Critical Behavior at the k-Satisfiability Threshold

Scott Kirkpatrick*
IBM TJ Watson Research Center
Yorktown Heights, NY 10598
kirk@watson.ibm.com

Bart Selman
AT&T Bell Laboratories
Murray Hill, NJ 07974
selman@research.att.com

Abstract

The satisfiability of random Boolean formulae with precisely k variables per clause is a popular testbed for the performance of search algorithms in artificial intelligence and computer science. For $k = 2$, formulae are almost aways satisfiable when the ratio of clauses to variables is less than 1; for ratios larger than 1, the formulae are almost never satisfiable. We present data showing a similar threshold behavior for higher values of k. We also show how finite-size scaling, a method from statistical physics, can be used to characterize size dependent effects near the threshold. Finally, we commment on the relationship between thresholds and computational complexity.

*This work was done while visiting the Racah Institute of Physics and Center for Neural Computation, Hebrew University, Jerusalem, 91904 Israel during calendar 1993. Email: kirk@fiz.huji.ac.il. S.K. also acknowledges the hospitality of the Salk Institute, where a portion of the work was carried out, with support from the McDonnell-Pew Foundation.

1 Threshold Phenomena and Random k-SAT

Properties of randomly generated combinatorial structures often exhibit sharp threshold phenomena. The first and best-known example of such threshold behavior was found in randomly generated graphs. Consider graphs created with N vertices and M edges selected at random from the $N(N-1)/2$ possibilities. Erdös and Rényi (1) studied the connected clusters formed when $N \to \infty$ with the ratio $\alpha = M/N$ held constant. They identified three regimes: when $\alpha < 1/2$, small isolated clusters of maximum size lnN are found; when $\alpha > 1/2$, a single "giant component" with size of order N absorbs many of the clusters; precisely at $\alpha_c = 1/2$, clusters of size $N^{2/3}$ are found. Subsequent work (2, 3) has made precise the "sharpness" of the threshold: its width is of order $N^{-1/3}$ about $\alpha_c = 1/2$. This is now recognized as the prototype of all "percolation" phenomena observed in disordered materials (4).

Recently, threshold phenomena have also been observed in randomly generated Boolean formulae. Mitchell *et al.* (5) consider the k-satisfiability problem (k-SAT). An instance of k-SAT is a Boolean formula in conjunctive normal form (CNF), *i.e.*, a conjunction (logical AND) of disjunctions or clauses (logical ORs), where each disjunction contains exactly k literals. A literal is a Boolean variable or its negation. The task is to determine whether there is an assignment to the variables such that all clauses evaluate to true. Here, we will use N to denote the number of variables and M for the number of clauses in a formula.

For randomly generated 2-SAT instances, it has been have shown analytically that for large N, when the ratio $\alpha = M/N$ is less than 1 the instances are almost all satisfiable, whereas for ratios larger than 1, almost all instances are unsatisfiable (6, 7). For $k \geq 3$, a rigorous analysis has proven to be elusive. Loose upper and lower bounds have been obtained, but there is yet no rigorous proof of the existence of a threshold (8). Experimental evidence, however, strongly suggests a threshold with $\alpha \approx 4.3$ for 3SAT (5,9,10).

One of the main reasons for studying randomly generated 3CNF formulae is for their use in the empirical evaluation of combinatorial search algorithms (10, 11, 12). 3CNF formulae are good candidates for the evaluation of such algorithms because determining their satisfiability is an NP-complete problem (13). It is therefore widely believed that there does not exist an efficient (polynomial time) algorithm for determining the satisfiability of 3CNF formulae. This also holds for larger values of k. For $k = 1$ or 2, the satisfiability problem can be solved efficiently (14).

One has to be careful however in the use of randomly generated formulae: simple heuristic methods can quite easily determine the satisfiability of many of such formulae. This has led to some overly strong claims in the literature about handling very large formulae. Computationally challenging test instances can be obtained by generating formulae at or near the threshold (5). Cheeseman *et al.* (15) have made a similar observation of increased computational cost for heuristic search at a boundary between two distinct phases or behaviors of a combinatorial model. We will provide a precise characterization of the dependency on N of the threshold phenomena for k-SAT with k ranging from 2 to 6.

