Asymptotic and Finite Size Parameters for Phase Transitions: Hamiltonian Circuit as a Case Study*

Jeremy Frank† and **Ian P. Gent**‡ and **Toby Walsh**‡

August 15, 1997

Keywords: analysis of algorithms, computational complexity, phase transitions, constrainedness, finite size scaling

Ensembles of random NP-hard problems often exhibit a phase transition in solvability with a corresponding peak in search cost [3]. Problem instances from such phase transitions are now used routinely to benchmark algorithms. To study such phase transitions, parameters have been derived either from asymptotic scaling results or from the constrainedness [6]. Using the Hamiltonian Circuit (HC) problem as a case study, we show that a simple re-scaling of the constrainedness gives a parameter whose relation with the asymptotic parameter is almost independent of problem size for small problems. As a result both approaches are equally able to model the phase transition. This justifies the use of the constrainedness parameter in NP problems where asymptotic results are currently unknown. In addition constrainedness can be used to compare different problem classes and as a meta-heuristic.

The HC problem is: given an undirected graph, decide whether there is an ordered sequence of nodes such that, for all i, (n_i, n_{i+1}) and (n_n, n_1) are edges in the graph. Here, we consider random graphs with a fixed number of edges, those edges distributed randomly through the graph. A key asymptotic result is known. For graphs generated with n nodes and e edges; if we fix $e/(n \log n)$, then as $n \to \infty$, there is almost certainly a circuit if $e/(n \log n) > \frac{1}{2}$, and almost certainly not if $e/(n \log n) < \frac{1}{2}$, as shown by Koršunov [12, 13]. This asymptotic result suggests the parameter $e/n \log n$ for studying phase transitions. This has been used as a

*This research was partly supported by a Strathclyde University grant and EPSRC grant GR/L/24014 supporting the third author. Authors are listed in alphabetical order.

†Caelum Research Corporation, NASA Ames Research Center, Mail Stop 269-1, Moffet Field, CA 94035-1000, USA. frank@ptolemy.arc.nasa.gov

‡Department of Computer Science, University of Strathclyde, Glasgow G1 1XH, United Kingdom. [ipg,tw]@cs.strath.ac.uk

parameter by Cheeseman, Kanefsky and Taylor [3], and by Frank and Martel [4]. Another domain in which an asymptotic result has been used to suggest a parameter is the Euclidean TSP, in which Gent and Walsh [9] showed the existence of a phase transition using a parameter based on a result of Beardwood *et al* [2].

In many NP-complete domains, more ad hoc methods have been used to construct appropriate parameters. These have recently been unified by the introduction of κ, a general constrainedness parameter [6]. In a random ensemble of problems each defined over a fixed state space S, this is defined by

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

where $\langle Sol \rangle$ is the expected number of solutions per problem in the ensemble, and $\mathcal{N}$ is the size of the problem, taken as $\log_2 |S|$. Subject to multiplication by appropriate constants, κ is identical to parameters previously used in satisfiability, graph colouring, constraint satisfaction, and number partitioning [6]. However, in neither the HC problem nor the Euclidean TSP is the relationship between κ and the asymptotic parameter known. This therefore leaves open a question, whether the use of κ is justified if it does not relate well to known asymptotic parameters. Since it is not known how to calculate κ in the Euclidean TSP, we address this question in the HC problem.

To calculate κ, we need ρ, the probability of a randomly selected circuit of nodes having every edge in the graph. There are e edges in the graph out of a possible $n(n-1)/2$. The probability that the first edge of any circuit is in the graph is $2e/n(n-1)$. That leaves $e-1$ edges and $(n(n-1)/2)-1$ places to put them. So the next edge is in the graph with chance $(e-1)/[(n(n-1)/2)-1]$. Then the next with chance $(e-2)/[(n(n-1)/2)-2]$. Since there are n edges under investigation we get an overall probability of

$$\rho = \prod_{i=0}^{n-1} \frac{(e-i)}{(n(n-1)/2)-i}$$

Since we may designate a starting point arbitrarily, and because we may take circuits in either direction, the number of distinct potential circuits is $(n-1)!/2$. So $\langle Sol \rangle$ is given by $\rho(n-1)!/2$ and $\mathcal{N}$ by $\log_2((n-1)!/2)$ yielding

$$\kappa = -\frac{\sum_{i=0}^{n-1} \log_2 \frac{(e-i)}{(n(n-1)/2)-i}}{\log_2((n-1)!/2)} \tag{1}$$

