Scaling Effects in the CSP Phase Transition*

Ian P. Gent[†], Ewan MacIntyre,[†] Patrick Prosser[†], and Toby Walsh[‡]

Research Report/95/179
Department of Computer Science, University of Strathclyde
April 18, 1995

Abstract

Phase transitions in constraint satisfaction problems (CSP's) are the subject of intense study. We identify an order parameter for random binary CSP's. There is a rapid transition in the probability of a CSP having a solution at a critical value of this parameter. The order parameter allows different phase transition behaviour to be compared in an uniform manner, for example CSP's generated under different regimes. We then show that within classes, the scaling of behaviour can be modelled by a tehnique called "finite size scaling". This applies not only to probability of solubility, as has been observed before in other NP-problems, but also to search cost, the first time this has been observed. Furthermore, the technique applies with equal validity to several different methods of varying problem size. As well as contributing to the understanding of phase transitions, we contribute by allowing much finer grained comparison of algorithms, and for accurate empirical extrapolations of behaviour.

---

*The fourth author is supported by a HCM Postdoctoral Fellowship. We thank the Department of Computer Science at the University of Strathclyde for CPU cycles, and Judith Underwood for her help.

[†]Department of Computer Science, University of Strathclyde, Glasgow G1 1XH, Scotland.
[‡]Mechanized Reasoning Group, IRST, Trento, and DIST, University of Genoa, Italy.

1 Introduction

A phase transition in random CSP problems has recently been the subject of intensive theoretical and empirical study [29, 13, 25, 26, 21, 22, 5]. Theory predicts approximately where the phase transition can be expected, but otherwise very little information is available as to what behaviour can be expected at different problem sizes and at different points with respect to the phase transition.

Compared to SAT, the model for generation of random CSP problems is complicated, and no order parameter is used in presenting data. Yet the identification of the order parameter for random SAT problems [2, 19] is fundamental to current research on phase transitions in SAT. In this paper we introduce a possible order parameter for CSP. The phase transition is always expected at the same value of this order parameter. Using extant data, we show that it can be much more easily understood by plotting with respect to the order parameter than by raw plots as previously presented.

In the rest of the paper, we show that the order parameter can be used to direct detailed and meaningful comparisons of significantly different methods of generating random binary CSP's. We show that our proposed order parameter gives meaningful results as we change the number of variables in our problems, change the number of values in the domain of variables, or vary both simultaneously. In each case the phase transition in probability seems to occur at similar values of the order parameter with changing problem size. Furthermore, median search cost seems to peak over the same range of the order parameter.

Using our results on these problems, we are able to show that the technique of "finite size scaling" can be applied to problems in CSP phase transitions. The result is an empirical prediction of how probability of solubility varies with problem size, and this prediction might be used to help derive experimental parameters for future experiments. We show this for each different method of varying problem size that we investigate, the first time finite size scaling has been applied to computational phase transition on changing domain size.

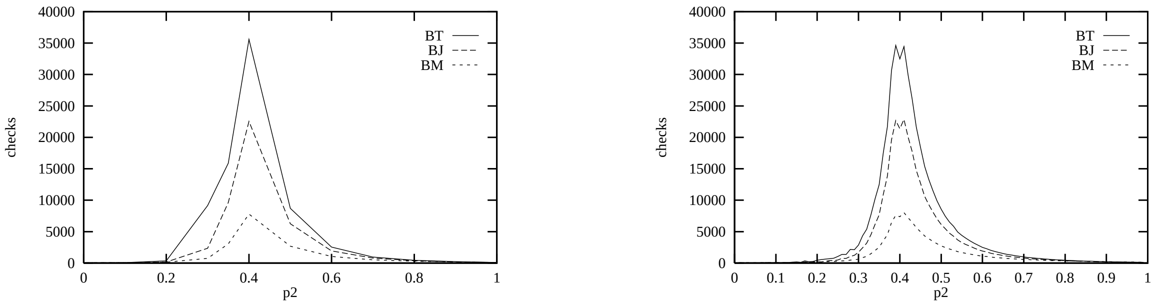
The interest in how the probability of solubility varies with problem size is largely due to the correlation with peak in search cost at that phase transition. We show how remarkable a correlation this is by demonstrating that finite size scaling, applied with parameters derived only from examination of probability data, seems to apply to search cost also. This is the first application we know of finite size scaling applied to a measure of search cost. The implications for the experimental analysis of algorithms are very significant.

2 Binary Constraint Satisfaction Problems

In the binary constraint satisfaction problem (CSP) we have a set of variables, where each variable has a domain of values, and a set of constraints acting between pairs of variables. The problem is then to assign values to variables, from their respective domains, such that the constraints are satisfied [3, 17, 27]. One way of addressing this problem is via systematic search, ie. algorithms based on backtracking search [20], and one of the objectives is to find a solution, or determine that none exists, with minimal search effort, where effort is measured as the number of compatibility checks performed between pairs of variables.

Given a CSP with n variables with uniform domain size of m, there will be m^n possible assignments of variables to values. The best known complete algorithms for CSP's are exponential in the worst case.

Numerous studies have been performed on random CSP's, in order to measure the performance of algorithms [4, 24, 28] and to investigate the nature of problems [21, 25]. Typically, random CSP are categorised using four parameters, namely $\langle n, m, p_1, p_2 \rangle$, where n is the number of variables, m is the uniform domain size, p_1 is the proportion of edges in the constraint graph (ie. the density of the constraint graph), and p_2 is the proportion of pairs of instantiations over a constraint that are disallowed (ie. the tightness of the constraints)[21, 25, 5]. [1] It has been observed that if n, m, and p_1 are held constant, there is a small range of values of p_2 where average search effort rapidly increases to a peak and then falls away, while at the same time the proportion of soluble problems drops to zero. That is, there is a phase transition [25, 21, 5].



(a) (b)

Figure 1: Gaschnig's experiments on random 10-queens, and ours on $\langle 10, 10, 1.0 \rangle$

Probably the earliest report of the complexity peak in CSP's is due to Gaschnig [7]. One of the studies in his thesis was on random 10-queens. In the n-queens problem n non-attacking queens have to be placed on an $n \times n$ chess board, and in the random 10-queens problem a solution (or proof that none exists) has to be found for $\langle 10, 10, 1.0, p_2 \rangle$. Figure 1(a) shows (we believe, for the first time) a plot of the results from Gaschnig's experiments for random 10-queens. Constraint tightness p_2 is varied in steps of 0.1 (with the exception of the point $p_2 = 0.35$) and 150 problems are generated at each point. Three curves are plotted, one for chronological backtracking (BT), one for backjumping (BJ), and one for backmarking (BM) [11, 7, 6].[2] The y-axis is the average number of consistency checks and the x-axis is p_2. Figure 1(a) clearly shows a peak in average search effort at $p_2 = 0.4$ for the three algorithms. [3] In Figure 1(b) the experiments are repeated, but p_2 is varied in

[1] That is, in a random CSP $\langle n, m, p_1, p_2 \rangle$ as defined in [21, 25] there will be exactly $p_1.n.(n-1)/2$ constraints, and each constraint will have exactly $p_2.m^2$ conflicts. A lisp version of such a problem generator, and supporting search algorithms etc., is available via anonymous FTP at site ftp.cs.strath.ac.uk in directory local/pat/csp-lab/

[2] Note that no variable or value ordering heuristics were used.