Our analysis shows that the threshold in k-SAT closely resembles the phase transition studied in spin glasses (16). We use finite-size scaling, a method from statistical physics in which observation of the width of the "critical region" of a transition highlights the "universal" behavior of quantities across the entire critical region. Such an analysis is usually justified by expressing all quantities rescaled to a "correlation length", with which two-point correlations decay exponentially at large distances when one is not at the critical point (4, 17). Below, we first illustrate the finite-size scaling approach on random graphs. We show that the method also works here, even though every two vertices are potentially neighbors, and there are neither metric nor correlation functions.

2 Finite-Size Scaling in the Random Graph Ensemble

The physical argument behind phenomenological scaling is that sufficiently close to a threshold or critical point, systems of all sizes are indistinguishable except for an overall change of scale. In the random graph ensemble, the clusters of size $N^{2/3}$ that occur close to its threshold in a random graph with $N = 100$ should simply look like coarse versions of the clusters found in a graph with $N = 10000$. However, to make the comparison, the narrow threshold observed for very large graphs must be expanded in scale to compare it with the broader threshold seen in small graphs. In the random graph ensemble, we know exactly how to do this. The parameterization $\alpha = 1/2 + \lambda N^{-1/3}$ is known to give simple formulae for properties close to threshold, capturing the main dependence on N. For graphical analysis of data from computer experiments, it is easier to plot data against the "rescaled α", $N^{1/3}(\alpha - 1/2)$, which will accomplish the same transformation.

But what to plot? A well-behaved quantity for this problem proves to be the sizes of the larger clusters, normalized to the size of the largest cluster in the same graph. When α is small, all clusters should be of the same general magnitude, so these ratios tend to unity. For $\alpha \gg 1/2$, normalizing by the giant cluster makes all the ratios tend to zero. Let L_k denote the size of the k-th largest cluster found in a particular graph, and $<>$ represent averaging over many samples of graphs with the same M and N. In Fig. 1a, we show $< L_2/L_1 >$ for graphs with $N = 100, 1000, 10000$, and 40000, plotted against α. 10000 samples were analyzed for each data point, giving roughly 1% accuracy throughout. The curves show the sharpening of the transition with increasing N and all intersect at the critical point, $\alpha = 1/2$. Plotting them against the "rescaled α", in Fig. 1b, we find that the averaged normalized cluster sizes follow universal forms through the critical regions, only separating when $|\alpha - 1/2| > N^{-1/3}$. The values of the size ratios at the critical point are independent of N, and appear to be new results. These are: $< L_2/L_1 >= 0.538 \pm .0015$, $< L_3/L_1 >= 0.382 \pm .001$, $< L_4/L_1 >= 0.302 \pm .0006$, $< L_5/L_1 >= 0.252 \pm .0006$ (averaged over 160,000 samples at each N).

We shall follow this procedure, determining the critical concentration by exact calculation if possible, by observation of the measured properties if not. For example, the intersection of all the lines in Fig. 1a identifies the critical point even if α_c were not known to be 1/2. We will determine the appropriate scaling power empirically in the analysis of k-SAT. Why does it work? It appears that power-law behavior in the size, N, replaces scaling with respect to a length. This is consistent with renormalization-group derivations of finite-size scaling (18), in which lengths occur only when raised to the appropriate power to yield the volume of the system, or N.

3 Experimental data

We have generated extensive data on the satisfiability of randomly generated k-CNF formulae with k ranging from 2 to 6. Fig. 2 shows the fraction of random k-SAT formulae that is *unsatisfiable* as a function of the ratio, α. For example, the left-most curve in Fig. 2 shows the fraction of formulae that is unsatisfiable for random 2CNF formulae with 50 variables over a range of values of α.

Each data point was generated using 10000 randomly generated formulae, giving roughly 1% accuracy. We used a highly optimized implementation (9) of the Davis-Putnam procedure (19).

This procedure performs a backtrack search through the space of possible truth assignments, and is the fastest known complete procedure for satisfiability testing on many classes of formulas (20).

Fig. 2 shows a clear threshold for each value of k. Except for the case $k = 2$, the curves cross at a single point and sharpen up with increasing N. For $k = 2$, the intersections between the curves for the largest values of N seem to be converging to a single point as well, although the curves for smaller N deviate.

From the perspective of performance evaluation for search algorithms, the point where 50% of the formulae are unsatisfiable is thought to be where the computationally hardest problems are found (5, 15). Note that the 50% point lies somewhat to the right of the scale-invariant point (the point where the curves cross each other), and shifts with N.

