Thresholds for colourability and satisfiability in random graphs and boolean formulae.

Michael Molloy

Abstract

We survey the progress on two fundamental problems in random graphs and random boolean formulae: How many edges must be added to a random graph until it is not a.s. k-colourable? How many clauses must be added to a random instance of k-SAT until it is not a.s. satisfiable?

1 Introduction

All the graphs considered in this paper are vertex-labelled and finite. Unless specified otherwise, they are always simple and undirected.

The *random graph process* runs as follows. Start with a collection of n vertices and no edges. For $i = 1$ to $\binom{n}{2}$ select an edge uniformly at random from amongst those edges which are not yet present and add it to the graph. (A random choice is *uniform* if all possible choices are equally likely.) Thus, our graph evolves in a random manner from a graph with no edges, to a "typical" graph with a few edges, and so on until we reach the complete graph. We use $G_{n,M}$ to denote a graph obtained by halting the random graph procedure on n vertices after step M. Equivalently, $G_{n,M}$ is a graph selected uniformly at random from amongst all graphs with n vertices and M edges.

We are are usually concerned with $G_{n,M}$ for large values of n and where M is a function of n. So most of our computations are asymptotic, thus allowing for a certain amount of approximation. For any graph property P, we say that $G_{n,M}$ *almost surely (a.s.)* has P if

$$\lim_{n\to\infty} \mathbf{Pr}(G_{n,M} \text{ has } P) = 1.$$

Another commonly studied random graph model is $G_{n,p}$, the graph on n vertices formed by making $\binom{n}{2}$ independent coin flips, one for each potential edge, each time adding the edge to the graph with probability p. These two models are in many respects equivalent. Roughly speaking, for most properties P, if $M \approx \binom{n}{2}p$ then $G_{n,M}$ a.s has P iff $G_{n,p}$ a.s. has P. Section 1.4 of [46] contains a more precise statement of this equivalence, but for now, suffice it to say that for all properties P considered in this paper, and for any constant $c > 0$,

$$\lim_{n\to\infty} \mathbf{Pr}(G_{n,M=cn} \text{ has } P) = \lim_{n\to\infty} \mathbf{Pr}(G_{n,p=2c/n} \text{ has } P).$$

This fact is often very convenient. For example, in this paper we are primarily concerned about the model $G_{n,M}$, but sometimes we will prove theorems about it indirectly by proving the corresponding theorems for $G_{n,p}$.

The field of random graph theory was founded by Erdős and Rényi in two seminal papers [33, 34]. In these papers, besides doing a very thorough analysis of many basic properties of a random graph, they asked the following questions, which were obviously inspired by what many people would consider to be the most natural questions to ask about graphs.

Under the random graph process,

A. How long does it typically take until G is connected?

B. For any fixed graph H, how long does it typically take until G contains H as a subgraph?

C. How long does it typically take until G is non-planar?

D. How long does it typically take until G has a perfect matching?

E. For any constant k, how long does it typically take until G has a K_k-subdivision?

F. How long does it typically take until G has a Hamilton path?

G. What is the distribution of the chromatic number of $G_{n,M}$? Specifically, what can be said when $M = cn$ for $c > \frac{1}{2}$?

Erdős and Rényi answered questions A, B and C in these two papers (see [56] for more on question C), but left the others open.

Question D was answered in 1966[35]. Question E was settled in 1979[8] (see also [57]). Question F was answered in 1983[52] (see also [18, 21]). Also, many other basic graph properties became well understood for the random graph, such as the clique number and independence number [20, 59]. But Question G remained elusive.

In the late 80's, Bollobás[16] and independently, Matula and Kucera[61], determined the asymptotics of $\chi(G_{n,M})$ when $M = cn^2$, for any constant $c > 0$, and Bollobás proof extended to $M > n^{2-\delta}$ for any $\delta < 1/3$. Later Łuczak[55] determined the asymptotics of $\chi(G_{n,M})$ when M/n is at least a large constant (see section 6 of this paper). To this date, though, we only have a very limited understanding of $\chi(G_{n,M=cn})$ when $c > \frac{1}{2}$ is a small constant.

Focussing on this elusive range of M, Erdős often asked the following refinement of question G (see, for example, his section in [13]).

Question 1.1 *(a) For each $k \geq 3$, is there a constant c_k^* such that for every $\epsilon > 0$, a.s. $\chi(G_{n,M=(c_k^*-\epsilon)n}) \leq k$ and a.s. $\chi(G_{n,M=(c_k^*+\epsilon)n}) > k$?*

(b) If these constants exist, determine them.

Most of the work towards this problem has been focussed on the search for c_3^*, as this is the first difficult case. Thus far, no one has proven that c_3^*

exists, although as we will see in Section 5, Achlioptas and Friedgut[4] came very close.

At first glance, one might think that the reason Question 1.1 is so difficult is that 3-colourability is NP-hard. However, focussing on computational complexity here can be misleading. Since 3-colourability is NP-hard, there is (probably) no simple characterisation of 3-colourable graphs. However, we cannot rule out the possibility that there is a simple characterisation of 3-colourable graphs which fails for only a very small number of graphs; i.e. a simple characterization which is a.s. valid on a random graph.

For example, Hamiltonicity is NP-complete. But it turns out that in the random graph process, a.s. the graph will get a Hamilton cycle as soon as it has minimum degree at least 2. I.e. "minimum degree at least 2" is a (surprisingly) simple characterization of Hamiltonian graphs which is a.s. valid. Analysing local properties of random graphs, such as vertex degrees, tends to be much easier than analysing global properties, and so this simple characterization is a tremendous help for determining the threshold for Hamiltonicity. (It turns out that a.s. the graph becomes Hamiltonian at around $M = \frac{1}{2}n \ln n + \frac{1}{2}n \ln\ln n + \Theta(n)$ [52].)[1]

Similarly, for n even, a.s. the graph has a perfect matching as soon as it has minimum degree at least 1. This a.s. valid characterization is much more helpful in answering Question D than Edmonds' polytime algorithm for finding a perfect matching[31] (and its underlying characterization which is valid for all graphs). Also, the graph is a.s. connected as soon as it has minimum degree at least 1, which is more helpful than the fact that we can test for connectivity very easily, eg. with a breadth-first search. (It turns out that a.s. the graph becomes connected and gets a perfect matching at around $M = \frac{1}{2}n \ln n + \Theta(n)$ [33, 35].)