[3] Gaschnig referred to the CSP as a SAP (satisficing assignment problem) and L as the degree of a constraint (the fraction of distinct pair tests that have the value true, ie. $L = 1 - p_2$). The plot of 1(a) uses the data in Figure 4.4.3-1, page 301 of [7]. Gaschnig noted the existence of a sharp peak at $L \approx 0.6$, pages 179 and 180.

steps of 0.01, and this confirms that the peak in average search effort does indeed occur at $p_2 = 0.4$. Furthermore, 52% of the $\langle 10, 10, 1.0, 0.4\rangle$ problems are soluble. It appears that Gaschnig failed to notice this phenomenon.

3 An Order Parameter for Binary CSP's

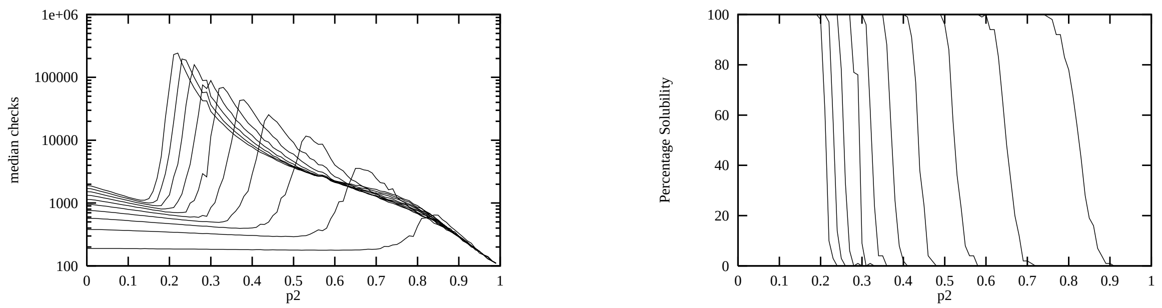
Given the random CSP $\langle n, m, p_1, p_2\rangle$ the expected number of solution is given by (1)

$$E(N) = m^n(1 - p_2)^{\frac{p_1 n(n-1)}{2}} \tag{1}$$

In [25] it is conjectured that average search effort will be greatest when an ensemble of problems have on average one solution, ie. $E(N) = 1$, and this will correspond to the crossover point where half the problems are soluble. An equivalent theory was independently developed by Williams and Hogg [29]. For given values of n, m, and p_1 there will be a critical value of constraint tightness, p_{2crit}, where average search effort will be a maximum, and this may be predicted via (2)

$$p_{2crit} = 1 - m^{-2/((n-1)p_1)} \tag{2}$$

For example, using (2) we can predict the critical value of constraint tightness for $\langle 10, 10, 1.0\rangle$, ie. Gaschnig's random 10-queens experiments, and that is $p_{2crit} = 0.400$, in full agreement with his observations.



(a) (b)

Figure 2: $\langle 20, 10, p_1\rangle$ (a) Median search effort against p_2, and (b) Percentage solubility against p_2

Figure 2 shows the median search effort for the CSP's $\langle 20, 10, p_1, p_2\rangle$ (ie. 20 variables, uniform domain size of 10, p_1 varying from 0.1 to 1.0 in steps of 0.1, p_2 varying from 0.01 to 0.99 in steps of 0.01). In Figure 2(a) 10 contours are given, the leftmost is for $p_1 = 1.0$ and the rightmost for $p_1 = 0.1$. The x-axis is constraint tightness, p_2, varying in steps of 0.01, and the y-axis is the *log* of the median search effort. At each value of p_1 and p_2 one hundred problems were sampled using the algorithm forward checking with conflict-directed backjumping (FC-CBJ) [20] allied to the fail first heuristic (FF) [23, 12]. What we see is that as the density of the constraint graph increases (ie. p_1 increases) the critical value of constraint tightness falls (ie. p_2 falls). Figure 2(b) shows how the solubility of problems varies with p_1 and p_2. Again 10 contours are given, the leftmost for $p_1 = 1.0$ and the rightmost for $p_1 = 0.1$.

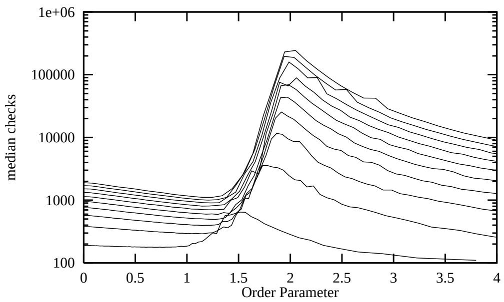
In some respects Figure 2 suggests that it might be difficult to compare CSP's of different size (ie. varying n and m) or structure (ie. varying p_1 and p_2) because their graphs translate along the x-axis, changing shape as they go. [4] What we would like to find is some parameter that characterises CSP regardless of size or structure, ie. an *order parameter*. Order parameters have been identified for 3-SAT, namely $\frac{clauses}{variables}$, and in 3-COL the average degree γ [2]. We can derive an order parameter τ for CSP's as follows. First, we rearrange (2) to get

$$(n-1)p_1 \log_m\left(\frac{1}{1-p_{2crit}}\right) = 2 \tag{3}$$

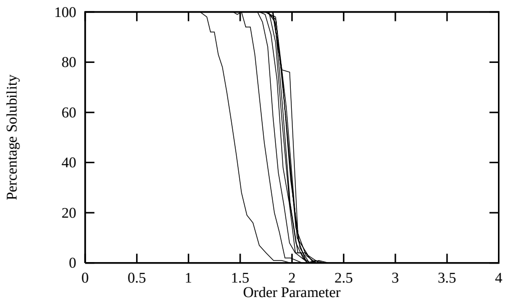
This gives us a prediction for the location of the crossover point expressed by a function of the random generation parameters taking a certain *constant* value, and that function no longer has a first order dependency on n, ie. it does not have the exponential behaviour of (1). This immediately suggests that the LHS of (3), may be a suitable order parameter τ for CSP's. Accordingly, we *define* the order parameter, which we call τ, by

$$\tau =_{def} (n-1)p_1 \log_m\left(\frac{1}{1-p_2}\right)$$

Note that we define τ in terms of p_2 instead of p_{2crit} in (3). This means that τ is defined for *all* values of n, m, p_1, and p_2, and so can be used to compare CSP's generated with different parameters. If the theory of equation (2) were exactly correct, then values of τ less than 2 would lead to soluble problems, while values more than 2 would give insoluble problems. We will see in the rest of this paper that this is a reasonable but not completely accurate prediction.



(a)



(b)

Figure 3: $\langle 20, 10, p_1 \rangle$ (a) Median search effort against the order parameter, and (b) Percentage solubility against the order parameter

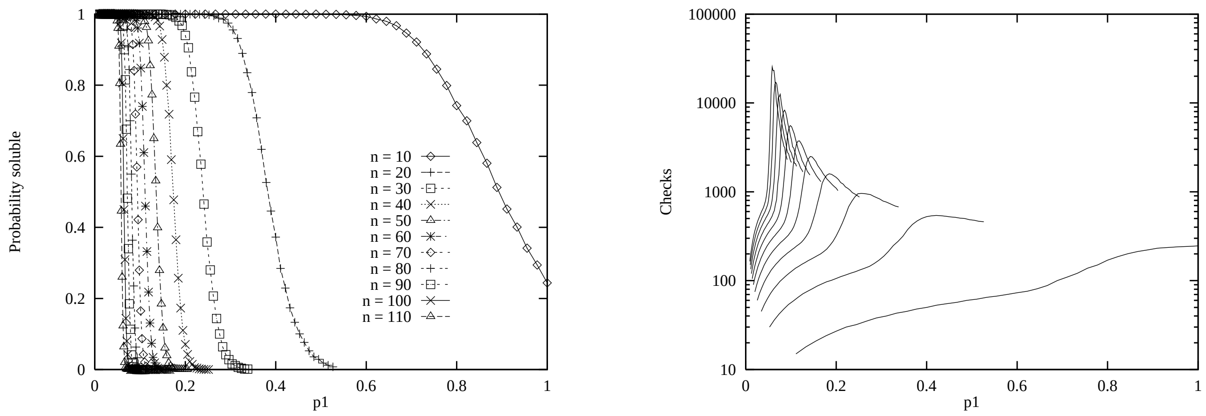
Figure 3 shows the same data as in Figure 2 but with the order parameter on the x-axis. In Figure 3(a) the contours of median search effort peak when $1.5 \leq \tau \leq 2.0$,

