Random Constraint Satisfaction: theory meets practice*

Ewan MacIntyre[1], Patrick Prosser[1], Barbara Smith[2], and Toby Walsh[1]

[1] The APES Research Group, Department of Computer Science, University of Strathclyde, Glasgow, United Kingdom. Email {em,pat,tw}@cs.strath.ac.uk
[2] The APES Research Group, School of Computer Studies, University of Leeds, United Kingdom. Email bms@scs.leeds.ac.uk

Abstract. We study the experimental consequences of a recent theoretical result by Achlioptas *et al.* that shows that conventional models of random problems are trivially insoluble in the limit. We survey the literature to identify experimental studies that lie within the scope of this result. We then estimate theoretically and measure experimentally the size at which problems start to become trivially insoluble. Our results demonstrate that most (but not all) of these experimental studies are luckily unaffected by this result. We also study an alternative model of random problems that does not suffer from this asymptotic weakness. We show that, at a typical problem size used in experimental studies, this model looks similar to conventional models. Finally, we generalize this model so that we can independently adjust the constraint tightness and density.

1 Introduction

One of the most exciting areas in AI in recent years has been the study of phase transition behaviour. In a seminal paper that inspired many later researchers, Cheeseman, Kanefsky, and Taylor demonstrated that the hardest search problems often occur around a rapid transition in solubility [2]. Problems from such transitions in solubility are routinely used to benchmark algorithms for many different NP-complete problems. Experimental results about phase transition behaviour have come thick and fast since the publication of [2]. For example, in random 3-SAT, the phase transition was quickly shown to occur when the ratio of clauses to variables is approximately 4.3 [14]. Unfortunately, theory has often proved more difficult. A recent result proves that the width of the phase transition in random 3-SAT narrows as problems increases in size [3]. However, we only have rather loose but hard won bounds on its actual location [4, 13]. For random constraint satisfaction problems, Achlioptas *et al.* recently provided a more negative theoretical result [1]. They show that the conventional random models are almost surely trivially insoluble for large enough problems. This paper studies the impact of this theoretical result on experimental studies.

* Supported by EPSRC awards GR/L/24014 and GR/K/65706. The authors wish to thank other members of the APES research group. We are especially grateful to Ian Gent who derived the expected number of cliques in a random graph.

2 Constraint satisfaction

A binary constraint satisfaction problem consists of a set of variables, each with a domain of values, and a set of binary constraints. Each constraint rules out a subset of the possible values for a pair of variables. Each assignment of values to variables ruled out is called a nogood. Associated with each problem is a constraint graph. This has variables as vertices and edges between variables that appear in nogoods. The constraint satisfaction decision problem is to decide if there is an assignment of values to variables so that none of the constraints are violated.

Four models of random problems are used in most experimental and theoretical studies. In each model, we generate a constraint graph G, and then for each edge in this graph, choose pairs of incompatible values. The models differ in how we generate the constraint graph and how we choose incompatible values. In each case, we can describe problems by the tuple $\langle n, m, p_1, p_2 \rangle$, where n is the number of variables, m is the uniform domain size, p_1 is a measure of the density of the constraint graph, and p_2 is a measure of the tightness of the constraints.

model A: with probability p_1, we select each one of the $n(n-1)/2$ possible edges in G, and for each edge with probability p_2 we pick each one of the m^2 possible pairs of values as incompatible;

model B: we uniformly select exactly $p_1 n(n-1)/2$ edges for G, and for each edge we uniformly pick exactly $p_2 m^2$ pairs of values as incompatible;

model C: with probability p_1, we select each one of the $n(n-1)/2$ possible edges in G, and for each edge we uniformly pick exactly $p_2 m^2$ pairs of values as incompatible;

model D: we uniformly select exactly $p_1 n(n-1)/2$ edges for G, and for each edge with probability p_2 we pick each one of the m^2 possible pairs of values as incompatible.

3 Phase transitions

Constraint satisfaction algorithms are now routinely benchmarked using random problems from one of these four models. To help unify experimental studies with different problems, Gent *et al.* [8] define the constrainedness, κ of an ensemble of combinatorial problems as,

$$\kappa =_{\text{def}} 1 - \frac{\log_2(\langle Sol \rangle)}{N}$$

where N is the log base 2 of the size of the state space, and $\langle Sol \rangle$ is the expected number of these states that are solutions. Since $0 \leq \langle Sol \rangle \leq 2^N$, it follows that $\kappa \in [0, \infty)$. If $\kappa \approx 0$ then problems are very under-constrained and soluble. It is usually very easy to find one of the many solutions. If $\kappa \approx \infty$ then problems are very over-constrained and insoluble. It is usually relatively easy to prove their insolubility. If $\kappa \approx 1$ then problems are on the "knife-edge" between solubility

and insolubility. It is often difficult to find solutions or prove the insolubility of such problems. This definition of constrainedness has been used to locate phase transitions behaviour both in NP-complete problems like constraint satisfaction, and in polynomial problems like enforcing arc consistency [5].