The thresholds move rapidly to the right with increasing k. This is because a clause with k distinct variables constrains only one of the 2^k truth assignments to the k variables, thus the constraints get weaker as k increases. Each of the M clauses reduces the space of possible truth assignments from 2^N by a factor of $(2^k - 1)/(2^k)$, neglecting interactions between the clauses. Raising this factor to the M^{th} power and asking when the number of satisfying configurations is reduced, on average, to 1, gives a plausible estimate of the threshold, identical to Chvátal's (6) upper bound: $\alpha_{ann} = (k - ln_2(2^k - 1))^{-1} \approx (ln2)2^k$. This is called an annealed estimate for α_c, because it ignores interactions between clauses, just as annealed theories of materials (16) average over many details of the disorder. This expression is also the first step in an analytic theory of k-SAT (21). We have marked $\alpha_{ann}(k)$ with arrows in the figures, and tabulate it in Table 1.

4 Results of Finite-Size Scaling Analysis

From Fig. 2, it is clear that the threshold "sharpens up" for larger values of N. Fig. 3a shows the effect in more detail for random 3CNF formulae. Both the threshold shift and the increasing slope in the curves of Fig. 3a can be accounted for by finite-size scaling. We plot the fraction of samples unsatisfied against the dimensionless rescaled variable,

$$y = N^{1/\nu}(\alpha - \alpha_c)/\alpha_c \quad .$$

Values for α_c and ν must be derived from the experimental data. First α_c is determined as the crossing point of the curves for large N in Fig. 3a. Then ν is determined to make the slopes match up through the critical region. In Fig. 3b we find that these two parameters capture both the threshold shift and the steepening of the curves, using $\alpha_c = 4.17$ and $\nu = 1.5$. We see that F, the fraction of unsatisfiable formulae, is given by $F(N, \alpha) = f(y)$, where the invariant function, f, is that graphed in Fig. 3b.

A description of the 50% threshold shift follows immediately. If we define y' by $f(y') = 0.5$, then $\alpha_{50} = \alpha_c(1 + y'N^{-1/\nu})$. From Fig. 3b we find that $\alpha_{50} \approx 4.17 + 3.1N^{-2/3}$. Crawford and Auton (9) fit their data on the 50% point as a function of N by arbitrarily assuming that the leading correction will be $O(1/N)$. They obtain $\alpha_{50} = 4.24 + 6/N$. However, the two expressions differ by only a few percent as N ranges from 10 to ∞.

Rescaling the data for random 4SAT, in Fig. 3c, leads to an even better fit to a single curve. We obtained similarly good fits for our rescaled data for the other values of k. In Table 1 we give the critical parameters obtained from this analysis. The error bars are subjective, and show the

k	α_{ann}	α_c	α'	ν
2	2.41	1.0	2.25	2.6±.2
3	5.19	4.17±.03	0.74	1.5±.1
4	10.74	9.7±.05	0.67	1.25±.05
5	21.83	21.0±.1	0.71	1.1±.05
6	44.01	43.5±.2	0.69	1.05±.05

Table 1: Critical parameters for random k-SAT.

range of each parameter over which the best fits were obtained. Note that ν appears to be tending to 1, and α_{ann} becomes an increasingly good approximation to α_c as k increases.

Finally, we note that the rescaled curves in Figs. 3b and 3c are very similar. In fact, combining the various rescaled curves in Figure 4, we find that the curves for $k \geq 3$ all roughly coincide, with the 50% point ≈ 0.7 for each. The curve for $k = 2$ is similar in form, but shifted to the right from the other ones, with $y' \approx 2.25$. Even "constant-probability SAT" (5), a model which rarely generates the "hard" search problems reported for 3-SAT, can be scaled onto a crossover function quite similar in form to Figs. 3b and 3c.