[4] Although all the curves in Figure 2(a) have the same *signature*, at high values of p_1 they are more defined.

close to the expected value of the order parameter. However, for increasing p_1 the phase transition occurs more sharply and at values of the order parameter nearer 2. That is, for denser constraint graphs, the prediction of Smith, Williams and Hogg for the location of the phase transition becomes more accurate. Finally, it is clear that as p_1 increases towards 1, so the peak mean search effort increases considerably. Only the last of these points could have been seen clearly from the graphs as presented in Figure 2, even though in this case we have presented *less* data – we cut off values of p_2 giving an order parameter greater than 4.0. These points were made in [21], but this had to be done by further analysis, and could not be read off directly from a simple plot as we did by looking at Figure 3.

We hope that our graphs argue for us the case that data should be presented with respect to the order parameter we propose. We show in the rest of the paper that there are further advantages to study of the order parameter. Indeed we will be able to make detailed numerical predictions based on it.

4 Changing Number of Variables



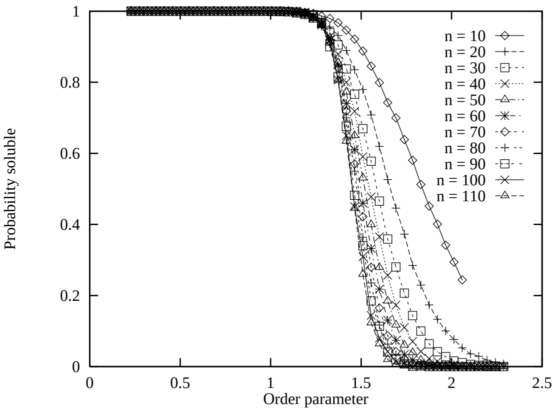
(a) Probability (b) Median Checks

Figure 4: $\langle n, 3, p_1, \frac{2}{9} \rangle$ plotted using p_1

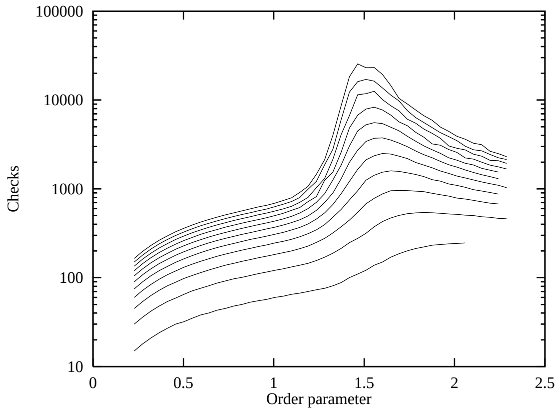
A set of experiments was carried out to investigate the effect of increasing the number of variables, ie. n, on the order parameter τ. In these experiments we vary the number of variables and the density of the constraint graph, while holding the domain size and tightness of constraints constant. We chose a domain size $m = 3$ and tightness of constraint $p_2 = \frac{2}{9}$, corresponding to the set of experiments reported in [5], and we will refer to them as $\langle n, 3, p_1, \frac{2}{9} \rangle$.[5] The search algorithm used for this set of experiments, and all subsequent experiments reported here, was FC-CBJ-FF, ie. forward checking with conflict-directed backjumping using the fail-first heuristic. A report on the implementation of this procedure is given in [18]. For each n from 10 to 110 in steps of 10 we tested problems from $p_1 = \frac{1}{n-1}$

[5]This corresponds to the experiments by Frost and Dechter with N and C varying, $K = 3$, and $T = \frac{2}{9}$. See Figure 1. in [5]. A problem with N variables and C constraints is exactly equivalent to a problem generated using our model and parameters $\langle \mathrm{N}, 3, 2\mathrm{C}/\mathrm{N}(\mathrm{N}-1), \frac{2}{9} \rangle$.

to $p_1 = \frac{10}{n-1}$ in steps of $\frac{1}{5(n-1)}$. (An exception is $n = 10$ where the maximum value of p_1 is $\frac{9}{n-1}$.) These parameters are equivalent to varying the average degree of nodes in the constraint graph from 1 to 10 in steps of 0.2. For each n from 10 to 70 inclusive we tested 10,000 randomly generated problems at each value of p_1, while for n from 80 to 110 we tested 1,000 problems for each p_1.



(a) Probability of solubility



(b) Median Checks

Figure 5: $\langle n, 3, p_1, \frac{2}{9} \rangle$ plotted using Order parameter

In Figure 4(a) we show how probability of solubility varies as n changes. Because the parameter that varies for each n is p_1, we plot probability of solubility (y-axis) against p_1 (x-axis). With increasing n, the phase transition occurs at smaller values of p_1, ie. the left most contour is for $n = 110$ and the right most contour is for $n = 10$. In Figure 4(b) we show how the median search cost changes with increasing n, on a logarithmic scale. As n increases the peak in median cost increases greatly and occurs at smaller values of p_1 and appears to coincide approximately with the transition in probability of solubility.

Just as we saw in §3, our data is much more easily understood in terms of the order parameter τ than in terms of the raw parameter p_1. In Figure 5(a) we show the probability of solubility for each problem size tested, plotted against the order parameter (x-axis). Our data covers the range of order parameter τ from 0.229 to 2.29. We can now see that the phase transition in probability always starts at a value of the order parameter τ slightly larger than 1. Comparison with equation (3) shows that this is considerably smaller than the value of 2 predicted by the theory of Smith, Williams and Hogg. Nevertheless, using their theory we have derived a parameter at a fixed value of which the phase transition seems to cluster. It is also clear that the sharpness of the phase transition tends to increase with increasing n. We show in §5 that this increasing sharpness can be characterised *precisely*.

Figure 5(b) plots median search cost against the order parameter for each value of n tested. The peak in cost covers a similar range to the phase transition in solubility. However, as n increases, the peaks in median search cost become more sharply defined and appear at smaller values of the order parameter.

5 Scaling of Probability

In this section we show that the probability of solubility for a given problem class scales in an astonishingly simple way. The same technique that we use has been used in other NP complete problem classes and so seems to be of very general validity. However, since the technique is borrowed from statistical physics we briefly review some analogies between phase transitions in physical and computational problems.

Similar phase transition phenomena occur in many physical systems [30]. For example, in a ferromagnet (a permanent magnet) a phase transition occurs at a critical temperature, the Curie temperature. Above this temperature, and in the absence of an external magnetic field, the ferromagnet has no magnetization. If, however, the ferromagnet is cooled then it becomes abruptly magnetized at the Curie temperature. Several other macroscopic properties like the magnetic susceptibility (the change in magnetization induced by an external field) also undergo a phase transition at the Curie temperature.

