2+*p*-COL

Toby Walsh
Cork Constraint Computation Center
University College Cork
Cork
Ireland
tw@4c.ucc.ie

Abstract

Like the well known 2+p-SAT problem, the 2+p-COL problem smoothly interpolates between P and NP by mixing together the polynomial 2-coloring problem and the NP-complete 3-coloring problem. As with 2+p-SAT, the polynomial subproblem can dominate the problem's solubility and the search complexity. The 2+p-COL problem class has, however, at least one very significant difference over the 2+p-SAT problem class. 2-SAT and 3-SAT (and thus 2+p-SAT) have sharp transitions in satisfiability. On the other hand, 3-COL has a sharp transition in solubility but 2-COL has a smooth transition. In the 2+p-COL problem, we therefore observe phase transition behavior in which there appear to be both smooth and sharp regions. We also show how this problem class can help to understand algorithm behavior by considering search trajectories through the phase space.

Introduction

Phase transition behavior has given new insight into what makes NP-complete problems hard to solve (Cheeseman, Kanefsky, & Taylor 1991; Mitchell, Selman, & Levesque 1992; Gomes & Selman 1997; Walsh 1999). Some of the most interesting phase transition results have come from the 2+p-SAT problem class introduced in (Monasson *et al.* 1999). This mixes together the polynomial 2-SAT and the NP-complete 3-SAT problem. This lets us explore in detail the interface between P and NP. Surprisingly, the polynomial 2-SAT subproblem dominates the satisfiability and cost to solve 2+p-SAT problems up to $p = 2/5$ (Monasson *et al.* 1998; Achlioptas *et al.* 2001). This is despite the problem being NP-complete for any fixed $p > 0$. Given the insight that 2+p-SAT has provided into computational complexity and algorithm performance, we decided to look more deeply into the interface between P and NP by means of five new problem classes (Walsh 2002). Here, we look in more detail at the results of the first of these investigations, the 2+p-COL phase transition. First, however, we review results concerning the 2+p-SAT problem class.

2+p-SAT

A random k-SAT problem in n variables has l clauses of length k drawn uniformly at random. A sharp transition in satisfiability has been proved for random 2-SAT at $l/n = 1$ (Chvatal & Reed 1992; Goerdt 1992), and conjectured for random 3-SAT at $l/n \approx 4.3$ (Mitchell, Selman, & Levesque 1992). Associated with this transition is a rapid increase in problem difficulty. The random 2-SAT transition is continuous as the backbone (the fraction of variables taking fixed values) increases smoothly. On the other hand, the random 3-SAT transition is discontinuous as the backbone jumps in size at the phase boundary (Monasson *et al.* 1998).

To study this in more detail, Monasson et al. introduced the 2+p-SAT problem class (Monasson *et al.* 1999). This interpolates smoothly from the polynomial 2-SAT problem to the NP-complete 3-SAT problem. A random 2+p-SAT problem in n variables has l clauses, a fraction $(1 - p)$ of which are 2-SAT clauses, and a fraction p of which are 3-SAT clauses. This gives pure 2-SAT problems for $p = 0$, and pure 3-SAT problems for $p = 1$. For any fixed $p > 0$, the 2+p-SAT problem class is NP-complete since the embedded 3-SAT subproblem can be made sufficiently large to encode other NP-complete problems within it.

By considering the satisfiability of the embedded 2-SAT subproblem and by assuming that the random 3-SAT transition is at $l/n \approx 4.3$, we can bound the location the random 2+p-SAT transition to:

$$1 \leq \frac{l}{n} \leq \min(\frac{1}{1-p}, 4.3)$$

Surprisingly, the upper bound is tight for $p \leq 2/5$ (Achlioptas *et al.* 2001). That is, the 2-SAT subproblem *alone* determines satisfiability up to $p = 2/5$. Asymptotically, the 3-SAT clauses do not determine if problems are satisfiable, even though they determine the worst-case complexity. Several other phenomena occur at $p = 2/5$ reflecting this change from a 2-SAT like transition to a 3-SAT like transition. For example, the transition shifts from continuous to discontinuous as the backbone jumps in size (Monasson *et al.* 1998). In addition, the average cost to solve problems appears to increase from polynomial to exponential both for complete and local search algorithms (Monasson *et al.* 1998; Singer, Gent, & Smaill 2000). Random 2+p-SAT problem thus look like polynomial 2-SAT problems up to $p = 2/5$ and NP-complete 3-SAT problems for $p > 2/5$.

