Resolution Complexity of Random Constraint Satisfaction Problems: Another Half of the Story

Yong Gao and Joseph Culberson

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

Abstract. Let $\mathcal{C}^{2,k,t}_{n,cn}$ be a random constraint satisfaction problem(CSP) of n binary variables, where $c > 0$ is a fixed constant and the cn constraints are selected uniformly and independently from all the possible k-ary constraints each of which contains exactly t tuples of the values as its restrictions. We establish upper bounds for the tightness threshold for $\mathcal{C}^{2,k,t}_{n,cn}$ to have an exponential resolution complexity. The upper bounds partly answers the open problems regarding the CSP resolution complexity with the tightness between the existing upper and lower bound [1].

1 Introduction

Phase transition and threshold phenomena in NP complete problems have been extensively investigated. Many problems such as propositional satisfiability (SAT), graph coloring, and the constraint satisfaction problem (CSP), have been shown to have a solubility threshold under various random models. Over the past ten years, much attention has been paid to the identification of the exact value of the threshold and/or the upper and lower bounds for the threshold [2–4]. Recently, research interest started to switch to analytical investigation of the links between the solubility threshold phenomena and the algorithmic complexity to solve these NP complete problems.

In the study of the phase transition of CSPs, many natural models of random CSPs have been proposed, but not all of them are guaranteed to exhibit a threshold. A detailed discussion of the random models of CSPs and their limitations can be found in [5–7].

In this paper, we consider $\mathcal{C}^{2,k,t}_{n,cn}$, a random binary CSP model defined on n binary variables where the $cn, c > 0$, constraints are selected uniformly and independently from all the possible k-ary constraints each of which excludes exactly t tuples of the values. In [6], it is shown that for any $c > 0$, if $t \geq 2^{k-1}$, then $\mathcal{C}^{2,k,t}_{n,cn}$ is flawed in the sense that it is almost always trivially unsatisfiable and can be checked in linear time. In [1], Mitchell shows that for $0 < t < k-1$, the resolution complexity of $\mathcal{C}^{2,k,t}_{n,cn}$ is almost surely exponential. A similar exponential complexity result has also been established in [8] under a different CSP random model. The main result of this paper is a set of tightness upper bounds for the threshold of exponential complexity of $\mathcal{C}^{2,k,t}_{n,cn}$. These upper bounds partly answer the open problems regarding the CSP resolution complexity with the tightness between the existing upper and lower bound [1, 6].

In the study of the resolution complexity of SAT, there has been much interest in the necessary clause density at which unsatisfiable SAT instances can be recognized polynomially [9, 10].

Currently, the best result shows that there are polynomial algorithms to certify unsatisfiable random k-SAT instances with at least $n^{k/2+o(1)}$ clauses [10]. Since a binary CSP is naturally equivalent to a SAT problem, our result shows that $\mathcal{C}_{n,cn}^{2,k,t}$ is an alternative random SAT model in which instances with $O(n)$ clauses can be recognized as unsatisfiable polynomially.

The rest of the paper is organized as follows. In the next section, we introduce basic concepts related to CSPs and their random models. In section 3, we present our results with some discussion. Section 4 is devoted to the proof of the results.

2 Preliminaries

Throughout this paper, we consider binary CSPs defined on n variables $x = (x_1, \cdots, x_n)$, each of which has $\mathcal{D} = \{0, 1\}$ as its domain. A k-ary relation over $\mathcal{D}$ is a map $R : \mathcal{D}^k \to \{0, 1\}$. The set $R^{-1}(0) = \{x \in \mathcal{D} : R(x) = 0\}$ is called the set of restrictions defined by the k-ary relation R and $|R^{-1}(0)|$ is called the *tightness* of the relation.

A binary CSP $\mathcal{C}$ consists of a set of binary variables $x = (x_1, \cdots, x_n)$ and a set of *constraints* $(C_1, \cdots, C_m)$. Each constraint C_i is specified by its scope, a subset of the variables x, and a relation R_{C_i} which gives a set of restrictions on the scope variables. The size of the scope of a constraint C is denoted by $|C|$. Associated with a CSP is the *constraint hypergraph* with vertices corresponding to the set of variables and edges corresponding to the set of constraint scopes.