So, despite the NP-completeness of 3-colourability, it is quite feasible that there is a fairly simple property which is a.s. equivalent to 3-colourability and that this will lead to an answer for Question 1.1. See Questions 4.1 and 7.1 for two approaches along these lines.

The fact that we don't know whether c_k^* exists, makes it awkward to discuss it formally. So instead, we define

$$\begin{aligned} c_k^- &= \sup\{c : \text{ a.s. } G_{n,M=cn} \text{ is } k\text{-colourable}\} \\ c_k^+ &= \inf\{c : \text{ a.s. } G_{n,M=cn} \text{ is not } k\text{-colourable}\}. \end{aligned}$$

Most researchers in this field think that for $k \geq 3$, c_k^* exists. If it does, then we would have $c_k^- = c_k^+ = c_k^*$. So it is common practice to define

$$c_k = c_k^-,$$

and to discuss c_k, which certainly does exist, instead of discussing c_k^*.

[1] $\Theta(z)$ is used to denote a function t of the same asymptotic order as z.

Besides random graphs, we will also be discussing random boolean formulae. The most commonly studied such model is a random instance of k-SAT. An instance of k-SAT is a boolean formula in Conjunctive Normal Form (CNF), in which every clause has size k. For the reader who is unfamiliar with these terms, we provide the following definitions:

We have several boolean variables, say $v_1, ..., v_n$. A *literal* is a variable v_i along with a sign, and is represented as "v_i" (positive) or "$\bar{v}_i$" (negative). Thus, there are $2n$ possible literals. In CNF, a *clause* is a collection of literals, no two corresponding to the same variable. They are joined by "OR", so the clause is satisfied iff at least one of the literals is true. The boolean formula is a collection of clauses, all joined by "AND", so the formula is satisfied iff every clause is satisfied. A boolean formula is *satisfiable* iff at least one of the 2^n possible truth assignments to its variables causes the formula to be satisfied.

Just as with random graphs, we define the *random k-SAT process.* We start with a collection of n variables and no clauses. For $i = 1$ to $2^k\binom{n}{k}$ we select a clause uniformly at random from amongst those clauses which are not yet present and add it to the formula. $F^k_{n,M}$ is the formula after M iterations, or equivalently, it is a formula chosen uniformly at random from amongst all instances of k-SAT with n variables and M clauses.

Alternatively, we define $F^k_{n,p}$ to be the random formula on n variables, where each of the $2^k\binom{n}{k}$ possible k-clauses is present independently with probability p. As was the case with random graphs, these two models are very similar. For any property P considered in this paper, and for any constant $d > 0$,

$$\lim_{n\to\infty} \mathbf{Pr}(F^k_{n,M=dn} \text{ has } P) = \lim_{n\to\infty} \mathbf{Pr}(F^k_{n,p=d\times k!/(2^k n^{k-1})} \text{ has } P).$$

(Note that the expected number of clauses in $F^k_{n,p=d\times k!/(2^k n^{k-1})}$ is $dn + o(n)$.)

There are many natural questions to ask about random graphs, but with random boolean formula, there is one obvious question which dwarfs all the other ones in importance: What is the probability that the formula is satisfiable? This question is usually asked in the following form:

Question 1.2 *(a) For each $k \geq 3$, is there a constant d^*_k such that for every $\epsilon > 0$, a.s. $F^k_{n,M=(d^*_k-\epsilon)n}$ is satisfiable and a.s. $F^k_{n,M=(d^*_k-\epsilon)n}$ is not satisfiable?*

(b) If these constants exist, determine them.

Again, we do not even know whether these constants exist, although Friedgut[39] has shown something very close (see section 5). So we define

$$d_k = d^-_k = \sup\{d : \text{ a.s. } F^k_{n,M=dn} \text{ is satisfiable}\}.$$

We define d^+_k in the obvious way, analagous to the definition of c^+_k. Most researchers in this field think that for $k \geq 3$, d^*_k exists. If it does, then it must be equal to d_k (and to d^+_k). So we discuss d_k instead of d^*_k.

Question 1.1 has received a great deal of attention from researchers in random graph theory. Question 1.2 has received a phenomenal amount of attention from researchers in a wide range of fields including probabilistic combinatorics, theoretical computer science, artificial intelligence and statistical physics. The vast majority of this focus has been on finding d_3^*. Sophisticated simulations and some very sophisticated but non-rigorous analysis using the so-called "replica method" has produced strong evidence that

Conjecture 1.3 $d_3^* \approx 4.2$.

(Of course, Conjecture 1.3 implicitly assumes that d_3^* exists.)

Questions 1.1 and 1.2 behave like twin problems. For example, most new techniques that help with one problem turn out to help with the other one. By now, most people who work on Question 1.1 will, at least occasionally, work on Question 1.2. However, the other direction is far from true - probably most people in, for example, statistical physics who work on Question 1.2 are not even aware of Question 1.1. This is unfortunate, since I think that a good understanding of both problems is valuable when attacking either one of them.

The purpose of this paper is to present an up-to-date survey of the status of these two questions, placing them side-by-side in order to emphasize their similarities and how certain techniques have proven fruitful for both.

For random graphs, the focus of this paper will be on $G_{n,M}$ for the range $M = O(n)$. Thus, we omit virtually all results on the chromatic number of $G_{n,M}$ for much higher values of M (which includes some beautiful work).

Throughout this paper, we use $\mathbf{Exp}(X)$ to denote the expected value of X. For an event A, we use $\overline{A}$ to denote the complement of A, i.e. the event that A does not occur. We use $\Pr(A|B)$ and $\mathbf{Exp}(X|B)$ to denote the probability of A and the expected value of X, both conditioned on the event B occuring. To avoid clutter, we adopt the convention of omitting floor and ceiling signs. Such instances can be formalized by inserting those signs where appropriate, along with the occasional negligible corrective term.