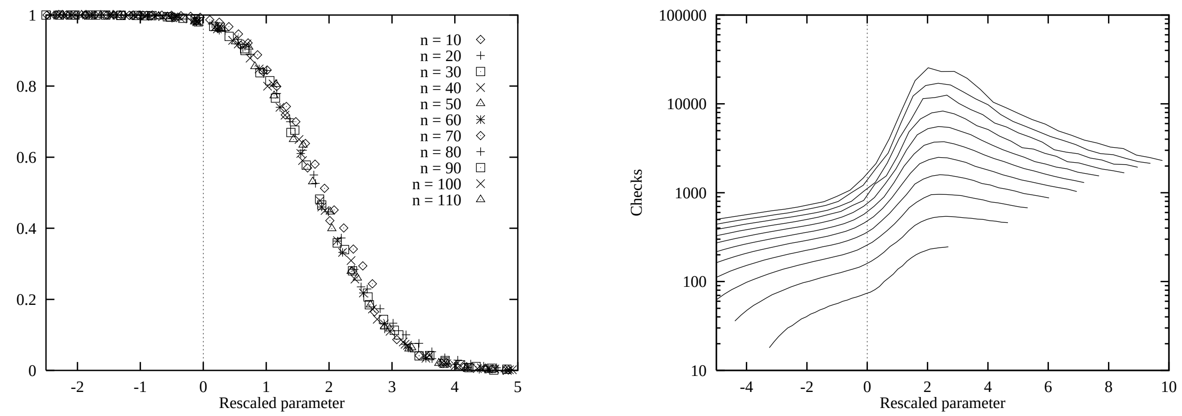
A simple model of a ferromagnet is the Ising model. This has N atoms arranged on a regular lattice. Each atom has a magnetic spin which can be either "up" or "down". The ferromagnet can therefore be in one of 2^{N} possible states. Magnetism is a short-range force promoting neighbouring spins to line up together. Correlations can, however, occur between more distant spins. At a high temperature, thermal fluctuations are large and spins are independent of each other. The ferromagnet therefore has no net magnetization. As the temperature is lowered towards the Curie temperature, spins become correlated over increasingly large distances. At the Curie temperature, spins are totally correlated – changing the spin of a single atom changes all other spins.

Several analogies can be made with binary CSP's. A CSP has n variables. Each variable can take one of m values. The CSP therefore has m^n possible variable-value pairs. Although interactions between variables are restricted to binary constraints, correlations can occur between the values of variables not directly connected via a binary constraint. Our order parameter, which is related to the expected number of solutions to the CSP, serves as a proxy for the temperature. If the order parameter is small then, as there are many models, variables can take values largely independently of each other. As the order parameter is increased, the values of variables become increasingly correlated. The theory of Smith, Williams, and Hogg, predicts only one expected model at the phase transition. The values of variables are therefore totally correlated with each other.

Statistical mechanics describes the behaviour of a ferromagnet in the thermodynamic limit when the volume and number of atoms goes to infinity. For finite systems, a heuristical technique called "finite-size" scaling have been developed to model phase transition phenomena [1]. Finite-size scaling also appears to be useful for modelling the behaviour of the phase transition in a variety of combinatorial problems including propositional satisfiability [15, 16, 8, 9], and the traveling salesman problem [10]. Around the phase transition, finite-size scaling predicts that problems of all sizes are indistinguishable except for a change of scale. This would suggest that,

$$Prob(solution) = f((\tau - \tau_c) \,.\, N^{1/\nu}) \tag{4}$$

where f is some fundamental function, τ is the order parameter, τ_c is the critical value of



(a) Probability of solubility (b) Median Checks

Figure 6: $\langle n, 3, p_1, \frac{2}{9} \rangle$ Plotted with rescaled parameter τ_{N}

the order parameter at the phase transition, and $N^{1/\nu}$ provides the change of scale.

If the prediction of (4) holds then there must be a "fixed point", a single value of the order parameter at which all different problem sizes give the same percentage solubility. This may appear not to be the case in Figure 5, except of course at 0 and 100% solubillity. However, examining our data more closely in the region of high percentage solubility and interpolating between points on the plot where necessary, we did observe a fixed point.[6] We found very similar behaviour at $\tau = 1.25$, where all n gave probabilities in the range (0.974,0.982) except for $n = 10$ which gave 0.991. Taking sample sizes into account, all probabilities were within two standard deviations of an estimate for probability of soluble of 0.976 except for $n = 10$ and $n = 20$. We take 1.25 for the fixed point and thus for the critical value τ_c. It is interesting to note that this is considerably smaller than the value 2 predicted by (3). This is however consistent with observations by Prosser [21, 22] that the prediction of theory seems to be less accurate in the case of sparse constraint graphs, such as these graphs are for $n > 20$ at the critical value.

Having selected τ_c, if (4) holds then there will be a single value ν to provide a fit to (4). Another way of seeing this is to rescale our data so that instead of plotting the raw order parameter τ, we define a rescaled parameter which depends on the order parameter and the problem size, N. We call this τ_{N} and in line with (4), using n for the problem size, define it by

$$\tau_{\mathrm{N}} =_{def} (\tau - \tau_c) \, . \, n^{1/\nu} \tag{5}$$

If the conjecture of (4) holds for the correct value of the exponent ν, we expect to see the probability curves for each n very closely aligned if we plot them against τ_{N}. If so, then the resulting curve gives us an empirical prediction of the function f. Having chosen τ_c, one can estimate ν empirically by assuming that (5) holds. Then for a given probability of solubility, we can observe the values of τ that give that probability for different values of n. Say that for n_1 and n_2 we observe the same probability at τ_1 and τ_2. Then from (5) we

[6]Although the probability plots are clearly curved, locally straight line interpolation seems to be acceptable.

expect that

$$(\tau_1 - \tau_c)\,.\,{n_1}^{1/\nu} = (\tau_2 - \tau_c)\,.\,{n_2}^{1/\nu}$$

Rearrangement gives us

$$\nu = \frac{\log(n_2/n_1)}{\log((\tau_1 - \tau_c)/(\tau_2 - \tau_c))} \tag{6}$$

We first estimated ν using this formula and the 50% solubility points, again using linear interpolation where necessary. The choice of 50% is because it is significantly different from the probability of 0.976 at the fixed point, giving sufficient range for the scaling to take effect. Using (6) for each of the 55 pairs of $10 \leq n_1 < n_2 \leq 110$, gave a median estimate for ν of 2.32 with a lower quartile estimate of 2.16 and an upper quartile of 2.51. Rescaling based on 25% solubility gives a very similar result, with a median estimate of 2.33. We thus choose $\nu = 2.3$. The fact that this choice gives a good fit to (4) is confirmed dramatically in Figure 6(a). (The vertical line represents $\tau_{\mathrm{N}} = 0$.) Under this scaling all probability curves are almost identical, except the curve for $n = 10$ which rests slightly above the rest. This suggests that probability of solubility in this model can be described by finite size scaling with parameters $\tau_c = 1.25$, $\nu = 2.3$, and f as seen in Figure 6(a).