We analysed the HC phase transition using κ to determine whether or not it is as good a parameter for this problem as it is for other problems it has been used for in the past [6, 8]. We implemented Martello's algorithm [14]: as this algorithm is for directed graphs, we added an edge in each direction for each edge in the undirected graph. We ran tests on random graphs from 5 to 30 nodes in steps of 1 (although for clarity we restrict graphed data to multiples of 5). At each number of nodes n we varied the number of edges e in the graph in steps of 1, excepting $n = 30$ where the step size was 6 edges. For each value of n and e we generated 1000 problems pseudo-randomly. Figure 1 shows the observed probability of a circuit existing plotted against κ. This graph shows behaviour very similar to that in many other classes when plotted against constrainedness or equivalent parameters, including satisfiability [11], constraint satisfaction [5], number partitioning [8], and the asymmetric TSP [6]. As in all these classes, the transition gets sharper with respect to κ as we increase the number of nodes. Figure 2 shows that, as in other problem classes, there is a critical value κ_c where the curves approximately intersect. The critical value κ_c is particularly interesting as it seems to be a point of scale-invariant behaviour. Figure 2 suggests that $\kappa_c = 0.7$, since at this value the probability is in the range 0.840 ± 0.015 except for $n = 5$. Using data at all values of n from 5 to 30, $\kappa_c = 0.7$ seems to be an accurate fixed point from $n = 9$ upwards. The very smallest sizes deviate slightly, but this deviation has been seen in many classes such as those just mentioned.

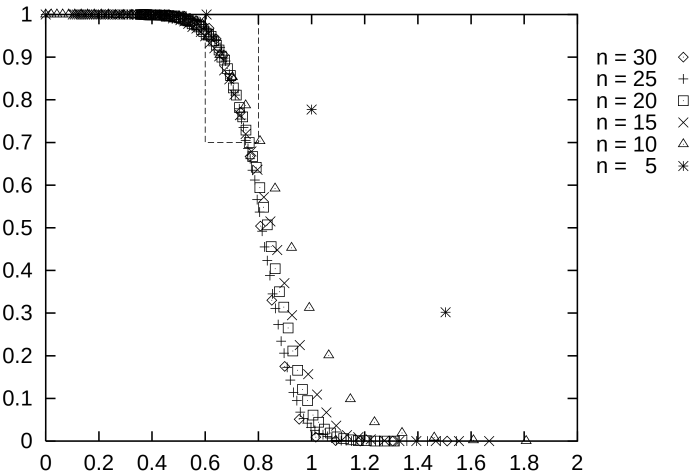

Figure 1: Probability of a circuit existing (y-axis) plotted against κ (x-axis). The boxed area is shown in more detail in Figure 2.

It is now standard to characterise the increasing sharpness of the curves in Figure 1 using the technique of finite size scaling [11, 5, 6, 8]. Given a critical value κ_c and a scaling constant ν, define $\gamma =_{def} \mathcal{N}^{1/\nu}(\kappa - \kappa_c)/\kappa_c$. We choose $\kappa_c = 0.7$ as described above and empirically choose ν to minimize the discrepancies between observed behaviour at each problem size. One technique for choosing ν is described by Gent *et al* [5], based on behaviour when probability of a circuit existing is 0.5. Each pair of values of n results in an estimate of ν. We obtained a median estimate of $\nu = 2.4 \pm 0.5$, with the errors indicating upper and lower quartiles in the estimates