Consider, for example, binary constraint satisfaction problems from model B. The state space has m^n states, one for each possible assignment of values to the n variables. Each of the $p_1 n(n-1)/2$ edges in the constraint graph rules out a fraction $(1-p_2)$ of the possible assignments of values to variables. Thus,

$$\langle Sol \rangle = m^n (1-p_2)^{p_1 n(n-1)/2} \qquad N = n \log_2(m)$$

Substituting these into the definition of constrainedness gives,

$$\kappa = \frac{n-1}{2} p_1 \log_m(\frac{1}{1-p_2})$$

Gent *el al.* [7] show experimentally that rapid transitions in solubility occur around $\kappa \approx 1$ for a selection of model B problems with between 10 and 110 variables and domains of sizes between 3 and 50. Problem hardness for a wide variety of algorithms tends to peak around these transitions.

4 The problem with random problems

Achlioptas *et al.* [1] identify a shortcoming of all four random models. They prove that if $p_2 \geq 1/m$ then, as n goes to infinity, there almost surely exists a *flawed* variable, one which has every value *unsupported*. A value for a variable is unsupported if, when the value is assigned to the variable, there exists an adjacent variable in the constraint graph that cannot be assigned a value without violating a constraint. A problem with a flawed variable cannot have a solution. They argue that therefore *"...the currently used models are asymptotically uninteresting except, perhaps, for a small region of their parameter space ..."* (when $p_2 < 1/m$). Further, they claim that *"...the threshold-like picture given by experimental results [with these models] is misleading, since the problems with defining parameters in what is currently perceived as the underconstrained region (because a solution can be found fast) are in fact overconstrained for large n (obviously, larger than the values used in experiments) ..."*. Note that this result does not apply to problems in which the constraints have certain types of structure. For example, if each constraint only allows variables to take different values then problems encode graph colouring, which has good asymptotic properties.

Achlioptas *et al.* [1] propose an alternative random problem class, model E which does not suffer from this asymptotic shortcoming, and which does not separate the generation of the constraint graph from the selection of the nogoods. In this model, we select uniformly, independently and with repetitions, $pm^2 n(n-1)/2$ nogoods out of the $m^2 n(n-1)/2$ possible. They prove that if a random instance generated using this model has less than $n/2$ nogoods then it almost surely has a solution (theorem 6, page 113). They conjecture that substantially

stronger bounds could be derived to increase the number of allowed nogoods. We note that model E is not entirely novel since Williams and Hogg study random problems with both a fixed number of nogoods picked uniformly, and with an uniform probability of including a nogood [15]. As Achlioptas *et al.* themselves remark [1], the expected number of repetitions in model E is usually insignificant (for instance, it is $O(1)$ when the number of nogoods is $\Theta(n)$), and repetitions are only allowed to simplify the theoretical analysis. The differences between model E and the models of Williams and Hogg are therefore likely to be slight.

5 Experimental practice

Achlioptas *et al.*'s result does not apply to random problems for which $p_2 < 1/m$. To study the practical significance of this restriction, we surveyed the literature from 1994 (when phase transition experiments with random constraint satisfaction problems first started to appear), covering all papers in the proceedings of CP, AAAI, ECAI and IJCAI which gave details of experiments on random constraint satisfaction problems. The results of this survey are summarized in Tables 1 and 2. An experimental study is deemed "inapplicable" if the problem sets tested include an ensemble of problems with $p_2 < 1/m$.

Conference	Inapplicable studies	Total studies
AAAI-94	2	3
ECAI-94	0	4
CP-95	3	4
IJCAI-95	1	5
AAAI-96	0	4
CP-96	3	5
ECAI-96	1	5
AAAI-97	2	4
CP-97	0	7
IJCAI-97	0	1
totals	12	42

Table 1. Summary of results of the literature survey.

Just over a quarter of papers include problems to which the results of [1] do not apply. The most common exception are random problems with $m = 3$ and $p_2 = 1/9$ or $2/9$. Model B is the most common model of generation, followed by model A. Whilst a significant number of papers use problems outside the scope of [1], nearly three quarters use problem sets that are vulnerable to these criticisms. In addition, all of the papers which included inapplicable problem sets also used some instances with $p_2 \geq 1/m$. In conclusion therefore, the results of [1] apply to most published experiments.