The implications of this result are significant. First, in this particular model it should help design future experiments. For example, should we wish a probability of solubility of 0.5, then we can interpolate the empirically predicted value of the rescaled parameter, which in this case occurs at $\tau_{\mathrm{N}} \approx 1.81$, this being the median value interpolated from the 11 values of n. We can rearrange (5) to find what value of τ gives a given value of the τ_{N} parameter for a given n. This is given by

$$\tau = \tau_c + \frac{\tau_{\mathrm{N}}}{n^{1/\nu}}$$

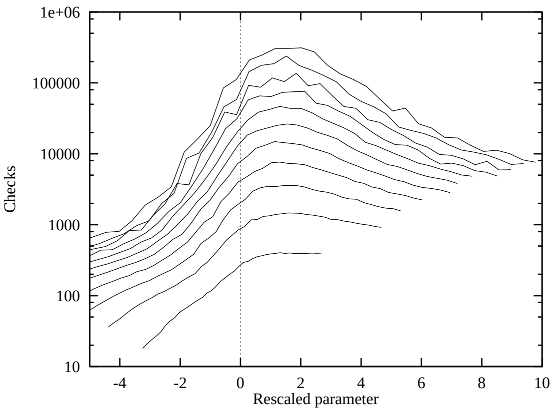
This suggests that in an $\langle n, 3, p_1, \frac{2}{9}\rangle$ problem, 50% solubility occurs when $\tau \approx 1.25 + (1.81/n^{\frac{1}{2.3}})$. We can unpack the definition of the order parameter to give the raw parameter p_1. In this problem class $p_1 = \log_3(\frac{9}{7})\tau/(n-1)$. We thus expect to see 50% solubility at

$$p_1 \approx (5.46 + 7.91/n^{\frac{1}{2.3}})/(n-1)$$

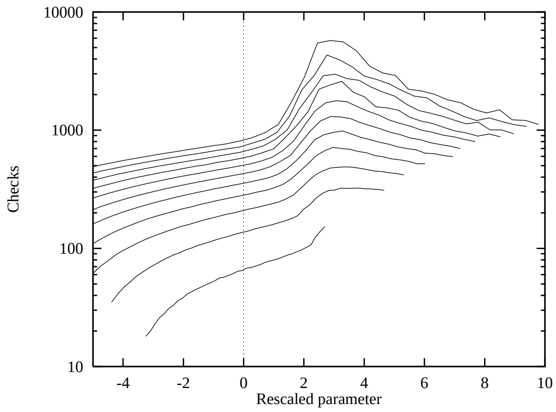
Fortunately we are able to test this prediction with published data, as [5] report the number of constraints observed at 50% solubility for this model. The number of constraints C is $p_1 n(n-1)/2$, so we predict that for 50% solubility,

$$\mathrm{C} \approx \frac{n}{2}(5.46 + 7.91/n^{\frac{1}{2.3}}) \tag{7}$$

[5] report results extend over much wider n than we have reported, up to $n = 275$. For this n, [5] reports 50% solubility at 847 constraints. Equation (7) predicts 846 constraints. This is remarkably accurate, just 1 constraint out in a 275 variable problem, when the largest n used to make the extrapolation was 110 variables. For smaller n, our prediction is not quite as accurate, but it is never more than 9 constraints out, which occurs at $n = 150$ with a prediction of 477 constraints but an observation of 468. Unlike data reported by [5], our



(a) 99% cost behaviour



(b) 10% of cost behaviour

Figure 7: $\langle n, 3, p_1, \frac{2}{9}\rangle$ plotted using rescaled parameter τ_{N}

data can of course also be used to interpolate for any other value of percentage solubility, and to extrapolate this to any problem size.

More significant still is the likelihood that we will see similar kinds of finite scaling in other randomly generated CSP's. This is likely because once similar kinds of scaling were observed in SAT problems [15] they were observed in many different classes of SAT problems [16, 8]. We expect that similar kinds of predictions made from examining only small problems should be available for large problems in many different classes of CSP's.

6 Finite Size Scaling of Search Cost

We showed in §5 that finite size scaling could be used to describe very accurately the change in probability of solubility as n varies. Since the main feature of CSP problems that interests us is how hard it is to solve these problems, it is natural therefore to ask if changes in behaviour of search cost can be similarly corrected using finite size scaling? The remarkable answer is that not only can this be done, but it seems to be achievable using the *identical* rescaled parameter τ_{N}.

In Figure 6(b) we show what happens if we plot the same data previously plotted in Figure 5(b) against τ_{N}. Instead of the peaks in search cost occurring at different values of the parameter as seen previously, the peak appears to be at very similar values for each different n. This strongly suggests that the same finite size scaling that is effective for probability of solubility also models accurately the behaviour of search cost. We believe that this is the first time that this observation has been made.

It seems that not only median, but other measures of search cost scale in exactly the same way. Figure 7(a) shows how the 99 percentile behaves against τ_{N}. That is, the graph plots at each point the cost that was exceeded by only 1% of problems. Just as with median behaviour, these contours line up very closely. Of course the 99 percentile is considerably worse than median behaviour, but we also note that the peaks in these curves occur at smaller values, peaking at $\tau_{\mathrm{N}} \approx 1$ compared with $\tau_{\mathrm{N}} \approx 2$ for median. Figure 7(b) shows behaviour of the 10 percentile, i.e. the cost exceeded by all but 10% of problems. Yet again the contours line up closely. This time the peaks are at a larger value, $\tau_{\mathrm{N}} \approx 3$.

It is particularly significant that we were able to use exactly the rescaled parameter τ_{N} with the *same* parameters τ_c and ν as used in §5. The values $\tau_c = 1.25$ and $\nu = 2.3$ were chosen to model scaling of probability of solubility, and this is an entirely algorithm independent feature of a problem. Yet the same parameters also accurately describe the scaling of search cost in a particular algorithm, FC-CBJ-FF. This would suggest that the finite size scaling of search cost behaviour that we have observed may be algorithm-independent. While the details of contours seen with different algorithms will vary, the scaling parameters may be identical in each case. Of course at this stage this is only speculation since we have only observed scaling with a single algorithm, but the implications for understanding the scaling of search cost are enormous.

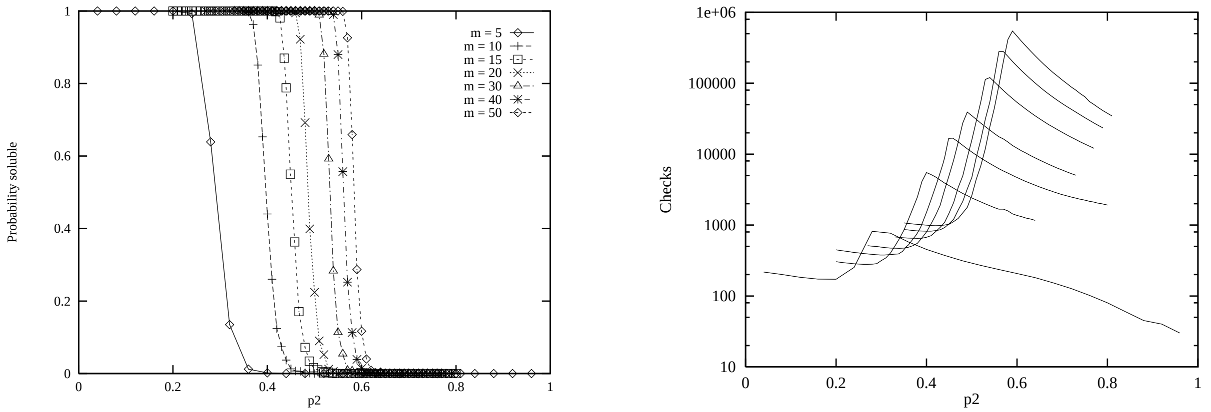
7 Changing Domain Size

To test our conjecture that very similar kinds of scaling would be seen with different random CSP classes, we tested the completely different model of $\langle 10, m, 1.0, p_2 \rangle$. Notice that in §4 we fixed m and p_2 while varying n and p_1: we now fix n and p_1 while varying m and p_2. Since we have fixed $n = 10$ and $p_1 = 1$, all constraint graphs we consider are simply 10-cliques, while before we typically looked at sparse constraint graphs.

