An Improved Upper Bound on the non-3-colourability threshold

Paul E. Dunne and Michele Zito*)
Department of Computer Science, University of Liverpool, Liverpool L69 7ZF, U.K.

Abstract

In this paper we derive an improved upper bound on the average vertex degree, δ, needed to ensure that: $\forall\, \varepsilon > 0$ and n sufficiently large, a random n-vertex graph with at least $(\delta + \varepsilon)\, n/2$ edges is almost certainly not 3-colourable.

Keywords: 3-Colouring, phase transition threshold.

1. Introduction

Experimental studies have indicated that a number of *NP*-complete decision problems exhibit the following behaviour: if X_n is the set of instances of f of size n, and $m(x)$ is some parameter of instances $x \in X_n$, then instances for which $m(x)$ is 'small' almost all have $f(x) = 1$ and instances for which $m(x)$ is 'large' almost all have $f(x) = 0$. Furthermore, there is an apparent *threshold*, θ_f, with which almost all instances having $m(x)/n < \theta_f$ are such that $f(x) = 1$; and almost all instances having $m(x)/n > \theta_f$ are such that $f(x) = 0$. Behaviour of this nature is known as a *phase transition*. For 3-SAT, taking the size of an instance, F, to be the number of propositional variables upon which F depends and m to be the number of clauses in F, the experiments of [4] suggest that a phase transition occurs at a threshold of $m/n \approx 4.24$; for the 3-colouring problem, using n as the number of vertices and m as the total graph degree, [7] reports an apparent phase transition in the vicinity of $m/n \approx 4.6$. Phase transition thresholds are of interest since their location gives information about the 'hardest' instances of a given size and thereby provide a potential testbed with which to assess heuristics to solve such problems. Furthermore, analytic proofs that instances of f with some value of m/n 'almost certainly' have $f(x) = 1$ often lead to efficient average case algorithms to solve f, e.g. the Hamiltonian Cycle algorithms of Angluin and Valiant [1].

Purely analytic studies of phase transition phenomena have largely concentrated on satisfiability problems. Goerdt [6] proved that a sharp threshold exists at $m/n = 1$ for 2-SAT. Results for 3-SAT are rather less complete, and it has yet to be confirmed by analysis that a threshold exists. Initial upper bounds on the clause-to-variable ratio m/n were obtained in [3] where it was demonstrated that 'almost all' m-clause, n-variable instances with $m/n > 5.19..$ are unsatisfiable. The best bounds to date are those of [8,9] where this is lowered to $4.598..$. All of these methods extend to give an upper bound for k-SAT in terms of k. A lower bound of 3.003 is announced in [5], improving the earlier bound of 2.99 given in [2].

In this paper we obtain a small improvement in the upper bound for the phase transition threshold in 3-Colouring n-vertex graphs. An elementary 'first-moment' method suffices to show that if $m(G)$ is the degree sum of an n-vertex graph G, then 'almost all' graphs with $m(G)/n > 5.419$ are not 3-vertex colourable. We develop the technique discovered by [8], in order to reduce this to 5.2057. The remainder of this paper is organised as follows. In the next section we give some basic definitions and notations used. In Section 3, we review the 'first moment' argument and outline the technique introduced in [8] to improve the 3-SAT threshold. In Section 4, we prove our main result.

2. Definitions and Notations

We consider the following model of random graphs: from the sample space of all $n(n-1)/2$ possible edges over n vertices, uniformly, independently, and with replacement select m edges to form the edges of the random graph G (thus an edge may be chosen more than once). Such models are common in phase transition analyses, moreover, [8] observe that results may be generalised to other random models. We denote by $G_{n,m}$ the sample space of n-vertex, m-edge graphs generated in this model, and subsequently G will denote a random graph chosen from this sample space.

