Artificial Intelligence 193 (2012) 1–17

Contents lists available at SciVerse ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

A general model and thresholds for random constraint satisfaction problems

Yun Fan [a], Jing Shen [b,*], Ke Xu [c]

[a] Department of Mathematics, Central China Normal University, Wuhan, 430079, China
[b] School of Science, Naval University of Engineering, Wuhan, 430033, China
[c] State Key Lab of Software Development Environment, Beihang University, Beijing, 100191, China

ARTICLE INFO

Article history:
Received 7 February 2012
Received in revised form 23 July 2012
Accepted 17 August 2012
Available online 21 August 2012

Keywords:
Constraint satisfaction problem
Phase transition

ABSTRACT

In this paper, we study the relation among the parameters in their most general setting that define a large class of random CSP models d-k-CSP where d is the domain size and k is the length of the constraint scopes. The model d-k-CSP unifies several related models such as the model RB and the model k-CSP. We prove that the model d-k-CSP exhibits exact phase transitions if $k \ln d$ increases no slower than the logarithm of the number of variables. A series of experimental studies with interesting observations are carried out to illustrate the solubility phase transition and the hardness of instances around phase transitions.

1. Introduction

The constraint satisfaction problem (CSP in short) is a central topic in the artificial intelligence community. It is interesting that many random CSPs exhibit phase transitions. Since the seminal work of Cheesman et al. [3], the link between the threshold phenomenon of CSPs and the computational hardness of CSPs has attracted a great interest of mathematicians, physicists and theoretical computer scientists. For the investigation of the threshold phenomenon of CSPs, numerous experimental studies [26,28,30] have been carried out, which results show strong evidences that the instances in the phase transition region are hard to solve. Therefore these CSPs with hard instances play an important role as they are used as benchmarks to test different CSP algorithms.

Random k-SAT is a special case of random CSPs, where each formula has precisely k literals per clause. In the past two decades, random k-SAT has been widely studied. The satisfiability threshold for random 2-SAT is obtained in [4,15]. Friedgut made tremendous progress in [13] towards establishing the existence of a sharp threshold for random k-SAT. It is theoretically attained in [21] and [9] that the updated lower bound and upper bound of the threshold point of random 3-SAT are 3.53 and 4.4898, respectively. However, the exact location of the phase transition of random k-SAT for $k \geqslant 3$ is still under investigation.

Much research has gone into the study of random models of CSPs with constant domain size at least 2, e.g. [1,6–8,17–19,24,28]. The initial standard models, named A, B, C and D [19,28], turn out to be flawed as they do not exhibit non-trivial satisfiability threshold when the length of constraint scopes and the size of domains are all fixed, see [1]. To get non-trivial phase transitions, researchers have been going on three directions. Some CSP models [17–19] restrict particular "structures" on constraint relations, so that the phase transitions are attained at a cost of more expenditure on generating random instances.

* Corresponding author.
E-mail addresses: yunfan02@yahoo.com.cn (Y. Fan), shendina@hotmail.com (J. Shen), kexu@nlsde.buaa.edu.cn (K. Xu).

0004-3702/$ – see front matter
http://dx.doi.org/10.1016/j.artint.2012.08.003

Instead of incorporating more structures in the relations, some researchers left the constraint relations without any restrictions and focused on allowing the domain size to grow up with the number of variables, e.g. [10,27,31]. In 2000, Xu and Li [31] constructed the well-known model RB, where the length of constraint scopes is fixed but the domain size of the instances is growing up about a power function of the number of variables. A few years later, Frieze and Molloy proposed two natural models of random binary CSPs with non-constant domain size, and determined how fast the domain size must grow with the number of variables to guarantee that the two models exhibit coarse threshold [14]. To ensure that these models can generate hard instances, the complexity of CSPs had been extensively investigated, e.g. [2,5,18,20,22,25,29,30,32]. For the practical side, benchmarks based on the model RB had been widely used in various kinds of algorithm competitions and in many research papers on algorithms.

From another point of view, inspired by the work of Frieze and Wormald [16], some CSP models, e.g. [11,12], allow the length of constraint scopes to grow moderately and the exact locations of phase transitions are gained. In [11] the authors proposed the so-called model k-CSP; differently from the model RB, the domain size of the model k-CSP is fixed while the length, denoted by k, of constrain scopes of the model k-CSP is growing up to a logarithm function of n (the number of variables); they located mathematically the exact phase transition point and demonstrated experimentally the performance of the model k-CSP in [11]. Because no restriction is put on constraint relations and the length k is growing very slowly while the domain size d is fixed, it is convenient to generate the instances of the model k-CSP; this is suitable for algorithmic practice.

In this paper we study a new random CSP model, d-k-CSP, where both the domain size d and the length of constrain scopes k are allowed to vary with the number of variables n. We show that the new model has a phase transition and the exact threshold point can be quantified precisely if $k \ln d \geqslant (1+\varepsilon)\ln n$ for an arbitrarily small positive real number ε. That is the major result of this paper.

The new model d-k-CSP covers obviously both the model RB and the model k-CSP as two specially extreme cases; and in the two extreme cases, i.e. either k is fixed or d is fixed, the above major result covers the well-known results in [31] and [11]. Moreover, the effective range of the model d-k-CSP is much more extensive than those of the model RB and the model k-CSP, it works well whenever $k \ln d$ increases no slower than $\ln n$. Thus it provides us a lot of various choices to deal with phase transitions of random CSPs; the various choices could meet various theoretical and practical requirements. For example, we can deduce at once from our major result that the model d-k-CSP with d growing up to $\ln n$ and k growing up to $\frac{\ln n}{\ln\ln n}$ has a phase transition and the exact threshold point can be located precisely; further, a series of experimental studies are carried out for that case and the results are reported in this paper.

The rest of the paper is organized as follows. In Section 2, we formulate the random model d-k-CSP precisely, and state our major result on the phase transition of the model d-k-CSP. Section 3 is contributed to a complete proof of our major result. The previous methods for the similar questions, e.g. the arguments in [11], are far from enough for the new extensive model; a new key idea to prove our major result is to divide the variant area of parameters, which we consider to estimate the satisfiability probability in the "second moment method" stage, by suitable curves into several subareas, so that in each subarea we can estimate the probability in different ways. In Section 4, we compare precisely the model d-k-CSP with some previous models and deduce the corresponding corollaries, and illustrate the vast effective range of the model d-k-CSP. In Section 5, we present the results of a series of experiments about the model d-k-CSP for the case $d = \ln n$. Finally we draw our conclusions in Section 6.

2. Random model *d*-*k*-CSP

In this paper, $\ln x = \log_e x$ denotes the natural logarithm function, and $\exp x = e^x$ denotes the natural exponential function. Denote by $|T|$ the cardinality of the set T.

Any instance of a *constraint satisfaction problem* is a triple $I = (X, D, C)$, where $X = (x_1, \ldots, x_n)$ is a sequence of n *variables*, $D = (D_1, \ldots, D_n)$ is a sequence of finite sets which are called the *domains*, and $C = (C_1, \ldots, C_t)$ with $C_i = (X_i, R_i)$ which are called the *constraints*; more precisely, $X_i = (x_{i_1}, \ldots, x_{i_{k_i}})$ are subsequences of X of length k_i for $i = 1, \ldots, t$, called the *constraint scopes*, and correspondingly, R_i are subsets of $D_{i_1} \times \cdots \times D_{i_{k_i}}$, called the *constraint relations*. Let $A = D_1 \times \cdots \times D_n$. Any element $(a_1, \ldots, a_n) \in A$ is viewed as an *assignment* to the variables $X = (x_1, \ldots, x_n)$ with values in the domains D, i.e. an evaluation $f(X) = (f(x_1), \ldots, f(x_n)) = (a_1, \ldots, a_n)$. If an assignment $(a_1, \ldots, a_n) \in A$ satisfies that $(a_{i_1}, \ldots, a_{i_{k_i}}) \in R_i$ for all $i = 1, \ldots, t$, then we say that $(a_1, \ldots, a_n)$ is a *solution* of the instance I; and at that case we say that the instance I is *satisfiable*.

Definition 2.1. Let n be the number of variables and t be the number of constraints; let $d = d(n)$ and $k = k(n)$ be two integer functions of the natural number n such that $d(n) > 1$ and $k(n) > 1$; let p be a positive constant with $0 < p < 1$. A random CSP model is said to be *d-k-CSP* if the instances are generated as follows:

- every cardinality $|D_i| = d$ for $i = 1, \ldots, n$;
- for $i = 1, \ldots, t$, the constraints are generated as follows:
  - the constraint scopes $X_i = (x_{i_1}, \ldots, x_{i_k})$ are randomly selected with repetition allowed from the subsequences of length k of the variable sequence $(x_1, \ldots, x_n)$;

– the constraint relations R_i are randomly selected with repetition allowed from the subsets of $D_{i_1} \times D_{i_2} \times \cdots \times D_{i_k}$ with cardinality $|R_i| = pd^k$.

Let Pr(SAT) denote the probability of a random instance of the model d-k-CSP being satisfiable. We have the following asymptotic property of the model d-k-CSP.