An assignment to the variables $x = (x_1, \cdots, x_n)$ is a solution to the CSP if it satisfies all the relations associated with the set of constraints. A CSP is called satisfiable if there is at least one satisfiable assignment. Throughout the rest of the paper, we assume that all the constraints of a CSP have the same scope size, and use the following notation:

1. n, the number of variables; m, the number of constraints;
2. k, the scope size of a constraint; t, the tightness of a constraint.

Consequently, the constraint hypergraph will be always k-uniform.

Definition 1. Random CSPs *Let $0 < t < 2^k$ be an integer and $c > 0$ a real number. The random model $\mathcal{C}_{n,m}^{2,k,t}, m = cn$, of CSPs specifies a random instance of CSPs by first selecting a set of $m = cn$ variable scopes randomly without replacement from the set of $\binom{n}{k}$ subsets of variables, and then for each scope, choosing a relation R over the scope variables uniformly from all the possible $\binom{2^k}{2^k-t}$ relations.*

The random CSP model $\mathcal{C}_{n,m}^{2,k,t}, m = cn$, can be generalized to allow for non-integer tightness t as follows. For an integer t, the constraints are constructed as usual. For a non-integer $t = t_0 + \alpha$, where t_0 is an integer and $0 < \alpha < 1$, a constraint selects a random set of restrictions of size t_0 with probability $1 - \alpha$ and a random set of restrictions of size $t_0 + 1$ with probability α.

3 Main Results

In this section, we present our main results with some discussion.

Theorem 1. *Let $\mathcal{C}_{n,cn}^{2,k,t}$ be a random CSP. Then, we have*

$$\lim_{n\to\infty} Pr\{\mathcal{C}_{n,cn}^{2,k,t} \textit{ is satisfiable }\} = 0$$

if c and t satisfy one of the following

1. *For $t = 2^{k-2} - 1 + \alpha, 0 < \alpha < 1$,*
$$c > \frac{\binom{2^k}{2^{k-2}}}{2k(k-1)\alpha} \tag{1}$$
2. *For $t = 2^{k-2} + j + \alpha, 0 < \alpha < 1, \quad 0 \le j \le 2^{k-1} - 2^{k-2} - 1$,*
$$c > \frac{1}{k(k-1)} \frac{\binom{2^k}{2^{k-2}}}{\binom{2^{k-2}+j}{2^{k-2}}} (1 + \alpha \frac{2^{k-2}}{j+1})^{-1}. \tag{2}$$

The theorem is proved by showing that for any tightness t satisfying 2, a random instance of $\mathcal{C}_{n,cn}^{2,k,t}$ almost surely implies an unsatisfiable 2-SAT subproblem. The intuition is that a constraint C with t restrictions is equivalent to a 3-CNF formula with t clauses defined on exactly three variables. If $t > 2^{k-2}$, there is a non-zero probability that these t clauses imply a 2-clause. As a result, if there are enough constraints, we will get enough implied 2-clauses to form an unsatisfiable 2-CNF formula in a form called the criss-cross loop. In fact, this situation has been shown to be true in [11] in a different context where the so-called NK landscape model is analyzed. An NK landscape model can be viewed as a special random CSP where the number of constraints is equal to the number of the variables and each constraint contains a unique variable as one of its scope variables.

Since the resolution complexity of an unsatisfiable 2-SAT problem is polynomial, we have

Corollary 1. *For any t and c satisfying the conditions in Theorem 1, the resolution complexity of $\mathcal{C}_{n,cn}^{2,k,t}$ is almost surely polynomial.*

	Resolution Complexity of $\mathcal{C}_{n,cn}^{2,k,t}$			
Scope Size	$2^{\Omega(n)}$ ([1])	Unknown	Polynomial for certain c (this paper)	Linear ([6])
3	$\{1\}$		$(1,4)$	$[4,8]$
4	$[1,2]$	$(2,3]$	$(3,8)$	$[8,16]$
5	$[1,3]$	$(3,7]$	$(7,16)$	$[16,32]$
k	$[1,k-2]$	$(k-1,2^{k-2}-1]$	$(2^{k-2}-1,2^{k-1})$	$[2^{k-1},2^k]$

Table 1. Ranges of tightness with different complexity

From Theorem 1, we can see that for a given tightness $2^{k-2}-1 < t < 2^{k-1}$, the resolution complexity for the random CSP $\mathcal{C}^{2,k,t}_{n,cn}$ is polynomial if the constraint-to-clause ratio is larger than a certain value. This partly answers the open problems regarding the resolution complexity of random CSP inside the tightness interval $k-2 < t < 2^{k-1}$ ([1]). For $k=3$, $c > \frac{7}{3}$, and integer tightness t, our results actually show that $t=2$ is the exact tightness threshold for the exponential resolution complexity. Table 1 shows the current status of the tightness interval of different resolution complexity. The first and the last columns are from [1].