3-COL is the decision problem that takes as input an n-vertex graph $G(V, E)$ with vertex set V and edges E. 3-COL$(G) = 1$ if each vertex of G can be assigned exactly one of three colours $\{r, b, g\}$ in such a way

*) The work reported in this paper was carried out with support from E.P.S.R.C. grant GR/L/77089

that no two adjacent vertices are assigned the same colour; such a colouring is called a *legal 3-colouring of* G. $\chi = c_1\, c_2 \cdots c_n \in \{ r, b, g \}^n$ denotes an arbitrary 3-colouring of n vertices. For $t \in \{ r, b, g \}$, χ_t denotes the set of vertices coloured t by χ. C_n is the set of all possible 3-colourings; and $good(G)$ the set of all legal 3-colourings of $G(V, E)$.

If $G \in G_{n,m}$, then $P[\, 3 - COL(G)\,]$ is the probability that a legal 3-colouring exists for G. The *expected* number of legal colourings for a random $G \in G_{n,m}$, is defined in the usual way by

$$E[\, |\, good(G) \,|\,] \;=\; \sum_{G \in G_{n,m}} P[\, G\,]\, |good(G)|$$

For $\rho \in \mathbf{R}^+$, we say that *almost all* $G \in G_{n,\,\rho n/2}$ are *non-3-colourable* if for $G \in G_{n,\rho n/2}$,

$$\lim_{n \to \infty} P[\, 3 - COL(G)\,] \;=\; 0.$$

Thus $\theta \in \mathbf{R}^+$ is an upper bound on the non-3-colourability threshold if: $\forall\, \rho > \theta$, almost all $G \in G_{n,\rho n/2}$ are non-3-colourable.

3. An elementary upper bound on the non-3-colourability threshold.

In this section we obtain the so-called 'first-moment' upper bound on the threshold and outline the method of [8] that will be developed to improve this.

Lemma 3.1: Let G be a random graph from $G_{n,m}$. Then

$$P[\, 3 - COL(G)\,] \;\le\; E[\, |\, good(G) \,|\,] \;=\; (\, 2/3\,)^m\, 3^n$$

Proof: For the first inequality simply observe that,

$$P[\, 3 - COL(G)\,] \;\le\; \sum_{\chi \in C_n} P[\, \chi \in good(G)\,]$$

$$= \sum_{G \in G_{n,m}} P[\, G\,]\, |good(G)| \;=\; E[\, |good(G)|\,]$$

For the second inequality, first observe that for a fixed colouring, χ, the probability that a randomly selected edge has both end-points coloured differently by χ is exactly 2/3. Thus,

$$E[\, |good(G)|\,] \;=\; \sum_{G \in G_{n,m}} P[G\,]\, |good(G)|$$

$$= \sum_{\chi \in C_n} P[\, \chi \in good(G)\,] \;=\; (\, 2/3\,)^m\, 3^n \qquad \square$$

Chvátal and Szemerédi [3] observe in the context of 3-SAT using a similar model of m clause random formulae over n propositional variables, that analyses in the style of Lemma 3.1, give

$$P[\, F \textit{ is satisfiable}\,] \;\le\; E[\, |\{\, \alpha : F(\, \alpha\,) = 1\, \}|\,] \;=\; (7/8)^m\, 2^n$$

Letting $m = \rho n$, an upper bound on the value θ such that a random F with $m/n > \theta$ is almost certainly unsatisfiable, is found by choosing ρ so that the expected number of satisfying assignments goes to 0, i.e. θ is the unique positive solution for ρ of $2(7/8)^\rho = 1$.

In the same way, from Lemma 3.1, an upper bound on the non-3-colourability threshold is given by the unique positive solution for ρ of the identity $3\,(2/3)^{\rho/2} = 1$, which is approximately 5.419. Thus if $\rho > 5.419$ then almost all graphs in $G_{n,\rho n/2}$ are not 3-colourable.