Conference	Author initials	Model	$\langle n, m \rangle$	$p_2 < 1/m$?
AAAI-94	[DF,RD]	B	$\langle 25-250, 3 \rangle$	$p_2 = 1/9, 2/9$
	[DF,RD]	B	$\langle 25-275, 3 \rangle$	$p_2 = 1/9, 2/9$
			$\langle 15-60, 6 \rangle$	$p_2 = 4/36$
			$\langle 15-35, 9 \rangle$	no
	[NY,YO,HH]	B	$\langle 20, 10 \rangle$	no
ECAI-94	[PP]	D	$\langle 20, 10 \rangle$, $\langle 20, 20 \rangle$, $\langle 30, 10 \rangle$	no
	[BMS]	B	$\langle 8, 10 \rangle$	no
	[DL]	B	$\langle 10, 20 \rangle$	no
	[DS,ECF]	A	$\langle 50, 8 \rangle$	no
CP-95	[IPG,EM,PP,TW]	B	$\langle 10-110, 3 \rangle$	$p_2 = 2/9$
			$\langle 10, 10 \rangle$, $\langle 20, 10 \rangle$, $\langle 10, 5-50 \rangle$, ...	no
	[JL,PM]	A	$\langle 10, 10 \rangle$	no
	[FB,PR]	B	$\langle 25, 3 \rangle$	$p_2 = 1/9$
			$\langle 35, 6 \rangle$, $\langle 50, 6 \rangle$	$p_2 = 4/36$
			$\langle 15, 9 \rangle$, $\langle 35, 9 \rangle$	no
	[FB,AG]	B	$\langle 25, 3 \rangle$	$p_2 = 2/9$
			$\langle 25, 6 \rangle$, $\langle 15, 9 \rangle$	no
IJCAI-95	[ECF,PDH]	A	$\langle 50, 8 \rangle$	no
	[DF,RD]	B	$\langle 125, 3 \rangle$	$p_2 = 1/9$
			$\langle 35, 6 \rangle$	$p_2 = 4/36$
			$\langle 250, 3 \rangle$, $\langle 50, 6 \rangle$, $\langle 35, 9 \rangle$, ...	no
	[PM,JL]	D	$\langle 10, 10 \rangle$, $\langle 20, 10 \rangle$, $\langle 30, 10 \rangle$	no
	[KK,RD]	B	$\langle 100, 8 \rangle$	no
	[BMS,SAG]	B	$\langle 20, 10 \rangle$, $\langle 50, 10 \rangle$	no
AAAI-96	[AC,PJ]	B	$\langle 16, 8 \rangle$, $\langle 32, 8 \rangle$	no
	[ECF,CDE]	B	$\langle 100, 6 \rangle$	no
	[IPG,EM,PP,TW]	B	$\langle 20, 10 \rangle$	no
	[KK,RD]	B	$\langle 100, 8 \rangle$, $\langle 125, 6 \rangle$, $\langle 150, 4 \rangle$	no
CP-96	[CB,JCR]	B	$\langle 35, 6 \rangle$	$p_2 = 4/36$
			$\langle 125, 3 \rangle$, $\langle 350, 3 \rangle$	$p_2 = 1/9$
			$\langle 35, 9 \rangle$, $\langle 50, 6 \rangle$, $\langle 50, 20 \rangle$, ...	no
	[DAC,JF,IPG,EM,NT,TW]	B	$\langle 20, 10 \rangle$	no
	[IPG,EM,PP,BMS,TW]	B	$\langle 20-50, 10 \rangle$	no
	[JL,PM]	B	$\langle 15, 5 \rangle$	$p_2 = 1/25 - 4/25$
			$\langle 10, 10 \rangle$	$p_2 = 1/100 - 9/100$
	[RJW]	A	$\langle 30, 5 \rangle$, $\langle 100, 5 \rangle$	$p_2 = 0.1$
ECAI-96	[JEB,EPKT,NRW]	B	$\langle 50, 10 \rangle$	no
	[BC,GV,DM,PB]	B	$\langle 50, 10 \rangle$, $\langle 20, 5 \rangle$	no
	[SAG,BMS]	B	$\langle 30-70, 10 \rangle$	no
	[ACMK,EPKT,JEB]	B	$\langle 30, 5 \rangle$	$p_2 = 0.12$
			$\langle 40, 5 \rangle$	$p_2 = 0.08$
			$\langle 60, 5 \rangle$	$p_2 = 0.04$
			$\langle 10, 5 \rangle$, $\langle 20, 5 \rangle$, $\langle 10, 10 \rangle$, ...	no
	[Jl,PM]	B	$\langle 10, 10 \rangle$	no
AAAI-97	[AM,SES,GS]	B	$\langle 6-12, 9 \rangle$	no
	[DRG,WKJ,WSH]	B	$\langle 10, 5 \rangle$	no
	[IPG,EM,PP,TW]	B	$\langle 10-120, 3 \rangle$	$p_2 = 2/9$
			$\langle 10, 10-100 \rangle$	no
	[DF,IR,LV]	B	$\langle 20, 4 \rangle$	$p_2 = 0.125$
			$\langle 150, 3 \rangle$	$p_2 = 0.222$
			$\langle 20-75, 6 \rangle$, $\langle 20, 10 \rangle$	no
CP-97	[IPG,JLU]	D	$\langle 10, 10 \rangle$	no
	[IR,DF]	B	$\langle 100, 8 \rangle$	no
	[DS,ECF]	B	$\langle 20, 20 \rangle$, $\langle 40, 20 \rangle$	no
	[BMS,SAG]	B	$\langle 10, 10 \rangle$	no
	[PG,JKH]	B	$\langle 50, 10 \rangle$, $\langle 100, 15 \rangle$, $\langle 250, 25 \rangle$, ...	no
	[ED,CB]	B	$\langle 100, 20 \rangle$	no
	[IPG,EM, PP, PS, TW]	B	$\langle 20-70, 10 \rangle$	no
IJCAI-97	[RD, CB]	B	$\langle 20, 10 \rangle$	no

Table 2. Parameters and models used in some previous studies of random constraint satisfaction problems.

6 Probability of flawed variables

As Achlioptas *et al.* themselves suggest [1], previous experimental studies will not have been greatly influenced by the existence of flawed variables since problem sizes are usually too small. Using the Markov inequality, they give a first moment bound on the probability of a flawed variable,

$$\Pr\{\text{problem has a flawed variable}\} \leq n(1-(1-p_2^m)^n)^m$$

For example, for the popular $\langle n, 10, 1, 1/2\rangle$ problem class, they calculate that the probability of a flawed variable is less than 10^{-5} even for n as large as 200. At what size of problem and sample do flawed variables start to occur?