Theorem 2.1. *Let the notations be as in Definition 2.1, and assume that* $t = r \cdot \frac{n \ln d}{-\ln p}$ *for a constant parameter* r. *If both the following conditions* (i) *and* (ii) *are satisfied*:

(i) *the limit* $\lim_{n\to\infty} \frac{1}{d(n)}$ *exists*;
(ii) $k \geqslant \frac{1}{p}$ *and there is a positive real number* ε *such that* $k \ln d \geqslant (1+\varepsilon)\ln n$;

then

$$\lim_{n\to\infty} \Pr(\mathrm{SAT}) = \begin{cases} 0, & r > 1; \\ 1, & r < 1. \end{cases}$$

The theorem is stated in an extensive version so that it covers several known or guessed cases as consequences which peoples are concerned with, we'll discuss them in Section 4. Here we just make some remarks on the statements of the theorem. The proof of the theorem will be provided in Section 3.

Remark 2.1. The condition (i) of the theorem makes a necessary restriction for the function $d(n)$ as $n \to \infty$, to avoid wild variances of $d(n)$ which we could not control. Since it is an asymptotic result, and $d(n)$ is positive integer-valued, "$\lim_{n\to\infty} \frac{1}{d(n)} > 0$" implies that there is an integer N such that $d(n)$ is constant for $n > N$. Thus, asymptotically speaking, the condition (i) implies that either $\lim_{n\to\infty} d(n) = \infty$ or $d = d(n)$ is constant.

The condition (ii) of the theorem is also understood in asymptotic sense. Precisely, the condition "$k \ln d \geqslant (1+\varepsilon)\ln n$" implies that there is an integer N such that $k \ln d \geqslant (1+\varepsilon)\ln n$ for all $n > N$. Similarly, the condition "$k \geqslant \frac{1}{p}$" implies that there is an integer N such that $k(n) \geqslant \frac{1}{p}$ for all $n > N$.

Remark 2.2. The positive real number ε in the condition (ii) of the theorem could be arbitrarily small; but, at the theoretical level, it should be fixed in the asymptotic process once it was given. In practice, however, we can ignore it completely to assume that $k \ln d > \ln n$, which is enough to guarantee the appearance of the phase transition; because: the number of the instances generated by the actual operations in Definition 2.1 is always finite.

Remark 2.3. In accordance with our present approach, the convergence speed of the limits in Theorem 2.1 has to be considered case by case, thus we do not mention it in the statements of the theorem. Tracking the proof in Section 3 below, though a bit complicated, we can see that:

(a) for the case where $r > 1$, the probability $\Pr(\mathrm{SAT}) \to 0$ exponentially with $n \to \infty$;
(b) for the case where $r < 1$, there are two subcases:
  (b.1) if $d(n)$ is constant, then the probability $\Pr(\mathrm{SAT}) \to 1$ exponentially with $n \to \infty$;
  (b.2) if $\lim_{n\to\infty} \frac{1}{d(n)} = 0$, then the probability $\Pr(\mathrm{SAT}) \to 1$ at a rate not slower than that of a power function of $\frac{1}{d}$ converging to 0.

3. A proof of Theorem 2.1

Keep the notations and assumptions of Theorem 2.1. We prove Theorem 2.1 in seven subsections. First we deal with the case "$r > 1$" in Section 3.1; and then we turn to the case "$r < 1$", i.e. the stage of the so-called "second moment method". After a preparation in Section 3.2 which reduces it to estimations of some functions of variables $(n, x) \in (1, \infty) \times [0, 1]$, we are faced with a complicated situation; the previous methods for the similar questions, e.g. the arguments in [11] where the area $(1, \infty) \times [0, 1]$ was divided by lines into subareas and the estimations were made in each subarea, are not powerful enough for the present case. The key idea to carry it forward is to divide the area by curves. We'll outline the further proofs in Section 3.3.

3.1. Expectation of the number of solutions and the case $r > 1$

Given any n, let $\mathcal{G}$ denote the set of all the instances generated by the model d-k-CSP with $X = (x_1, \ldots, x_n)$ and $D = (D_1, \ldots, D_n)$. Then $\mathcal{G}$ is a probability space with equal probability for all samples. For $I \in \mathcal{G}$, let $\mathrm{Sol}(I)$ denote the set of solutions of the instance I.

For any $\mathbf{a} = (a_1, \ldots, a_n) \in A$, let

$$S_{\mathbf{a}} = \begin{cases} 1, & \text{if } \mathbf{a} \in \mathrm{Sol}(I); \\ 0, & \text{otherwise;} \end{cases}$$

which is a 0–1 random variable over the probability space $\mathcal{G}$. Then

$$S = \sum_{\mathbf{a} \in A} S_{\mathbf{a}}$$

is a non-negative integer random variable over the probability space $\mathcal{G}$.

The random variable S is obviously the number of solutions of the random instance $I \in \mathcal{G}$; in particular, $\Pr(S > 0) = \Pr(\mathrm{SAT})$, the probability for the random instance I being satisfiable.

It is easy to see that the expectation $\mathrm{E}(S_{\mathbf{a}}) = p^t$, and

$$\mathrm{E}(S) = \sum_{\mathbf{a} \in A} \mathrm{E}(S_{\mathbf{a}}) = p^t d^n = d^{\tau n}, \quad \text{where } \tau = 1 - r; \tag{1}$$

so

$$\lim_{n \to \infty} \mathrm{E}(S) = \begin{cases} 0, & r > 1; \\ \infty, & r < 1. \end{cases}$$

By Markov's inequality, we get that

$$\lim_{n \to \infty} \Pr(\mathrm{SAT}) = 0 \quad \text{if } r > 1.$$

3.2. Preparation for the case $r < 1$

In the rest of Section 3 we always assume that $r < 1$, and prove that $\lim_{n \to \infty} \Pr(\mathrm{SAT}) = 1$ by proving that $\lim_{n \to \infty} \frac{1}{\Pr(S>0)} = 1$.

As a preparation of the proof, in this subsection we apply the following conditional expectation inequality to upper bound $\frac{1}{\Pr(S>0)}$:

$$\Pr(S > 0) \geqslant \sum_{\mathbf{a} \in A} \frac{E(S_{\mathbf{a}})}{E(S | S_{\mathbf{a}} = 1)},$$

where $\mathrm{E}(S | S_{\mathbf{a}} = 1)$ denotes the conditional expectation of S assuming that $S_{\mathbf{a}} = 1$ occurs, see [23, Theorem 6.10]. In fact, in our case this inequality is equivalent to the so-called second moment method, see [11, Appendix A].

Let $\mathbf{a} = (a_1, \ldots, a_n) \in A$. For any $\mathbf{b} = (b_1, \ldots, b_n) \in A$, considering the number, denoted by m, of such indices i that $a_i = b_i$, we have that

$$\mathrm{E}(S_{\mathbf{b}} | S_{\mathbf{a}} = 1) = \left(\frac{\binom{m}{k}}{\binom{n}{k}} + \left(1 - \frac{\binom{m}{k}}{\binom{n}{k}} \right) \frac{pd^k - 1}{d^k - 1} \right)^t,$$

where $\binom{m}{k}$ stands for binomial coefficient. Denote $\sigma_{m,n} = \binom{m}{k} / \binom{n}{k}$. So

$$\begin{aligned} E(S | S_{\mathbf{a}} = 1) &= \sum_{\mathbf{b} \in A} E(S_{\mathbf{b}} | S_{\mathbf{a}} = 1) \\ &= \sum_{m=0}^{n} \binom{n}{m} (d-1)^{n-m} \left(\sigma_{m,n} + (1 - \sigma_{m,n}) \frac{pd^k - 1}{d^k - 1} \right)^t \\ &= \sum_{m=0}^{n} \binom{n}{m} (d-1)^{n-m} \left(p + (1-p) \frac{d^k \sigma_{m,n} - 1}{d^k - 1} \right)^t, \end{aligned}$$

which is independent of the choice of $\mathbf{a}$; hence

$$\Pr(S > 0) \geqslant \frac{\sum_{\mathbf{a}' \in A} \mathrm{E}(S_{\mathbf{a}'})}{\sum_{m=0}^{n} \binom{n}{m} (d-1)^{n-m} \left(p + (1-p) \frac{d^k \sigma_{m,n} - 1}{d^k - 1} \right)^t}.$$

Combining it with the formula (1), we have

$$\frac{1}{\Pr(S > 0)} \leqslant \sum_{m=0}^{n} R_m, \quad \text{with } R_m = \frac{\binom{n}{m} (d-1)^{n-m} \left(p + (1-p) \frac{d^k \sigma_{m,n} - 1}{d^k - 1} \right)^t}{d^n p^t}. \tag{2}$$

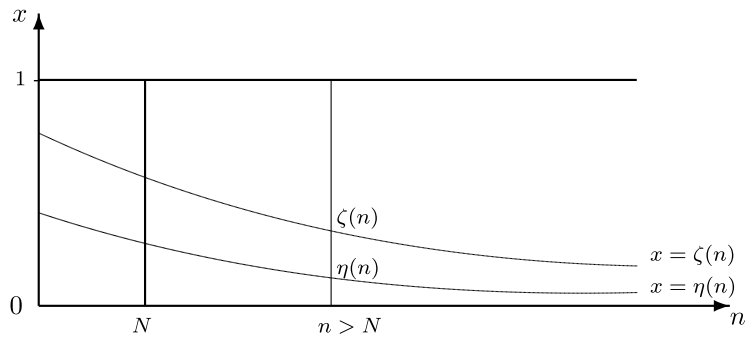

Fig. 1. The idea of the proof for the case $r < 1$.