1.1 A few tools

It is a common misperception amongst combinatorialists that if one wishes to work in random graph theory then it is necessary to first become an expert in probability. This is not at all true. An understanding of the basics, such as expected value and variance, and familiarity with a few concentration inequalities will get you well on your way. (If you wish to work with "the probabilistic method" then you also need to learn the Lovász Local Lemma.)

The first tool that one needs is:

Linearity of Expectation:

$$\mathbf{Exp}(X_1 + ... + X_t) = \mathbf{Exp}(X_1) + ... + \mathbf{Exp}(X_t).$$

Proof For any outcome ω of a random experiment, we denote by $X_i(\omega)$ the corresponding value of X_i. Recall that the expected value of X_i is $\sum_\omega \mathbf{Pr}(\omega) \times X_i(\omega)$. Linearity of Expectation follows immediately from this as

$$\sum_\omega \mathbf{Pr}(\omega) \times (X_1(\omega) + \ldots + X_t(\omega)) = \sum_{i=1}^{t} \left(\sum_\omega \mathbf{Pr}(\omega) \times X_i(\omega) \right).$$

□

We give an explicit illustration of the use of Linearity of Expectation in the proof of Theorem 3.1. Our next tool is:

Markov's Inequality *For any non-negative random variable X, and any $t > 0$,*

$$\mathbf{Pr}(X \geq t) \leq \frac{\mathbf{Exp}(X)}{t}.$$

Proof (for discrete variables) Recall that $\mathbf{Exp}(X)$ can also be expressed as $\mathbf{Exp}(X) = \sum_i i \times \mathbf{Pr}(X = i)$. Therefore, since X is always positive, $\mathbf{Exp}(X) \geq \sum_{i \geq t} i \times \mathbf{Pr}(X = i) \geq t \times \mathbf{Pr}(X \geq t)$. □

The following fact, which follows immediately from setting $t = 1$ in Markov's Inequality, is very useful.

Corollary 1.4 *If* $\mathbf{Exp}(X) = o(1)$ *then a.s.* $X = 0$.

It is often tempting to think that when $\mathbf{Exp}(X)$ is very large, then a.s. X is large. This is not always true, as the following example indicates.

Consider a lottery with a jackpot of 4^n pounds. The probability of winning the jackpot is 2^{-n}, and there are no other prizes. Your expected winnings are 2^n, but with very high probability, your winnings will be 0.

I refer to this type of situation as a *jackpot phenomenon*. For a more interesting illustration, recall that a.s. the random graph process becomes Hamiltonian when $M \approx \frac{1}{2} n \ln n + \frac{1}{2} \ln \ln n$. By the equivalency of our two models, this implies that a.s. $G_{n,p=\ln n/n}$ is non-Hamiltonian. However, if the reader computes the expected number of Hamilton cycles in $G_{n,p=10/n}$, she will discover a surprising jackpot phenomenon. This will also provide a good exercise in applying the Linearity of Expectation: Let X be the number of Hamilton cycles. Consider the $\frac{1}{2}(n-1)!$ undirected cyclic arrangements of the vertices, $\sigma_1, \ldots, \sigma_{\frac{1}{2}(n-1)!}$ and for each $1 \leq i \leq \frac{1}{2}(n-1)!$ let $X_i = 1$ if σ_i is a Hamilton cycle and $X_i = 0$ otherwise. (X_i is an example of an *indicator variable*.) Note that $X = \sum_{i=1}^{\frac{1}{2}(n-1)!} X_i$ and apply Linearity of Expectation to compute $\mathbf{Exp}(X)$.

In light of these *jackpot phenomena*, it is often useful to prove that with high probability X is very close to $\mathbf{Exp}(X)$. If this is true, we say that X is highly *concentrated*.

Simple Concentration Bound *Let X be a random variable determined by n independent trials $T_1, ..., T_n$, and satisfying*

$$\text{changing the outcome of any one trial can affect } X \text{ by at most } c, \tag{1.1}$$

then

$$\mathbf{Pr}(|X - \mathbf{Exp}(X)| \geq t) \leq 2\mathrm{e}^{-\frac{t^2}{2c^2 n}}.$$

(1.1) rules out the possibility that a single trial can determine a huge jackpot, along with some more subtle variants of this problem. The constant "2" in the exponent can be improved somewhat (it can be moved from the denominator to the numerator), but in this field, the values of the constants in such bounds are rarely important. As stated, the Simple Concentration Bound is a special case of Azuma's Inequality and (with a slight weakening of some constants) of Talagrand's Inequality, two of the most powerful concentration bounds commonly used in combinatorics See [58] for a good survey.

Finally, the following bound, which can be easily verified using Stirling's formula, will often be useful. For $a \geq b \geq 1$,

$$\binom{a}{b} < \left(\frac{ea}{b}\right)^b.$$

2 k=2

The reason that Questions 1.1 and 1.2 both specify $k \geq 3$ is that the case $k = 2$ is very well understood, and the case $k = 1$ is trivial.

Obviously, a random graph with even one edge is not 1-colourable. An instance of 1-SAT is satisfiable iff there are no two conflicting 1-clauses, i.e. if there is no variable v for which (v) and $(\bar{v})$ are both clauses. It is not hard to show that the first pair of conflicting 1-clauses will occur at around $M = \Theta(n^{1/2})$ (this is essentially the birthday problem).

For 2-colourability, Erdős and Rényi[34] proved the following:

Theorem 2.1 *(a) For $0 < c < \frac{1}{2}$, $\mathbf{Pr}(\chi(G_{n,M=cn}) = 2) = \zeta + o(1)$ for some constant $\zeta = \zeta(c)$ with $0 < \zeta < 1$.*

(b) For $c > \frac{1}{2}$, a.s. $\mathbf{Pr}(\chi(G_{n,M=cn}) > 2)$.