The key idea underlying the improved 3-SAT bound in [8] is rather than counting the number of satisfying assignments overall for F, the number of satisfying assignments of a particular form are counted. Letting $S^{\#}(F)$ denote the set of 'special' satisfying assignments for F, a 'Markov style' inequality of the form $P[\, \textit{Random F is satisfiable}\,] \le E[\, |S^{\#}(\, F\,)|\,]$ is demonstrated. A straightforward combinatorial argument then gives an upper estimate on $E[\, |S^{\#}(F)|\,]$. Choosing the clause to variable ratio so that this estimate goes to 0, yields an upper bound on the unsatisfiability threshold as before. The 'special' satisfying assignments for a formula F are defined in [8] by: $\alpha \in S^{\#}(\, F\,)$ if $F(\, \alpha\,) = 1$ and all assignments β obtained by changing a single 0 value in α to the value 1, have $F(\beta) = 0$.

In the next section we adapt this method to $3-COL$. While obtaining an expression to bound the expected number of special graph colourings is relatively straightforward, the analysis of this to derive the threshold bound turns out to be rather more involved than the corresponding stage of the 3-SAT analysis in [8].

4. Improving the non-3-colourability threshold.

We define the set of *special colourings* of G, denoted $good^{\#}(G)$, as follows.

Let χ be a 3-colouring. $\chi \in good^{\#}(G)$ if both (a) and (b) below hold:

a) $\chi \in good(G)$.

b) If μ is a colouring obtained from χ by changing the colour of a single vertex in χ_r, then $\mu \notin good(G)$.

Thus, if $\chi \in good^{\#}(G)$ then changing the colour of a single red vertex turns a legal 3-colouring into an illegal one.

Lemma 4.1:

a) G is 3-colourable if and only if $good^{\#}(G) \neq \emptyset$.

b) $P[\,3-COL(G)\,] \;\leq\; E[\,|good^{\#}(G)|\,]$.

c) $E[\,|good^{\#}(G)|\,] = (2/3)^m \sum_{\chi \in C_n} P[\,\chi \in good^{\#}(G) \mid \chi \in good(G)\,]$

Proof: a) Since $good^{\#}(G) \subseteq good(G)$, if $good^{\#}(G)$ is non-empty then G must have some legal 3-colouring. On the other hand, suppose that G has a legal 3-colouring, χ say. If $\chi \notin good^{\#}(G)$, then by definition, a new legal 3-colouring of G can be formed by changing the colour of some vertex in χ_r. Eventually one must obtain a colouring in $good^{\#}(G)$ since there are only finite number of vertices in χ_r.

b) This is similar to the proof of the first inequality in Lemma 3.1.

c)

$$
\begin{aligned}
E[\,|good^{\#}(G)|\,] &= \sum_{G \in G_{n,m}} P[G]\,|good^{\#}(G)| \\
&= \sum_{\chi \in C_n} P[\,\chi \in good^{\#}(G)\,] \\
&= \sum_{\chi \in C_n} P[\,\chi \in good^{\#}(G) \mid \chi \in good(G)\,]\,P[\,\chi \in good(G)\,] \\
&= (2/3)^m \sum_{\chi \in C_n} P[\,\chi \in good^{\#}(G) \mid \chi \in good(G)\,] \qquad \Box
\end{aligned}
$$

Theorem 4.1: Let $m = \theta n/2$, and G be chosen from $G_{n,m}$. We use $\Phi(n)$ to denote,

$$\Phi(n) =_{def} \sum_{\chi \in C_n} P[\,\chi \in good^{\#}(G) \mid \chi \in good(G)\,].$$

Then,

$$\Phi(n) \;\leq\; \sum_{k=0}^{n} \binom{n}{k} 2^k \left(1 - \exp\left(\frac{-\theta}{4-(3k/n)}\right) + O\left(\frac{1}{n}\right)\right)^{2(n-k)}$$

Proof: Fix a colouring χ, a vertex $v \in \chi_r$ and assume that $\chi \in good(G)$. The m selections for the edges of G must be drawn from a set of exactly

$$\binom{n}{2} - \left\{ \binom{|\chi_r|}{2} + \binom{|\chi_b|}{2} + \binom{|\chi_g|}{2} \right\}$$

edges.

