Phase Transitions and Algorithmic Complexity

Stefan Boettcher and Michelangelo Grigni
Emory University, Atlanta, GA 30322

Gabriel Istrate and Allon G. Percus
Los Alamos National Laboratory, Los Alamos, NM 87545

September 6, 2000

The most challenging instances of computational problems are often found near a critical threshold in the problem's parameter space [1], where certain characteristics of the problem change dramatically. Monasson *et al.* [2] have conjectured a correlation between computational complexity and physical "phase transition" behavior at this threshold: computationally "hard" problems (running time exponential in problem size) display a discontinuous phase transition under an order parameter; computationally "easy" ones (polynomial in problem size) display a continuous phase transition. They showed that this holds for a class of problems called K-satisfiability (K-SAT), which depending on the value of K can be either hard or easy. However, to argue convincingly for the existence of such a correlation, further evidence is needed. Here, we study two distinctly different graph-based problems, one hard and one easy, demonstrating numerically that in both cases the conjecture holds.

A random graph [3] is constructed out of a set of N vertices by assigning an edge to each pair of vertices with probability p. The two problems we consider on the graph are 3-coloring and graph bi-partitioning. The goal of 3-coloring (3-COL) is to label each vertex with any one of three colors so as to minimize the number of edges connecting two vertices of the same color (the "cost"). If the average connectivity pN per vertex is very small, almost all vertices have fewer than 3 neighbors and a zero-cost coloring exists. But at our observed critical value $(pN)_{\rm crit} \approx 4.73$ (see also Refs. [4, 1]), there is a sudden transition to a regime with almost no perfect colorings. For 3-COL, finding a zero-cost solution (if one exists) is NP-complete [5], and thus computationally hard in the worst case. In the graph bi-partitioning problem (GBP), the goal is to divide the N vertices into two sets of *exactly equal* size $N/2$ so as to minimize the number of edges connecting vertices in opposite sets (the "cost"). At low connectivity, the graph consists of many small disconnected clusters that are easily bi-partitioned, again at zero cost. At the critical value $(pN)_{\rm crit} = 2\ln 2$, a single giant cluster of size $> N/2$ emerges [6], making the cost positive. For the GBP, finding a zero-cost solution is computationally easy [5]. Unlike K-SAT, these graph problems exhibit a symmetry under relabeling of vertices. In physical models, different symmetries often cause distinct critical behaviors [7], making 3-COL and the GBP fertile testing grounds for the phase transition/complexity conjecture.

The order parameter used in K-SAT was the "backbone" fraction f, the fraction of the N variables that are frozen to the same value over all minimum-cost (not necessarily zero-cost) solutions. In our graph problems, however, due to the relabeling symmetry, no vertex's value can be frozen in this way. Furthermore, whereas in K-SAT the cost contributions arise from the individual variable values, in 3-COL and the GBP they arise from the arrangement of *pairs*. Take the backbone to be those pairs whose two members, over all optimal solutions, potentially contribute a positive cost: having always the *same* coloring for 3-COL, or being always in *opposite* partitions for the GBP. We define f as the fraction of such pairs in the graph. An analogy with the notion of a "spine", recently introduced for K-SAT [8] as a mathematically tractable generalization of the backbone, suggests that this definition is the correct one.

We consider a large number of instances of random graphs, of sizes up to $N = 1024$ and over a range of pN values near the critical threshold. For each instance, we wish to determine the backbone fraction f. For 3-COL, we use a rapid optimization heuristic called *extremal optimization* (EO) [9] that, based on a previous numerical study [10] as well as testbed comparisons with an exact algorithm [11], is expected to yield an excellent approximation of f. For the GBP, we use an exact algorithm.

Figure 1 shows f around the critical region. A simple argument [2] demonstrates that below the critical point the backbone fraction always vanishes for large N. Above the critical point, for 3-COL (Fig. 1a) f never vanishes, so the backbone is discontinuous. This is in close resemblance to K-SAT at $K = 3$ [2]; indeed, both typically are computationally hard at the threshold. For the GBP (Fig. 1b), on the other hand, the backbone is continuous at the critical point, vanishing on both sides for large N. This mirrors the behavior of K-SAT at $K = 2$ where, as for the GBP, finding a zero-cost solution is computationally easy.