A graph is 2-colourable iff it has no odd cycles. For $0 < c < \frac{1}{2}$, the expected number of odd cycles in $G_{n,M=cn}$ is $C + o(1)$ for a constant $C = C(c) > 0$. Furthermore, it turns out that the probability there are no odd cycles is $\zeta = e^{-C} + o(1)$. To see part (b) of the theorem, we must look at what is the most fundamental result in random graph theory, also from [34]:

Theorem 2.2 *(a) For $c < \frac{1}{2}$, a.s. $G_{n,M=cn}$ has the following structure: (i) the largest component has size $O(\ln n)$; and (ii) no component has more than one cycle, and all but $O(1)$ components are trees.*

(b) *For $c > \frac{1}{2}$, there are constants $b(c) > a(c) > 0$ such that a.s. $G_{n,M=cn}$ has a component with $a(c)n + o(n)$ vertices and $b(c)n + o(n)$ edges (hence it has many cycles). This is referred to as the giant component. All other components are unicyclic and have size at most $O(\ln n)$.*

It is not hard to show that once the giant component appears, a.s. at least one of its many cycles will be odd.

The case for 2-SAT is similar, the main difference being that the change is much sharper. Independently, Chvátal and Reed[27], Fernandez de la Vega[36] and Goerdt[43] each proved:

Theorem 2.3 *(a) For $c < 1$, a.s. $F^2_{n,M=cn}$ is satisfiable.*

(b) For $c > 1$, a.s. $F^2_{n,M=cn}$ is not satisfiable.

To see why this should be true, take any instance of 2-SAT, F, and form a graph $G(F)$ whose vertices are the $2n$ literals of F, and whose edges are the pairs of literals which form 2-clauses in F. For $d < 1$, $G(F^2_{n,M=dn})$ will a.s. not have a giant component and its small simple components will be very easy to satisfy. We can form a digraph $D(F)$ on the same vertex set by representing the clause $(\ell_1 \vee \ell_2)$ by the two directed edges $\bar{\ell}_1 \to \ell_2$ and $\bar{\ell}_2 \to \ell_1$. The significance of $D(F)$ is that $(\ell_1 \vee \ell_2)$ is equivalent to the statements "$\bar{\ell}_1$ implies ℓ_2" and "$\bar{\ell}_2$ implies ℓ_1". For $d > 1$, one can show that $D(F^2_{n,M=dn})$ will a.s. have a directed cycle containing both x and $\bar{x}$ for some variable x. Clearly, this means that the formula is unsatisfiable.

A great deal of work has gone into studying $G_{n,M}$ when $M = \frac{1}{2}n + o(n)$, culminating in [45]. $F^2_{n,M=cn}$ when $M = n + o(n)$ is much more complicated. Some recent pioneering work can be found in [19].

3 Upper Bounds

3.1 k-colourability

The first upper bound on the threshold for k-colourability, was obtained by Erdős[32] (in a somewhat different form) by focussing on large stable sets in the random graph.

Theorem 3.1 $c_k \leq k \ln k + k$

Proof We will work in the model $G_{n,p}$.

Since each colour class is a stable set, it is easy to see that if G is k-colourable, then G has an stable set of size at least n/k.

So we let X denote the number of stable sets of size n/k in $G_{n,p=2c/n}$, and compute $\mathbf{Exp}(X)$. Enumerate the subsets of n/k vertices as $I_1, .., I_{\binom{n}{n/k}}$.

Define the random variable X_i to be equal to 1 if I_i is an stable set, and 0 otherwise. $X = \sum X_i$, and $\mathbf{Exp}(X_i) = \mathbf{Pr}(X_i = 1) = (1-p)^{\binom{n/k}{2}}$. Therefore,

$$\begin{aligned}
\mathbf{Exp}(X) &= \sum \mathbf{Exp}(X_i) \\
&= \binom{n}{n/k}(1-p)^{\binom{n/k}{2}} \\
&< \left(\frac{\mathrm{e}n}{n/k}\right)^{n/k} \mathrm{e}^{-p\times\frac{1}{2}(n/k)((n/k)-1)} \\
&= \exp((n/k)\times(\ln(\mathrm{e}k) - (c/k) - O(n^{-1}))) \\
&= o(1),
\end{aligned}$$

when $c > k\ln k + k$. Thus, by Corollary 1.4, a.s. $X = 0$ and so there are no stable sets of size n/k. □

Remark In the future, we won't be so explicit in our use of Corollary 1.4 nor in our use of indicator variables when applying the Linearity of Expectation.

For example, Theorem 3.1 yields $c_3 \leq 6.29...$, $c_4 \leq 9.54...$, and $c_5 \leq 13.04...$. By being a little more careful about the inequality in the calculations above, we get a slight improvement to $c_3 \leq 5.72...$, $c_4 \leq 8.99...$, $c_5 \leq 12.51...$. For large values of k, this bound turns out to be asymptotically optimal (see section 6), but for small values of k, eg. $k = 3$, it is rather poor, as we shall see.

Luc Devroye (see [26]) showed that we can obtain much better bounds on c_k by counting the expected number of k-colourings directly, rather than focussing on large stable sets:

Theorem 3.2 $c_k \leq \ln k / \ln(k/(k-1))$.

This reduces the bound of Theorem 3.1 by roughly $2k$, yielding, for example, $c_3 \leq 2.71...$, $c_4 \leq 4.81...$, $c_5 \leq 7.21...$.