In order to simplify the notation we denote this quantity by $e(n, \chi_b, \chi_g)$, noting that it is equal to

$$n(|\chi_b| + |\chi_g|) - (|\chi_b|^2 + |\chi_b||\chi_g| + |\chi_g|^2)$$

If, in addition, $\chi \in good^{\#}(G)$, then there must be at least one vertex $u \in \chi_b$ *and* at least one vertex $w \in \chi_g$ such that $\{ v, u \}$ and $\{ v, w \}$ are both edges selected for G. The probability of there being at least one such edge in both cases is

$$\left(1 - \left(1 - \frac{|\chi_b|}{e(n, \chi_b, \chi_g)} \right)^m \right) \left(1 - \left(1 - \frac{|\chi_g|}{e(n, \chi_b, \chi_g)} \right)^m \right)$$

Such edges must be present for each distinct $v \in \chi_r$. Thus, $P[\, \chi \in good^{\#}(G) \,|\, \chi \in good(G)\,]$ is bounded above by,

$$\left(\left(1 - \left(1 - \frac{|\chi_b|}{e(n, \chi_b, \chi_g)} \right)^m \right) \left(1 - \left(1 - \frac{|\chi_g|}{e(n, \chi_b, \chi_g)} \right)^m \right) \right)^{|\chi_r|} \tag{4.1}$$

Letting $k = |\chi_b| + |\chi_g|$ this probability is maximised when $|\chi_b| = |\chi_g| = k/2$. Substituting these in (4.1) and simplifying yields,

$$\left(1 - \left(1 - \frac{1}{2n - (3k/2)} \right)^m \right)^{2(n-k)} \tag{4.2}$$

Recalling that $m = \theta\, n/2$, and that

$$1 - \left(1 - \frac{1}{2n - (3k/2)} \right)^m \;\leq\; 1 - \exp\left(\frac{-m}{2n - (3k/2)} \right) + O\left(\frac{m}{(2n - (3k/2))^2} \right)$$

It follows that (4.2) is at most

$$\left(1 - \exp\left(\frac{-\theta}{4 - (3k/n)} \right) + O\left(\frac{1}{n} \right) \right)^{2(n-k)}$$

Now observing that there are $\binom{n}{k} 2^k$ ways of allocating the sets χ_b and χ_g so that $|\chi_r| = n - k$, we get our final upper bound

$$\Phi(n) \;\leq\; \sum_{k=0}^{n} \binom{n}{k} 2^k \left(1 - \exp\left(\frac{-\theta}{4 - (3k/n)} \right) + O\left(\frac{1}{n} \right) \right)^{2(n-k)} \tag{4.3} \quad \square$$

The remainder of the argument concentrates on deriving an upper bound for the summation in (4.3). Recall we have, so far, established that with $m = \theta n/2$, for a randomly chosen m-edge, n-vertex graph,

$$E\,[\, |good^{\#}(G)| \,] \;\leq\; (2/3)^{\theta n/2} \sum_{k=0}^{n} \binom{n}{k} 2^k \left(1 - \exp\left(\frac{-\theta}{4 - (3k/n)} \right) + O\left(\frac{1}{n} \right) \right)^{2(n-k)}$$

$$= \sum_{k=0}^{n} \exp\left(\log\left((2/3)^{\theta n/2}\, g(n, k, \theta) \right) \right)$$

(where $g(n, k, \theta)$ is the general summation term in (4.3))

$$= \sum_{k=0}^{n} \exp\left(n\, h(n, k, \theta) \right) \tag{4.4}$$

where,

$$h(n, k, \theta) \;=\; (\theta/2) \log(2/3) + (k/n) \log 2 \tag{4.5}$$

$$+ (1/n) \log \binom{n}{k} + 2\,(1 - (k/n))\, \log\left(1 - \exp(-\theta/(4 - (3k/n))) \right)$$