By making a few simplifying assumptions, we can estimate the probability of a flawed variable with reasonable accuracy. This estimate might be used to determine parameters for experimental studies. Our first assumption is that each variable is connected to exactly $p_1(n-1)$ others. In practice, some variables have a greater degree, whilst others have a lesser degree. Fortunately, our experiments show that this mean-field approximation does not introduce a large error into the estimate. We also assume independence between the probabilities that the different variables have at least one unflawed value. The probability that there are no flawed variables is then simply the product of the probabilities that the variables have at least one unflawed value. For model A problems, we have,

$$\begin{aligned}
&\Pr\{\text{problem has a flawed variable}\}\\
&= 1-\Pr\{\text{there are no flawed variables}\}\\
&= 1-(\Pr\{\text{a variable has at least one unflawed value}\})^n\\
&= 1-(1-\Pr\{\text{every value for the variable is flawed}\})^n\\
&= 1-(1-(\Pr\{\text{a value for the variable is flawed}\})^m)^n\\
&= 1-(1-(\Pr\{\text{value inconsistent with every value of an adjacent variable}\})^m)^n\\
&= 1-(1-(1-\Pr\{\text{value consistent with a value of every adjacent variable}\})^m)^n\\
&= 1-(1-(1-(\Pr\{\text{value consistent with a value of an adjacent variable}\})^{p_1(n-1)})^m)^n\\
&= 1-(1-(1-(1-\Pr\{\text{value inconsistent with every value of adjacent variable}\})^{p_1(n-1)})^m)^n\\
&= 1-(1-(1-(1-(\Pr\{\text{value inconsistent with a value of adjacent variable}\})^m)^{p_1(n-1)})^m)^n
\end{aligned}$$

For model A, the probability that a given value is inconsistent with every value of an adjacent variable is p_2. Hence, we obtain the estimate,

$$\Pr\{\text{problem has a flawed variable}\} = 1-(1-(1-(1-p_2^m)^{p_1(n-1)})^m)^n$$

A similar derivation can be made for model B problems. In this model each constraint is picked uniformly from the ${}_{m^2}C_{p_2 m^2}$ possible binary constraints. If we assign a value to one of the variables involved in a constraint, then ${}_{m^2-m}C_{p_2 m^2-m}$ of the possible constraints have nogoods that rule out all the values for the other variable. Hence, the probability that a particular value for a variable is inconsistent with every value for an adjacent variable is given by,

$$\Pr\{\text{value inconsistent with every value of adjacent variable}\} = \binom{m^2-m)}{p_2 m^2-m} \Big/ \binom{m^2}{p_2 m^2}$$

Thus, for model B problems, we obtain the estimate,

$$\Pr\{\text{problem has a flawed variable}\} = 1-(1-(1-(1-\binom{m^2-m}{p_2 m^2-m}/\binom{m^2}{p_2 m^2})^{p_1(n-1)})^m)^n$$

Note that we have assumed independence between the probabilities that the m different values for a given variable are flawed. The probability that every value for a variable is flawed is then simply the product of the probabilities that each individual value is flawed. Whilst this independence assumption is valid for model A, it is not strictly true for model B.

7 Problem size

We can use these estimates for the probability of flawed variables to determine when flawed variables will start to occur in experimental studies. To test the accuracy of these estimates and to compare them with the simpler first moment bound, we generated random problems from the popular model B and calculated the fraction with a flawed variable. Since flawed variables are more likely in dense constraint graphs, we generated problems with complete constraint graphs (i.e. with $p_1 = 1$). As in other studies (e.g. [12, 6]), we also generated a separate set of problems in which the average degree of the vertices in the constraint graph is kept constant. That is, we vary p_1 as $1/(n-1)$. As we argue in Section 9, the constraint tightness at the phase transition then remains roughly constant. Keeping the average degree constant also reduces the probability of flawed variables occurring. In Table 3, we give the results for $\langle n, 10, 1, 1/2\rangle$ and $\langle n, 10, 19/(n-1), 1/2\rangle$ with n from 200 to 4000. In this (and indeed all the subsequent experiments) our estimate for the probability of a problem having a flawed variable is very close to the observed fraction of problems with flawed variables, and much closer than the first moment bound to the observed fraction of flawed variables.

With complete constraint graphs, flawed variables are observed in samples of 1000 when the problems have 500 or more variables. This is beyond the size of problems typically solved with systematic procedures but potentially within the reach of approximation or local search algorithms. By comparison, with constraint graphs of constant average degree, flawed variables are not observed in samples of 1000 even when the problems have thousands of variables. Because of the greater homogeneity of model B problems, we expect flawed variables to be less likely than in model A. Our estimates for the probability of a flawed variable support this conjecture. For example, for $\langle 1000, 10, 1, 1/2\rangle$ problems, our estimate for the probability that a model A problem has a flawed variable is 0.99986 whilst for a model B problem it is 0.275.