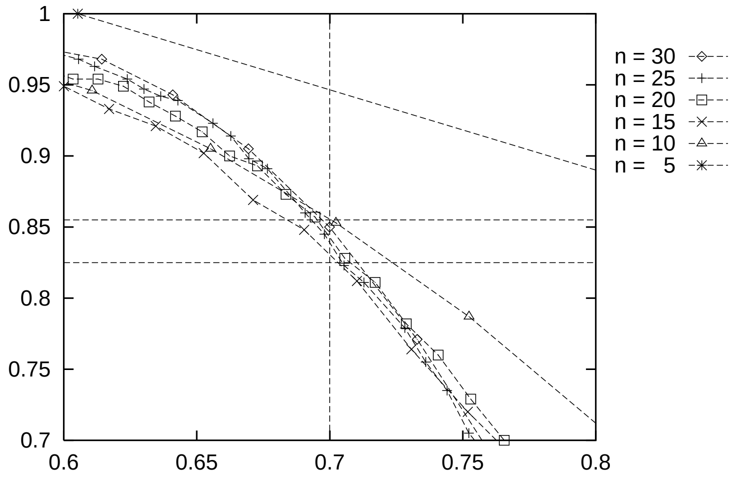

Figure 2: Zoom in on the area outlined in Figure 1 with probability (y-axis) plotted against κ (x-axis). Observed data points for each value of n have been joined by lines to guide interpolation. Additionally, a vertical line shows $\kappa = 0.7$ and horizontal lines show the range of probabilities 0.840 ± 0.015

of ν when all data from $n = 5$ to $n = 30$ was used. Using $\gamma = \mathcal{N}^{1/2.4}(\kappa - 0.7)/0.7$ and $\mathcal{N} = \log_2((n-1)!/2)$ yields the plot seen in Figure 3. That all curves are very similar, again apart from the smallest n, shows that finite size scaling can empirically correct the constrainedness parameter very successfully.

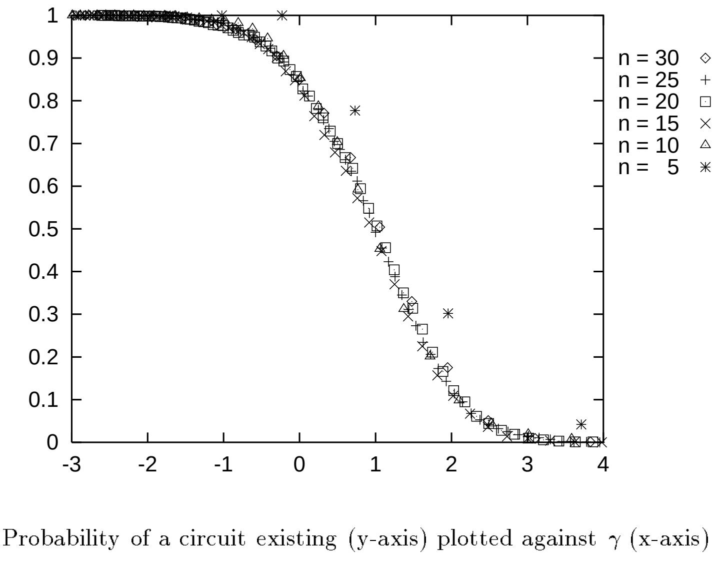

Figure 3: Probability of a circuit existing (y-axis) plotted against γ (x-axis)

The hardest problems to solve appear at the phase transition in the probability of a circuit existing [3]. The rescaled constrainedness parameter shows this correlation particularly clearly. Figure 4 shows the maximum number of leaf nodes used in each data set. We report worst case behaviour rather than mean or median because most problems were solved after searching just one leaf node, so the mean is dominated by the worst case behaviour. This behaviour is similar to the occurrence of 'exceptionally hard problems' in domains such as graph colouring [10], satisfiability [7], and constraint satisfaction [16]. The hardest problems are always in the region of γ where some graphs have circuits and some do not, and generally in the middle of this region. Random problems generated with more uniform graphs might produce less variable behaviour and greater median problem difficulty, an effect which has been noted in satisfiability and constraint satisfaction [7, 16].

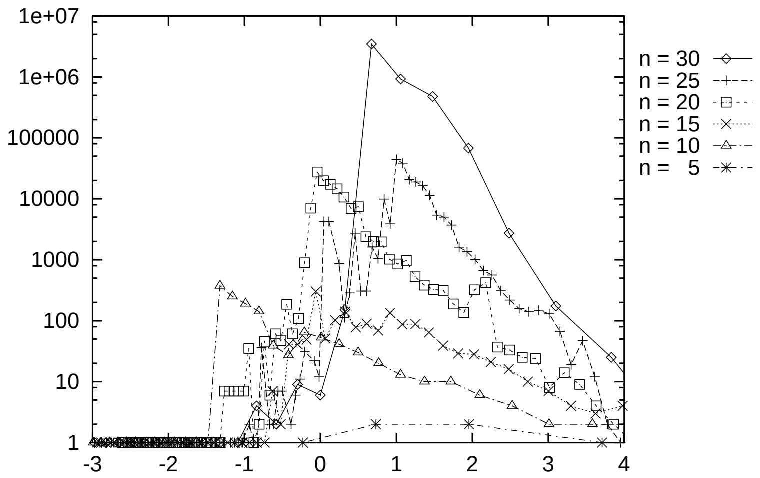

Figure 4: Worst case number of leaf nodes searched (y-axis) plotted against γ (x-axis).