The 2+p-SAT problem class helps us understand the performance of the DLL algorithm for solving 3-SAT (Cocco & Monasson 2001). At each branch point in its backtracking search tree, the DLL algorithm has a mixture of 2-SAT and 3-SAT clauses. We can thus view each branch as a trajectory in the 2+p-SAT phase space. For satisfiable problems solved without backtracking (i.e. $l/n < 3$), trajectories stay within the satisfiable part of the phase space. For satisfiable problems that require backtracking (i.e. $3 < l/n < 4.3$), trajectories cross the phase boundary separating the satisfiable from the unsatisfiable phase. The length of the trajectory in the unsatisfiable phase gives a good estimate of the amount of backtracking needed to solve the problem. Finally, for unsatisfiable problems, trajectories stay within the unsatisfiable part of the phase space. The length of the trajectory again gives a good estimate of the amount of backtracking needed to solve the problem.

2+p-COL

A random k-COL problem consists of n vertices, each with k possible colors, and e edges drawn uniformly and at random. Like k-SAT, k-COL is NP-complete for $k \geq 3$ but polynomial for $k = 2$. To interpolate smoothly from P to NP, we introduce the random 2+p-COL problem class in which random graphs have a fraction $(1 - p)$ of their vertices with 2 colors, and a fraction p with 3 colors. Note that the vertices with 2 colors are fixed at the start and cannot be chosen freely. In addition, the 2 color vertices all have the same 2 colors available. Like 2+p-SAT, the 2+p-COL problem class is NP-complete for any fixed $p > 0$.

2+p-COL has one major difference to 2+p-SAT. Whilst 2-SAT, 3-SAT and 3-COL all have sharp transitions, 2-COL has a smooth transition (Achlioptas 1999). The probability that a random graph is 2-colorable is bounded away from 1 as soon as the average degree is bounded away from 0, and drops gradually as the average degree is increased, only hitting 0 with the emergence of the giant component (and an odd length cycle). Hence 2-colorability does not drop sharply around a particular average degree (as in 3-colorability) but over an interval that is approximately: $0 < e/n < 1/2$.

Phase transition

In Figure 1, we see how the random 2+p-COL phase transition varies as we increase p and n. At $p \approx 0.8$, the 2+p-COL transition appears to sharpen significantly. In Figure 2, we look more closely at $p = 0.8$. For $p = 0.8$, there is a region up to around $e/n \approx 1.8$ in which the transition appears smooth and like that of 2-COL. The nature of the transition then appears to change to a sharp 3-COL like transition, with the probability of colorability dropping rapidly from around 90% to 0%. We thus appear to have both smooth and sharp regions.

Achlioptas defines the location of the colorability phase transition as the point where graphs can no longer always be colored. We define δ_{2+p} as the largest ratio e/n at which 100% of problems are colorable:

$$\delta_{2+p} = \sup\{\frac{e}{n} \mid \lim_{n \mapsto \infty} \text{prob}(2+p\text{-colorable}) = 1\}$$

From (Achlioptas 1999), $\delta_2 = 0$ and $1.923 < \delta_3 < 2.522$. For any fixed $p < 1$, a random 2+p-COL problem contains

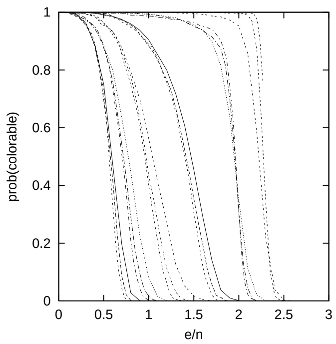

Figure 1: Probability that 2+p-COL problem is colorable (y-axis) against e/n (x-axis). Plots are for p=0 (leftmost), 0.2, 0.4, 0.6, 0.8 and 1.0 (rightmost) for 100, 200 and 300 vertex graphs (increasing sharpness).

