# The 50% Point in Constraint-Satisfaction Problems

Ian P. Gent and Patrick Prosser
Department of Computer Science
University of Strathclyde
Glasgow G1 1XH
United Kingdom
ipg@cs.strath.ac.uk, pat@cs.strath.ac.uk

Research Report/95/180
Department of Computer Science, University of Strathclyde
18 April 1995

**Abstract**

A beautifully simple linear relationship has been observed for the 50 percent solubility point in random 3-SAT problems. We show that a similar relationship holds for binary constraint satisfaction problems. We are able to do this by a detailed re-examination of data published by other authors.

Phase transitions have recently been observed in many different classes of randomly generated NP problems [1]. As an order parameter is varied, an easy-hard-easy pattern is often observed, with the hard problems orders of magnitude harder than the easy ones. The hard region is correlated with a transition from mainly soluble problems to mainly insoluble problems. This correlation makes identification of the point where 50% of problems are soluble of particular interest.

One of the problem classes in which phase transitions have been most studied is random 3-SAT. For example, [2] observed a beautifully simple linear relationship between the number of clauses needed for the crossover point, and the number of variables. They found that the number of clauses for 50% solubility is predicted very accurately by the formula $4.24N + 6.21$, where $N$ is the number of variables.

More recently, there have been several studies of phase transitions in binary constraint satisfaction problems. For example, [3] report phase transitions as the number of constraints is varied for a given number of variables. They generate problems with $N$ variables, each variable taking one of $K$ possible values. $C$ out of the $N(N-1)/2$ possible constraints between pairs of variables are chosen at random, and for each one a fraction $T$ of the $K^2$ possible pairs of values are ruled out. Naturally, Frost and Dechter queried whether a simple linear relationship existed between $N$ and $C$ at the crossover point. While noting that "the ratio of $C$ to $N$ appears to be asymptotically linear", they also report that "we did not observe a precise linear relationship" between $N$ and $C$. We re-examined these authors' data and noticed that despite these comments there was indeed a simple linear relationship of remarkable precision.

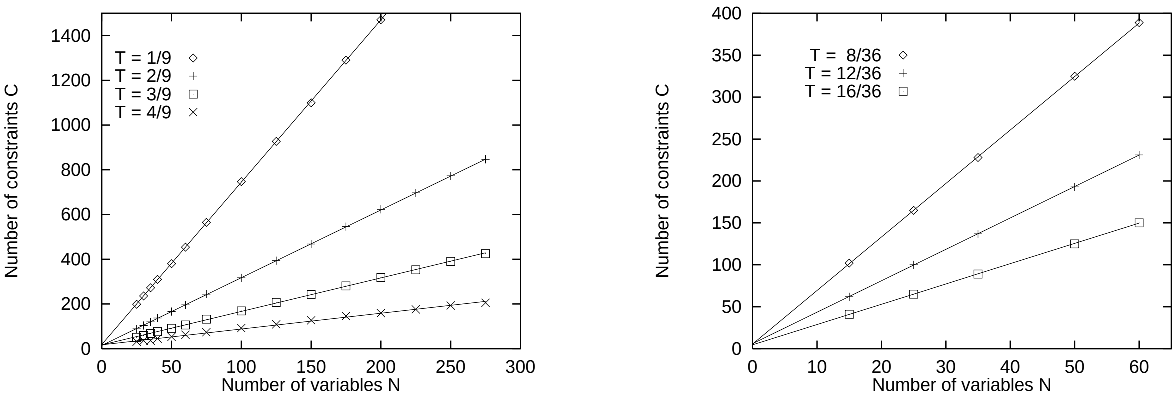
We refer to Figure 1 in [3]. This reports the number of constraints $C$ observed at the 50% solubility point for each $N$ tested in a variety of different problem classes. Most data is presented for $K = 3$ and $T = \frac{1}{9}$, $\frac{2}{9}$, $\frac{3}{9}$, and $\frac{4}{9}$, so we examine these first. If a relationship of the simple form found by [2] exists, we expect that for each $T$, the values of $C$ will fit a formula of the form

$$C = \alpha + \beta N \tag{1}$$

For each $T$, the values $\alpha$ and $\beta$ can be calculated by least squares estimation. Doing this, we obtained the values $\alpha = 18.6$, $\beta = 7.26$ for $T = \frac{1}{9}$. For $T = \frac{2}{9}$ we got $\alpha = 14.3$, $\beta = 3.03$; for $T = \frac{3}{9}$ we got $\alpha = 16.5$, $\beta = 1.50$; and for $T = \frac{4}{9}$ we got $\alpha = 17.4$, $\beta = 0.707$. Figure 1(a) plots as points the values of $C$ at 50% solubility reported by [3] for each $N$ and $T$. These are *not* connected by lines. The straight lines instead represent the predictions of (1) using the appropriate values for $\alpha$ and $\beta$ reported above. The fit is remarkable. Nowhere is an error of more than 8 constraints observed. For $T = \frac{1}{9}$ the mean absolute error is just 1.6; for $T = \frac{2}{9}$ it is 1.1; for $T = \frac{3}{9}$ it is 1.8; and for $T = \frac{4}{9}$ it is 2.9.

We repeated the exercise for data given by [3] for $K = 6$, although only five data points are given. Least squares estimates of $\alpha$ and $\beta$ are: 5.56 and 6.38 for $T = \frac{8}{36}$; 5.95 and 3.75 for $T = \frac{12}{36}$; and 4.56 and 2.42 for $T = \frac{16}{36}$. This is shown in Figure 1(b). The fit is even better, as is expected given less data to fit. In no case does this predict 50% solubility more than *one* constraint out.

To conclude, we have confirmed [3]'s observation of linear behaviour of the 50% solubility point in binary constraint satisfaction problems. However, we have shown that this holds not only asymptotically but remarkably precisely for all problem sizes observed to date.




(a) $K = 3$ (b) $K = 6$

Figure 1: Linear relationship between $N$ and $C$ for 50% Solubility

This extends results of [2] from SAT problems to an entirely new domain. It is natural to ask if similar results may hold in other NP-complete problem classes. Finally, we wish to applaud [3]'s inclusion of raw data without which this reinterpretation of their results would have been impossible. We wonder how much more data, both published and unpublished, will repay reinterpretation.

# References

[1] 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.

[2] J.M. Crawford and L.D. Auton. Experimental results on the crossover point in satisfiability problems. In *Proceedings of the Eleventh National Conference on Artificial Intelligence*, pages 21–27. AAAI Press/The MIT Press, 1993.

[3] D. Frost and R. Dechter. In search of the best search: an empirical evaluation. In *Proceedings AAAI-94*, pages 301–306, 1994.