Define a function for $(n, x) \in (1, \infty) \times [0, 1]$ as follows

$$R(n,x) = B(n,x)W(n,x) = \frac{\binom{n}{nx}(d-1)^{n-nx}\left(p + (1-p)\frac{(dx)^k - 1}{d^k - 1}\right)^t}{d^n p^t}, \tag{3}$$

where

$$B(n,x) = \binom{n}{nx}\left(\frac{1}{d}\right)^{nx}\left(1 - \frac{1}{d}\right)^{n(1-x)}, \qquad W(n,x) = \left(1 + \frac{1-p}{p} \cdot \frac{(dx)^k - 1}{d^k - 1}\right)^t.$$

Lemma 3.1.

(i) $R_m \leqslant R(n, \frac{m}{n})$ *for all* $m = 0, 1, \ldots, n$.
(ii) $R_m \leqslant B(n, \frac{m}{n})$ *for all* $m = 0, 1, \ldots, k-1$.

Proof. (i) For $m \geqslant k$, since $n \geqslant m$, and $\sigma_{m,n} = \frac{m(m-1)\cdots(m-k+1)}{n(n-1)\cdots(n-k+1)}$, we have $\frac{m-i}{n-i} \leqslant \frac{m}{n}$ for $i = 0, 1, \ldots, k-1$; so $\sigma_{m,n} \leqslant (\frac{m}{n})^k$. On the other hand, for $m = 0, 1, \ldots, k-1$, we have $\sigma_{m,n} = 0 \leqslant (\frac{m}{n})^k$.

(ii) For $m = 0, 1, \ldots, k-1$, from that $\sigma_{m,n} = 0$ we have that

$$p + (1-p)\frac{d^k \sigma_{m,n} - 1}{d^k - 1} = \frac{pd^k - 1}{d^k - 1} \leqslant p. \qquad \square$$

3.3. Outline of the proof for the case $r < 1$

To prove that $\lim_{n\to\infty} \Pr(\mathrm{SAT}) = 1$, by the inequality (2) it is enough to show that $\lim_{n\to\infty} \sum_{m=0}^{n} R_m = 1$, i.e. for any given positive real number θ there is an integer N such that $\sum_{m=0}^{n} R_m < 1 + \theta$, $\forall n > N$.

From now on till the end of this section, we always assume that $\delta = \theta/3$, and consider the area $(1, \infty) \times [0, 1]$ of the n–x plane. The key idea to achieve the above objective is to design an integer N and two curves

$$\begin{cases} x = \eta(n), \\ x = \zeta(n), \end{cases}$$

where $\zeta(n)$, $\eta(n)$ are two functions of n satisfying that

$$0 < \eta(n) \leqslant \zeta(n) < 1, \quad \forall n > N,$$

such that for any $n > N$ we have that

$$\begin{cases} \sum_{0 \leqslant \frac{m}{n} < \eta(n)} R_m < 1 + \delta; \\ n(\zeta(n) - \eta(n))R(n,x) < \delta, \quad \forall x \text{ with } \eta(n) \leqslant x < \zeta(n); \\ nR(n,x) < \delta, \quad \forall x \text{ with } \zeta(n) \leqslant x \leqslant 1. \end{cases} \tag{4}$$

In other words, we divide the area $(1, \infty) \times [0, 1]$ of the n–x plane by the curves into three areas, as illustrated in Fig. 1, so that we can deal with $R(n, x)$ for each area in different ways. But note that it may happen that $\eta(n) = \zeta(n)$ somewhere, for that n there is no x satisfying $\eta(n) \leqslant x < \zeta(n)$.

The inequalities of (4) are enough to complete the proof of Theorem 2.1 for the case $r < 1$, because they imply that for any $n > N$ we have

$$\begin{aligned}\sum_{m=0}^{n} R_m &= \sum_{0\leqslant \frac{m}{n}<\eta(n)} R_m + \sum_{\eta(n)\leqslant \frac{m}{n}<\zeta(n)} R_m + \sum_{\zeta(n)\leqslant \frac{m}{n}\leqslant 1} R_m \\ &< 1+\delta + \sum_{\eta(n)\leqslant \frac{m}{n}<\zeta(n)} \frac{\delta}{n(\zeta(n)-\eta(n))} + \sum_{\zeta(n)\leqslant \frac{m}{n}\leqslant 1} \frac{\delta}{n} \\ &\leqslant 1+\delta + \frac{n(\zeta(n)-\eta(n))\cdot\delta}{n(\zeta(n)-\eta(n))} + \frac{n(1-\zeta(n))\cdot\delta}{n} \\ &\leqslant 1+3\delta = 1+\theta.\end{aligned}$$

We will construct the function $\eta(n)$ in the next subsection such that the first inequality of (4) holds. Then, by Remark 2.1, we consider two cases: either $d(n)$ is a constant or $\lim_{n\to\infty} d(n)=\infty$. To proceed with the proof for the two cases, we need some auxiliary results to explore the function $\zeta(n)$, which will be provided in Section 3.5. After that, we complete the proof for the two cases in Sections 3.6 and 3.7 respectively.

3.4. Construction of $\eta(n)$

Proposition 3.1. *Let λ be a real number such that $0<\lambda<\frac{\varepsilon}{1+\varepsilon}$, and set*

$$\eta(n)=\max\left\{\frac{k(n)}{n}, \frac{1}{d(n)^{1-\lambda}}\right\}.$$

Then there exists an integer N_η such that

$$\sum_{0\leqslant \frac{m}{n}<\eta(n)} R_m < 1+\delta, \quad \forall n > N_\eta.$$

Proof. Given any integer $n>0$, assume that $\frac{m}{n}<\eta(n)$. If $\frac{k(n)}{n}\geqslant\frac{1}{d(n)^{1-\lambda}}$, then $m<k$ and, by Lemma 3.1, we have $R_m\leqslant B(n,\frac{m}{n})$.

Otherwise, $0<\frac{k(n)}{n}<\frac{1}{d(n)^{1-\lambda}}$ and $\eta(n)=\frac{1}{d(n)^{1-\lambda}}$. For $0\leqslant x\leqslant\frac{1}{d^{1-\lambda}}$, recalling that $t=\frac{rn\ln d}{-\ln p}$ and noting that $\frac{(dx)^k-1}{d^k-1}\leqslant x^k$, we can bound the function $W(n,x)$ in (3) as follows:

$$\begin{aligned}W(n,x) &= \exp\left(\frac{rn\ln d}{-\ln p}\ln\left(1+\frac{1-p}{p}\cdot\frac{(dx)^k-1}{d^k-1}\right)\right) \\ &\leqslant \exp\left(\frac{rn\ln d}{-\ln p}\cdot\frac{1-p}{p}\cdot\frac{(dx)^k-1}{d^k-1}\right) \\ &\leqslant \exp\left(\frac{rn\ln d}{-\ln p}\cdot\frac{1-p}{p}\cdot x^k\right) = \exp\left(\frac{r(1-p)}{-p\ln p} n x^k \ln d\right);\end{aligned}$$

and, by the condition that $k\ln d\geqslant(1+\varepsilon)\ln n$, we have

$$nx^k\ln d \leqslant d^{-(1-\lambda)k} n\ln d \leqslant \left(d^{\ln n/\ln d}\right)^{-(1-\lambda)(1+\varepsilon)} n\ln d = n^{-(\varepsilon-\lambda-\varepsilon\lambda)}\ln d;$$

since $\frac{k}{n}<\frac{1}{d^{1-\lambda}}$, i.e. $\ln d<\frac{\ln n-\ln k}{1-\lambda}$, we get

$$W(n,x)<\exp\left(\frac{r(1-p)}{-(1-\lambda)p\ln p}\cdot\frac{\ln n-\ln k}{n^{\varepsilon-\lambda-\varepsilon\lambda}}\right).$$

Note that the right hand side of the inequality is a real number larger than 1.

Summarizing the above, for any $n>0$ and any m with $0\leqslant\frac{m}{n}<\eta(n)$, by Lemma 3.1 we have

$$R_m\leqslant B\left(n,\frac{m}{n}\right)\cdot\exp\left(\frac{r(1-p)}{-(1-\lambda)p\ln p}\cdot\frac{\ln n-\ln k}{n^{\varepsilon-\lambda-\varepsilon\lambda}}\right).$$

As $0<\lambda<\frac{\varepsilon}{1+\varepsilon}$, we have $\varepsilon-\lambda-\varepsilon\lambda>0$, hence $\lim_{n\to\infty}\frac{\ln n-\ln k}{n^{\varepsilon-\lambda-\varepsilon\lambda}}=0$. Thus there is an integer N_η such that

$$R_m < B\left(n,\frac{m}{n}\right)\cdot(1+\delta), \quad \forall n>N_\eta,\ \forall m \text{ with } 0\leqslant\frac{m}{n}<\eta(n).$$

Hence, for any $n>N_\eta$ we have

$$\sum_{0\leqslant \frac{m}{n}<\eta(n)} R_m < \sum_{0\leqslant \frac{m}{n}<\eta(n)} B\left(n, \frac{m}{n}\right)\cdot(1+\delta)$$
$$< (1+\delta)\cdot \sum_{0\leqslant \frac{m}{n}\leqslant 1} \binom{n}{m}\left(\frac{1}{d}\right)^m\left(1-\frac{1}{d}\right)^{n-m}$$
$$= (1+\delta)\cdot\left(\frac{1}{d}+\left(1-\frac{1}{d}\right)\right)^n = 1+\delta. \qquad \square$$

Remark 3.1. As Proposition 3.1 handles the first inequality of (4), to complete the proof of Theorem 2.1, we need to construct the function $\zeta(n)$ satisfying the last two inequalities of (4); by Remark 2.1 we continue the proof in two cases.