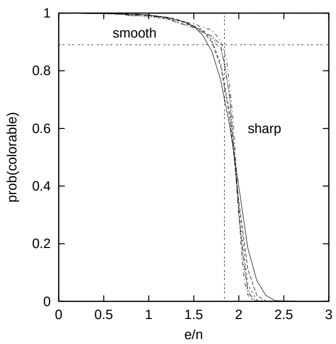

Figure 2: Probability that 2+p-COL problem is colorable (y-axis) against e/n (x-axis) for $p = 0.8$. Plots are for n=50, 100, 150, 200 and 250 vertex graphs.

a 2-COL subproblem that grows in size with n and has average degree bounded away from 0. This subproblem has a probability of being 2-colorable that asymptotically is less than 1. Hence, the random 2+p-COL problem has a probability of being 2+p-colorable that asymptotically is also less than 1, and $\delta_{2+p} = 0$ for all $p < 1$.

We can define a dual parameter γ_{2+p}, which is the smallest ratio e/n at which 0% of problems are colorable:

$$\gamma_{2+p} = \inf\{\frac{e}{n} \mid \lim_{n \mapsto \infty} \text{prob}(2+p\text{-colorable}) = 0\}$$

Since colorability is a monotonic property (adding edges can only ever turn a colorable graph into an uncolorable graph), $\gamma_{2+p} \geq \delta_{2+p}$. Note that δ_{2+p} marks the start of the phase transition whilst γ_{2+p} marks its end. The start stays fixed at $\delta_{2+p} = 0$ for all $p < 1$ and jumps discontinuously to γ_3 at $p = 1$. The end appears to behave more smoothly, increasing smoothly as we increase p. From (Achlioptas 1999), $\gamma_2 \approx \frac{1}{2}$, and $1.923 < \gamma_3 < 2.522$. For a sharp transition like 3-coloring, $\gamma_3 = \delta_3$. As with 2+p-SAT:

$$\gamma_2 \leq \gamma_{2+p} \leq \min(\gamma_3, \frac{\gamma_2}{1-p})$$

In Figure 3, we have estimated experimentally the location of γ_{2+p} by analysing data for graphs with up to 300 vertices and compared the observed experimental location of the (end of the) phase transition with the upper and lower bounds. As with 2+p-SAT, the upper bound (which looks just at the 2-COL subproblem) appears to be tight for $p < 0.8$.

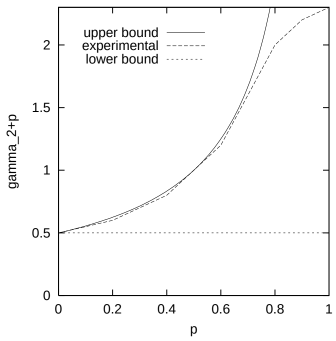

Figure 3: The location of the end of the 2+p-COL phase transition, γ_{2+p} (y-axis) against p (x-axis) for $p = 0$ to 1 in steps of 1/10.

Search cost

The cost of 2+p-coloring also increases around $p \approx 0.8$. To solve 2+p-COL problems, we encode them into SAT and use zchaff, which is currently the fastest DLL procedure. Our results are thus algorithm dependent and should be repeated with other solvers. Note that the encodings of 2+p-COL problems into SAT give 2+p-SAT problems (but does not sample uniformly as the 3-clauses are only ever positive). In Figure 4, we see that there is a change in the search cost around $p \approx 0.8$ where we appear to move from polynomial to exponential search cost. This is despite 2+p-COL being NP-complete for all fixed and non-zero p. However, this is perhaps not so surprising as up to $p \approx 0.8$, the polynomial 2-COL subproblem *alone* appears to determine asymptotically if the problem is colorable. Beyond this point, the NP-complete 3-COL subproblem contributes to whether the problem is colorable or not.

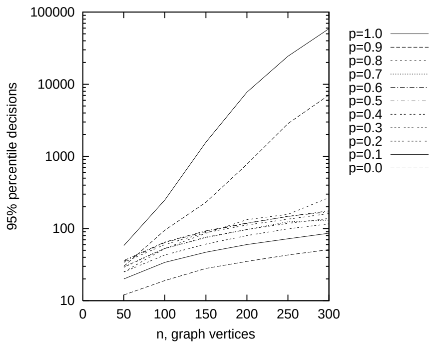

Figure 4: 95% percentile in the search cost to solve 2+p-COL problems at the phase boundary (y-axis logscale) against number of vertices (x-axis). Plots are for p=0 to 1 in steps of 1/10 (increasing hardness)