5 Conclusions

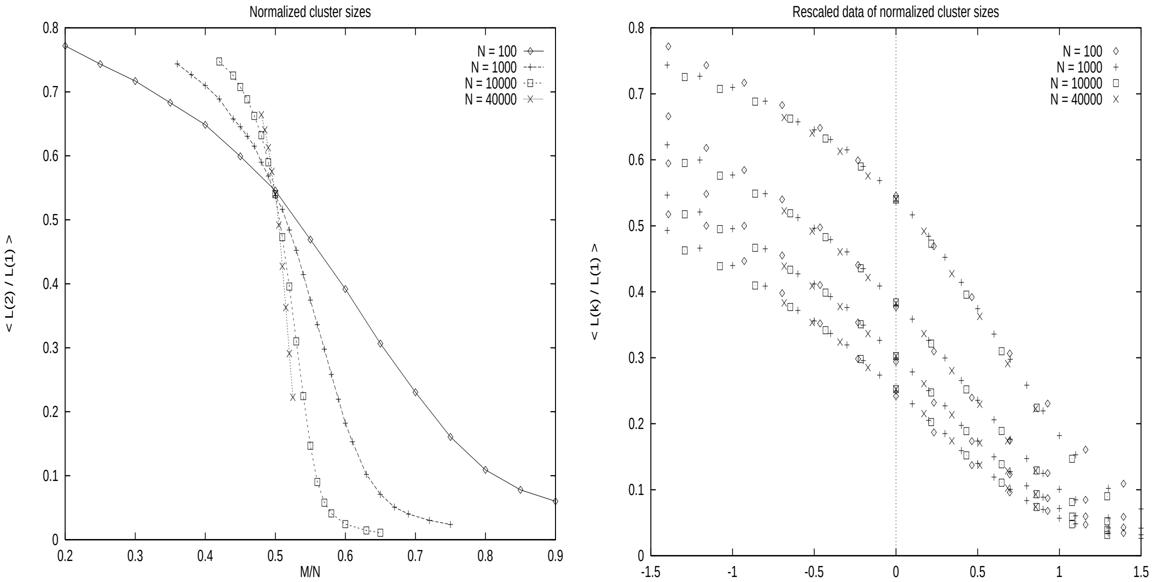
We have shown how finite-size scaling methods from statistical physics can be used to model the threshold phenomenon in randomly generated k-SAT problems. Given the good fit of our scaling analysis, we conjecture that this method can also be of use in characterizing phase transitions in other combinatorial problems of interest to the computer science community.

Several authors have attempted to relate computational compexity to the characteristics of phase transitions in models of disordered systems. Fu and Anderson (22) have proposed spin glasses (magnets with 2-spin interactions of random sign) as having inherent exponential complexity. Also, computationally hard graph coloring problems and 3SAT problems were found at or near phase transitions (5, 15). Huberman and colleagues (23) were first to focus on the diverging correlation length seen at continuous phase transitions as the root of computational complexity. In fact, both effects can play important roles, but are not sufficient and may not even be necessary.

There are NP-complete problems (*e.g.*, travelling salesman or max-clique) which lack a clear phase boundary at which "hard problems" cluster. Percolation thresholds are phase transitions, yet the cost of exploring the largest cluster never exceeds N steps. Similarly, finding a satisfying assignment, or proving its non-existence, for any 2-SAT instance, can be done in linear time. The spin glass Hamiltonians studied in (22) are similar to our 2-SAT formulae, but the questions studied are different and computationally much harder. Finding an assignment which falsifies the minimum number of clauses is like finding the ground state in a spin glass phase, and does not reduce to a single search on the directed graph of Aspvall (14). For 2-SAT, in fact, finding such "ground states" is NP-hard. Therefore, if both diverging correlations (diverging in size if no lengths are defined) and a "spin-glass" phase occurs, we expect search to be exponentially difficult.