- d is a constant.
- $\lim_{n\to\infty} d(n) = \infty$.

3.5. Auxiliary results for the case $r < 1$

Before going on to construct the function $\zeta(n)$ for the two cases, we show some results about the areas of $(1,\infty)\times[0,1]$ of the n–x plane divided by straight lines.

Recall that the binomial coefficients $\binom{n}{m}$ can be bounded from above by the so-called natural entropy function $H(x) = -x\ln x - (1-x)\ln(1-x)$ for $x\in[0,1]$ as follows:

$$\binom{n}{m} < \exp\big(nH(m/n)\big);$$

the entropy function $H(x)$ is differentiable and satisfies that $H(x) = H(1-x)$, $0\leqslant H(x)\leqslant \ln 2$ and $H(0) = H(1) = 0$.

Proposition 3.2. *For the function $R(n,x)$ defined in (3), there are a real number ρ with $0<\rho<1$ and an integer N_ρ such that*

$$nR(n,x) < \delta, \quad \forall n > N_\rho,\ \forall x\in[\rho,1].$$

Proof. Note that $p + (1-p)\frac{(dx)^k-1}{d^k-1} < 1$, $(d-1)^{n(1-x)} < d^{n(1-x)}$ and $e^{nH(x)} = d^{nH(1-x)/\ln d}$. Denote $p^t d^n = d^{\tau n}$ with $\tau = 1-r$ as in the formula (1). By the formula (3) we have

$$nR(n,x) \leqslant \frac{n\binom{n}{nx}(d-1)^{n(1-x)}}{d^n p^t} < \frac{ne^{nH(x)}(d-1)^{n(1-x)}}{d^{\tau n}}$$
$$< \frac{n}{d^{\frac{\tau n}{2}}}\cdot\frac{d^{n\left(\frac{H(1-x)}{\ln d}+(1-x)\right)}}{d^{\frac{\tau n}{2}}}.$$

As $\lim_{x\to 1} H(1-x) = 0$, there is a real number ρ with $0<\rho<1$ such that

$$\frac{H(1-x)}{\ln d} + (1-x) < \frac{\tau}{2}, \quad \forall x\in[\rho,1];$$

so there is an integer N_ρ such that

$$nR(n,x) < \delta, \quad \forall n > N_\rho,\ \forall x\in[\rho,1]. \qquad \square$$

In the rest of this section we always assume that ρ and N_ρ satisfy the condition in Proposition 3.2.

Lemma 3.2. *Let μ be any real number with $0<\mu<1$. If $k\geqslant 1/p$, then*

$$\frac{\ln(1+\frac{1-p}{p}x^k)}{x} \leqslant -\ln p, \quad \forall x\in[\mu,1].$$

Proof. Let

$$g(x) = \frac{\ln(1+\frac{1-p}{p}x^k)}{x},$$

and

$$g'(x) = \frac{\frac{(1-p)kx^k}{p+(1-p)x^k} - \ln(1+\frac{1-p}{p}x^k)}{x^2}.$$

We have

$$\ln\left(1+\frac{1-p}{p}x^k\right) - \frac{(1-p)kx^k}{p+(1-p)x^k} < \frac{1-p}{p}x^k - \frac{(1-p)kx^k}{p+(1-p)x^k}$$
$$= (1-p)x^k\left(\frac{1}{p} - \frac{k}{p+(1-p)x^k}\right) \leqslant (1-p)x^k\left(\frac{1}{p}-k\right).$$

When $k \geqslant \frac{1}{p}$, then $g'(x) \geqslant 0$, i.e. $g(x)$ is an increasing function. The maximum value of $g(x)$ in $[\mu, 1]$ is $g(1) = -\frac{1}{\ln p}$. Thus the lemma is proved. □

Remark 3.2. For the function $R(n,x)$ in (3), since $\frac{(dx)^k-1}{d^k-1} < x^k$ and $t = \frac{rn\ln d}{-\ln p}$, we have that

$$\ln W(n,x) < \ln\left(\left(1+\frac{1-p}{p}x^k\right)^t\right) = \frac{rn\ln d}{-\ln p}\ln\left(1+\frac{1-p}{p}x^k\right).$$

Also, $\ln B(n,x) < nH(x) + nx\ln d^{-1} + n(1-x)\ln(1-d^{-1})$. So we get an estimation for $R(n,x)$ as follows:

$$\ln R(n,x) < n\left(H(x) + x\ln\frac{1}{d} + (1-x)\ln\left(1-\frac{1}{d}\right) + \frac{r\ln d}{-\ln p}\ln\left(1+\frac{1-p}{p}x^k\right)\right). \tag{5}$$

Proposition 3.3. *Let μ be any real number with $0 < \mu < 1$. If $\lim_{n\to\infty} d(n) = \infty$, then there is an integer N_μ such that*

$$nR(n,x) < \delta, \quad \forall n > N_\mu,\ \forall x \in [\mu, 1].$$

Proof. By the estimation (5) we have

$$\ln\big(nR(n,x)\big) < n\left(\frac{\ln n}{n} + H(x) + (1-x)\ln\left(1-\frac{1}{d}\right) + x\ln\frac{1}{d} + \frac{r\ln d}{-\ln p}\ln\left(1+\frac{1-p}{p}x^k\right)\right),$$

where $\frac{\ln n}{n} + H(x) + (1-x)\ln(1-\frac{1}{d})$ is bounded from above for all $x \in [\mu, 1]$; further, because $r-1$ is a negative constant and $\ln d \to \infty$, we have

$$x\ln\frac{1}{d} + \frac{r}{-\ln p}\ln\left(1+\frac{1-p}{p}x^k\right)\ln d = \left(-1 + \frac{r}{-\ln p}\cdot\frac{\ln(1+\frac{1-p}{p}x^k)}{x}\right)x\ln d$$
$$\leqslant \left(-1 + \frac{r}{-\ln p}\cdot(-\ln p)\right)x\ln d$$
$$= (r-1)x\ln d \leqslant (r-1)\mu\ln d \underset{n\to\infty}{\longrightarrow} -\infty,$$

where Lemma 3.2 is used for the second line. Noting that x does not appear in $(r-1)\mu\ln d$, we have a uniform convergence $\lim_{n\to\infty}\ln(nR(n,x)) = 0$ for all $x \in [\mu, 1]$, which completes the proof. □

Lemma 3.3. *Assume that $a \in (0,1)$. Let*

$$h(x,a) = H(x) + x\ln a + (1-x)\ln(1-a), \quad x \in [0,1].$$

Then $h(x,a)$ is a non-positive differentiable function which is strictly increasing in $(0,a)$, and is strictly decreasing in $(a,1)$.

Proof. See, for example, [11, Lemma 3.1]. □

Remark 3.3. Since $\ln(1+\frac{1-p}{p}x^k) < \frac{1-p}{p}x^k$, with the function $h(x,a)$ defined as above, from (5) we can get another estimation of $R(n,x)$ as follows:

$$\ln\big(nR(n,x)\big) < n\left(\frac{\ln n}{n} + h\left(x,\frac{1}{d}\right) + \frac{r(1-p)}{-p\ln p}x^k\ln d\right). \tag{6}$$

3.6. The case "d is a constant"

In this subsection we assume that $r < 1$ and d is a constant. As pointed out in Remark 3.1, the following proposition completes the proof of Theorem 2.1 for this case.

Proposition 3.4. *Let λ, $\eta(n)$ and N_η be as in Proposition 3.1, ρ and N_ρ be as in Proposition 3.2. Assume that d is a constant. Set $\zeta(n) = \eta(n)$ for all n. Then there exists an integer $N \geqslant N_\eta$ such that*

$$nR(n,x) < \delta, \quad \forall n > N,\ \forall x \text{ with } \zeta(n) \leqslant x \leqslant 1.$$

Proof. If $\frac{1}{d^{1-\lambda}} \geqslant \rho$, then $\zeta(n) = \eta(n) \geqslant \frac{1}{d^{1-\lambda}} \geqslant \rho$ and, by taking $N = \max\{N_\eta, N_\rho\}$, we are done by Proposition 3.2. In the following we assume that $\frac{1}{d^{1-\lambda}} < \rho$. Consider the estimation (6) of $R(n,x)$. Let $-b = h(\frac{1}{d^{1-\lambda}}, \frac{1}{d})$. Note that $0 < \frac{1}{d} < \frac{1}{d^{1-\lambda}} < \rho < 1$. By Lemma 3.3, the number b is positive and

$$h\left(x, \frac{1}{d}\right) \leqslant -b, \quad \forall x \in \left[\frac{1}{d^{1-\lambda}}, \rho\right].$$

Since $k \ln d \geqslant (1+\varepsilon)\ln n$ and d is a constant, we have $k \to \infty$; so for $x \leqslant \rho < 1$ we can take an integer N_k such that

$$\frac{\ln n}{n} < \frac{b}{3} \quad \text{and} \quad \frac{r(1-p)}{-p\ln p} x^k \ln d < \frac{b}{3}, \quad \forall n > N_k.$$

By the inequality (6), there is an integer $N_k' > N_k$ such that

$$\ln\big(nR(n,x)\big) < -bn/3, \quad \forall n > N_k',\ \forall x \in \left[\frac{1}{d^{1-\lambda}}, \rho\right].$$

Since $-bn/3 \longrightarrow -\infty$, there is an $N_k'' \geqslant N_k'$ such that

$$nR(n,x) < \delta, \quad \forall n > N_k'',\ \forall x \in \left[\frac{1}{d^{1-\lambda}}, \rho\right].$$

Take $N = \max\{N_k'', N_\rho, N_\eta\}$. Since $\frac{1}{d^{1-\lambda}} \leqslant \zeta(n)$, combining the above inequality with Proposition 3.2, we have

$$nR(n,x) < \delta, \quad \forall\, n > N,\ \forall x \text{ with } \zeta(n) \leqslant x \leqslant 1. \qquad \square$$

3.7. The case "$\lim_{n\to\infty} d(n) = \infty$"

In this subsection we always assume that $r < 1$ and $\lim_{n\to\infty} d(n) = \infty$, and we complete the proof of Theorem 2.1 for this case.