In the HC problem, the parameter suggested by asymptotic analysis also seems to capture behaviour successfully at small sizes. Figure 5 shows the result of plotting probability of a circuit existing against $e/(n \log n)$. Low values of this parameter represent insoluble problems with few edges. The fit as problem size changes is of broadly similar quality to that seen in Figure 3. This is surprising: in other domains where asymptotic parameters have been used, such as the Euclidean TSP [9], finite size scaling has still been necessary to yield a parameter which gives such scale-invariant behaviour. Why this parameter is so good for the HC problem remains an interesting open question. Note however that the plotted curve cannot be representative of limiting behaviour as it is inconsistent with Koršunov's result of a limiting value of $e/(n \log n) = 0.5$.

To summarise, we have shown that two techniques for devising a parameter for the study of phase transitions both give good results in the case of Hamiltonian Circuits at small sizes. Which should we choose? We suggest there is little or nothing to choose between the two approaches at these problem sizes. To illustrate this, Figure 6 shows how $e/(n \log n)$ varies with γ for n up to 30 in the region of the phase transition. At each n we find that $e/(n \log n)$ is almost exactly the same function of γ in this region. Even for $n = 5$ the difference of about 0.1 from the

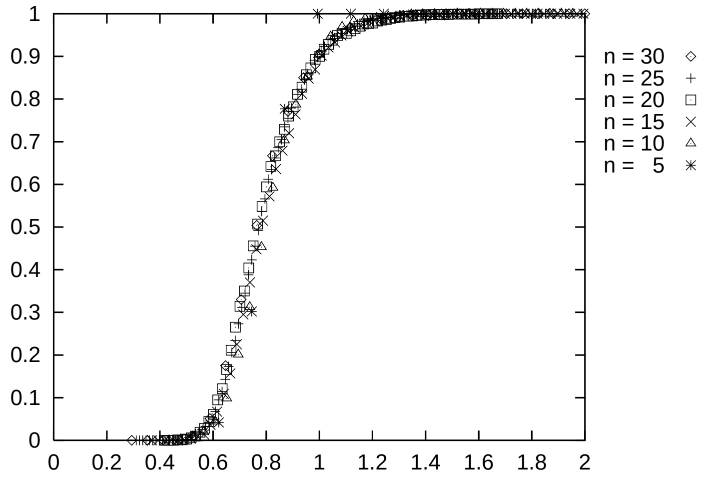

Figure 5: Probability of a circuit existing (y-axis) plotted against $e/(n \log n)$ (x-axis)

other plots in $e/(n \log n)$ at its worst corresponds to only about 0.8 of an edge. The relationship between the two parameters is almost independent of problem size for small n. Neither parameter can be seen as superior to the other for analysis of experimental data at these problem sizes.

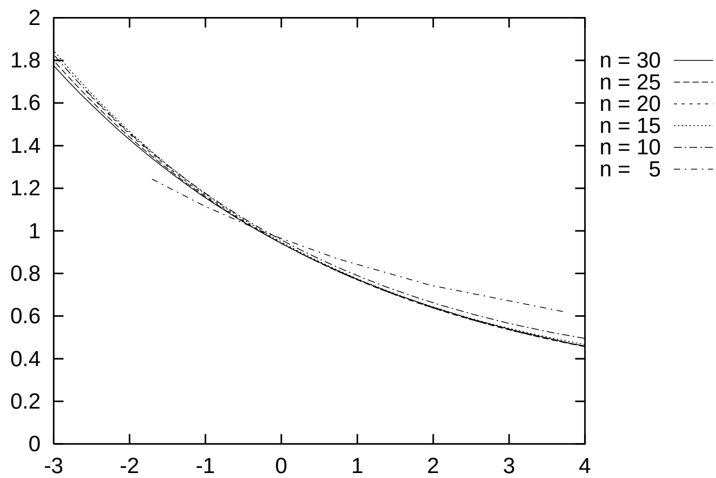

Figure 6: Relationship between γ (y-axis) and $e/(n \log n)$ (x-axis) for varying n.

