A Sharp Threshold for k-Colorability

Dimitris Achlioptas[1,*] **Ehud Friedgut**[2]
[1]*Department of Computer Science, University of Toronto, Toronto, Ontario, Canada M5S3G4; e-mail: optas@cs.toronto.edu*
[2]*Institute of Mathematics, The Hebrew University of Jerusalem, Givat Ram, Jerusalem, Israel; e-mail: ehudf@math.huji.ac.il*

Received 19 September 1997; accepted 4 February 1998

ABSTRACT: Let k be a fixed integer and $f_k(n,p)$ denote the probability that the random graph $G(n,p)$ is k-colorable. We show that for $k \geq 3$, there exists $d_k(n)$ such that for any $\epsilon > 0$,

$$\lim_{n\to\infty} f_k\left(n, \frac{d_k(n)-\epsilon}{n}\right) = 1, \quad \text{and} \quad \lim_{n\to\infty} f_k\left(n, \frac{d_k(n)+\epsilon}{n}\right) = 0.$$

As a result we conclude that for sufficiently large n the chromatic number of $G(n, d/n)$ is concentrated in one value for all but a small fraction of $d > 1$. Random Struct. Alg., 14, 63–70, 1999

Key Words: random graphs; coloring; sharp thresholds

1. INTRODUCTION

Let $G(n,p)$ denote the random graph on n vertices where each edge appears independently with probability $p = p(n)$ [4]. We will say that $G(n,p)$ has a property A *almost surely* (a.s.) if the probability it has A tends to 1 as n tends to infinity.

Correspondence to: D. Achlioptas
* Supported by a PGS B Scholarship from the Natural Sciences and Engineering Research Council of Canada.

Let $\chi(G)$ denote the chromatic number of a graph G. Shamir and Spencer [11] proved that for every constant $\alpha > \frac{1}{2}$, if $p = O(n^{-\alpha})$ then the chromatic number of $G(n,p)$ is almost surely concentrated in some fixed number of values. That is, there exists a function $t = t(n,p)$ and a constant $s = s(\alpha)$, which is at most the smallest integer strictly larger than $(2\alpha + 1)/(2\alpha - 1)$, such that a.s. $t \le \chi(G(n,p)) \le t + s$. A further step in this direction was made by Łuczak [8] who showed that if $a > \frac{5}{6}$, then $s(\alpha) = 2$, i.e., $\chi(G(n,p))$ is a.s. *two point* concentrated. It is not difficult to see that the two point width of the concentration interval is best possible for general p. Finally, Alon and Krivelevich [1] proved two point concentration for $\alpha = \frac{1}{2} + \delta$, for any $\delta > 0$.

In a recent paper [5] the second author proved a necessary and sufficient condition for a monotone graph property to have a sharp threshold (for a definition see Section 2). In this paper, we use this condition to prove that k-colorability has a sharp threshold, for every constant $k \ge 3$. The main result we present is the following.

Theorem 1.1. *Let $f_k(n,p)$ denote the probability that the random graph $G(n,p)$ is k-colorable. For every constant $k \ge 3$, there exists $d_k(n)$ such that for any $\epsilon > 0$,*

$$\lim_{n\to\infty} f_k\left(n, \frac{d_k(n) - \epsilon}{n}\right) = 1, \quad \textit{and} \quad \lim_{n\to\infty} f_k\left(n, \frac{d_k(n) + \epsilon}{n}\right) = 0.$$

Note that the theorem does not hold for $k = 2$. To see this, consider $G(n, c/n)$, when $c \in (0,1)$. The probability that the random graph contains an odd cycle, and is thus non-2-colorable, is bounded away from 0 for any $c > 0$, but it is also bounded away from 1 for any $c < 1$. In the terminology we will introduce shortly, non-2-colorability has a coarse threshold.

Remark 1.2. We believe that for all $k \ge 3$, $d_k(n)$ converges. However, we are not able to prove this and it remains an interesting open problem.

Since for $d \ge 1$, a.s. $3 \le \chi(G(n, d/n)) = O(d)$ the existence of the limit of $d_k(n)$ would imply *one point* concentration of $\chi(G(n, d/n))$ for all, but a vanishing fraction of $d \ge 1$. Nonetheless, Theorem 1.1 implies the following:

Corollary 1.3. *Let $\epsilon > 0$ be some constant. For a given n, define a number $x \ge 1$ to be bad if the chromatic number of $G(n, x/n)$ is not determined with probability greater than $1 - \epsilon$, that is, $\forall k$, $\Pr[\chi(G(n, x/n)) = k] \le 1 - \epsilon$. For $y > 1$ define $X_\epsilon^y = \{x$: $1 \le x \le y$, and x is bad$\}$.*

For every $\epsilon > 0$, and every y,

$$\lim_{n\to\infty} m(X_\epsilon^y) = 0,$$

where m is the Lebesgue measure.