Keep λ, $\eta(n)$, N_η as in Proposition 3.1, and ρ, N_ρ as in Proposition 3.2.

Reviewing the inequality (5), since $(1-x)\ln(1-d^{-1}) < 0$, we can relax it as: $\ln R(n,x) < n(H(x) + x\ln\frac{1}{d} + \frac{r(1-p)}{-p\ln p}x^k \ln d)$; i.e.

$$\ln R(n,x) < nx\ln d\left(\frac{H(x)}{x\ln d} - 1 + \frac{r(1-p)}{-p\ln p}x^{k-1}\right). \tag{7}$$

First we have an estimation for the term $\frac{H(x)}{x\ln d}$, which will be used to construct the function $\zeta(n)$.

Lemma 3.4. *There is an integer N_α such that*

$$\frac{H(x)}{x\ln d} < 1 - \lambda + \frac{\lambda}{6}, \quad \forall n > N_\alpha,\ \forall x \text{ with } \eta(n) \leqslant x \leqslant 1. \tag{8}$$

Proof. By the assumption we have $0 \leqslant 1 - x < 1$, hence

$$\frac{H(x)}{x\ln d} = \frac{-\ln x}{\ln d} + \frac{-(1-x)\ln(1-x)}{x\ln d} \leqslant \frac{-\ln x}{\ln d} + \frac{-\ln(1-x)}{x\ln d}.$$

As $\eta(n) \geqslant \frac{1}{d^{1-\lambda}}$, for $x \geqslant \eta(n)$ we have $x \geqslant \frac{1}{d^{1-\lambda}}$, i.e. $-\ln x \leqslant (1-\lambda)\ln d$; so

$$\frac{H(x)}{x\ln d} \leqslant 1 - \lambda + \frac{-\ln(1-x)}{x\ln d}.$$

Further, since $\lim_{x\to 0}\frac{-\ln(1-x)}{x}=1$, the function $\frac{-\ln(1-x)}{x}$ for $x\in(0,1]$ is bounded from above; and $d\to\infty$ in the present case, we have a uniform convergence:

$$\frac{-\ln(1-x)}{x\ln d}\xrightarrow{n\to\infty}0,\quad \forall x\in(0,1],$$

from which the conclusion follows. □

Recall that $k\ln d\geqslant(1+\varepsilon)\ln n$. As $\eta(n)\geqslant\frac{k}{n}$, for $x\geqslant\eta(n)$ we have $x\geqslant k/n$, i.e. $k\leqslant xn$; hence

$$\ln n<k\ln d<xn\ln d,\quad \forall n>N_\eta,\ \forall x \text{ with } \eta(n)\leqslant x\leqslant 1.$$

Take a real number β such that $1>\beta>1-\frac{\lambda}{6}$, then

$$\frac{(1-\beta)\ln n}{xn\ln d}<\frac{\lambda}{6},\quad \forall x \text{ with } \eta(n)\leqslant x\leqslant 1. \tag{9}$$

Proposition 3.5. *Let N_α be as in (8), and β be as in (9) above. Set*

$$\zeta(n)=\max\left\{\eta(n),\frac{1}{n^\beta}\right\}.$$

Then $\zeta(n)\geqslant\eta(n)$ and there is an integer $N_\zeta\geqslant N_\eta$ such that

$$n\big(\zeta(n)-\eta(n)\big)R(n,x)<\delta,\quad \forall n>N_\zeta,\ \forall x \text{ with } \eta(n)\leqslant x<\zeta(n).$$

Proof. Since $\lim_{n\to\infty}\frac{\lambda k\ln d}{2}=\infty$, there is an integer N_e such that

$$\exp\left(-\frac{\lambda k\ln d}{2}\right)<\delta,\quad \forall n>N_e. \tag{10}$$

Therefore, there is an integer $N_\zeta\geqslant\max\{N_\eta,N_\alpha,N_e\}$ such that

$$\frac{r(1-p)}{-p\ln p}\left(\frac{1}{n^\beta}\right)^{k-1}<\frac{\lambda}{6},\quad \forall n>N_\zeta. \tag{11}$$

Now we show that N_ζ fits the requirement.

Assume that $n>N_\zeta$. If $\eta(n)\geqslant\frac{1}{n^\beta}$, then $\zeta(n)=\eta(n)$ and there is no x satisfying $\eta(n)\leqslant x<\zeta(n)$.

Otherwise, $0<\eta(n)<\frac{1}{n^\beta}$ hence $\zeta(n)=\frac{1}{n^\beta}$. By the inequality (7) we have

$$\ln\big(n^{1-\beta}R(n,x)\big)<xn\ln d\left(\frac{(1-\beta)\ln n}{xn\ln d}+\frac{H(x)}{x\ln d}-1+\frac{r(1-p)}{-p\ln p}x^{k-1}\right). \tag{12}$$

But, for $x<\zeta(n)=\frac{1}{n^\beta}$ we get from (11) that

$$\frac{r(1-p)}{-p\ln p}x^{k-1}<\frac{r(1-p)}{-p\ln p}\left(\frac{1}{n^\beta}\right)^{k-1}<\frac{\lambda}{6},\quad \forall n>N_\zeta,\ \forall x<\zeta(n). \tag{13}$$

Combining (12) with (9), (8) and (13), for $n>N_\zeta$ and $\eta(n)\leqslant x<\zeta(n)$ we have

$$\ln\big(n^{1-\beta}R(n,x)\big)<(xn\ln d)\cdot(-\lambda/2)=-\frac{xn\lambda\ln d}{2};$$

however, $x\geqslant\eta(n)\geqslant\frac{k}{n}$, i.e. $xn\geqslant k$, so $\ln(n^{1-\beta}R(n,x))<-\frac{\lambda k\ln d}{2}$; further, it is clear that $\zeta(n)-\eta(n)\leqslant\frac{1}{n^\beta}$; so we get that

$$n\big(\zeta(n)-\eta(n)\big)R(n,x)\leqslant n^{1-\beta}R(n,x)<\exp\left(-\frac{\lambda k\ln d}{2}\right).$$

By (10), we reach the desired inequality:

$$n\big(\zeta(n)-\eta(n)\big)R(n,x)<\delta,\quad \forall n>N_\zeta,\ \forall x \text{ with } \eta(n)\leqslant x<\zeta(n).\qquad \square$$

As pointed out in Remark 3.1, Proposition 3.5 and the following proposition complete the proof of Theorem 2.1 for the present case.

Proposition 3.6. *Keep the notations in Proposition 3.5. Then there is an integer $N \geqslant N_\zeta$ such that*

$$nR(n,x) < \delta, \quad \forall n > N,\ \forall x \text{ with } \zeta(n) \leqslant x \leqslant 1.$$

Proof. From the inequality (7), we have

$$\ln\big(nR(n,x)\big) < xn\ln d\left(\frac{\ln n}{xn\ln d} + \frac{H(x)}{x\ln d} - 1 + \frac{r(1-p)}{-p\ln p}x^{k-1}\right). \tag{14}$$

As $\zeta(n) \geqslant 1/n^\beta$, for $x \geqslant \zeta(n)$ we have $x \geqslant 1/n^\beta$, hence

$$\frac{\ln n}{xn\ln d} \leqslant \frac{n^\beta \ln n}{n\ln d} = \frac{\ln n}{n^{1-\beta}\ln d} \xrightarrow{n\to\infty} 0;$$

so there is an integer N_β such that

$$\frac{\ln n}{xn\ln d} < \frac{\lambda}{6}, \quad \forall n > N_\beta,\ \forall x \text{ with } \zeta(n) \leqslant x \leqslant 1. \tag{15}$$

Since $\frac{r(1-p)}{-p\ln p}$ is a constant, we have a real number μ with $0 < \mu < 1$ such that

$$\frac{r(1-p)}{-p\ln p}x^{k-1} < \frac{\lambda}{6}; \quad \forall x \leqslant \mu. \tag{16}$$

Let $N' = \max\{N_\zeta, N_\beta\}$; by (14), (15), (8) and (16), for $n > N'$ and $\zeta(n) \leqslant x \leqslant \mu$ we have

$$\ln\big(nR(n,x)\big) < (xn\ln d)(-\lambda/2) \leqslant -\frac{\lambda k\ln d}{2};$$

which implies that $nR(n,x) < \exp(-\frac{\lambda k\ln d}{2})$. Citing (10) again, we get

$$nR(n,x) < \delta, \quad \forall n > N',\ \forall x \text{ with } \zeta(n) \leqslant x \leqslant \mu. \tag{17}$$

Finally, since $d \to \infty$ in the present case, we can cite Proposition 3.3 to have an integer $N \geqslant N'$ such that

$$nR(n,x) < \delta, \quad \forall n > N,\ \forall x \text{ with } \mu \leqslant x \leqslant 1. \tag{18}$$

The proposition follows from (17) and (18). □

4. Related CSP models

Keep the notations in Theorem 2.1. In this section, we consider some special cases of the size d of domains and the length k of constraint scopes, and get the related CSP models which were studied before; as a comparison of Theorem 2.1 with those obtained before, we further describe the activity range of the theorem with illustrations.