The existence of upper bounds characterized by unsatisfiable 2-SAT subproblems raises concerns that $\mathcal{C}^{2,k,t}_{n,cn}$ might be still flawed even if the tightness t is less that 2^{k-1}. However, this is not the case. Using a random hypergraph argument and the fact that a 2-clause cycle is satisfiable, it can be shown that for any fixed $t \leq 2^{k-1}-1$, $\mathcal{C}^{2,k,t}_{n,cn}$ does have a phase transition with a threshold lower bounded by $\frac{1}{k(k-1)}$.

Theorem 2. *For any fixed $t \leq 2^{k-1}-1$ and $c < \frac{1}{k(k-1)}$, $\mathcal{C}^{2,k,t}_{n,cn}$ is almost surely satisfiable.*

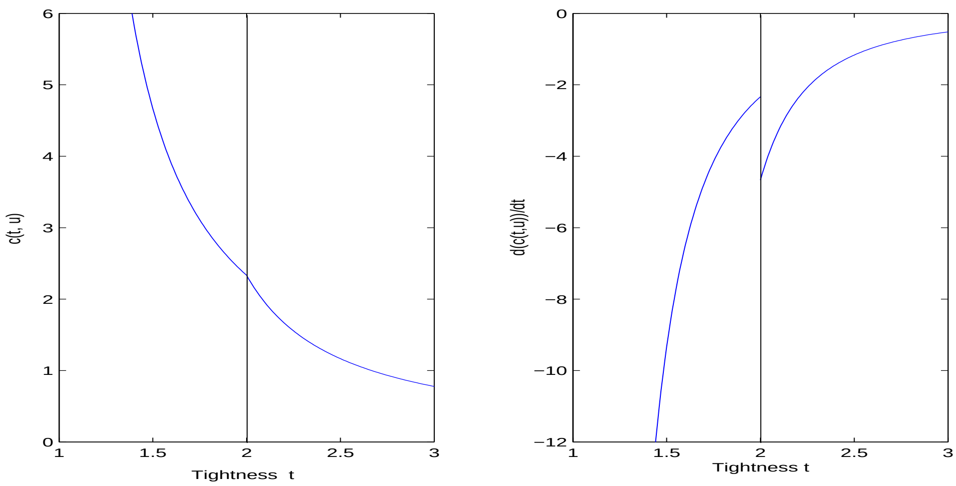

Fig. 1. The upper bound $c_3(t,u)$ for the threshold $c_3(t)$ as a function of tightness t. Left figure: the function itself. Right figure the derivative of the function.

Having established that $\mathcal{C}^{2,k,t}_{n,cn}$ has a phase transition, it is obvious that the tightness t serves almost the same role as the parameter p in the (2+p)-SAT to model the gradual changing from the first order transition to the second order transition. For each fixed tightness $1 \leq t \leq 2^{k-1}-1$, let $c_k(t)$ be the constraint-to-variable ratio threshold of the satisfiability transition. When $t=1$, we get the k-SAT model, and hence, $c_k(1)$ is exactly the k-SAT threshold. As t gradually increases, $c_k(t)$ decreases to a limit value larger than or equal to $\frac{1}{k(k-1)}$, continuously or discontinuously. Theorems 1 and 2 indicate that for random CSPs, it is possible to have any types of easy-hard complexity pattern if we can pick an appropriate tightness and ratio relation. The property of the threshold as a function of the clause-to-variable ratio and the tightness deserves further investigation and the behavior of the upper bounds in Theorem 1, as depicted in Figure 1, is suggestive.

4 Proof of the Results

4.1 Proof of Theorem 1

First, we need some definitions that are used to characterize unsatisfiable 2-SAT problems.