Note that the bad xs are exactly those for which x/n belongs to the threshold interval for the property of non-k-colorability for some k.

2. PRELIMINARIES

Let us introduce some terminology. A property is a nontrivial, proper subset of the set of all labeled graphs. We only consider properties that are invariant (closed) under graph automorphisms. We say that A is a monotone property if whenever a graph G has (is in) A and G is a subgraph of a graph H then H also has A. For a monotone graph property A let

$$\mu(p) = \mu(A, n, p) = \Pr[G(n,p) \text{ has property } A].$$

Note that $\mu(p)$ is a polynomial in p, monotone in [0, 1], such that $\mu(0) = 0$ and $\mu(1) = 1$. Hence, for fixed n and $0 < \tau < 1$ we can define p_0, p_c, p_1 by

$$\mu(p_0) = \tau,$$
$$\mu(p_c) = \tfrac{1}{2},$$
$$\mu(p_1) = 1 - \tau.$$

We call $[p_0, p_1]$ the *threshold interval* and define $\delta = \delta(\tau) = p_1 - p_0$. We say that a property A has a *sharp threshold* if for all $\tau > 0$ the ratio $\delta(\tau)/p_c$ tends to 0 as n tends to infinity. If for some $\tau > 0$, the ratio $\delta(\tau)/p_c$ is bounded away from 0 we say that A has a *coarse threshold*. Clearly, these two cases are not exhaustive as the ratio δ/p_c could "oscillate" with n.

In [5] the second author gives necessary and sufficient conditions for a graph property to have a coarse threshold. Roughly speaking, a graph property with a coarse threshold can be approximated by the property of having a subgraph isomorphic to a graph from a fixed list of graphs. Another necessary condition for a property to have a coarse threshold, aimed at helping prove "by contradiction" that the threshold is in fact sharp, is Theorem 2.1, which we use. Note that if A has a coarse threshold, i.e., there exists a constant C such that for all n, $\delta/p_c \geq 1/C$ then for every n there exists p^* within the threshold interval such that

$$p^* \cdot \frac{d\mu}{dp}\bigg|_{p=p^*} \leq C.$$

For given n, p, and a fixed graph H, let $\mathrm{Ex}(H,p) = \mathrm{Ex}(H,p,n)$ denote the expected number of copies of H in $G(n,p)$. Recall that a *balanced* graph is one for which the average degree is no smaller than that of any of its subgraphs.

Theorem 2.1 [5]. *Let $\delta > 0$. There exist functions $B(\epsilon, C)$, $b_1(\epsilon, C)$, $b_2(\epsilon, C)$ such that for all n, p^*, C, and ϵ, and any monotone graph property A such that $\delta < \mu(p^*) < 1 - \delta$ and $p^* \cdot d\mu/dp\,|_{p=p^*} \leq C$ here exists a graph H with no more than B edges such that*

- *H is balanced*
- $b_1 < \mathrm{Ex}(H, p^*) < b_2$
- *Let $\Pr[A \mid H]$ denote the probability that $G \in G(n, p^*)$ belongs to A conditional on the appearance of a specific copy of H in G. Then*

$$\Pr[A \mid H] > 1 - \epsilon.$$

Note that conditioning on the appearance of, say, a triangle in $G(n,p)$ is not the same as conditioning on the appearance of three *specific* edges $(i,j),(j,k),(k,i)$.

3. THE PROOF

The idea of the proof is as follows. For fixed k, assuming a not sharp threshold for non-k-colorability we have from Theorem 2.1 the existence of a graph H such that adding it to $G(n,p)$ changes the probability of k-colorability significantly. We compare the effect of such a graph to the effect of adding a large set of random edges to $G(n,p)$. We show that the random edges a.s. have no effect on the k-colorability of the graph, but their effect is comparable to the effect of H, thus reaching a contradiction.

Let us start by considering a slightly different model of random graphs. In this model we first choose each edge independently with probability p, exactly as in $G(n,p)$. Then we "add" M random edges, i.e., we pick uniformly at random M pairs of vertices, and for each one of them we add the corresponding edge if it does not already exist (otherwise we do nothing). For a monotone graph property A, let $\mu^{+}(p,M)=\mu^{+}(A,n,p,M)$ denote the probability that the resulting graph, $G(n,p,M)$, has property A and recall that $\mu(p)$ denotes the probability that $G(n,p)$ has A. The following observation is simple but very useful. Let $N=\binom{n}{2}$.