References

1. P. Erdös and A. Rényi, *Publ. of the Math. Inst. of the Hung. Acad. Sci.*, **5**, 17 (1960).
2. B. Bollobás, *Random Graphs* (Academic Press, London, 1985), and J. Spencer, *Ten Lectures on the Probabilistic Method* (Soc. for Ind. and Appl. Math., Philad., 1987).
3. S. Janson, D.E. Knuth, T. Łuczak, B. Pittel, *Random Structures and Algorithms*, **4**, 231 (1993).
4. D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor and Francis, London, 1992).
5. D. Mitchell, B. Selman, and H.J. Levesque, *Proc. 10th Natl. Conf. on Artif. Intel.* (AAAI Press, Menlo Park, 1992), p. 459.
6. V. Chvátal and B. Reed, *Proc. Ann. ACM Symp. on the Theo. of Comput.* (Assoc. for Comput. Mach., New York, 1992), p. 620.
7. A. Goerdt, *Proc. 17th Int. Symp. on the Math. Found. of Comp. Sc.* (Springer-Verlag, Berlin, 1990).
8. A.Z. Broder, A.M. Frieze, E. Upfal, *Proc. 4th Ann. ACM-SIAM Symp. on Discr. Alg.,* (Soc. for Ind. and Appl. Math., Philad., 1993), p. 322.
9. J.M. Crawford and L.D. Auton, *Proc. 11th Natl. Conf. on Artif. Intel.* (AAAI Press, Menlo Park, 1993), p. 21.
10. T. Larrabee and Y. Tsuji, *Proc. AAAI Symp. on AI and NP-hard problems*, 112 (1993).
11. B. Selman, H.J. Levesque, D. Mitchell, *Proc. 10th Natl. Conf. on Artif. Intel.* (AAAI Press, Menlo Park, CA, 1992), p. 440.
12. A.P. Kamath, N.K. Karmarkar, K.G. Ramakrishan, M.G.C. Resende, *Annals of O.R.*, **25**, 43 (1990).
13. S.A. Cook, *Proc. 3rd Ann. ACM Symp. on the Theo. of Comput.,* 151 (1971). (Assoc. for Comput. Mach., New York, 1971), p. 151.
14. B. Aspvall, M.F. Plass, R.E. Tarjan, *Inform. Process. Let.*, **8**, 121 (1979).
15. P. Cheeseman, B. Kanefsky, W.M. Taylor, *Proc. 12th Int. Conf. on Artif. Intel.* (Morgan Kaufmann, San Mateo, CA, 1991), p. 331.
16. M. Mézard, G. Parisi, M.A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1986).
17. S. Kirkpatrick, and R.H. Swendsen, *Comm. ACM*, **28**, 363 (1985).
18. M.N. Barber, in *Phase Transitions and Critical Phenomena*, **8**, C. Domb and J. L. Lebowitz (Eds.) (Academic Press, London, 1983), p. 145.
19. M. Davis and H. Putnam, *J. Assoc. Comput. Mach.*, **7**, 201 (1960).
20. O. Dubois, P. Andre, Y. Boufkhad, J. Carlier, *Proc. 2nd DIMACS Implementation Challenge*, Piscateway, NJ (1993).
21. S. Kirkpatrick, G. Györgyi, N. Tishby, L. Troyansky, to appear in Advances in Neural Information Processing Systems **6**, 1993, and unpublished work.
22. Y. Fu, in *Lectures in the Sciences of Complexity*, D. Stein (Ed.) (Addison-Wesley, 1989), p. 815.
23. S.H. Clearwater, B.A. Huberman, T. Hogg, *Science*, **254**, 1181 (1991).



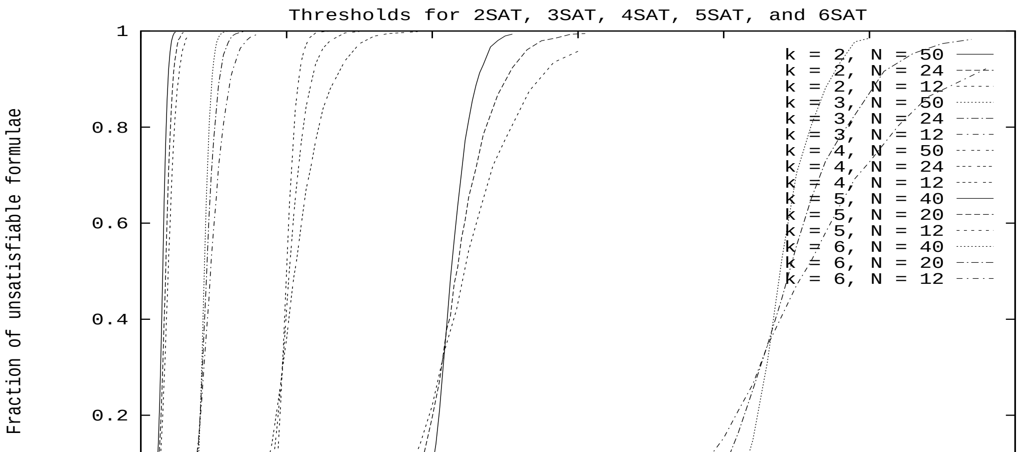
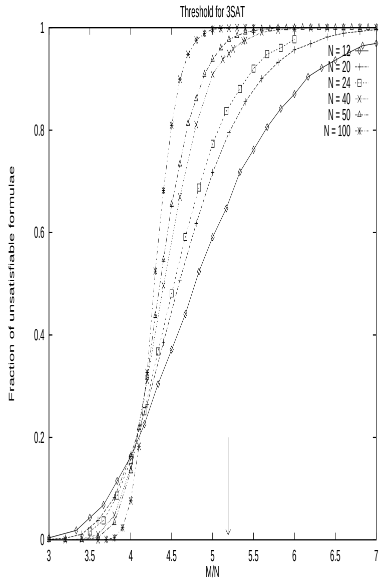
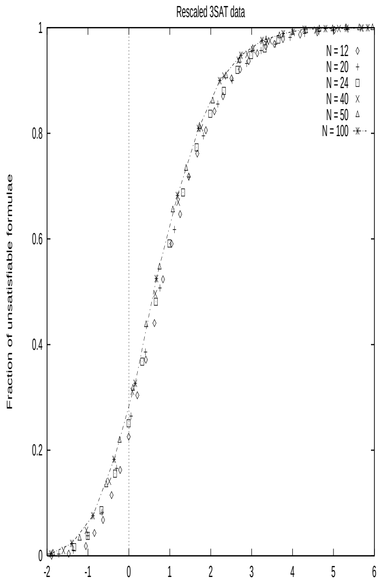
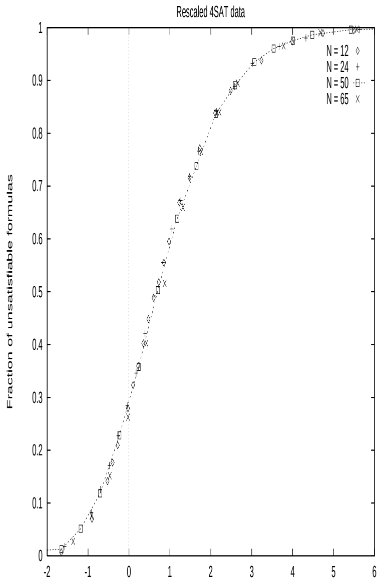