Definition 2. *Given a vertex set $U = \{u_0, u_1, \cdots, u_{3p+1}\}$ with the size $|U| = l = 3p + 2$, a k-criss-cross loop (k-cc-loop)is a k-uniform hypergraph $\mathcal{L}(U, E)$ with the set of hyperedges $E = \{E_1, \cdots, E_l\}$ defined as*

$$\begin{aligned} &E_i = (u_i, u_{i-1}) \cup L_i, 1 \leq i \leq p \text{ or } p+2 \leq i \leq 3p, \\ &E_{p+1} = (u_0, u_{p+1}) \cup L_{p+1}, E_{3p+1} = (u_0, u_p) \cup L_{3p+1}, \\ &E_{3p+2} = (u_0, u_{p+1}) \cup L_{3p+2} \end{aligned}$$

where $L_i, 1 \leq i \leq 3p+2$ is a sequence of vertex subsets of size $k-2$ such that $\{u_i, 1 \leq i \leq l-1\}$ and $L_i, 1 \leq i \leq l$ are mutually disjoint.

In a hypergraph of k-cc-loop, there are exactly two cycles touched at the special vertex u_0. This construct was first proposed by Franco in [9] and is closely related to the notion of simple cycle used in the study of the phase transition of random 2-SAT.

Definition 3. *Let $\mathcal{L}(U, E)$ be a k-cc-loop and $\mathcal{C}$ be a set of constraints each of which corresponds to an hyperedge of $\mathcal{L}(U, E)$. We say $\mathcal{C}$ is a reducible k-cc-loop on $\mathcal{L}(U, E)$ if each constraint C_i implies a 2CNF clause defined on the cyclic variables such that the resulting 2CNF clauses form two contradictory cycles, making the formula unsatisfiable.*

Lemma 1. *Let $\mathcal{C}_{n,cn}^{2,k,t}$ be a random CSP. Let $V = \{v_0, v_1, \cdots, v_{3p+1}\}$ be an ordered subsequence of variables and $L_i, 1 \leq i \leq l$ be an ordered sequence of subsets of variables that are mutually disjoint and disjoint with V. Then, the probability that $\mathcal{C}_{n,cn}^{2,k,t}$ contains a reducible k-cc-loop defined by $V, L_i, 1 \leq i \leq l$, is*

$$\frac{1}{4}\left(\frac{2rck!}{n^{k-1}}\right)^l O(1),$$

where r is such that

1. *For $t = 2^{k-2} - 1 + \alpha$ with $0 < \alpha < 1$,*

$$r = \frac{1}{\binom{2^k}{2^{k-2}}}(1 + 2^{k-2}\alpha),$$

2. *For $t = 2^{k-2} + j + \alpha$ with $0 \le \alpha < 1$ and $0 \le j \le 2^{k-1} - 2^{k-2} - 1$,*

$$r = \frac{\binom{2^{k-2}+j}{2^{k-2}}}{\binom{2^k}{2^{k-2}}}(1 + \alpha\frac{2^{k-2}}{j+1})$$

Proof. Let $N = \binom{n}{k}$ be the number of possible hyperedges. Then the probability that $\mathcal{C}^{2,k,t}_{n,cn}$ contains the sequence of constraints defined on $V, L_i, 1 \le i \le l$ is

$$\frac{1}{\binom{N}{cn}}\binom{N-l}{cn-l}. \tag{3}$$

For a given sequence $(u_0, u_1, \cdots, u_{l-2})$ of literals of the variables $(v_0, v_1, \cdots, v_{l-2})$ and a constraint C containing two variables v_i and v_j as its scope variables, we calculate the probability that C implies the clause $u_i \vee u_j$. We focus on the second case, i.e., $t = 2^{k-2} + j + \alpha$ with $0 \le \alpha < 1$ and $0 \le j \le 2^{k-1} - 2^{k-2} - 1$, and the first case of $t = 2^{k-2} - 1 + \alpha$ can be obtained similarly. Recall that a constraint selects a restriction set of size $t = 2^{k-2} + j$ with probability $1 - \alpha$ and of size $t = 2^{k-2} + j + 1$ with probability α.