Proof Let X be the number of k-colourings of $G_{n,M=cn}$. There are k^n partitions of the n vertices into k colours; consider any one of them, with parts $A_1, ..., A_k$. This partition is a valid k-colouring iff the graph chooses no edges from amongst the $\sum \binom{|A_i|}{2} \geq k \times \binom{n/k}{2}$ potential edges between two vertices in the same part. This happens with probability at most:

$$\binom{\binom{n}{2} - k\times\binom{n/k}{2}}{cn} \Big/ \binom{\binom{n}{2}}{cn} = \left(\frac{k-1}{k}\right)^{cn+o(n)}.$$

Therefore,

$$\mathbf{Exp}(X) < k^n \left(\frac{k-1}{k}\right)^{cn+o(n)} = \exp(n(\ln k - c\ln(k/(k-1)) + o(1))) = o(1),$$

for $c > \ln k / \ln(k/(k-1))$. □

The inequalities in the above calculation are fairly tight, and so with more care, one can establish that for $c > \ln k/\ln(k/(k-1))$, the expected number of k-colourings in $G_{n,M=cn}$ is exponentially high. It is tempting to guess that this implies there will a.s. be at least one k-colouring, and so the threshold c_k is equal to $\ln k/\ln(k/(k-1))$. However, things are not that simple; a fairly easy calculation shows that, in fact, $c_k < \ln k/\ln(k/(k-1))$; i.e. there are values of c for which the expected number of k-colourings is exponentially high, but a.s. there are no k-colourings. The reason for this apparent paradox is that the number of k-colourings exhibit a jackpot phenomena - if there is at least one k-colouring, then there will probably be an exponentially high number of k-colourings, and this causes the expected number to be deceptively high.

Theorem 3.3 $c_k < \ln k/\ln(k/(k-1))$.

Proof The equivalence of our two random graph models will be helpful here. Choose a random graph $G = G_{n,p=2c/n}$. Let Z be the number of vertices of degree 0 in G. $\mathbf{Exp}(Z) = n(1-p)^{n-1} \approx e^{-2c}n$. Define $\zeta = \frac{1}{2}e^{-2\ln k/\ln(k/(k-1))} < \frac{1}{2}e^{-2c}$ and let A be the event that $Z \geq \zeta n$. A straightforward application of the Simple Concentration Bound shows that Z is highly concentrated and so $\mathbf{Pr}(A) = 1 - o(1)$. Now consider some $G' = G_{n,M=cn}$, let Z' be the number of degree 0 vertices in G' and let A' be the event that $Z' \geq \zeta n$. By the equivalency of our models, $\mathbf{Pr}(A') = 1 - o(1) > \frac{1}{2}$ for n sufficiently large.

Let X be the number of k-colourings of G'. If $X > 0$ then $X \geq k^{Z'}$, since we are free to colour the Z' degree 0 variables any way we choose. Therefore, for n sufficiently large,

$$\begin{aligned}
\mathbf{Exp}(X) &\geq \mathbf{Pr}(A') \times \mathbf{Exp}(X|A') \\
&> \frac{1}{2}\mathbf{Exp}(X|A') \\
&= \frac{1}{2}[\mathbf{Pr}(X = 0|A') \times 0 + \mathbf{Pr}(X > 0|A') \times \mathbf{Exp}(X|X > 0 \wedge A')] \\
&\geq \frac{1}{2}\mathbf{Pr}(X > 0|A') \times k^{\zeta n}.
\end{aligned}$$

Choose $c < \ln k/\ln(k/(k-1))$ such that $\mathbf{Exp}(X) \leq \frac{1}{2}k^{\frac{1}{2}\zeta n}$, which we can do since $\mathbf{Exp}(X) = (t(c) + o(1))^n$ where $t(c)$ is continuous in c. Then, by the above inequalities, $\mathbf{Pr}(X > 0|A') \leq k^{-\frac{1}{2}\zeta n}$. Therefore,

$$\mathbf{Pr}(X > 0) \leq \mathbf{Pr}(A') \times k^{-\frac{1}{2}\zeta n} + \mathbf{Pr}(\overline{A'}) = o(1).$$

Thus, a.s. $\chi(G_{n,M=cn}) > k$ despite the fact that the expected number of k-colourings of $G_{n,M=cn}$ is exponentially high! $\square$

In [63], Molloy and Reed improved the bound on c_3 further to obtain:

Theorem 3.4 $c_3 \leq 2.571...$

Proof Outline The proof consisted of two steps: (1) They showed that the analysis in the proof of Theorem 3.2 extends to random graphs on a fixed degree sequence. I.e., suppose G is chosen uniformly at random from amongst all graphs on n vertices with a specified degree sequence that has average degree greater than $2\ln 3/\ln(3/2) = 2.709....$ Then the expected number of 3-colourings of G is o(1).

(2) They showed that by repeatedly deleting all vertices of degree less than 3 from $G_{n,p=c/n}$, $c \geq 2.571..$, we will a.s. arrive at a (large) subgraph H of average degree greater than 2.709.... Furthermore, H is uniformly random with respect to its degree sequence. Hence, the expected number of 3-colourings of H is $o(1)$, and so a.s. it has no 3-colouring. Thus, a.s. G has no 3-colouring. □

In the next section, we will discuss a recent technique developed independently by Kirousis et al[51] and by Dubois and Boufkhad[28] which was first used to obtain an upper bound on d_3. This technique was applied, in increasing degrees of precision, to obtain the following upper bounds on c_3:

Theorems 3.5 *(a)* $c_3 \leq 2.602$ (Dunne and Zito[30]).

(b) $c_3 \leq 2.522$ (Achlioptas and Molloy[6])

(c) $c_3 \leq 2.495$ (independently due to Kaporis et al[49] and McDiarmid and Fountoulakis[37]).

To date, the last of these is the best known upper bound on c_k. It is not hard to show that this bound is not tight and that sufficient effort will produce a better bound. We will discuss these bounds and their proofs in the next subsection, after presenting the technique responsible for them.

We close this section by noting an interesting fact, first mentioned in [26]. Performing the calculations in the proof of Theorem 3.2 using the $G_{n,p}$ model, rather than the $G_{n,M}$ model yields a worse bound: Each of the k^n partitions of the vertex set is a valid k-colouring with probability at most $(1-p)^{k\times\binom{n/k}{2}}$, since none of the at least $k \times \binom{n/k}{2}$ edges within a part can be chosen. Thus, the expected number of k-colourings is at most $k^n(1-\frac{2c}{n})^{\frac{n^2}{2k}} \approx \exp(-n(\ln k - c/k))$ which is o(1) for $c > k\ln k$. By being more careful, one can check that these calculations are asymptotically tight in that for $c < k\ln k$, the expected number of k-colourings is exponentially high. (The main observation we need is that the vast majority of the k^n partitions have all parts of size very near n/k.)