We complete the proof of our bound as follows. Observe that for any fixed n and θ the function $h(n,k,\theta)$ attains a maximal value for some k with $0 \le k \le n$. For a given value θ we isolate the region in which this maximum occurs. First we show for any given θ that we can choose positive values α and ε with $0 < \alpha + \varepsilon < 1$ such that, for n sufficiently large, $h(n,k,\theta) < h(n,k+1,\theta)$ whenever $0 \le k < \alpha n$ and $h(n,k,\theta) > h(n,k,\theta)$ whenever $(\alpha+\varepsilon)n \le k < n$, i.e. for a given n and θ the function $h(n,k,\theta)$ increases up to (at least) $k = \alpha n$ and decreases from (at most) $k = (\alpha+\varepsilon)n$, where α, ε depend only on θ (and n being sufficiently large). The 'critical' region of the summation in (4.4) to be considered is when $\alpha n \le k \le (\alpha+\varepsilon)n$. Thus, the final step of the proof simply involves finding α, ε, and θ for which $h(n,k,\theta) < 0$ for all $\alpha n \le k \le (\alpha+\varepsilon)n$. From the proceeding analyses this establishes that $\exp(n\, h(n,k,\theta)) < \exp(-\delta n)$ for some $\delta > 0$ which is sufficient to show that upper bound of (4.4) approaches 0 as n approaches infinity.

In order to keep the expressions involved simple, we employ the following notations:

$$\Delta(n,k,\theta) =_{def} h(n,k+1,\theta) - h(n,k,\theta)$$

$$\gamma(n,k,\theta) = 1 - \exp\left(\frac{-\theta}{4-(3k/n)}\right)$$

Lemma 4.2: $\forall\, \theta > 0, \exists\, \alpha, \varepsilon > 0$ such that $0 < \alpha + \varepsilon < 1$ and

$$\Delta(n,k,\theta) > 0 \quad \text{if } 0 \le k < \alpha n$$
$$\Delta(n,k,\theta) < 0 \quad \text{if } (\alpha+\varepsilon)n \le k < n$$

Proof: From (4.5) we have,

$$\Delta(n,k,\theta) = \frac{1}{n}\left(\log\left(\frac{2(n-k)}{k+1}\right) - 2\log\gamma(n,k+1,\theta)\right) + 2\left(1-\frac{k}{n}\right)\log\left(\frac{\gamma(n,k+1,\theta)}{\gamma(n,k,\theta)}\right)$$

Since, $\gamma(n,k,\theta) < \gamma(n,k+1,\theta) < 1\ \forall\, 0 \le k < n$, if we consider $0 \le k < \alpha n$ and $(\alpha+\varepsilon)n \le k < n$ then we have

$$\Delta(n,k,\theta) \ge \frac{1}{n}\left(\log\left(\frac{2(1-\alpha)}{\alpha+1/n}\right) - 2\log(\gamma(n,\alpha n+1,\theta)\right) \tag{4.6}$$

$$> 0$$

for some α depending on θ.

Similarly, for $(\alpha+\varepsilon)n \le k < n$

$$\Delta(n,k,\theta) \le \frac{1}{n}\left(\log\left(\frac{2(1-\alpha-\varepsilon)}{\alpha+\varepsilon+1/n}\right) - 2\log(\gamma(n,(\alpha+\varepsilon)n,\theta)\right)$$
$$+ 2(1-\alpha-\varepsilon)\log\left(\frac{\gamma(n,(\alpha+\varepsilon)n+1,\theta)}{\gamma(n,(\alpha+\varepsilon)n,\theta)}\right)$$

$$< 0$$

(for some $\varepsilon > 0$ depending on α and θ and n sufficiently large). □

Lemma 4.3: For any $\alpha > n/2$ and $\varepsilon > 0$ such that $\alpha + \varepsilon < 1$, if $\alpha n \le k \le (\alpha+\varepsilon)n$, then

$$h(n,k,\theta) \le (\theta/2)\log(2/3) + (\alpha+\varepsilon)\log 2 + \log(\alpha^{-1}(1/\alpha-1)^{\alpha-1})$$
$$+ 2(1-\alpha-\varepsilon)\log(1-\exp(-\theta/(4-3\alpha-3\varepsilon)))$$