As we are dealing with binary constraints, it is easy to see that the constraint C implies the claus $u_i \vee u_j$ if and only if the set of restrictions contains the set of 2^{k-2} binary vectors $(u_i, u_j, *)$ with $*$ being any binary vectors in $\{0,1\}^{k-2}$. Therefore, the probability that C implies the clause $u_i \vee u_j$ is

$$\begin{aligned} r &= \frac{\binom{2^k-2^{k-2}}{j}}{\binom{2^k}{2^{k-2}+j}}(1-\alpha) + \frac{\binom{2^k-2^{k-2}}{j+1}}{\binom{2^k}{2^{k-2}+j+1}}\alpha \\ &= \frac{\binom{2^{k-2}+j}{2^{k-2}}}{\binom{2^k}{2^{k-2}}}(1 + \alpha\frac{2^{k-2}}{j+1}). \end{aligned} \tag{4}$$

Since there are $l-2$ ways to select the literal sequences (both of the positive and negative literals of special variable v_0 have to appear) and the constraints select their restrictions independently, the probability that the sequence of constraints is a reducible k-cc-loop is

$$r^l 2^{l-2}. \tag{5}$$

The lemma is proved by combining (3), (4), and (5). □

Lemma 2. *Let $t = 2^{k-2} + \alpha, 0 \le \alpha < 1$. The expected number of reducible k-cc-loops in the random CSP $\mathcal{C}^{2,k,t}_{n,cn}$ is*

$$\frac{1}{4n}(2rck(k-1))^l O(1)$$

where r the same as in Lemma 1.

Proof. Let $V = \{v_0, v_1, \cdots, v_{3p+1}\}$ be an ordered subsequence of variables and $L_i, 1 \le i \le l$, be an ordered sequence of subsets of variables that are mutually disjoint and disjoint with V.

From lemma 1, the probability that the CSP contains a reducible k-cc-loop on V and L_i is

$$\frac{1}{4}\left(\frac{2rck!}{n^{k-1}}\right)^l O(1)$$

The number of ways of choosing the ordered sequence (V, L_i) is

$$\begin{aligned}
&\binom{n}{l-1}(l-1)!\prod_{i=0}^{l-1}\binom{n-l+1-(k-2)i}{k-2} \\
&= \binom{n}{l-1}(l-1)!\frac{1}{((k-2)!)^l}\frac{(n-l+1)!}{(n-l+1-l(k-2))!} \\
&\sim n^l n^{l(k-2)} = n^{l(k-1)}
\end{aligned}$$

where the term $\prod_{i=0}^{l-1}\binom{n-l+1-(k-2)i}{k-2}$ is the total number of ways to choose the sequence L_i. □

Proof of Theorem 1. Assume that $t = 2^{k-2} + \alpha$ with $0 < \alpha < 1$. Let $p = \ln^2 n$ so that $l = O(\ln^2 n)$. Let A_l be the number of reducible k-cc-loops contained in $\mathcal{C}^{2,k,t}_{n,cn}$. We need to show that $Pr\{A_l > 0\} > 0$ for sufficiently large n. Lemma 2 tells us that

$$\lim_{n\to\infty} \mathcal{E}\{A_l\} = \infty.$$

We go with the second moment method. To do so, we claim that

$$var(A_l) = o(\mathcal{E}\{A_l\}^2).$$

For an ordered sequence of variables and its associated sequence of subsets of variables $\mathcal{L} = (V = (v_0, \cdots, v_{l-2}), L_i, 1 \le i \le l$, let $I_{\mathcal{L}}$ be the indicator function of the event that $\mathcal{C}^{2,k,t}_{n,cn}$ contains a reducible k-cc-loop on $\mathcal{L}$. Then, $A_l = \sum_{\mathcal{L}} I_{\mathcal{L}}$ with the sum over all the possible choices of $\mathcal{L}$. We have

$$var(A_l) = \sum_{\mathcal{L}} var(I_{\mathcal{L}}) + \sum_{\mathcal{L}\neq\mathcal{M}} (\mathcal{E}[I_{\mathcal{L}}I_{\mathcal{M}}] - \mathcal{E}[I_{\mathcal{L}}]\mathcal{E}[I_{\mathcal{M}}]).$$

By the proof of lemma 2,

$$\mathcal{E}^2[A_l] = (\frac{1}{4n}(2rck(k-1))^l)^2 O(1).$$

It is easy to see that $\sum_{\mathcal{L}} var(I_{\mathcal{L}}) = o(\mathcal{E}^2[A_l])$. We will prove that

$$\sum_{\mathcal{L}\neq\mathcal{M}} \mathcal{E}[I_{\mathcal{L}}I_{\mathcal{M}}] = o(\mathcal{E}^2[A_l]). \tag{6}$$

Let $\mathcal{L}_1 = (V_1, L^1_i, 1 \le i \le l)$ and $\mathcal{L}_2 = (V_2, L^2_i, 1 \le i \le l)$ be two ordered sequences of variables and associated sequences of subsets of variables. Consider two sets of hyperedges obtained in the same way as that in Definition 2. We say that $\mathcal{L}_1$ and $\mathcal{L}_2$ share q hyperedges if the two sets of hyperedges have q hyperedges in common.