Search trajectories

2+p-COL can be used in a similar way to 2+p-SAT to study 3-coloring algorithms. At each branching decision in a coloring algorithm, some vertices have three colors available, whilst others only have two. If any vertex has only a single color available, we are not at a branching point as we can commit to this color and simplify the problem. The algorithm thus has a sequence of 2+p-COL problems to solve. Under a number of different branching heuristics, it can be shown that these subproblems sample uniformly from the random 2+p-COL problem class. We can therefore view each branch in the algorithm's search tree as a trajectory in the 2+p-COL phase space. In Figure 5, we plot a number of trajectories for 3-coloring graphs with Brelaz's algorithm (Brelaz 1979).

The trajectories are qualitatively very similar to those of the DLL algorithm in the 2+p-SAT phase space (Cocco & Monasson 2001). For $e/n \leq 1.5$, problems are solved without backtracking. Trajectories trace an arc that stays within the 'colorable' part of the phase space. On the other hand, the algorithm backtracks for problems with $e/n \geq 2$. For graphs with $e/n = 2$, the trajectory starts in the 'colorable' part of the phase space and crosses over into the 'uncolorable' part of the phase space. The algorithm then backtracks till we return to the 'colorable' part of the phase space. This sort of knowledge might be used both to model

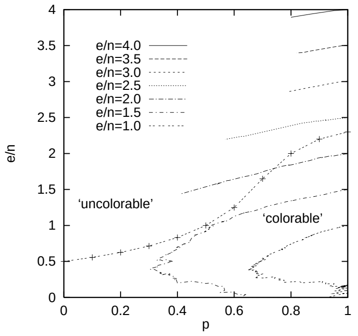

Figure 5: Trajectories in the 2+p-COL phase space for Brelaz's graph coloring algorithm on 3-COL problems with $n = 300$ and e/n from 1 to 4. The crossed line gives the experimental observed values of γ_{2+p}. The region marked 'uncolorable' is where graphs are asymptotically not 2+p-colorable, whilst the region marked 'colorable' is where graphs are asymptotically sometimes 2+p-colorable.

algorithms and to improve them. For example, we could develop a randomization and restart strategy (Gomes, Selman, & Kautz 1998) which restarts when we estimate to have branched into an insoluble part of the phase space.

Other variants

2+p-COL

In the problems studied so far, the 2-colorable nodes always have the same two colors. We could, however, look at the related 2+p-COL problem in which the 2-colorable nodes have *any* two of the three colors available to the 3-colorable nodes. These two colors are chosen at random and fixed before search begins. We obtained similar results with this variant of the 2+p-COL problem. However, the shift from a smooth 2-COL like transition to a sharp 3-COL like transition now appears to occur somewhere around $p \approx 0.96$ (compared to $p \approx 0.8$ previously). In addition to the transition becoming sharp, (median) problem difficulty increases rapidly at the 2+p-colorability transition for $p \geq 0.96$. We observe a large peak in problem difficulty around $e/n \approx 2.5$ for $p \geq 0.96$.

2COL2SAT

We also studied a problem class which interpolates between a smooth and a sharp transition where both the problems are polynomial. A 2-COL problem can be encoded as a constraint satisfaction problem on 0/1 variables in which the constraints are all disequalities. On the other hand, a 2-SAT problem can also be encoded as a constraint satisfaction problem on 0/1 variables in which the constraints are all clauses. We can therefore interpolate between these two problem classes merely by changing the type of constraints between the 0/1 variables from disequalities to clauses. In 2COL2SAT, we generate a constraint satisfaction problem in n 0/1 variables in which a fraction $(1 - p)$ of the e constraints are disequalities, and a fraction p are clauses. This maps between the smooth 2-COL transition and the sharp 2-SAT transition. At all times, however, the problem remains polynomial. The phase transition shifts to larger e/n as p increases. This is to be expected as disequalities are roughly twice as constraining as clauses. For $0 < p < 1$, the transition appears to have both smooth and sharp regions. This supports the picture seen with 2+p-COL where we also had transitions with both smooth and sharp regions. This problem class also demonstrates that the change from a smooth to a sharp transition alone in not the cause of a change from P (polynomial) type average case problem difficulty to NP type average case problem difficulty.