From this problem class, we tested problems for m =5, 10, 15, 20, 30, 40, and 50. Except for $m = 5$, where p_2 can only vary in steps of $1/m^2 = 0.04$ we varied p_2 in steps of at most 0.01, covering at least a region of the order parameter from 1 to 4. We tested 1000 problems at each value of p_2. Figure 8 shows how probability of solubility and median search effort varies with m. We give one contour in each figure for each m, plotted against p_2. As m increases, the phase transition occurs at larger values of p_2, and the peak in problem difficulty grows.

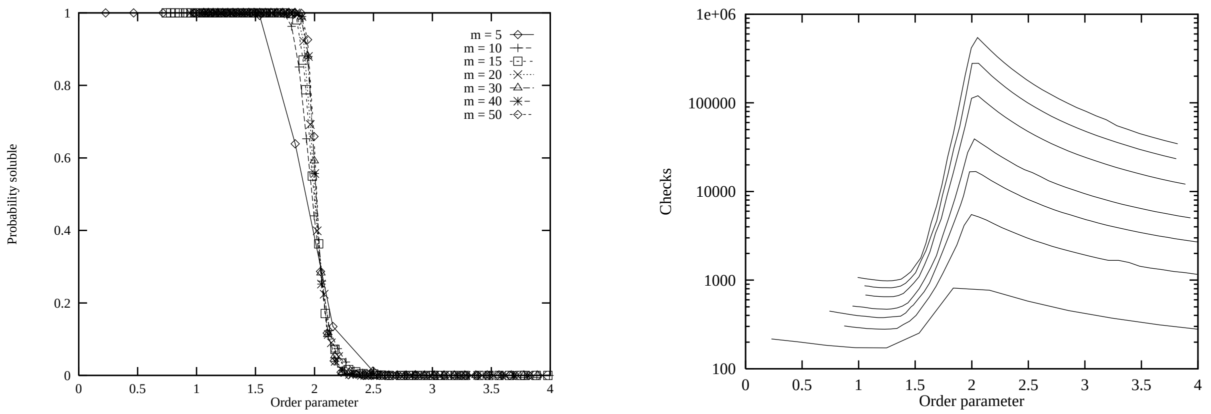
In §3 we proposed an order parameter for binary CSP's. If it is to be useful, it should aid comparison of our data for this problem class both with changing m and with our earlier data for $\langle n, 3, p_1, \frac{2}{9} \rangle$. This is confirmed by Figure 9 which shows our data replotted against the order parameter τ. It can be seen that the probability phase transition and worst case median behaviour always occurs at similar values of τ. We observed a fixed point in probability of solubility at $\tau_c = 2.05$ where solubility was always 0.30 ± 0.01. This is much closer to the expected critical value of $\tau = 2$ than we saw in the previous problem class. Both the location of the fixed point τ_c and the probability of solubility at that point are significantly different from the values $\tau_c \approx 1.25$ and 0.976 solubility that we saw in $\langle n, 3, p_1, \frac{2}{9} \rangle$. However, just as in that class, both the probability transition and the peaks in median behaviour become sharper with increasing m.

Finite size scaling may seem to be an inappropriate techique, because we are no longer changing the number of variables, n. However, we are certainly varying the problem size N by changing the number of values each variable takes, m. It is important to take account of m when considering the problem size. For example, to specify a *solution* to a binary CSP requires $n \log_2(m)$ bits, as the value each variable takes can be specified in $\log_2(m)$ bits. So, properly, this is the measure we should have used in §5 when considering finite



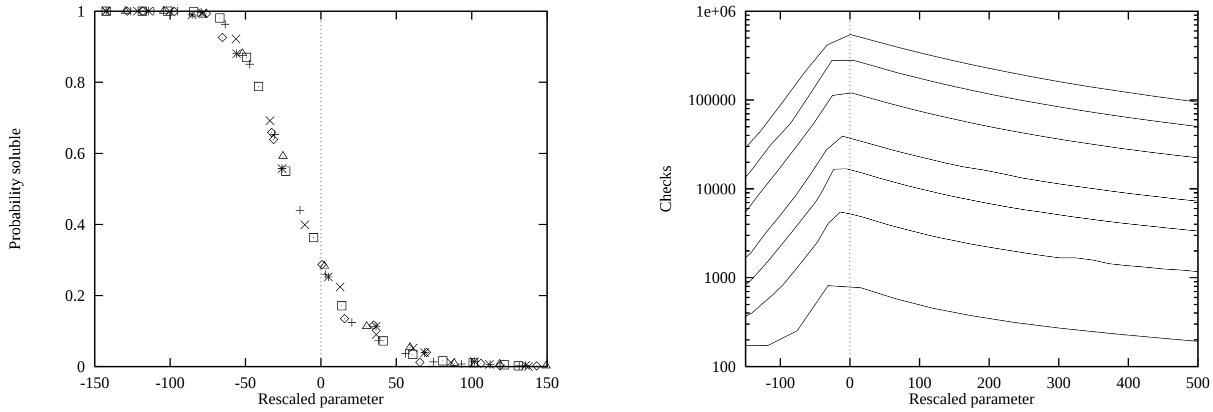
(a) Probability of Solubility (b) Median Checks

Figure 8: $\langle 10, m, 1.0, p_2 \rangle$ plotted using p_2



(a) Probability of solubility (b) Median Checks

Figure 9: $\langle 10, m, 1.0, p_2 \rangle$ plotted using order parameter τ



(a) Probability of Solubility (b) Median Checks

Figure 10: $\langle 10, m, 1.0, p_2 \rangle$ plotted using rescaled parameter τ_{N}

size scaling. However, m was constant at 3 and so does not affect the parameter ν that we derived there. Here, we are varying m, so we redefine the rescaled order parameter as

$$\tau_{\mathrm{N}} =_{def} (\tau - \tau_c) \, . \, (n \log_2(m))^{1/\nu} \tag{8}$$

Having done this we can proceed as before, equation (6) becoming

$$\nu = \frac{\log(n_2 \log_2(m_2)/n_1 \log_2(m_1))}{\log((\tau_1 - \tau_c)/(\tau_2 - \tau_c))} \tag{9}$$

We estimated ν from equation (9) using the 50% probability point. We simply chose this as it is significantly different from the probability of 0.3 at the fixed point. The median estimate for ν was 0.63 with lower and upper quartiles of 0.55 and 0.68 respectively. These estimates of $\tau_c \approx 2.05$, $\nu \approx 0.63$ give an extremely good fit to a prediction of finite size scaling. This is seen in Figure 10 (a) which shows our data for probability of solubility plotted against the rescaled parameter τ_{N}.

Exactly as we saw in §6, we can use the identical parameters τ_c and ν to rescale contours of median cost. This is seen in Figure 10(b). As before the contours line up very closely, suggesting that finite size scaling can be applied to search cost in this problem class.