Assume that $\mathcal{L}_1$ and $\mathcal{L}_2$ share q hyperedges. Similar to the proof of lemma 1, we have

$$\begin{aligned}\mathcal{E}[I_{\mathcal{L}_1}|I_{\mathcal{L}_2}] &\leq \frac{1}{\binom{N-l}{cn-l}}\binom{N-2l+q}{cn-2l+q}r^{l-q}2^{l-q-2} \qquad (7)\\ &= \frac{1}{4}\left(\frac{2rck!}{n^{k-1}}\right)^{l-q} O(1) \qquad (8)\end{aligned}$$

Therefore, from lemma 1

$$\mathcal{E}[I_{\mathcal{L}_1}I_{\mathcal{L}_2}] = \left(\frac{2rck!}{n^{k-1}}\right)^{2l-q} O(1).$$

To prove (6), we need to count the number of pairs of k-cc-loops sharing q hyperedges. The idea of the counting is similar to those used in [9]. The following concepts about the cycle nodes in a k-cc-loop are required. Let $\mathcal{L}$ be a k-cc-loop and S a set of hyperedges in $\mathcal{L}$. We call a cycle node appearing in $\mathcal{L}$

1. *fixed* if it belongs to at least two hyperedges in S;
2. *limited* if it belongs to one hyperedges in S; and
3. *free* if it does not appear in any edges in S.

We need to consider two different cases: (1) The set of shared hyperedges is connected; and (2) The set of shared hyperedges has $h \geq 2$ connected components. In each of the cases, we also need to distinguish how many of the 4 special hyperedges (hyperedges containing the special node v_0) are shared.

Write A_q for the total number of pairs of k-cc-loops sharing q and $A_q(S)$ for the total number of pairs of k-cc-loops sharing a given set S of q hyperedges.

Case 1: (The shared hyperedges are connected) Let S be such a set of hyperedges with $|S| = q$. We consider three situations:

1. (Each node appears in S are incident to at most two hyperedges of S). In this case, S makes $q-1$ cycle nodes fixed and 2 cycle nodes limited in any k-cc-loops containing S. Therefore, the total number of pairs of k-cc-loops containing S is

$$\begin{aligned}|A_q(S)| &\leq \left(lk^2n^{(l-1-(q-1)-2)}\binom{n}{k-2}^{l-q}\right)^2\\ &= \frac{l^2k^4(n^{l-q-2}n^{(k-2)(l-q)})^2}{(k-2)!^{2(l-q)}}\\ &= \frac{l^2k^4}{n^4((k-2)!)^{2(l-q)}}n^{2(k-1)(l-q)}. \qquad (9)\end{aligned}$$

where the term l is for the number of possible positioning of S in a k-cc-loop. As the number of sets of hyperedges like S is less than

$$\binom{n}{q-1}(q-1)!\binom{n}{k-1}^2\binom{n}{k-2}^{q-2} = nn^{(k-1)q}\frac{1}{((k-2)!)^q},$$

the total number of pairs of k-cc-loops sharing q hyperedges in this case is less than

$$|A_q(S)| \cdot \text{ number of } S \leq \frac{l^2k^4}{n^3((k-2)!)^{2l-q}}n^{2(k-1)l}n^{-(k-1)q}. \tag{10}$$

2. (One node v appear in three or more hyperedges in S and $q = |S| < p+3$) In this case, depending on the number of hyperedges that v appears (3 or 4), S makes $q-2$ (or $q-3$) cycle nodes fixed and 3 (or (4)) cycle nodes limited in any k-cc-loops containing S. It follows that

$$\begin{aligned} |A_q(S)| &\leq \left(k^3n^{l-1-(q-2)-3}\binom{n}{k-2}^{l-q}\right)^2 \\ &= \frac{k^6}{n^4((k-2)!)^{2(l-q)}}n^{2(k-1)(l-q)} \end{aligned} \tag{11}$$

The number of such S is at most

$$\binom{n}{q-2}\binom{n}{k-1}^3\binom{n}{k-2}^{q-3} \leq n^{q-2}n^{(k-2)(q-3)}n^{3(k-1)} = nn^{(k-1)q}\frac{1}{((k-2)!)^q}.$$