Even though the backbone is defined in terms of minimum-cost solutions, its behavior appears to correlate with the complexity of finding a zero-cost solution at the threshold. This is because instances there have low cost: determining this cost is only polynomially more difficult than determining whether a zero-cost solution exists. Interestingly, a recent report [12] has proposed a version of the spine [8] defined exclusively in terms of zero-cost graphs. This quantity also displays a discontinuous transition for 3-COL, however it is an (uncontrolled) upper bound on the backbone and may *not* generally correlate with zero-cost complexity. In fact, the authors speculate that at the 3-COL threshold, although the spine is discontinuous, the backbone is *continuous*. Our study contradicts this speculation, instead confirming the link between backbone behavior and typical-case computational complexity.

References

[1] Cheeseman, P., Kanefsky, B. & Taylor, W.M. in *Proc. 13th Int. Joint Conf. on Artificial Intelligence (IJCAI–91)* (eds Mylopoulos, J. & Reiter, R.) 331–337 (Morgan Kaufmann, San Mateo, California, 1991).

[2] Monasson, R., Zecchina, R., Kirkpatrick, S., Selman, B. & Troyansky, L. *Nature* **400**, 133–137 (1999).

[3] Bollobás, B. *Random Graphs* (Academic Press, London, 1985).

[4] Hogg, T. *Artificial Intelligence* **81**, 127–154 (1996).

[5] Garey, M.R. & Johnson, D.S. *Computers and Intractability: A Guide to the Theory of NP-Completeness* (W. H. Freeman, New York, 1979).

[6] Erdős, P. & Rényi, A. *Publication of the Mathematical Institute of the Hungarian Academy of Science* **5**, 17–61 (1960).

[7] Goldenfeld, N. *Lectures on Phase Transitions and the Renormalization Group* (Addison-Wesley, Reading, 1992).

[8] Bollobás, B., Borgs, C., Chayes, J.T., Kim, J.H. & Wilson, D.B. Technical Report MSR-TR-99-41, available at *http://xxx.lanl.gov/abs/math.CO/9909031*.

[9] Boettcher, S. & Percus, A.G. *Artificial Intelligence* **119**, 275–286 (2000).

[10] Boettcher, S. *J. Phys. A: Math. Gen.* **32**, 5201–5211 (1999).

[11] Trick, M.A. *color.c graph coloring code*, available at *http://mat.gsia.cmu.edu/COLOR/solvers/trick.c*.

[12] Culberson, J. & Gent, I. Technical Report APES-19-2000, available at *http://www.apes.cs.strath.ac.uk*.

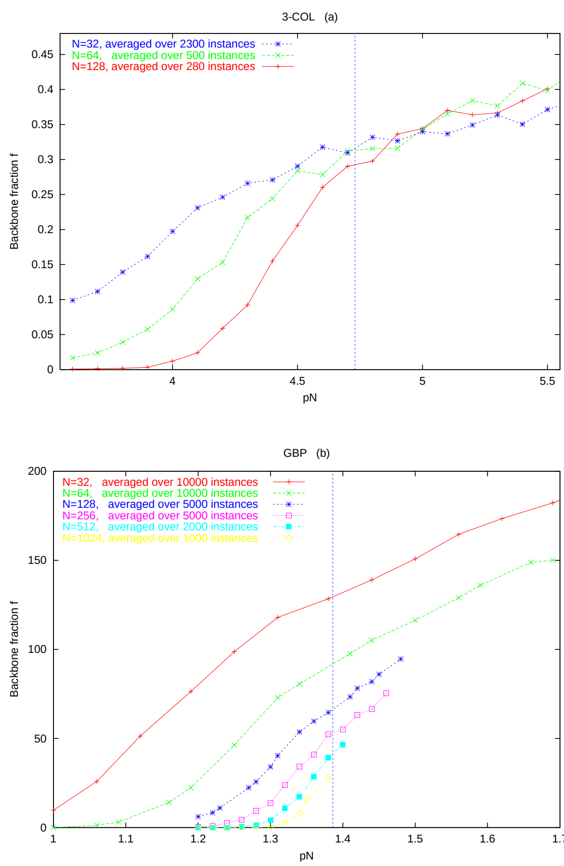

Figure 1: Plot of the estimated backbone fraction (a) for 3-COL and (b) for the GBP, on random graphs, as a function of connectivity pN. For 3-COL, the systematic error based on benchmark comparisons with random graphs is consistently under 0.005; for the GBP, f is found by exact enumeration. The critical points at $pN \approx 4.73$ for 3-COL and $pN = 2\ln 2$ for the GBP are shown by a vertical line.