Lemma 3.1. *For any monotone graph property A, if $M=o(\sqrt{Np})$ then $|\mu(p)-\mu^{+}(p,M)|=o(1)$.*

Proof. Keeping in mind that the number of edges in $G(n,p)$ has binomial distribution with variance $\sigma^2=Np(1-p)$ it seems plausible that for $M=o(\sqrt{Np(1-p)})$, adding M random edges cannot make much of a difference, however a direct calculation of this is somewhat cumbersome. Instead, we find the property which is most sensitive to changes in the value of p and use it to prove our bound. Let A_k denote the number of graphs on n vertices that have property A and have precisely k edges. Then

$$\mu(p)=\sum A_k p^k(1-p)^{N-k}, \tag{1}$$

and hence

$$\frac{d\mu}{dp}=\frac{1}{p(1-p)}\sum A_k p^k(1-p)^{N-k}(k-Np). \tag{2}$$

Upon inspecting (2), we see that for any $p\in[0,1]$ every graph with $k<Np$ edges has negative contribution to $d\mu/dp$ at p while every graph with $k\geq Np$ has nonnegative. Thus, the monotone property with the greatest derivative at p is the property B of "having at least Np edges." For a monotone property A and an edge e, let $I_e(A,n,p)$ denote the probability of the following: after all other edges appear in the random graph on n vertices independently and with probability p, the appearance or nonappearance of e determines whether the graph has property A. Russo's formula (see [9, 10]) gives

$$\frac{d\mu}{dp}=\sum I_e(A,n,p). \tag{3}$$

By symmetry, $I_e(B,n,p)$ is identical for all edges e and equal to the probability that prior to determining the status of e, exactly $\lceil Np-1 \rceil$ edges have appeared. Since the number of edges in $G(n,p)$ has binomial distribution it follows that $I_e(B) = O(1/\sqrt{Np})$ and thus, by (3), that for any monotone property $d\mu/dp = O(\sqrt{N/p})$. Thus, if $|p-p'| = o(\sqrt{p/N})$ then $|\mu(p) - \mu(p')| = o(1)$.

Assume now that the lemma is false and, hence, for some choice of $M = o(\sqrt{Np})$ and $\epsilon > 0$, we have $|\mu(p) - \mu^+(p,M)| > \epsilon$, for all sufficiently large n. Let us set $p' = p + \lambda M/N$ and assume the dynamic point of view of the random graph process (i.e., $G(n,p) \subseteq G(n,p')$). For λ sufficiently large, but independent of n, $G(n,p')$ has at least M more edges than $G(n,p)$, with probability greater than $1-\epsilon/2$. On the other hand, for any constant λ, if $M = o(\sqrt{Np})$ then $|p-p'| = o(\sqrt{p/N})$ and hence,

$$|\mu(p) - \mu^+(p,M)| < \frac{\epsilon}{2} + |\mu(p) - \mu(p')| = \frac{\epsilon}{2} + o(1),$$

a contradiction. ■

The following lemma compares the effect of certain constraints placed on the coloring of a random graph with the effect of adding a set of random edges.

Lemma 3.2. *Let A be the property of non-k-colorability, $0 < \tau < 1$ a constant and p such that $\mu(p) \le 1 - \tau$. Assume that for a list of colors $c_1, \dots, c_M$, where each $c_i \in \{1, \dots, k\}$, the following is true: if we pick vertices $w_1, \dots, w_M$ uniformly at random, the probability that $G(n,p)$ has a k-coloring where each w_i receives a color other than c_i is not greater than $\tau/2$. Then for $M' = M2^{2^M}$, $\mu^+(p,M') \ge 1 - \tau/2$.*

Proof. The experiment we consider consists of first choosing $G(n,p)$ and then choosing, sequentially, the M vertices (constraints). Saying that, subject to the constraints, the probability of k-colorability is at most $\tau/2$, means that conditioning on $G(n,p)$ being k-colorable, the constraints spoil k-colorability with probability at least $\frac{1}{2}$. Let $\mathscr{E}_i$ denote the event that there is no legal k-coloring consistent with the first i constraints. Thus, conditioning on $G(n,p)$ being k-colorable, i.e., $\overline{\mathscr{E}_0}$,

$$\Pr[\mathscr{E}_{M-1}] + \Pr[\mathscr{E}_M | \overline{\mathscr{E}_{M-1}}] \Pr[\overline{\mathscr{E}_{M-1}}] \ge \tfrac{1}{2}. \tag{4}$$