Then, the total number of pairs of k-cc-loops sharing S like this is at most

$$\frac{k^6}{n^3((k-2)!)^{2l-q}}n^{2(k-1)l}n^{-(k-1)q}. \tag{12}$$

3. (One node v appears in three or more hyperedges in S and $q = |S| \geq p+3$) In this case, in addition to the above, we need to consider the situation where S itself forms a cycle. Then, S makes $q-1$ fixed cycle nodes and 1 limited cycle node. The total number of k-cc-loop pairs sharing S like this is at most

$$\frac{k^2}{n^2((k-2)!)^{2l-q}}n^{2(k-1)l}n^{-(k-1)q}. \tag{13}$$

Case 2: (The shared hyperedges form $h \geq 2$ connected components) Again, let S be such a set of hyperedges with $|S| = q$. In this case, the total number of sets of shared hyperedges is more than that in Case 1. But this is compensated by the decreasing of free cycle nodes—the total number of fixed cycle nodes is $q-h$ ($q-h-2$ or $q-h-3$, depending on the number of special hyperedges in S), while the total number of limited cycle nodes is at least $2h$. As a result, the total number of pairs of k-cc-loops sharing a set of disconnected hyperedges is less than the bounds we get in Case 1.

In summary, the total number of pairs of k-cc-loops sharing a set of q hyperedges is such that

$$|A_q| \leq \begin{cases} \frac{l^2k^4}{n^3(k-2)!)^{2l-q}} n^{2(k-1)l} n^{-(k-1)q}, \text{ if } q \leq p+2 \\ \frac{k^2}{n^2(k-2)!)^{2l-q}} n^{2(k-1)l} n^{-(k-1)q}, \text{ if } q > p+2. \end{cases} \tag{14}$$

Summing over all the $0 \leq q < l$ gives the desired result of (6).

4.2 Proof of Theorem 2

The proof of Theorem 2 is based on the concepts and results of hypertrees and unicycles in random hypergraphs.

Definition 4. *([12]) Let $\mathcal{G}$ be a k-uniform hypergraph with r vertices and s edges. The excess of $\mathcal{G}$ is defined to be*

$$ex(\mathcal{G}) = (k-1)s - r.$$

Generalizing the concepts of trees and cycles in graphs, we call a connected hypergraph $\mathcal{G}$ (1) a hypertree if $ex(\mathcal{G}) = -1$; (2) unicyclic if $ex(\mathcal{G}) = 0$.

Consider the random k-uniform constraint hypergraph $\mathcal{G}(n, m)$ associated with $\mathcal{C}_{n,cn}^{2,k,t}$. From [12], for $c < \frac{1}{k(k-1)}$, $\mathcal{G}(n, m)$ almost surely consists of hypertrees and unicyclic components. In this case, an instance of the random CSP is satisfiable if and only if the subproblems corresponding to the components of the constraint hypergraph are all satisfiable. A subproblem corresponding to a hypertree is satisfiable [5]. In the following, we prove that a subproblem corresponding to a unicyclic component is also satisfiable if the tightness of the constraint is less than 2^{k-1}. We break up the task into three lemmas.

Lemma 3. *For any uncyclic k-uniform hypergraph $\mathcal{G}$ with the edge set $E = (E_1, \cdots, E_t)$, we have*

$$|E_i \cap E_j| \leq 2, \forall 1 \leq i, j \leq t.$$

Proof. Assume that $a = |E_i \cap E_j| > 2$. Let

$$G' = (V, E - \{E_i\}).$$

Then, G' has at most $k - a + 1$ connected components $\{G_1, \cdots, G_{k-a+1}\}$. Since a connected hypergraph has at least an excess of -1, we have

$$ex(\mathcal{G}) = ex(G_1) + \cdots + ex(G_{k-a+1}) + (k-1) \geq a - 2 > 0.$$

A contradiction to the unicyclicness of $\mathcal{G}$.

Due to Lemma 3, we only need to consider unicycles in which edges have at most size 2 intersection.

Lemma 4. *Let $\mathcal{C}$ be a CSP such that*

1. *Its constraint graph $\mathcal{G}(V, E)$ is unicyclic ;*
2. *The tightness t is less than 2^{k-1}; and*
3. *There are a pair of hyperedges E_i and E_j with $|E_i \cap E_j| = 2$.*

Then, $\mathcal{C}$ is satisfiable.