Thus, for $\ln k/\ln(k/(k-1)) < c < k\ln k$, the expected number of k-colourings in $G_{n,p=2c/n}$ is exponentially high, while in $G_{n,M=cn}$ it is exponentially low. An intuitive reason for this is as follows: in the $G_{n,p}$ model, there is a chance that the number of edges is much fewer than $p\binom{n}{2}$. The probability of this happening is very small, but if it does happen, then it will typically result in a very large number of k-colourings. Thus, again we have a jackpot phenomenon which produces a significant increase in the expected number.

3.2 Satisfiability

As we will see, most of the techniques used in subsection 3.1 yield analogous bounds on d_k. The analogue of Theorem 3.2 was first shown by Franco and Paull[38]:

Theorem 3.6 $d_k \leq \ln 2/\ln(2^k/(2^k-1))$.

For example, this gives $d_3 \leq 5.19...$, $d_4 \leq 10.74...$, $d_5 \leq 21.80...$.

Proof The proof is very similar to that of Theorem 3.2. We work in the model $F^k_{n,M=dn}$. There are 2^n possible truth assignments to the n variables. For any one truth assignment, the probability that it satisfies the random formula is $\left(\frac{2^k-1}{2^k}\right)^{dn}$, since each of the dn clauses has a $\frac{2^k-1}{2^k}$ probability of being signed in an acceptable way. Thus, the expected number of truth assignments is $2^n\left(\frac{2^k-1}{2^k}\right)^{dn} = \exp(n(\ln 2 - d\ln(2^k/(2^k-1))))$, which is $o(1)$ for $d > \ln 2/\ln(2^k/(2^k-1))$. $\Box$

Again, the same reasoning as in the proof of Theorem 3.3 yields that d_k is actually less than this bound, because variables which appear in no clauses help to create a jackpot phenomenon. Kamath et al[48] carried out this reasoning very carefully, also accounting for other contributions of small degree variables to the jackpots, and showed that the resulting drop in the bound is significant:

Theorem 3.7 $d_3 \leq 4.758$.

More recently, Kirousis et al[51] and, independently, Dubois and Boufkhad[28] introduced a very elegant new technique to (partially) account for the jackpot phenomenon. Define $X(f)$ to be the number of satisfying solutions to a formula f. Define $X'(f)$ to be equal to 1 if $X(f) > 0$ and 0 if $X(f) = 0$. Thus the expected number of solutions is the weighted average over all possible f of $X(f)$, while the probability of satisfiability is the weighted average of $X'(f)$. So a key to overcoming the jackpot phenomenon would be to compute an expected value which is closer to the average of X' than to the average of X.

They suggested that if f has at least one satisfying assignment, then we focus on the lexicographically maximum such assignment, i.e. the maximum satisfying assignment under the ordering where $A_1 > A_2$ if the two assignments agree on variables $v_1, ..., v_r$ and if $A_1(v_{r+1}) = T$ while $A_2(v_{r+1}) = F$. Of course, this maximum is unique. Therefore $X'(f)$ is equal to the number of lexicographically maximum satisfying assignments of f, and so the expected number of such assignments is equal to the probability of satisfiability.

Unfortunately, computing this expected number seems very difficult. So instead, we consider the expected number of assignments which are close to being maximum. We say that a satisfying assignment is i-maximal if one cannot obtain a lexicographically greater satisfying assignment by changing at most i variables, and we define $X_i(f)$ to be the number of i-maximal satisfying assignments of f. Kirousis et al [51] showed that just by focusing on $\mathbf{Exp}(X_1)$, we can get a good improvement on our upper bound, with a relatively small amount of effort:

Theorem 3.8 *$d_3 \leq 4.667$.*

Proof We will work in the model $F^3_{n,M=dn}$. Consider any truth assignment A with a variables set True and $n-a$ variables set False. As we saw earlier, the probability that A satisfies our random formula is $\left(\frac{7}{8}\right)^{dn}$. We will bound the probability that A is 1-maximal, conditional on the event that it is satisfying.

Note that A is 1-maximal iff for every variable v set False, changing v to True will result in a non-satisfying assignment. I.e., there must be at least one clause of the form $(\overline{v} \wedge x \wedge y)$ where x and y are two literals which are negated in A; we call such a clause a *blocker* for v. Consider any v which is set False. There are $\binom{n-1}{2}$ possible blockers for v. Conditioning on A being satisfying, simply reduces the choices of our random clauses to the $7\binom{n}{3}$ clauses which A satisfies. Thus, the probability, q, that v has a blocker satisfies

$$q = 1 - \binom{7\binom{n}{3} - \binom{n-1}{2}}{dn} \Big/ \binom{7\binom{n}{3}}{dn} = 1 - \mathrm{e}^{-3d/7} + o(1).$$

Now consider another variable u that is set False. Conditioning on v having at least one blocker only decreases the chances of u having at least one blocker since the number of clauses is fixed, and no clause can block both v and u (see [51] for a more careful elaboration on this point). Therefore, the probability that both v and u have blockers is at most q^2. Similarly, the probability that all a False variables have blockers is at most q^a, and so the expected number of 1-maximal satisfying solutions is at most

$$\begin{aligned}
\sum_{a=0}^{n} \binom{n}{a} \left(\frac{7}{8}\right)^{dn} \left(1 - e^{-3d/7} + o(1)\right)^a &= \left(\frac{7}{8}\right)^{dn} \sum_{a=0}^{n} \binom{n}{a} \left(1 - e^{-3d/7} + o(1)\right)^a \\
&= \left(\frac{7}{8}\right)^{dn} \left(2 - e^{-3d/7} + o(1)\right)^n \\
&= \exp[n(1 - d\ln(8/7) + \ln(2 - e^{-3d/7} + o(1))],
\end{aligned}$$

which is $o(1)$ for $d \geq 4.667$. (The second equality comes from the identity $\sum_{a=0}^{n} \binom{n}{a} x^a = (1+x)^n$.) □