With constraint graphs of constant average degree, we can estimate when we expect to observe flawed variables. If $p_1 = \gamma/(n-1)$ and a fraction f of problems contain flawed variables then, by rearranging our estimates for the probability of a flawed variable, the number of variables n_f in model A problems is,

$$n_f = \frac{\log(1-f)}{\log(1-(1-(1-p_2^m)\gamma)^m)}$$

n	sample size	fraction with flawed variables	estimate for $\Pr\{\text{flawed variable}\}$	1st moment bound
200	10^6	0.000000	0.000000	0.000006
500	10^4	0.0005	0.0006	0.0370
1000	10^3	0.272	0.275	> 1
1200	10^3	0.753	0.755	> 1
1500	10^3	1.000	0.999	> 1
2000	10^3	1.000	1.000	> 1
4000	10^3	1.000	1.000	> 1

(a) $\langle n, 10, 1, 1/2\rangle$

n	sample size	fraction with flawed variables	estimate for $\Pr\{\text{flawed variable}\}$	1st moment bound
200	10^3	0.000	0.000	0.000
500	10^3	0.000	0.000	0.037
1000	10^3	0.000	0.000	> 1
1500	10^3	0.000	0.000	> 1
2000	10^3	0.000	0.000	> 1
4000	10^3	0.000	0.000	> 1

(b) $\langle n, 10, 19/(n-1), 1/2\rangle$

Table 3. The impact of flawed variables on model B problems with a domain size of 10 and: **(a)** complete constraint graphs; **(b)** constraint graphs of constant average degree.

And in model B problems,

$$n_f = \frac{\log(1-f)}{\log(1-(1-(1-\left(\begin{smallmatrix} m^2-m \\ p_2 m^2 - m \end{smallmatrix}\right)/\left(\begin{smallmatrix} m^2 \\ p_2 m^2 \end{smallmatrix}\right))^\gamma)^m)}$$

For instance, for model B problems with similar parameters to those of Table 3 (i.e. $m = 10$, $\gamma = 19$ and $p_2 = 1/2$), $n_{1/1000} \approx 3.2 * 10^{17}$ and $n_{1/2} \approx 2.2 * 10^{19}$. That is, problems need more than 10^{17} variables before we start to observe flawed variables in samples of 1000 problem instances, and more than 10^{19} variables before half contain a flawed variable. As a consequence, at this domain size, constraint tightness, and degree of the constraint graph, experimental studies can safely ignore flawed variables.

With smaller domain sizes, we expect flawed variables to be more prevalent. To test this hypothesis, we generated problems with $m = 3$, $p_2 = 1/m$ and either complete constraint graphs or constraint graphs of constant average degree. Note that, for model B, $p_2 = 1/m$ is the smallest possible value which gives flawed variables. If $p_2 < 1/m$ then at least one value for each variable *must* be supported as each constraint rules out strictly less than m possible values. Note also that these problems have the same domain size and same constraint tightness as 3-colouring problems. Table 4 gives the results for $\langle n, 3, 1, 1/3\rangle$ and $\langle n, 3, 19/(n-1), 1/3\rangle$ with $n = 10$ to 2000. With complete constraint graphs, flawed variables occur with a significant frequency in problems with as few as 20 variables. This

is despite p_2 being given the minimal possible value. With constraint graphs of constant average degree, although flawed variables occur in problems with as few as 20 variables, their frequency increases much more slowly with n. We need a thousand or more variables to ensure that problems almost always include a flawed variable. By comparison, with complete constraint graphs, we need just 60 or so variables. Some of the experiments surveyed in Section 5 used random problems containing hundreds of variables with $m = 3$ and $p_2 = 1/3$. Flawed variables may therefore have had a significant impact on these experiments.

n	sample size	fraction with flawed variables	estimate for Pr{flawed variable}	1st moment bound
10	10^3	0.006	0.011	0.311
20	10^3	0.143	0.156	> 1
30	10^3	0.504	0.536	> 1
40	10^3	0.869	0.882	> 1
50	10^3	0.987	0.990	> 1
60	10^3	1.000	1.000	> 1

(a) $\langle n, 3, 1, 1/3 \rangle$

n	sample size	fraction with flawed variables	estimate for Pr{flawed variable}	1st moment bound
20	10^3	0.143	0.156	> 1
50	10^3	0.318	0.345	> 1
100	10^3	0.524	0.571	> 1
200	10^3	0.796	0.816	> 1
500	10^3	0.986	0.985	> 1
1000	10^3	0.999	1.000	> 1
2000	10^3	1.000	1.000	> 1

(b) $\langle n, 3, 19/(n-1), 1/3 \rangle$

Table 4. The impact of flawed variables on model B problems with a small domain size and: **(a)** complete constraint graph; **(b)** constraint graph of constant average degree.

8 Model E

At the sizes typically used in previous experimental studies, how does model E differ from the conventional models? To explore this issue, we compared problems from model E with $n = 20$ and $m = 10$ against problems of a similar size from the popular model B. As we argue in the next section, model E quickly gives problems with complete constraint graphs. We therefore used model B problems with $p_1 = 1$ as a comparison. For model B, we generated 1000 problems at each value of p_2 between 1/100 and 99/100. For model E, we generated 1000 problems at each value of p from 1/190 to 500/190 in steps of 1/190. Note that model E allows for repetitions when selecting nogoods so p can be greater than 1.