Proof. Let $G^{'} = (V, E - \{E_i\})$. Since $|E_i \cap E_j| = 2$. There should be exact $k-1$ connected components in $G^{'}$ such that (1) one of the component contains the intersection $E_i \cap E_j$, and each of the rest of the components contains exact one vertex from $E_i - E_j$; and (2)each of the connected components has an excess of -1. Otherwise, $\mathcal{G}$ would have an excess larger than 0. The satisfiability of the CSP can be shown by first satisfying the constraint corresponding to the hyperedge E_i and then satisfy other constraints. This is possible because for the tightness $t < 2^{k-1}$, there is always at least one assignment that satisfies E_i and E_j simultaneously.

Now, we are in a position to deal with the situation where hyperedges have an intersection with a size at most 1.

Lemma 5. *Let $\mathcal{C}$ be a CSP such that*

1. *Its constraint graph $\mathcal{G}(V, E)$ is unicyclic ;*
2. *The tightness t is less than 2^{k-1}; and*
3. *For any pair of hyperedges E_i and E_j, we have with $|E_i \cap E_j| \leq 1$.*

Then, $\mathcal{C}$ is satisfiable.

Proof. In this case, the constraint hypergraph $\mathcal{G}(V, E)$ contains one cycle $F = (F_1, \cdots, F_l)$ of the form

$$|F_i \cap F_{i+1}| = 1, 1 \leq i \leq i-1, |F_i \cap F_1| = 1.$$

and some additional hypetree branches attached to the cycle. If there is a partial assignment to the variables satisfying the constraints in the cycle, then we can always extend it to satisfy the hypetree branches. To see there exists such a partial assignment, let $y_i = F_i \cap F_{i+1}$ and $y_n = F_n \cap F_1$. Consider the two possible assignments 0 and 1 to y_1. If we assign $y_1 = 0$ or 1, we can find assignments to $y_i, 2 \leq i \leq n-1$ to satisfies $E_1, \cdots, E_{n-1}$. Assume that y_n is forced to take the value a_0 for the assignment $y_1 = 0$ and a_1 for the assignment $y_1 = 1$. Since there are at most $2^{k-1} - 1$ restrictions to the variables in E_1, we know at least one of the pairs $(y_1 = 0, y_n = a_0)$ and $(y_1 = 1, y_n = a_1)$ can satisfies the constraint corresponding to E_1. This shows the existence of a partial assignment that satisfies the constraints corresponding to the cycle hyperedges.

References

1. Mitchell, D.: Resolution complexity of random constraints. In: Proceedings Principles and Practices of Constraint Programming - CP 2002, Springer (2002) 295–309

2. Kirousis, L., P.Kranakis, D.Krizanc, Y.Stamation: Approximating the unsatisfiability threshold of random formulas. Random Structures and Algorithms 12(3) (1994) 253–269
3. Achlioptas, D.: A survey of lower bounds for random 3-sat via differential equations. Theoretical Computer Science **265** (2001) 159–185
4. Dubois, O.: Upper bounds on the satisfiability threshold. Theoretical Computer Science **265** (2001) 187–197
5. Molloy, M.: Models and thresholds for random constraint satisfaction problems. In: Proceedings of the 34th ACM Symposium on Theory of Computing, ACM Press (2002) 209 – 217
6. 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
7. Gent, I., MacIntyre, E., Prosser, P., Smith, B., T.Walsh: Random constraint satisfaction: Flaws and structure. CONSTRAINTS **6** (2001) 345–372
8. Xu, K., Li, W.: Many hard examples in exact phase transitions with applications to generating hard satisfiable instances. personal communication (2003)
9. Franco, J., Gelder, A.V.: A perspective on certain polynomial-time solvable classes of satisfiability. Discrete Applied Mathematics **125** (2003) 177–214
10. Friedman, J., Goerdt, A.: Recognizing more unsatisfiable random k-sat instances efficiently. In: Lecture Notes in Computer Science 2076, Springer (2001) 310–321
11. Gao, Y., Culberson, J.: An analysis of phase transition in NK landscapes. Journal of Artificial Intelligence Research **17** (2002) 309–332
12. Karonski, M., Luczak, T.: The phase transition in a random hypergraph. J. Comput. Appl. Math. **142** (2002) 125–135