Carrying out this analysis for general $k > 3$ yields that d_k is at most the unique positive solution of

$$\left(\frac{2^k - 1}{2^k}\right)^d (2 - e^{-kd/(2^k-1)}) = 1.$$

Of course, we gave a little bit away in our analysis when we upper bounded the probabilty that all False variables have blockers by what the value would be if the events of having blockers were independent. By being more exact there, Kaporis et al [50] improved this bound slightly to $d_3 \leq 4.64$. This was obtained independently in a similar (but much more complicated) manner by Dubois and Boufkhad[28]. An even better improvement can be obtained by focussing on $\mathbf{Exp}(X_2)$. In this way, Kirousis et al[51] obtained $d_3 \leq 4.598$. Janson et al[47], by being a little more precise, improved it slightly to:

Theorem 3.9 *$d_3 \leq 4.596$.*

This is the best bound to appear at the time this survey was written. Dubois et al[29] reported that they obtained the better bound of $d_3 \leq 4.506$, but their details are not yet available. They outline their approach as follows.

Given a boolean formula, F, the *degree* of a variable x is $(d^+(x), d^-(x))$, the number of appearances in F of x and $\bar{x}$. Recall that in the previous chapter, the first moment analysis of $G_{n,M=cn}$ yielded a much better bound than that of $G_{n,p=2c/n}$. The reason is that by conditioning on the number of edges being equal to the expected number, we remove the jackpots contributed by graphs with a small number of edges. Dubois et al report that a similar improvement can be obtained by conditioning on the event that for each (i, j), the number of variables with degree (i, j) is approximately equal to what we expect. This removes, for example, the jackpots contributed by formulas with a larger than expected number of variables of degree 0. So they consider a random formula with a fixed degree sequence such that the number of variables with degree (i, j) is equal to the expected number of such variables in $F^3_{n,M=dn}$, and they bound $\mathbf{Exp}(X_1)$ in that formula. The computations are reportedly very complicated.

Having described the technique introduced in [51], we can now discuss Theorems 3.5. Theorem 3.5a was proved by considering the analogue of X_1 for 3-colourings when we rank the colours $R > B = G$. Theorem 3.5b was proved by considering the same variable when we rank the colours $R > B > G$, and follow the proof of Theorem 3.8. Theorem 3.5c was proved by being a little more precise in the calculations, in the same way that the bound in Theorem 3.5c can be improved to 4.64.

We close this section with a question. It is not hard to show that for any fixed i, the bound on d_k obtained from $\mathbf{Exp}(X_i)$ is not optimal. Define $d_k^{(i)}$ to be equal to that bound, i.e. the value for d at which $\mathbf{Exp}(X_i)$ shifts from being exponentially low to exponentially high. Kirousis et al have asked:

Question 3.10 *Does $d_k = \lim_{i\to\infty} d_k^{(i)}$?*

Intuitively, this asks whether all the jackpots inherent in our expected value computations can be accounted for by clusters of many satisfying solutions which differ in O(1) variables.

4 Lower Bounds

4.1 Colouring

As described in the introduction, the first lower bound on the threshold for 3-colouring is $c_3 \geq \frac{1}{2}$. Most improvements on this lower bound came from focussing on the appearance of what is known as a *3-core*. Every graph G with $\chi(G) \geq k+1$ must have a $(k+1)$-critical subgraph, i.e. a subgraph H such that $\chi(H) = k+1$ and for any proper subgraph $H^{'} \subseteq H$, $\chi(H^{'}) \leq k$. It is well-known, and easy to prove (see [69]) that any $(k+1)$-critical graph must have minimum degree at least k. Thus, any graph which cannot be k-coloured must contain a non-empty subgraph with minimum degree at least k. The *k-core* of a graph G, is defined to be the largest such subgraph, if at least one such subgraph exists.

It is easy to see that the k-core is unique, and can be determined by the following simple procedure: repeatedly delete a vertex of degree less than k from the graph until no such vertices remain. Since a deletable vertex cannot become undeletable, the order in which the vertices are deleted is irrelevant, and the procedure will always leave the same subgraph - the k-core.

For $k \geq 3$, we define

$$\lambda_k = \lambda_k^- = \sup\{\lambda : \text{a.s. } G_{n,M=\lambda n} \text{ has no } k\text{-core}\}.$$

We define λ_k^+ in the obvious way, analagous to the definition of c_k^+ and d_k^+.

Recall that for some of the properties discussed in section 1, the threshold for the property coincided with the threshold for a much simpler, necessary but not always sufficient, property. For example, the threshold for a random graph to have a perfect matching is the same as the threshold for it to have no isolated vertices, and the threshold for the appearance of a Hamilton cycle is the same as the threshold for it to have minimum degree at least 2. Inspired by this trend, Bollobás[18] asked if $c_k = \lambda_k$. In fact, he asked the following somewhat stronger question:

Question 4.1 *Consider $k \geq 3$. Run the random graph process until the first k-core appears. Is the resulting graph a.s. not k-colourable?*

Łuczak[53] showed that when the 3-core appears, it must be very large, of size at least $\frac{1}{5000}n$. In fact, he showed that a.s. a random graph in the relevant range of probabilities will not contain any smaller subgraph with even *average* degree at least 3.