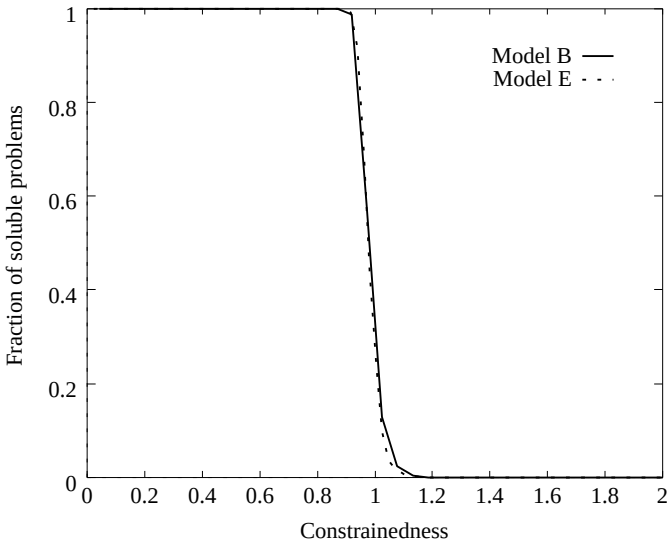

Fig. 1. Fraction of soluble problems against constrainedness, κ

To aid comparison, we estimated the constrainedness, κ of the generated problems. We have found κ a useful measure for comparing algorithm performance across a wide variety of different problem classes [9]. Since the nogoods in model E are selected independently and with repetitions, κ is approximately proportional to p. In Figure 1, we plot the fraction of soluble problems against the constrainedness. In both models, we see a rapid transition between soluble and insoluble problems at around $\kappa \approx 1$ as predicted. Associated with this transition is a peak in search cost. In Figure 2, we plot the median consistency checks performed by the forward checking algorithm with conflict-directed backjumping and the fail-first heuristic (FC-CBJ-FF). The search cost for the two models is very similar, depending almost entirely on their constrainedness and size. The only slight difference is that at very small values of p, model E problems do have complete constraint graphs and are easier to solve. We discuss the size of the constraint graph in more detail in the next section.

9 Constraint graph

Some of the experimental studies listed in Section 5 keep p_1 constant as n increases. Even if problem and sample sizes are small enough that flawed variables are unlikely, this may not be a very good idea. The transition between soluble and insoluble problems occurs around $\kappa \approx 1$. That is, when $-\frac{n-1}{2} p_1 \log_m(1-p_2) \approx 1$. If we fix m and p_1 then p_2 decreases as we increase n. Eventually p_2 is less than $1/m^2$ and, in model B at least, we are unable to generate any non-empty constraints. For instance, with $p_1 = 1$, $m = 3$ and $\kappa \approx 1$, p_2 is smaller than $1/m^2$

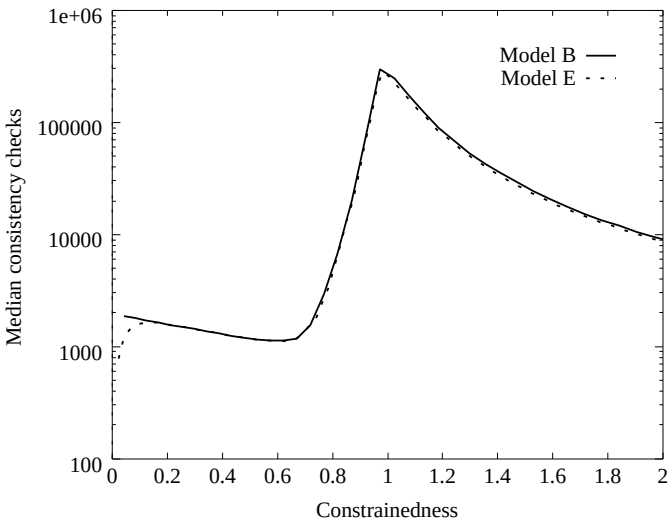

Fig. 2. Median search cost for FC-CBF-FF against constrainedness, κ

for n larger than about 20. In other words, even though flawed variables cannot occur since $p_2 < 1/m$, we cannot run an experiment at the phase transition with $m = 3$ and $p_1 = 1$ for n larger than about 20.

It may be better experimental practice to maintain the topology of the constraint graph by keeping the average degree constant. That is, to vary p_1 as $1/(n-1)$. If $\kappa \approx 1$ and $p_1 = \gamma/(n-1)$, then $p_2 \approx 1 - m^{-2/\gamma}$ which is constant. Hence the phase transition is expected to occur at a roughly constant value of p_2 as n varies. Experimental data for small n supports this conclusion. For example, Figure 2 of [12] shows that the transition in solubility for model B problems with $m = 10$ and $p_1 = 4.9/(n-1)$ occurs at $p_2 \approx 1 - m^{-2/4.9} \approx 0.6$ as n increases from 10 to 50. Of course, since $p_2 \geq 1/m$, such problems contain flawed variables and are trivially insoluble for large enough n. However, as we argued before, n needs to be so large that our experiments can safely ignore this fact. For instance, for $m = 10$, $p_1 = 4.9/(n-1)$, and $p_2 = 0.6$, we calculate that $n_{1/1000} \approx 5.6 * 10^{13}$ and $n_{1/2} \approx 3.8 * 10^{16}$. That is, problems need more than 10^{13} variables before we expect to observe flawed variables in samples of 1000 problem instances, and more than 10^{16} variables before half contain a flawed variable.