Proof: Immediate by using Stirling's approximation in $\binom{n}{\alpha n}$ in the definition of $h(n,k,\theta)$ and simplifying. □

Finally we obtain our main result,

Theorem 4.2: $\forall\ \rho > 5.2057$, almost all $G \in G_{n,\,\rho n/2}$ are not 3-vertex colourable. □

Proof: Let $\alpha = 0.696139$, $\varepsilon = 0.00001$, and $\theta = 5.20566$. It is easily confirmed by direct calculation that for $n \ge 10^{20}$,

$\Delta(n, k, \theta) > 0 \quad \forall\ 0 \le k < \alpha n$

$\Delta(n, k, \theta) < 0 \quad \forall\ (\alpha + \varepsilon)n \le k < n$

$h(n, k, \theta) < 0 \quad \forall\ \alpha n \le k \le (\alpha + \varepsilon)n$

Combining these with Lemma 4.2, Lemma 4.3 and expression (4.4) we see that $\forall\ \rho > 5.20566$, for G chosen randomly with n-vertices and $\rho n/2$ edges,

$$E[\,|good^{\#}(G)|\,] \le \sum_{k=0}^{n} \exp(n\,h(n, k, \theta))$$

$$\le \sum_{k=0}^{n} \exp(-\delta n) \underset{n}{\rightarrow} 0 \qquad \textit{for some } \delta>0\ \square$$

5. Conclusions

In this paper a new upper bound on the non-3-colourability threshold of 5.2057 has been derived. This compares with the elementary ‘first-moment’ bound of 5.419. Our methods derive from the techniques discovered by Kirousis *et al.* [8] for bounding the unsatisfiability threshold in 3-SAT. We observe, in passing, that our approach may be extended in a natural way to derive corresponding thresholds for non-k-colourability. In [9], Kirousis *et al.* have developed the method of [8] to further reduce the unsatisfiability threshold, this however, is achieved at the expense of a considerably more complicated probabilistic analysis and it is not clear, whether an analogous argument can be developed for colouring thresholds.

6. References

[1] D.Angluin and L.G. Valiant. Fast probabilistic algorithms for Hamiltonian Circuits and Matchings. *Journal of Computer and System Sciences*, 18, (1979), 82-93

[2] M.T. Chao and J. Franco: Probabilistic analysis of a generalization of the unit-clause literal selection heuristics for the k-satisfiable problem. *Information Science*, 51, (1990), 289-314

[3] V. Chvátal and E. Szemerédi: Many hard examples for resolution; *Jnl. of the ACM*, 35, (1988), 759-768

[4] J. Crawford and L. Auton. Experimental results on the crossover point in satisfiability problems. *Proc. of AAAI-93*, (1993), 21-27

[5] A.M. Frieze and S. Suen: Analysis of simple heuristics for random instances of 3-SAT; Manuscript, 1993

[6] A. Goerdt. A threshold for unsatisfiability, *Proc. 17th Symposium on Mathematical Foundations of Computer Science*, (Eds: I. Havel, V. Krbec), (1992), Springer-Verlag LNCS Vol. 629, 264-274

[7] T. Hogg and C. Williams. The hardest constraint problems: a double phase transition. *Artificial Intelligence*, 69, (1994), 359-377

[8] L.M. Kirousis, E.Kranakis, and D. Krizanc. A better upper bound for the unsatisfiability threshold. *Proc. DIMACS Workshop Satisfiability Problem: Theory and Applications*, (March 1996)

[9] L.M. Kirousis, E.Kranakis, and D. Krizanc. Approximating the unsatisfiability threshold of random formulas. *Proc. 4th Annual European Symposium on Algorithms, ESA ‘96*, (September 1996)