Conclusions

We have studied in more detail the interface between P and NP by means of the 2+p-COL problem. This smoothly interpolates between the polymomial 2-coloring problem and the NP-complete 3-coloring problem. The behavior of the 2+p-COL problem appears to be dominated by the embedded polynomial 2-COL subproblem up to $p \approx 0.8$. 3-COL has a sharp transition in solubility but 2-COL has a smooth transition. In the 2+p-COL problem, we observe phase transition behavior for $p < 0.8$ in which there appear to be both smooth and sharp regions.

What important lessons can be learnt from this study? First, it appears that we can have transitions with are both smooth and sharp. Problems like 2+p-COL let us study in detail how transitions sharpen and the large impact this has on search cost. Second, problem classes like these can help us understand algorithm behavior. For instance, we can view Brelaz's 3-coloring algorithm as searching trajectories in the 2+p-COL phase space. And finally, given the insights gained from studying the interface between P and NP, it is may be worth looking at the interface between other complexities classes. For example, we might study the interface between NP and PSpace.

Acknowledgements

This research was supported in full by an EPSRC advanced research fellowship. The author wishes to thank the members of the APES research group, as well as Carla Gomes, Bart Selman, and Christian Bessiere.

References

Achlioptas, D.; Kirousis, L.; Kranakis, E.; and Krizanc, D. 2001. Rigorous results for (2+p)-SAT. *Theoretical Computer Science* 265(1-2):109–129.

Achlioptas, D. 1999. *Threshold phenomena in random graph colouring and satisfiability*. Ph.D. Dissertation, Department of Computer Science, University of Toronto.

Brelaz, D. 1979. New methods to color the vertices of a graph. *Communications of ACM* 22:251–256.

Cheeseman, P.; Kanefsky, B.; and Taylor, W. 1991. Where the really hard problems are. In *Proceedings of the 12th IJ-*

CAI, 331–337. International Joint Conference on Artificial Intelligence.

Chvatal, V., and Reed, B. 1992. Mick gets some (the odds are on his side). In *Proceedings of the 33rd Annual Symposium on Foundations of Computer Science*, 620–627. IEEE.

Cocco, S., and Monasson, R. 2001. Trajectories in phase diagrams, growth processes and computational complexity: how search algorithms solve the 3-satisfiability problem. *Physical Review Letters* 86(8):1654–1657.

Goerdt, A. 1992. A theshold for unsatisfiability. In Havel, I., and Koubek, V., eds., *Mathematical Foundations of Computer Science*, Lecture Notes in Computer Science. Springer Verlag. 264–274.

Gomes, C., and Selman, B. 1997. Problem structure in the presence of perturbations. In *Proceedings of the 14th National Conference on AI*, 221–226. American Association for Artificial Intelligence.

Gomes, C.; Selman, B.; and Kautz, H. 1998. Boosting combinatorial search through randomization. In *Proceedings of 15th National Conference on Artificial Intelligence*, 431–437. AAAI Press/The MIT Press.

Mitchell, D.; Selman, B.; and Levesque, H. 1992. Hard and Easy Distributions of SAT Problems. In *Proceedings of the 10th National Conference on AI*, 459–465. American Association for Artificial Intelligence.

Monasson, R.; Zecchina, R.; Kirkpatrick, S.; Selman, B.; and Troyansky, L. 1998. Determining computational complexity for characteristic 'phase transitions'. *Nature* 400:133–137.

Monasson, R.; Zecchina, R.; Kirkpatrick, S.; Selman, B.; and Troyansky, L. 1999. 2+p SAT: Relation of typical-case complexity to the nature of the phase transition. *Random Structures and Algorithms* 15(3-4):414–435.

Singer, J.; Gent, I.; and Smaill, A. 2000. Local search on random 2+p-SAT. In *Proceedings of the 14th ECAI*. European Conference on Artificial Intelligence.

Walsh, T. 1999. Search in a small world. In *Proceedings of 16th IJCAI*. International Joint Conference on Artificial Intelligence.

Walsh, T. 2002. From P to NP: COL, XOR, NAE, 1-in-k, and Horn SAT. In *Proceedings of the 17th National Conference on AI*. American Association for Artificial Intelligence.