One shortcoming of model E is that it generates complete constraint graphs for quite small values of p, even though each constraint contains just a few nogoods. It is hard therefore to generate sparse constraint graphs with tight constraints. By comparison, in models A to D we can independently adjust the constraint tightness and density. In model E, we randomly select $pm^2n(n-1)/2$ nogoods independently and with repetitions. By a coupon collector's argument, we expect a complete constraint graph when $p \approx \log(n(n-1)/2)/m^2$. For ex-

ample, for $n = 20$, $m = 10$, we just need $p \approx 0.052$ before we expect a complete constraint graph. With a larger number of nogoods, there is a very small probability that the constraint graph is not complete. Assuming independence between the nogoods, we can estimate this probability as follows,

$$
\begin{aligned}
&\Pr\{\text{constraint graph is complete}\} \\
&= \Pr\{\text{all pairs of vertices are joined}\} \\
&= (\Pr\{\text{two given vertices are joined}\})^{n(n-1)/2} \\
&= (1 - \Pr\{\text{two given vertices are not joined}\})^{n(n-1)/2} \\
&= (1 - \Pr\{\text{no nogoods mention the two variables}\})^{n(n-1)/2} \\
&= (1 - (\Pr\{\text{a given nogood does not mention the two}\})^{pm^2 n(n-1)/2})^{n(n-1)/2} \\
&= (1 - (1 - \Pr\{\text{a given nogood does mention the two}\})^{pm^2 n(n-1)/2})^{n(n-1)/2}
\end{aligned}
$$

As there are $n(n-1)/2$ possible pairs of variables that a nogood could mention, the probability that any nogood does not mention two given variables is simply $2/n(n-1)$. Hence,

$$\Pr\{\text{constraint graph is complete}\} = (1 - (1 - \frac{2}{n(n-1)})^{pm^2 n(n-1)/2})^{n(n-1)/2}$$

For example, for $n = 20$ and $m = 10$, the probability that the constraint graph is incomplete is less than 10^{-2} when $p = 1/m$, and less than 10^{-18} when $p = 1/2$.

We can generalize model E to tackle this problem by reversing the usual process of generating a constraint graph and then selecting nogoods within it. In model F, we select uniformly, independently and with repetitions, $p_1 p_2 m^2 n(n-1)/2$ nogoods out of the $m^2 n(n-1)/2$ possible. We then generate a constraint graph with exactly $p_1 n(n-1)/2$ edges and throw out any nogoods that are not between connected vertices in this graph. Note that model E is a special case of model F in which $p_1 = 1$. Using similar arguments to [1], we can show that model F (like model E) is not trivially insoluble as we increase problem size. In addition, by setting p_1 small but p_2 large, we can generate sparse constraint graphs with tight constraints. We leave it as an open question if there are models with good asymptotic properties which admit problems with a few tight constraints, but which do not throw out nogoods.

10 Non-random problems

Random problems provide a plentiful and unbiased source of problems for benchmarking. However, we must be careful that our algorithms do not become tuned to solve random problems and perform poorly on real problems. All of the models discussed here generate simple binary constraints, but real problems can contain structures that occur very rarely in these models. For example, in a graph colouring problem derived from a real exam time-tabling problem at Edinburgh University, Gent and Walsh found a 10 clique of nodes with only 9 colours available [10]. This was in a 59 node graph with 594 edges. The presence of this clique dominated the performance of their graph colouring algorithm. Random graphs of similar size and density are very unlikely to contain such a large clique.

The probability that m given nodes in a random graph with n nodes and e edges are connected by the right $m(m-1)/2$ edges to form a m-clique is,

$$\Pr\{m \text{ given nodes form a } m\text{-clique}\} = \prod_{i=0}^{\frac{m(m-1)}{2}-1} \frac{e-i}{\frac{n(n-1)}{2}-i}$$

Multiplying this probability by ${}_nC_m$, the number of sets of m nodes in a n node graph, we get the expected number of m-cliques. By the Markov inequality, this gives a bound on the probability of the graph containing a m-clique,

$$\Pr\{m\text{-clique in graph of } n \text{ nodes \& } e \text{ edges}\} \leq \frac{n!}{m!(n-m)!} \prod_{i=0}^{\frac{m(m-1)}{2}-1} \frac{e-i}{\frac{n(n-1)}{2}-i}$$

For $n = 59$, $m = 10$ and $e = 594$, the probability of clique of size 10 or larger is less than 10^{-10}. It is thus very unlikely that a random graph of the same size and density as the graph in the exam time-tabling problem would contain a regular structure like a 10-clique. However, cliques of this size are very likely in the real data due to the module structure within courses.

As another example, Gomes *et al.* have proposed quasigroup completion as a constraint satisfaction benchmark that models some of the structure found in time-tabling problems [11]. Quasigroup completion is the problem of filling in the missing entries in a Latin square, a multiplication table in which each entry appears once in every row and column. An order n quasigroup problem can be formulated as n-colouring a graph with n^2 nodes and $n^2(n-1)$ edges. The edges form $2n$ cliques, with each clique being of size n and representing the constraint that each colour appears once in every row or column. For example, an order 10 quasigroup has 20 cliques of size 10 in a 100 node graph with 900 edges. With a random graph of this size and edge density, the probability of a clique of size 10 or larger is less than 10^{-20}. It is thus unlikely that a random graph of this size and density would contain a regular structure like a 10-clique, let alone 20 of them linked together. The random models are thus unlikely to generate problems like the exam time-tabling problem or quasigroup completion.