Note that $\Pr[\mathscr{E}_M | \overline{\mathscr{E}_{M-1}}]$ is the expected value of the fraction of vertices that must receive color c_M after the first $M-1$ constraints are imposed. Moreover, an edge between any two such vertices spoils k-colorability. Hence, if instead of the last constraint we add an edge at random, the probability of non-k-colorability would be

$$\lambda = \Pr[\mathscr{E}_{M-1}] + \left(\Pr[\mathscr{E}_M | \overline{\mathscr{E}_{M-1}}]\right)^2 (1 - o(1)) \Pr[\overline{\mathscr{E}_{M-1}}].$$

Subject to (4), λ is minimized when $\Pr[\mathscr{E}_{M-1}] = 0$, and hence $\lambda \ge \frac{1}{4} - o(1)$. Clearly, adding the random edge *before* the $M-1$ constraints has the same effect. Thus, we can repeat the argument M times and since the probability of non-k-colorability is essentially squared in each iteration, we get that, conditioning on $\overline{\mathscr{E}_0}$, after adding M random edges the probability of non-k-colorability is at least $2^{-2^M} - o(1)$.

Hence, adding $M2^{2^M}$ edges yields noncolorability with probability at least $\frac{1}{2}$, i.e., for $M' = M2^{2^M}$, $\mu^+(p, M') \geq 1 - \tau/2$. ∎

Proof of Theorem 1.1. Let A be the property of non-k-colorability and assume, for the sake of contradiction, that A does not have a sharp threshold. That is, there exists a constant C and an infinite sequence of integers $\{n_i\}$ where for each of them there is $p^* = p^*(n_i)$ inside A's threshold interval such that

$$p^* \cdot \frac{d\mu}{dp}\bigg|_{p=p^*} \leq C. \tag{5}$$

In the following, asymptotics are meant with respect to $\{n_i\}$ yet, for the sake of simplicity, we retain the standard asymptotic parlance. Let $\mu(p^*) = 1 - \tau$, for some $\tau > 0$. If in Theorem 2.1 we take $\epsilon = \tau/3$, it follows that there exists a fixed graph H, such that $b_1 < \mathrm{Ex}(H, p^*) < b_2$, for some positive constants b_1, b_2 and such that

$$\Pr[A|H] > 1 - \frac{\tau}{3}. \tag{6}$$

That is, the appearance of a specific copy of H lowers the probability of $G(n, p^*)$ being k-colorable from τ to less than $\tau/3$, or stated differently, conditioning on $G(n, p)$ being k-colorable the additional copy of H spoils this with probability at least $\frac{2}{3}$.

The first thought that comes to mind is that this does not lead to a contradiction since H may be a non-k-colorable graph. However, from the characterization of H given in Theorem 2.1 we may assume that this is not the case. To see this, first note that by an application of the first moment method [3] it follows that $G(n, d/n)$ is a.s. non-k-colorable for $d > 2k \log k$, and hence $p^* = O(n^{-1})$. Moreover, observe that any non-k-colorable graph T must contain a subgraph S of minimum degree at least k. Since $p^* = O(n^{-1})$, H is balanced, and $\mathrm{Ex}(H, p^*) > b_1$, every subgraph of H must have average degree at most 2, so we may assume H is k-colorable (or even 3-colorable). (This point also reflects why non-2-colorability has a coarse threshold. For $k = 2$, we cannot assume that H is 2-colorable as, for example, K_3 has positive probability of appearing in the vicinity of p_c.)

Now, fix any k-coloring $V_1, \ldots, V_k$ of the vertices of H and consider the graph R formed by taking a copy of K_k and joining all the vertices in V_i to all the vertices of K_k other than i. Clearly, R contains H as a subgraph and is uniquely k-colorable. By positive correlation of increasing events the probability of k-colorability conditioned on the appearance of a specific copy of R is even smaller than the probability conditioned on the appearance of a specific copy of H. (The probability space of all graphs containing a fixed copy of some graph with the measure induced by the conditional probability is a product space in which the FKG inequalities hold, (see [6]). Thus, without loss of generality, we can condition on R appearing in the random graph instead of H. Finally, consider the random graph $G_R(n, p^*)$ on n vertices where we first add edges deterministically so that vertices $v_1, \ldots, v_r$ induce a copy of R and then we have every other edge appear with probability p^*.