First, fixing the domain size d, from Theorem 2.1 we have the following corollary at once.

Corollary 4.1. *If d is a constant and there is a positive real number ε such that $k\ln d \geqslant (1+\varepsilon)\ln n$, then*

$$\lim_{n\to\infty}\Pr(\mathrm{SAT}) = \begin{cases} 0, & r > 1; \\ 1, & r < 1. \end{cases} \qquad \square$$

Obviously, the model d-k-CSP with constant domain size is just the model k-CSP studied in [11]. In fact, Corollary 4.1 is an improvement to the main result of [11], since its assumption "$k\ln d \geqslant (1+\varepsilon)\ln n$" is weaker than the corresponding assumption of the main theorem in [11].

On the other hand, if we fix the length k of constraint scopes, then we get the following consequence at once.

Corollary 4.2. *If k is a constant such that $k \geqslant \frac{1}{p}$ and $d \geqslant n^{(1+\varepsilon)/k}$ for a positive real number ε, then*

$$\lim_{n\to\infty}\Pr(\mathrm{SAT}) = \begin{cases} 0, & r > 1; \\ 1, & r < 1. \end{cases} \qquad \square$$

Particularly, we can take k to be a constant such that $k \geqslant \frac{1}{p}$ and $d = n^\alpha$ for a number $\alpha > \frac{1}{k}$, then the hypothesis of Corollary 4.2 is satisfied; and revise $t = r\cdot\frac{n\ln d}{-\ln p} = r\cdot\frac{\alpha n\ln n}{-\ln p}$ as $\frac{t}{n\ln n} = -\frac{r\alpha}{\ln p}$, one can rewrite the above phase transition of

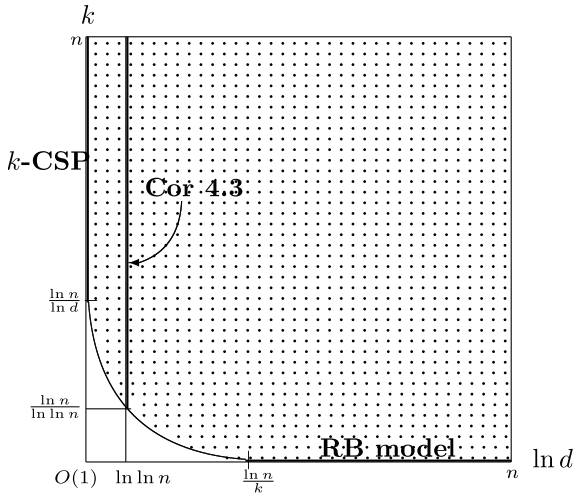

Fig. 2. The range where d-k-CSP works.

satisfiability as:

$$\lim_{n\to\infty} \Pr(\mathrm{SAT}) = \begin{cases} 0, & \frac{t}{n\ln n} > -\frac{\alpha}{\ln p}; \\ 1, & \frac{t}{n\ln n} < -\frac{\alpha}{\ln p}. \end{cases}$$

Therefore, the model d-k-CSP with constant length k of constrain scopes is just the model RB studied in [31]. Xu and Li showed that the model RB had a lot of hard instances and all instances at the threshold point had exponential tree-resolution complexity, see [30,32]. By relating the hardness of the model RB, there are also a lot of hard instances to solve in the model d-k-CSP.

However, Theorem 2.1 says much more than the two special cases, as it depends on a trade-off between the growing of d and the growing of k to guarantee that the exact phase transition points can be mathematically determined. For a series $\{a_n\}$, as usual, $a_n = O(1)$ stands for that $\lim_{n\to\infty} a_n < \infty$. For series $\{a_n\}$ and $\{b_n\}$ such that $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = \infty$, we denote $\Omega(a_n) < \Omega(b_n)$ if $\lim_{n\to\infty} \frac{a_n}{b_n} < 1$. In this notation Theorem 2.1 says that, roughly speaking, the model d-k-CSP has a phase transition and the threshold point can be exactly quantified whenever $\Omega(k\ln d) > \Omega(\ln n)$. We illustrate with Fig. 2 the range where the model d-k-CSP has an exact phase transition.

In Fig. 2, the horizontal axis is scaled by the growing speed of $\ln d$ while the vertical axis is scaled by the growing speed of k. Of course, both the axis start from $O(1)$, i.e. the constants. Since $k \leqslant n$, the vertical axis ends at n, i.e. $\Omega(n)$. And, if $\Omega(d(n)) > \Omega(e^n)$, then $d(n)$ is an exponentially growing variable with n, which is too large for the CSPs; so it is reasonable to end the horizontal axis at n too. The dotted area depicts where we have $\Omega(k\ln d) > \Omega(\ln n)$; in other words, it is just the range where Theorem 2.1 works. In particular, the points of the horizontal axis correspond exactly the case that k is constant; thus the line segment on the horizontal axis from $\frac{\ln n}{k}$ (for a fixed k) to n is just the range where the model RB works. Similarly, the line segment on the vertical axis from $\frac{\ln n}{\ln d}$ (for a fixed d) to n is just the range where the model k-CSP works.

As an example other than Corollary 4.1 and Corollary 4.2, we take the function $d(n) = \ln n$, then the model d-k-CSP has a phase transition which can be precisely quantified provided $\Omega(k(n)) > \Omega(\frac{\ln n}{\ln\ln n})$. It is just the following corollary and illustrated in Fig. 2 also with a line segment.

Corollary 4.3. *If $d = \ln n$ and there is a positive real number ε such that $k\ln\ln n \geqslant (1+\varepsilon)\ln n$, then*

$$\lim_{n\to\infty} \Pr(\mathrm{SAT}) = \begin{cases} 0, & r > 1; \\ 1, & r < 1. \end{cases} \qquad \square$$

According to Corollary 4.3, Table 1 gives the domain size and the corresponding minimal value of k satisfying the condition against n. It is noteworthy that the domain size and the length k of constraint scopes increase very slowly with the number of variables; please compare it with Tables 2 and 3. Therefore the domain size and the length k of constraint scopes are not big for the practically applied models.

Table 1
The value d and the minimal value of k for Corollary 4.3.

n	10	50	100	500	1000	5000	10 000
$d = \lceil \ln n \rceil$	3	4	5	7	7	9	10
$k = \lceil \frac{\ln n}{\ln d} \rceil$	3	3	4	4	4	4	5

Table 2
$d = 3$ and the minimal value of k for Corollary 4.1.

n	10	50	100	500	1000	5000	10 000
$d = 3$	3	3	3	3	3	3	3
$k = \lceil \frac{\ln n}{\ln 3} \rceil$	3	4	5	6	7	8	9

Table 3
$k = 3$ and the minimal value of d for Corollary 4.2.

n	10	50	100	500	1000	5000	10 000
$d = \lceil n^{1/3} \rceil$	3	4	5	8	10	18	22
$k = 3$	3	3	3	3	3	3	3

5. Experimental results for the case $d = \ln n$[1]

Experiments have been done to study the behavior of the model d-k-CSP with fixed length k of constraint scopes, see [30], and experiments for the model k-CSP, i.e. the model d-k-CSP with fixed domain size d, are reported in [11].

We have performed a series of experiments on the model d-k-CSP with $d = \ln n$, i.e. Corollary 4.3. In this section, we report the experimental results. The platform we have used for our experimentation is called Abscon (see http://www.cril.univ-artois.fr/~lecoutre).

Each random instance generated by the model d-k-CSP is characterized by a 5-tuple (k, n, d, r, p) of parameters, where k denotes the length of constraint scopes, n the number of variables, d the uniform domain size, $r = \frac{-t \ln p}{n \ln d}$ a measure of the constraint density, p a measure of the constraint tightness. At each setting of (k, n, d, r, p), 50 instances are generated.

According to Corollary 4.3, the instances of the model d-k-CSP change from being soluble to insoluble when the constraint density r is varied accordingly. Figs. 3 and 4 depict the solubility phase transition and the easy-hard-easy phase transition for $d = \ln n$, $p = 0.6$, $n \in \{40, 60, 80\}$ and $k = 3$. In Fig. 3 it clearly appears that the solubility phase transition happens around the theoretical threshold point $r = 1$, which illustrates that the theoretical result is in close agreement with the empirical result. On the other hand, the hard instances are found at the neighborhood of the phase transition point $r = 1$, as shown in Fig. 4. We remark that the vertical axis for CPU time in Fig. 4 uses a log scale $\ln T$ where T is the search time (seconds), because the search cost of solving the instances in the model d-k-CSP grows too fast when n is growing up, so that we could not depict the three curves for $n \in \{40, 60, 80\}$ in one and the same coordinate system if the CPU time axis was scaled by any multiple of seconds. Table 4 lists the correspondence between CPU time T (seconds) and $\ln T$.