The constrainedness parameter κ is useful for comparisons between problem classes and as the basis for a meta-heuristic. The general parameter is identical to those used in many other domains, so features such as the location of the phase transition can be compared directly. For example, we note that the critical value $\kappa = 0.7$ is smaller than that reported for 3-SAT, where the phase transition occurs at $\kappa = 0.82$ [6]. A consequence of this is that critically constrained HC problems must have a higher average number of solutions than critically constrained 3-SAT problems for equal problem sizes $\mathcal{N}$. This suggests that simple stochastic search procedures may work very well in random HC problems since they can be expected to perform better with increasing number of solutions, and procedures such as WSAT are known to perform very well in 3-SAT problems at the phase transition [15]. The theoretical results of Angluin and Valiant [1] support this suggestion, as do the experimental results observed by Frank and Martel [4]. Perhaps more significantly, the meta-heuristic of minimizing κ in backtracking search seems to capture the widely accepted informal heuristic of picking the most constrained choice out of those available [6]. Recalling the definition of κ in Equation (1), minimizing κ is equivalent to selecting an edge that maximises the remaining number of edges in the graph, given that each choice effectively reduces the number of nodes n in the remaining graph by 1. As has been suggested in constraint satisfaction problems [6], we can view the heuristic used in Martello's algorithm [14] as a proxy for this heuristic. Martello's heuristic is to pick the edge which minimizes the minimum of out-degree and in-degree of the selected node: this naturally tends to leave as many edges as possible in the graph. It would be interesting to explore further how successful the minimize-κ heuristic is for HC and how it relates to other heuristics for HC.

Our work has two consequences. First, as just discussed, there are significant benefits of the use of the general parameter κ for Hamiltonian Circuits. Second, our results show the robustness of the constrainedness parameter even in a domain where an asymptotic result is known. There are many NP-complete domains where no asymptotic result is currently known and so only the constrainedness parameter is available. We suggest that in such domains the constrainedness parameter can be used with some confidence.

References

[1] D. Angluin and L. G. Valiant. Fast probabilistic algorithms for Hamiltonian circuits and matchings. *Journal of Computer and System Sciences*, 18(2):155–193, 1979.

[2] J. Beardwood, J.H. Halton, and J.M. Hammersley. The shortest path through many points. *Proceedings of the Cambridge Philosophical Society*, 55:299–327, 1959.

[3] P. Cheeseman, B. Kanefsky, and W.M. Taylor. Where the really hard problems are. In *Proceedings of the 12th IJCAI*, pages 331–337, 1991.

[4] J. Frank and C. Martel. Phase transitions in random graphs. In *Proceedings, Workshop on Studying and Solving Really Hard Problems, CP-95*, 1995.

[5] I.P. Gent, E. MacIntyre, P. Prosser, and T. Walsh. Scaling effects in the CSP phase transition. In *Principles and Practice of Constraint Programming*, pages 70–87. Springer, 1995.

[6] I.P. Gent, E. MacIntyre, P. Prosser, and T. Walsh. The constrainedness of search. In *Proceedings of AAAI-96*, pages 246–252, 1996.

[7] I.P. Gent and T. Walsh. Easy problems are sometimes hard. *Artificial Intelligence*, pages 335–345, 1994.

[8] I.P. Gent and T. Walsh. Phase transitions and annealed theories: Number partitioning as a case study. In *Proceedings of ECAI-96*, pages 170–174, 1996.

[9] I.P. Gent and T. Walsh. The TSP phase transition. *Artificial Intelligence*, 88:349–358, 1996.

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

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

[12] A.D. Korshunov. The main properties of random graphs with a large number of vertices and edges. *Russian Math. Surveys*, pages 121–198, 1985.

[13] A.D. Koršunov. Solution of a problem of Erdös and Rényi on Hamiltonian cycles in nonoriented graphs. *Soviet Math. Dokl.*, 17(3):760–764, 1976.

[14] S. Martello. An enumerative algorithm for finding Hamiltonian circuits in a directed graph. *ACM Transactions on Mathematical Software*, 9:131–138, 1983.

[15] B. Selman, H. Kautz, and B. Cohen. Noise Strategies for Improving Local Search. In *Proceedings of the 12th National Conference on AI*, pages 337–343. American Association for Artificial Intelligence, 1994.

[16] B.M. Smith and S.A. Grant. Sparse constraint graphs and exceptionally hard problems. In *Proceedings of IJCAI-95*, pages 646–651, 1995.