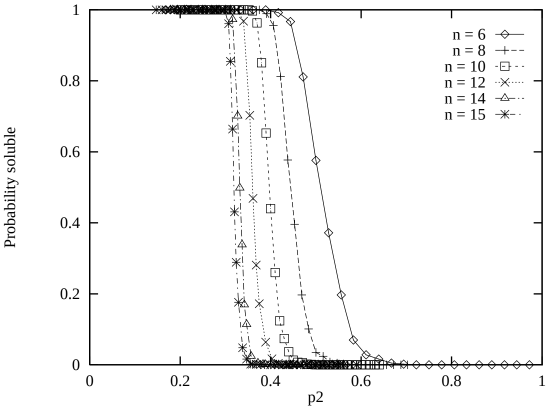
8 Changing Number of Variables and Domain Size

We have established that $n \log_2(m)$ provides a good measure of problem size when varying m. Finally, we ask if it also provides a good measure of problem size when varying n and m together? To test this, we return to our starting point in this paper, namely Gaschnig's random n-queens model. In the terms of §2 these are $\langle n, n, 1.0, p_2 \rangle$. Thus we vary both the number of variables and domain size, in this case keeping them identical. In our experiments we tested n = 6, 8, 10, 12, 14, and 15. We varied p_2 in steps depending on n, and tested 1000 problems at each value of p_2. Figure 11 shows probability of solubility and median search cost against p_2, one contour being given for each n. As n increases the phase transition occurs at smaller values of p_2, and search cost increases.

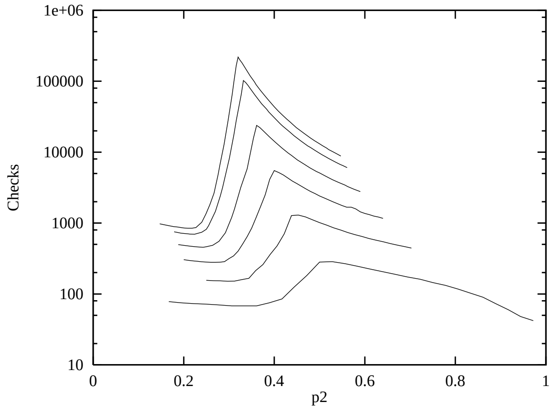
Once again, the use of our order parameter enables us to compare our data as n changes, and to contrast our data for this problem class with data from previous problem classes. Figure 12 shows the same data replotted against the order parameter τ. As in previous cases, we see the probability phase transition occurring over a similar range of τ, as do the peaks in median search cost. The curves become sharper with increasing n.

The fixed point in probability can be seen particularly clearly in Figure 12(a) at $\tau \approx 2$ Examining this data more closely we observed a fixed point in probability of solubility at $\tau_c = 1.98$ where it was always 0.51 ± 0.03. Note that the standard deviation at probability 0.5 in a sample of 1000 is 0.015, so all results were within 2 standard deviations of 0.51. As in §7, this is very close to the predicted critical value of $\tau = 2$, and again we note that we are looking at constraint graphs which are cliques. The critical value τ_c and the fixed point in probability of solubility are different to the previous two cases.

As in §7 we define the rescaled parameter τ_{N} by equation (8), using (9) to estimate ν. In this case we could not estimate ν using the 50% solubility point as it is too close

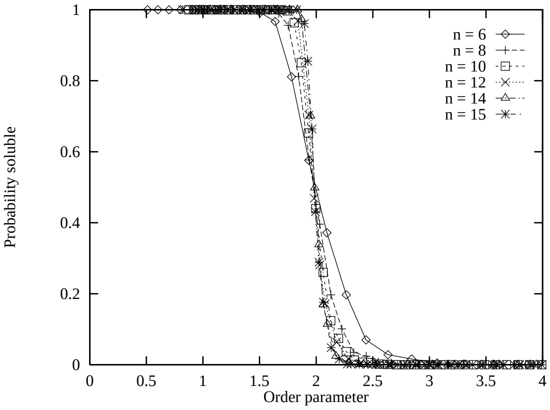


(a) Probability of solubility

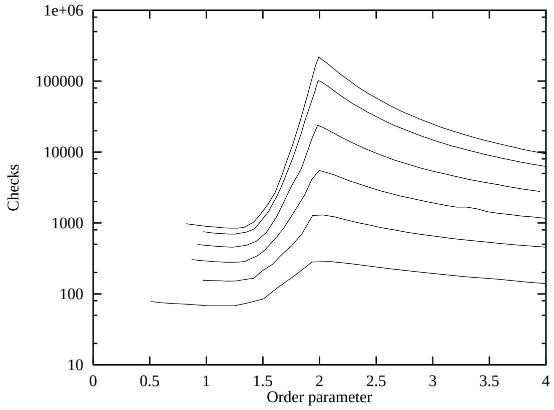


(b) Median Checks

Figure 11: $\langle n, n, 1.0, p_2 \rangle$ plotted using p_2

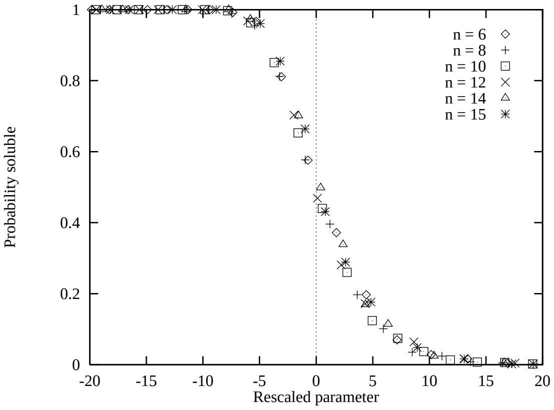


(a) Probability of solubility

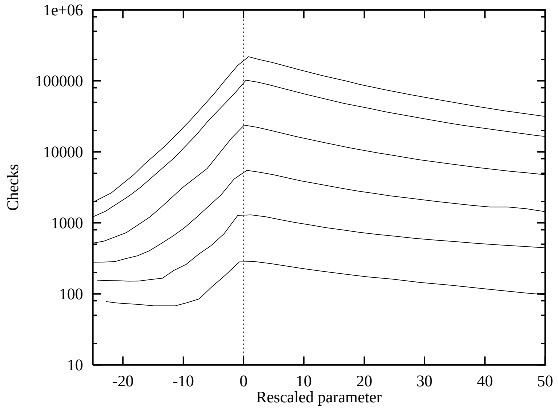


(b) Median Checks

Figure 12: $\langle n, n, 1.0, p_2 \rangle$ plotted using order parameter τ



(a) Probability of Solubility



(b) Median Checks

Figure 13: $\langle n, n, 1.0, p_2 \rangle$ plotted using rescaled parameter τ_{N}

to the fixed point probability of 0.51. Using 90% probability for estimation we obtained a median estimate of 1.02 with upper and lower quartiles of 0.78 and 1.09, while using 10% probability for estimation these values were 1.09, 1.03 and 1.45 respectively. Estimates of $\tau_c \approx 1.98$, $\nu \approx 1.0$ give an extremely good fit to a prediction of finite size scaling. This is seen in Figure 13 which shows our data for probability of solubility and median search cost plotted against the rescaled parameter $\tau_{\rm N}$. Again we point out the remarkable fact that finite size scaling can be applied to search cost using parameters derived solely from examination of probability data.

One of the standard becnchmark for CSP algorithms has been the n-queens, classified by Smith and Dyer as the problem class $\langle n, n, 1.0, (7n-2)/3n^2\rangle$. Using the rescaled order parameter above $\tau_{\rm N}$ for the n-queens problem, we see that as n increases $\tau_{\rm N}$ decreases; for 10-queens $\tau = 1.0$ and $\tau_{\rm N} = -32$, for 100-queens $\tau = 0.51$ and $\tau_{\rm N} = -979$, and for 1000-queens $\tau = 0.34$ and $\tau_{\rm N} = -16,300$. Therefore, as n increases the n-queens problem should become an easier instance of the class of problems $\langle n, n, 1.0\rangle$. This is in full agreement with [26].