Fig. 5 shows the computational complexity of solving the instances of the model d-k-CSP around the theoretical threshold $r = 1$ when n is varied from 20 to 80 in steps of 5. According to Corollary 4.3, k satisfies the condition $k \geqslant (1 + \varepsilon)\ln n / \ln \ln n$ for an arbitrary positive real ε, so we take $k = 3$ when n is varied from 20 to 80 in steps of 5. Table 5 gives the corresponding values of d satisfying the condition of Corollary 4.3 against n. Note that the vertical scale uses a log scale in Fig. 5, cf. Table 4. We summarize experimental results for $d = \ln n$, $k = 3$, $p = 0.6$ and $r \in \{0.98, 1, 1.02\}$, and observe that the curves in Fig. 5 look like the polygonal lines, which illustrate that the complexity of solving the instances with values d and k around the phase transition point grows exponentially with n. By Corollary 4.3 we know that d and k increase very slowly with n, making it feasible to use this class of models to generate random CSP instances in practice.

We also consider the effect of different length of constraint scopes to the solubility phase transition and the hardness phase transition. Fig. 6 shows the solubility phase transition for $d = \ln n$, $p = 0.6$, $n = 30$ and $k \in \{3, 4, 5\}$. Similarly, Fig. 7 indicates the hardness phase transition for $d = \ln n$, $p = 0.6$, $n = 30$ and $k \in \{3, 4, 5\}$. Figs. 6 and 7 present the solubility phase transition and the hardness phase transition for the different values of k both happen around the theoretical threshold point $r = 1$. It is observed that the solubility phase transition of the model d-k-CSP becomes sharper and solving the instances needs more time when the length k of constraint scopes increases. Therefore the solubility phase transition and the hardness phase transition are affected by the length of constraint scopes.

[1] In practice, $\ln n$ should be rounded to the nearest integer.

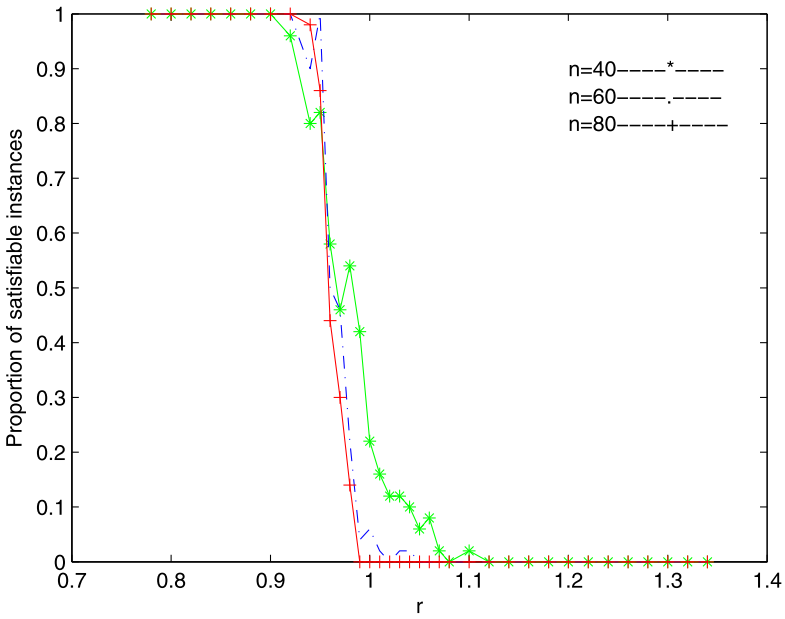

Fig. 3. The solubility phase transition for d-k-CSP$(3, n, \ln n, r, 0.6)$.

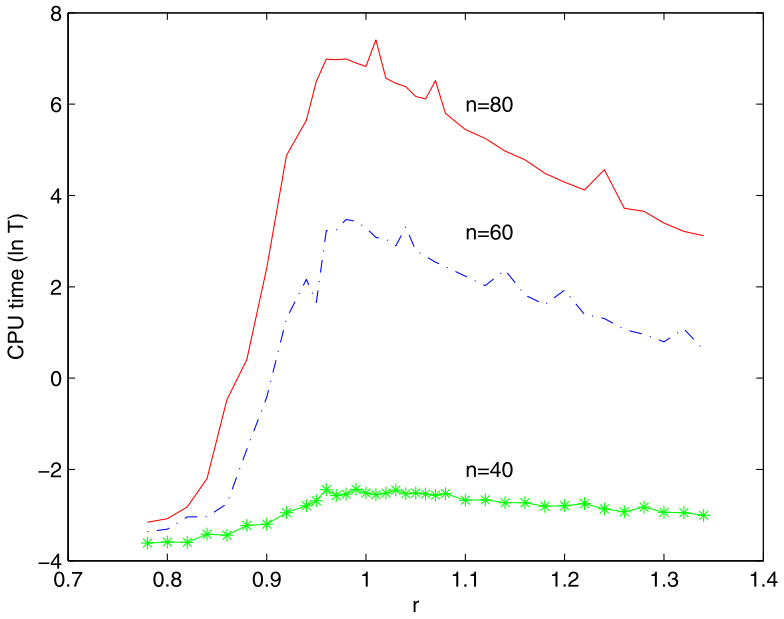

Fig. 4. Mean search cost of solving instances in d-k-CSP$(3, n, \ln n, r, 0.6)$.

Table 4
The correspondence between CPU time T (seconds) and $\ln T$.

T	0.002	0.018	0.135	1	7.389	54.60	403.4	2981
$\ln T$	−6	−4	−2	0	2	4	6	8

Table 5
The corresponding value of d against n for Corollary 4.3.

n	20	25	30	35	40	45	50	55	60	65	70	75	80
d	3	4	4	4	4	4	4	5	5	5	5	5	5

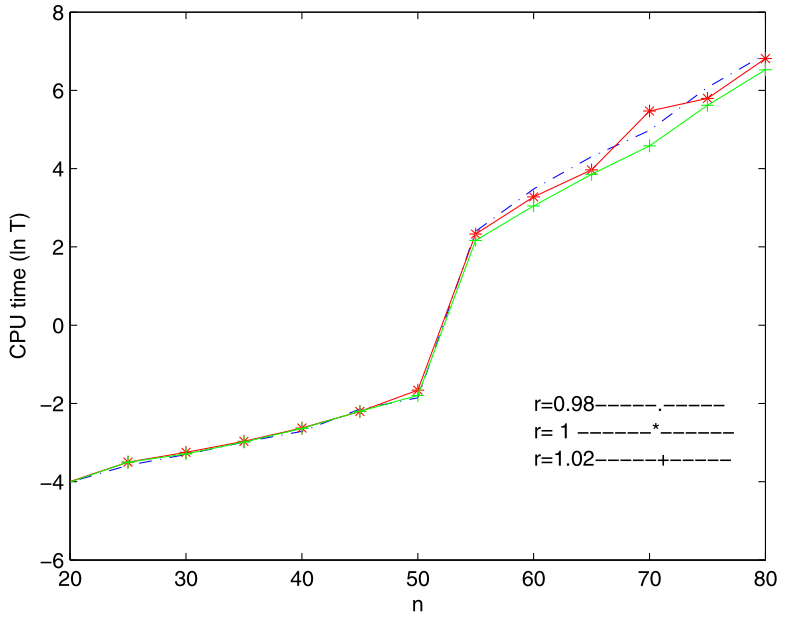

Fig. 5. Mean search cost of solving instances in d-k-CSP$(3, n, \ln n, r, 0.6)$.

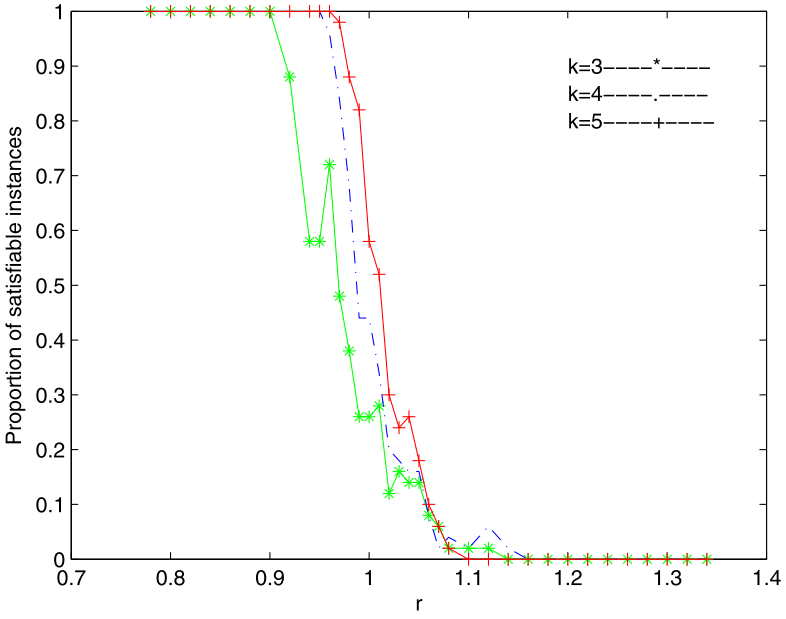

Fig. 6. The solubility phase transition for d-k-CSP$(k, 30, \ln n, r, 0.6)$.

6. Conclusions

In this paper, we consider such a type of random CSPs: for any given positive integer n, any instance with n variables has n domains of the same size d, and has t constraints with all constraint scopes of the same length k and all constraint relations of the same tightness p. We studied a general random model, called the model d-k-CSP, where the tightness p is fixed, but the domain size d and the length k of the constraint scopes are allowed to vary with the number n of variables. Various cases of the parameters which peoples in the artificial intelligence community are interested in are included, e.g. the model RB (with growing d and fixed k) and the model k-CSP (with fixed d and growing k). The core concept of the general-model-building is that some certain relations among the parameters that define instances of the CSPs could guarantee the presence of phase transitions.