Fig. 1a: The ratio of the size of second largest cluster to that of the largest cluster.
Fig. 1b: Rescaled cluster data. ($(k-1)^{\mathrm{st}}$ curve from top gives rescaling for $< L_k/L_1 >$ data.)
Fig. 2: Fraction of unsatisfiable formulae for 2-, 3-, 4-, 5-, and 6-SAT.







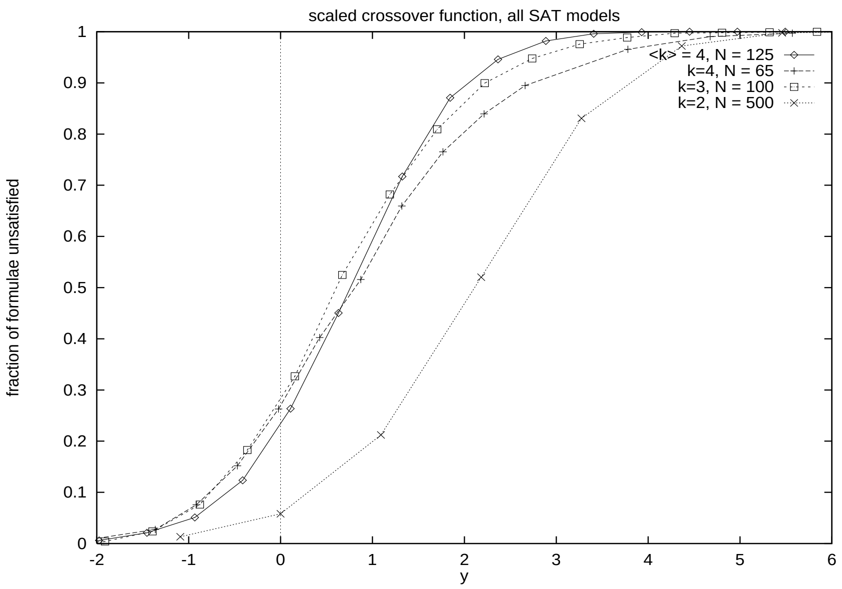


Fig. 3a: Threshold data for 3-SAT, $N = 12$ to 65.
Fig. 3b: Rescaled 3-SAT data using $\alpha_c = 4.17$, $\nu = 1.5$.
Fig. 3c: Rescaled 4-SAT data using $\alpha_c = 9.7$, $\nu = 1.25$.
Fig. 4: Rescaled crossover functions for $k = 2, 3, 4$, and $< k >= 4$ (see (5)).