Thus, by (6), the probability that $G_R(n, p^*)$ is non-k-colorable is greater than $1 - \delta/3$. We will show that this leads to a contradiction.

The subgraph induced by $v_{r+1}, \ldots, v_n$ is k-colorable with probability at least τ (we say at least since this subgraph has fewer than n vertices). Also, since r is a constant and $p^* = O(n^{-1})$, we have that a.s. there are no edges added between vertices $v_1, \ldots, v_r$. If ν is the number of edges between $v_1, \ldots, v_r$ and $v_{r+1}, \ldots, v_n$ then $\mathbf{E}[\nu]$ is bounded and hence there exists a constant M such that $\Pr(\nu \leq M) < \frac{1}{6}$. Moreover, it is straightforward to show that these edges a.s. do not have any endpoints in common besides $v_1, \ldots, v_r$. From this it follows that in $G(n - r, p^*)$ if we choose at random M vertices $w_1, \ldots, w_M$ from $v_{r+1}, \ldots, v_n$ and for each one of them forbid a color (that of its unique neighbor in $v_1, \ldots, v_r$), this decreases the probability of having a k-coloring consistent with these constraints from at least τ to no more than $\tau/2$. By Lemma 3.2, adding $m2^{2^M} = O(1)$ random edges instead, would reduce the probability by the same amount. Since

$$p^* \binom{n-r}{2} = \Omega(n)$$

this contradicts Lemma 3.1. ∎

3.1. Concluding Remarks

When $k = k(n) \to \infty$ as $n \to \infty$, Noga Alon and Michael Krivelevich pointed out to us that, while a sharp threshold can be shown using the results in [7], our approach yields a very short proof of this fact. Let us construct the random graph containing a specific copy of H by first including every edge with probability p^* and then adding a copy of H on a randomly chosen set of vertices. It is well-known that if $k \to \infty$ as $n \to \infty$, and $p = p(n)$ is such that $\Pr[\chi(G(n,p)) \geq k] \geq \epsilon$ then a.s. the largest independent set of $G(n,p)$ has size $o(n)$. Thus, a.s. every k-coloring of $G(n, p^*)$ has color classes of size $o(n)$. As a result, when we add H a.s. all its vertices end up in distinct color classes implying that the addition of H can only decrease the probability of k-colorability by $o(1)$.

The immediate open problem is determining whether $d_k(n)$ converges. Equivalently, one could also consider the existence of $\lim_{n\to\infty} f_k(n, d/n)$ when $d \neq \sup\{d \mid \lim_{n\to\infty} f_k(n, d/n) = 1\}$. Much more ambitiously, one could also try to determine the value of $\lim_{n\to\infty} d_k(n)$. Also of interest is to try and apply these techniques to the analogous question for choosability.

ACKNOWLEDGMENTS

We thank Noga Alon, Michael Krivelevich, Michael Molloy, and the three anonymous referees for numerous helpful suggestions.

REFERENCES

1. N. Alon and M. Krivelevich, On the concentration of the chromatic number of random graphs, Combinatorica 17 (1997), 303–313.
2. B. Bollobás, The chromatic number of random graphs, Combinatorica 8 (1988), 49–55.
3. V. Chvátal, Almost all graphs with 1.44n edges are 3-colorable, Random Structures Algorithms 2 (1991), 11–28.
4. P. Erdös and A. Rényi, On the evolution of random graphs, Publ Math Inst Hungarian Acad Sci 5 (1960), 17–61.
5. E. Friedgut, Sharp thresholds for graph properties and the k-sat problem, J Amer Math Soc, to appear.
6. C.M. Fortuin, P.W. Casteleyn, and J. Ginibre, Correlation inequalities on some partially ordered sets, Commun Math Phys 22 (1971), 89–103.
7. T. Łuczak, The chromatic number of random graphs, Combinatorica 11 (1991), 45–54.
8. T. Łuczak, A note on the sharp concentration of the chromatic number of random graphs, Combinatorica 11 (1991), 295–297.
9. G.A. Margulis, Probabilistic characteristics of graphs with large connectivity, Problemy Pederachi Informatsii 10 (1974), 101–108.
10. L. Russo, On the critical percolation probabilities, Z Wahrsch Verw Gebiete 56 (1981), 129–139.
11. E. Shamir and J.H. Spencer, Sharp concentration of the chromatic number of random graphs $G_{n,p}$, Combinatorica 7 (1987), 121–129.