint Satisfaction. PhD thesis, Department of Computer Science, University of Toronto, Canada (2002)
15. Franco, J., Gelder, A.V.: A perspective on certain polynomial-time solvable classes of satisfiability. Discrete Applied Mathematics **125** (2003) 177–214
16. Ben-Sasson, E., Wigderson, A.: Short proofs are narrow - resolution made simple. Journal of ACM **49** (2001)
17. Bollobas, B.: Random Graphs. Cambridge University Press (2001)