11 Conclusions

We have performed a detailed study of the experimental consequences of a recent theoretical result of Achlioptas *et al.* [1]. This result shows that, as we increase problem size, the conventional models of random problems almost surely contain a flawed variable and are therefore trivially insoluble. Our survey of previous experimental studies shows that most meet the restriction on their result that $p_2 \geq 1/m$. Fortunately, most (but not all) of these studies use too few variables and too large domains to be affected by the result. As expected, flawed variables occur most often with dense constraint graphs and small domains. With constraint graphs of fixed average degree and large domains, flawed variables

can be safely ignored. Achlioptas *et al.* propose an alternative random model which does not suffer from this asymptotic problem. We show that, at a typical problem size used in experiments, this model gives problems of a similar solubility and hardness to conventional models. However, it has a small experimental weakness since we cannot run tests with sparse but tight constraints. We therefore generalized the model so that we can independently adjust the constraint tightness and density. Finally, we showed that some of the structures that occur in real problems like large cliques are very rare in these random models.

What general lessons can be learnt from this study? First, experiments can benefit greatly from theory. Flawed variables are likely to have occurred in a small but significant number of previous experimental studies. A simple arc consistency algorithm would therefore have solved some of these problems. Experimental practice can now be improved to ensure that either we use an alternative model without this asymptotic problem, or we use a conventional model but choose parameters so that flawed variables are unlikely. Second, theory can benefit greatly from experiments. Theory provided an estimate for the probability of problems having flawed variables. Experiments quickly determined the accuracy of this estimate. Third, we must continue to improve and extend our random models so that we can generate a wide range of hard problems with which to test our algorithms.

References

1. D. Achlioptas, L.M. Kirousis, E. Kranakis, D. Krizanc, M.S.O. Molloy, and C. Stamatiou. Random constraint satisfaction: A more accurate picture. In *Proceedings of Third International Conference on Principles and Practice of Constraint Programming (CP97)*, pages 107–120, 1997.
2. P. Cheeseman, B. Kanefsky, and W.M. Taylor. Where the really hard problems are. In *Proceedings of the 12th IJCAI*, pages 331–337. International Joint Conference on Artificial Intelligence, 1991.
3. E. Friedgut. Sharp thresholds for graph properties and the k-SAT problem, 1998. Unpublished manuscript.
4. A. Frieze and S. Suen. Analysis of two simple heuristics on a random instance of k-SAT. *Journal of Algorithms*, 20:312–355, 1996.
5. I.P. Gent, E. MacIntyre, P. Prosser, P. Shaw, and T. Walsh. The constrainedness of arc consistency. In *3rd International Conference on Principles and Practices of Constraint Programming (CP-97)*, pages 327–340. Springer, 1997.
6. I.P. Gent, E. MacIntyre, P. Prosser, B.M. Smith, and T. Walsh. An empirical study of dynamic variable ordering heuristics for the constraint satisfaction problem. In *2nd International Conference on Principles and Practices of Constraint Programming (CP-96)*, pages 179–193, 1996.
7. I.P. Gent, E. MacIntyre, P. Prosser, and T. Walsh. Scaling effects in the CSP phase transition. In *1st International Conference on Principles and Practices of Constraint Programming (CP-95)*, pages 70–87. Springer-Verlag, 1995.
8. I.P. Gent, E. MacIntyre, P. Prosser, and T. Walsh. The constrainedness of search. In *Proceedings of the 13th National Conference on AI*, pages 246–252. American Association for Artificial Intelligence, 1996.

9. I.P. Gent, E. MacIntyre, P. Prosser, and T. Walsh. The scaling of search cost. In *Proceedings of the 14th National Conference on AI*, pages 315–320. American Association for Artificial Intelligence, 1997.
10. I.P. Gent and T. Walsh. Phase transitions from real computational problems. In *Proceedings of the 8th International Symposium on Artificial Intelligence*, pages 356–364, 1995.
11. C. Gomes and B. Selman. Problem structure in the presence of perturbations. In *Proceedings of the 14th National Conference on AI*, pages 221–226. American Association for Artificial Intelligence, 1997.
12. S. Grant and B.M. Smith. The phase transition behaviour of maintaining arc consistency. Research Report 95.25, School of Computer Studies, University of Leeds, 1995. A revised and shortened version appears in Proceedings of 12th ECAI, pages 175–179, 1996.
13. L.M. Kirousis, E. Kranakis, and D. Krizanc. Approximating the unsatisfiability threshold of random formulas. In *Proceedings of the 4th Annual European Symposium on Algorithms (ESA'96)*, pages 27–38, 1996.
14. D. Mitchell, B. Selman, and H. Levesque. Hard and Easy Distributions of SAT Problems. In *Proceedings of the 10th National Conference on AI*, pages 459–465. American Association for Artificial Intelligence, 1992.
15. C. Williams and T. Hogg. Exploiting the deep structure of constraint problems. *Artificial Intelligence*, 70:73–117, 1994.