It is quite remarkable that exactly the same kind of scaling, that is finite size scaling, should be so accurate for three entirely different methods of varying problem size we have considered in this paper. We have varied n only in $\langle n, 3, p_1, \frac{2}{9}\rangle$, we have varied m only $\langle 10, m, 1.0, p_2\rangle$, and we have varied n and m together in $\langle n, n, 1.0, p_2\rangle$. In each case, the same equation (4) has been shown to be directly applicable, with only the parameters τ_c and ν and the function f varying between problem classes.

9 Conclusions

When presenting the results of experiments on randomly generated CSP's $\langle n, m, p_1, p_2\rangle$ graphs have typically been plotted with either p_1 or p_2 as one of the axis, ie. one of the problem generation parameters has been used. This tends to give a distorted view of the data, as contours rarely line up. We have proposed an order parameter τ for randomly generated CSP's, where τ characterises CSP's regardless of size. This order parameter τ is derived from a theory that predicts that on average the hardest problems will occur when the expected number of solutions $E(N) = 1$.

Analysing the empirical data for experiments with number of variables n varying, we observed a single value of τ where problems of different sizes have the same percentage solubility, ie. a fixed point τ_c. Finite-sized scaling, a heuristical technique from statistical mechanics, was then applied to give a rescaled order parameter $\tau_{\rm N}$. Replotting the data with respect to $\tau_{\rm N}$ brought the picture into sharp focus; the solubility contours lie one on top of the other, and the peaks in median search effort coincide. Furthermore, we were able to use the rescaled order parameter to estimate the critical number of constraints at the crossover point, ie. 50% solubility, for larger values of n, and these were in close agreements with results reported elsewhere; ie. we have given some evidence of the predictive power of $\tau_{\rm N}$. The same rescaling technique was then applied to data from experiments with domain size m varying, and experiments with domain size m and number of variables n varying together. In both cases problem size was taken as $N = nlog_2(m)$, and in both cases the data was again brought into sharp focus. This suggests that the technique may be quite general.

One of the surprises of this investigation is that a finite scaling of the order parameter based on the solubility of problems has carried over to a scaling of search cost. The rescaled parameter $\tau_{\rm N}$ models the solubility of the problem (a problem-dependent property) *and* the behaviour of search cost (something that we might expect to be an algorithm-dependent property). The other surprise has been that finite size scaling has been so accurate for three very different classes of problems (ie. n varying, m varying, n and m varying together).

Obviously this work represents a starting point. In the future, we would like to know the detailed scaling parameters as problems are varied in more ways than we could consider in this paper. It would be very valuable if we could *size* problems with respect to graph density p_1 in order to rescale the data in §3. Finally, we note that the techniques applied in this paper effectively repair a theory which we showed to be inaccurate to a slight degree. However this repair is empirical. If our results could be used to help develop a more refined and accurate theory, it would be a pleasing validation of the empirical science of algorithms, as called for by Hooker [14].

References

[1] Michael N. Barber. Finite-size scaling. In *Phase Transitions and Critical Phenomena, Volume 8*, pages 145–266. Academic Press, 1983.

[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] R. Dechter, Constraint Networks, in *Encyclopedia of Artificial Intelligence* (Wiley, New York, 2nd ed., 1992) 276-286.

[4] R. Dechter and I. Meiri, Experimental evaluation of preprocessing algorithms for constraint satisfaction problems, *Artif. Intell.* **68**(2) (1994) 211-242.

[5] D. Frost and R. Dechter, In search of the best search: an empirical evaluation, *Proceedings AAAI-94,* Seattle, WA (1994) 301-306.

[6] J. Gaschnig, A general backtracking algorithm that eliminates most redundant tests, *Proceedings IJCAI-77* Cambridge, MA (1977) 457.

[7] J. Gaschnig, Performance measurement and analysis of certain search algorithms, Tech. Rept. CMU-CS-79-124, Carnegie-Mellon University, Pittsburgh, PA (1979).

[8] Ian P. Gent and Toby Walsh. The SAT phase transition. In *Proceedings of ECAI-94*, pages 105–109, 1994.

[9] Ian P. Gent and Toby Walsh. The satisfiability constraint gap. Research Paper 702, Dept. of Artificial Intelligence, Edinburgh, June 1994. To appear in *Artificial Intelligence* special issue on phase transitions in problem spaces.

[10] I.P. Gent and T. Walsh. The TSP phase transition. Research report, Department of Computer Science, University of Strathclyde, 1995. Submitted to First Workshop on AI and OR.

[11] S.W. Golomb and L.D. Baumert, Backtrack programming. *JACM* **12** (1965) 516-524.

[12] R.M. Haralick and G.L. Elliott, Increasing Tree Search Efficiency for Constraint Satisfaction Problems, *Artif. Intell.* **14** (1980) 263-313.

[13] T. Hogg and C. Williams. The hardest constraint problems: A double phase transition. *Artificial Intelligence*, 69:359–377, 1994.

[14] J.N. Hooker, Needed: An empirical science of algorithms, *Operations Research* **42** (2) (1994) 201-212.

[15] S. Kirkpatrick, G. Györgyi, N. Tishby, and L. Troyansky. The statistical mechanics of k-satisfaction. In *Advances in Neural Information Processing Systems 6*, pages 439–446. Morgan Kaufmann, 1994.

[16] S. Kirkpatrick and B. Selman. Critical behavior in the satisfiability of random boolean expressions. *Science*, 264:1297–1301, May 27 1994.

[17] V. Kumar, Algorithms for constraint satisfaction problems: a survey, *AI magazine* 13(1) (1992) 32-44.

[18] E. MacIntyre, Really hard problems, *Final Year Report for the B.Sc. degree in Computer Science*, Department of Computer Science, University of Strathclyde, Scotland, 1994.

[19] David Mitchell, Bart Selman, and Hector Levesque. Hard and easy distributions of SAT problems. In *Proceedings, 10th National Conference on Artificial Intelligence.* AAAI Press/The MIT Press, pages 459-465, 1992.

[20] Patrick Prosser. Hybrid algorithms for the constraint satisfaction problem. *Computational Intelligence*, 9:268–299, 1993.

[21] Patrick Prosser. Binary constraint satisfaction problems: Some are harder than others. In *Proceedings of ECAI-94*, pages 95–99, 1994.

[22] Patrick Prosser. An empirical study of phase transitions in binary constraint satisfaction problems. Research Report AISL-49-94, Department of Computer Science, University of Strathclyde, 1994 Revised version to appear in *Artificial Intelligence*

[23] P.W. Purdom, Search rearrangement backtracking and polynomial average time, *Artif. Intell.* **21** (1983) 117-133.

[24] D. Sabin and E.C. Freuder, Contradicting conventional wisdom in constraint satisfaction, *Proceedings ECAI-94,* Amsterdam, The Netherlands (1994) 125-129.

[25] Barbara M Smith. Phase transition and the mushy region in constraint satisfaction problems. In *Proceedings of ECAI-94*, pages 100–104, 1994.

[26] Barbara M. Smith and Martin E. Dyer. Locating the phase transition in binary constraint satisfaction problems. *Artificial Intelligence*, To appear.

[27] E.P.K. Tsang, Foundations of Constraint Satisfaction, *Academic Press*, 1993.

[28] E.P.K Tsang, J. Borrett, and A.C.M Kwan, An attempt to map the performance of a range of algorithm and heuristic combinations, to appear in *Proceedings AISB-95* (1995)

[29] C.P. Williams and T. Hogg. Exploiting the deep structure of constraint problems. *Artificial Intelligence*, 70:73–117, 1994.

[30] Kenneth G. Wilson. Problems in physics with many scales of length. *Scientific American*, 241:140–157, 1979.