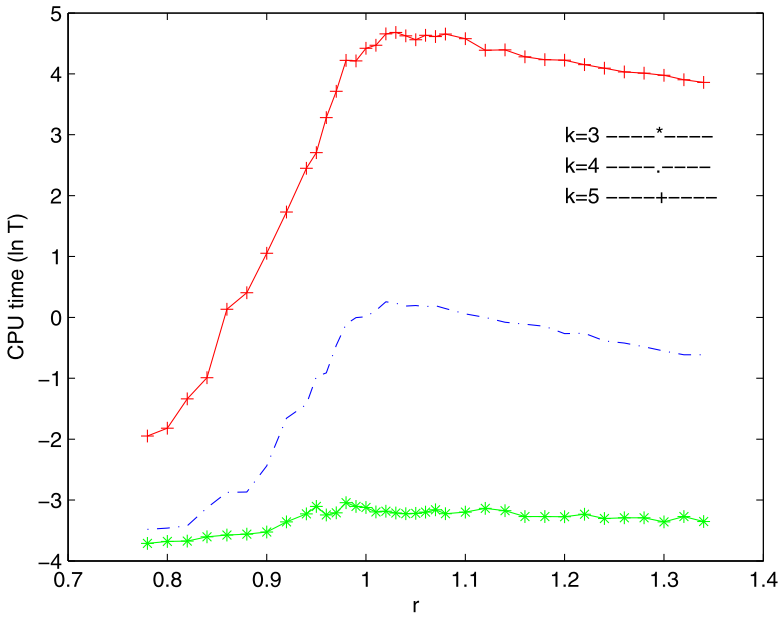

Fig. 7. Mean search cost of solving instances in d-k-CSP(k, 30, $\ln n$, r, 0.6).

From the new research perspective, we proved that for the model d-k-CSP a satisfiability phase transition threshold can be exactly quantified if $k \ln d$ increases no slower than $\ln n$.

And, to prove the theoretical result mathematically, in the stage of the "second moment method" we developed a new idea to divide the whole area, which we considered to estimate the satisfiability probability, by curves (instead of by lines as before), and in each divided area we made the estimations in different ways. As the success of the new methodology for the model d-k-CSP, it may be interesting to explore any applicability of the new approach to solve related open problems in the future.

Moreover, for the model d-k-CSP of the case $d = \ln n$, a series of experimental studies are carried out and the results are reported in this paper. The results not only illustrate the satisfiability phase transition which is consistent with the theoretical result, but also demonstrate that the hardness phase transition appears at the satisfiability phase transition point. Just like some well-known cases, we can get a lot of hard instances with the model d-k-CSP of the case $d = \ln n$.

Because of the extensively effective range of the model d-k-CSP, it provides a lot of choices to generate random CSP instances with a phase transition in practice. Though the product $k \ln d$ should increase no slower than $\ln n$, quite a part of the choices provided by the model d-k-CSP makes the domain size d and the length k of the constraint scopes growing up with n very slowly. The case of $d = \ln n$ is a new example of such choices. Thus, the model d-k-CSP is useful for testing CSP solvers, since such choices provided by the model can generate asymptotically non-trivial CSP instances with small domain size and small length of constraint scopes in an easy and natural way.

Acknowledgements

Thanks are given to NSFC for the support through Grant Nos. 11171370 and 60973033, and the National 863 Program (Grant No. 2012AA011005). It is also the authors' great pleasure to thank the anonymous referees for their many helpful comments.

References

[1] D. Achlioptas, L. Kirousis, E. Kranakis, D. Krizanc, M. Molloy, Y. Stamatiou, Random constraint satisfaction: A more accurate picture, in: Proceedings of the Third International Conference on Principles and Practice of Constraint Programming, in: LNCS, vol. 1330, Springer, 1997, pp. 107–120.
[2] E. Ben-Sasson, A. Wigderson, Short proofs are narrow – resolution made simple, Journal of the ACM 48 (2001) 149–169.
[3] P. Cheesman, B. Kanefsky, W. Taylor, Where the really hard problems are, in: Proceedings of IJCAI-91, IJCAII, 1991, pp. 331–337.
[4] V. Chvátal, B. Reed, Miks gets some (the odds are on his side), in: Proceedings of the 33rd IEEE Symposium on Foundations of Computer Science, IEEE Computer Society Press, 1992, pp. 620–627.
[5] V. Chvátal, E. Szemerédi, Many hard examples for resolution, Journal of the ACM 35 (1988) 208–759.
[6] N. Creignou, H. Daudé, Random generalized satisfiability problems, in: Proceedings of SAT, Elsevier, 2002.
[7] N. Creignou, H. Daudé, Generalized satisfiability problems: minimal elements and phase transitions, Theoretical Computer Science 302 (2003) 417–430.
[8] N. Creignou, H. Daudé, Combinatorial sharpness criterion and phase transition classification for random CSPs, Information and Computation 190 (2004) 220–238.
[9] J. Díaz, L. Kirousis, D. Mitsche, X. Pérez-Giménez, On the satisfiability threshold of formulas with three literals per clause, Theoretical Computer Science 410 (2009) 2920–2934.

[10] M. Dyer, A. Frieze, M. Molloy, A probabilistic analysis of randomly generated binary constraint satisfaction problems, Theoretical Computer Science 290 (2003) 1815–1828.
[11] Y. Fan, J. Shen, On the phase transitions of random k-constraint satisfaction problems, Artificial Intelligence 175 (2011) 914–927.
[12] A. Flaxman, A sharp threshold for a random constraint satisfaction problem, Discrete Mathematics 285 (2004) 301–305.
[13] E. Friedgut, Sharp thresholds of graph properties, and the k-sat problem, with an appendix by Jean Bourgain, Journal of the American Mathematical Society 12 (1999) 1017–1054.
[14] A. Frieze, M. Molloy, The satisfiability threshold for randomly generated binary constraint satisfaction problems, Random Structures & Algorithms 28 (2006) 323–339.
[15] A. Frieze, S. Suen, Analysis of two simple heuristics on a random instance of k-sat, Journal of Algorithms 20 (1996) 312–355.
[16] A. Frieze, N. Wormald, Random k-sat: A tight threshold for moderately growing k, in: Proceedings of the Fifth International Symposium on Theory and Applications of Satisfiability Testing, Springer, 2002, pp. 1–6.
[17] Y. Gao, J. Culberson, Consistency and random constraint satisfaction models with a high constraint tightness, in: Proceedings of the Tenth International Conference on Principles and Practice of Constraint Programming (CP-2004), Springer, 2004, pp. 17–31.
[18] Y. Gao, J. Culberson, Consistency and random constraint satisfaction models, Journal of Artificial Intelligence Research 28 (2007) 517–557.
[19] I. Gent, E. Macintyre, P. Prosser, B. Smith, Random constraint satisfaction: Flaws and structure, Constraints 6 (2001) 345–372.
[20] W. Jiang, T. Liu, T. Ren, K. Xu, Two hardness results on feedback vertex sets, in: Frontiers in Algorithmics and Algorithmic Aspects in Information and Management—Joint International Conference (FAW–AAIM), Springer, 2011, pp. 233–243.
[21] A. Kaporis, L. Kirousis, E. Lalas, The probabilistic analysis of a greedy satisfiability algorithm, in: The 10th Annual European Symposium on Algorithms, Springer, 2002, pp. 574–585.
[22] T. Liu, X. Lin, C. Wang, K. Su, K. Xu, Large hinge width on sparse random hypergraphs, in: Proceedings of the 22nd International Joint Conference on Artificial Intelligence, IJCAI/AAAI, 2011, pp. 611–616.
[23] M. Mitzenmacher, E. Upfal, Probability and Computing: Randomized Algorithm and Probabilistic Analysis, Cambridge Univ. Press, Cambridge, 2005.
[24] M. Molloy, Models and thresholds for random constraint satisfaction problems, in: Proceedings of the 34th ACM Symposium on Theory of Computing, ACM Press, 2001, pp. 209–217.
[25] M. Molloy, M. Salavatipour, The resolution complexity of random constraint satisfaction problems, SIAM Journal on Computing 37 (2007) 895–922.
[26] P. Prosser, An empirical study of phase transitions in binary constraint satisfaction problems, Artificial Intelligence 81 (1996) 81–109.
[27] B. Smith, Constructing an asymptotic phase transition in random binary constraint satisfaction problems, Theoretical Computer Science 265 (2001) 265–283.
[28] B. Smith, M. Dyer, Locating the phase transition in binary constraint satisfaction problems, Artificial Intelligence 81 (1996) 155–181.
[29] C. Wang, T. Liu, P. Cui, K. Xu, A note on treewidth in random graphs, in: 5th Annual International Conference on Combinatorial Optimization and Applications (COCOA), Springer, 2011, pp. 491–499.
[30] K. Xu, F. Boussemart, F. Hemery, C. Lecoutre, Random constraint satisfaction: Easy generation of hard (satisfiable) instances, Artificial Intelligence 171 (2007) 514–534.
[31] K. Xu, W. Li, Exact phase transitions in random constraint satisfaction problems, Journal of Artificial Intelligence Research 12 (2000) 93–103.
[32] K. Xu, W. Li, Many hard examples in exact phase transitions, Theoretical Computer Science 355 (2006) 291–302.