Theorem 4.2 *Consider any $0 < c < 3$. A.s. $G_{n,M=cn}$ has no subgraph with fewer than $\frac{1}{5000}n$ vertices and with average degree at least 3.*

Proof We will work in the model $G_{n,p}$. Consider any $4 \leq a \leq \frac{1}{5000}n$. If a graph G has a subgraph on a vertices with at least $\frac{3}{2}a$ edges, then by deleting edges, we see that G has a subgraph on a vertices with exactly $\lceil \frac{3}{2}a \rceil$ edges. The expected number of such subgraphs in $G_{n,p=2c/n}$ is at most

$$\begin{aligned}
\binom{n}{a}\binom{\binom{a}{2}}{\lceil \frac{3}{2}a \rceil} p^{\lceil \frac{3}{2}a \rceil} &\leq \left(\frac{en}{a}\right)^a \left(\frac{e\binom{a^2}{2}}{\lceil \frac{3}{2}a \rceil}\right)^{\lceil \frac{3}{2}a \rceil} \left(\frac{2c}{n}\right)^{\lceil \frac{3}{2}a \rceil} \\
&\leq \left(\frac{8e^5c^3a}{3^3n}\right)^{a/2} \\
&< \left(\frac{3}{4}\right)^{a/2}.
\end{aligned}$$

Furthermore, this expected number is of order $\left(O\left(\frac{a}{n}\right)\right)^{a/2}$ which is less than $O\left(n^{-2/3}\right)$ for $a < n^{1/3}$. Therefore, the expected number of such subgraphs is at most

$$\sum_{a=1}^{n^{1/3}} O(n^{-2/3}) + \sum_{a=n^{1/3}}^{n/1000} \left(\frac{3}{4}\right)^{a/2} = o(1),$$

as required. □

A very similar proof will show that for any c and $\epsilon > 0$ there is an $\alpha = \alpha(c, \epsilon) > 0$ such that a.s. $G_{n,M=cn}$ has no subgraph with fewer than αn vertices and with average degree at least $2 + \epsilon$.

Łuczak used this fact to obtain the first improvement on the trivial lower bound $c_3 \geq \frac{1}{2}$:

Theorem 4.3 $c_3 \geq \lambda_3 > .50005$.

Proof The size of the giant component in $G_{n,M=cn}$ is well understood (see [15, 13, 46]). For $c \leq .50005$, the giant component a.s. has fewer than $\frac{n}{5000}$ vertices, and so, by Theorem 4.2, it a.s. does not contain a 3-core. A.s., the other components are unicyclic and so they don't contain a 3-core either. □

Chvátal[26] used an intricate analysis to determine the precise value of c where the expected number of subgraphs with minimum degree 3 shifts from $o(1)$ to exponentially high.

Theorem 4.4 *For $\psi = 1.442...$ we have the following:*

(a) for $c < \psi$, the expected number of nonempty subgraphs in $G_{n,M=cn}$ with minimum degree 3 is $o(1)$;

(b) for $c > \psi$, this expected number is at least z^n for some $z = z(c) > 1$.

As an immediate corollary, he obtains $c_3 \geq \lambda_3 \geq \psi$.

Perhaps not surprisingly, a jackpot phenomenon occurs here, and λ_3 is in fact much higher than ψ. Molloy and Reed[63] showed that $c_3 \geq \lambda_3 \geq 1.675$. They also showed that $\lambda_3 \leq 1.7941$. Then, Pittel, Spencer and Wormald[67] determined the exact value of λ_k. Setting $\pi_i(x) = 1 - \mathrm{e}^{-x} \sum_{j=0}^{i-1} \frac{x^j}{j!}$, the probability that a Poisson variable with mean x is at least i, their result is:

Theorem 4.5
$$\lambda_k = \lambda_k^- = \lambda_k^+ = \min_{\gamma>0} \frac{\gamma}{2\pi_{k-1}(\gamma)}.$$

For example, $\lambda_3 = 1.675...$, $\lambda_4 = 2.57...$, $\lambda_5 = 3.40...$.

To do this, they used a system of differential equations to analyse the behaviour of a procedure which repeatedly deletes a randomly chosen vertex of degree less than k until no such vertices remain.

This result yields that, asymptotically, $\lambda_k = \frac{1}{2}k + \mathrm{o}(k)$. (It is an easy exercise to show that $\lambda_k \leq k-1$, since any graph with average degree at least $2(k-1)$ must contain a k-core.) Łuczak[55] showed that $c_k = \Theta(k \ln k)$ and so for high values of k, the answer to Bollobás' Question 4.1 is negative, i.e. $c_k \neq \lambda_k$. Molloy[64] noted that for $k \geq 4$, $c_k \neq \lambda_k$. To do so, he simply applied the following inequality[2], taken from [22] (see also [26]).

$$c_k \leq c_{k+1} - \frac{1}{2}\ln(2c_{k+1}+1). \tag{4.1}$$

Using the bound $c_3 \geq \lambda_3 \geq 1.675$, this yields $c_4 \geq 2.58... > \lambda_4$, and similarly for each $k > 4$. Inequality (4.1) is proven in [22] as follows:

Consider any $G = G_{n,p=2c/n}$. We build a stable set I using a simple greedy procedure whereby we run through the vertices in an arbitrary sequence. When we arrive at a vertex v, we add v to I, and remove all neighbours of v from the sequence. It turns out that a.s. $|I| \geq (\ln(2c+1)/(2c))n - o(n)$. The key observation is that the outcome of this procedure only depends on the edges from I to the rest of the graph. Therefore, after building I, we have exposed no information about the edges in $G-I$, and so the graph induced by $G-I$ is distributed exactly like $G_{n',p}$ where $n' = n - |I| = (1-\ln(2c+1)/(2c))n - o(n)$. If $G-I$ is k-colourable, then G is $(k+1)$-colourable. Thus, if $c > c_{k+1}$ then $G-I$ is not a.s. k-colourable; i.e. $G_{n',p}$ is not a.s. k-colourable. Therefore, $p > 2c_k/n'$ and so $2c_k < n' \times 2c/n = 2c - \ln(2c+1) + o(1)$. This yields (4.1).

Of course, using (4.1) to show $c_4 > \lambda_4$ requires the bound $c_3 \geq \lambda_3$ as a "base" and so there was no way to adjust this argument to show $c_3 > \lambda_3$. Eventually, Achlioptas and Molloy [1, 5] used a different technique to show $c_3 > \lambda_3$, and obtain what is, to date, the best known lower bound on c_3:

Theorem 4.6 $c_3 \geq 1.923$.

[2]There is a confusing typographical error in [64] - the "+1" term is missing from the inequality.

To prove this theorem, they use a set of differential equations to analyse the following procedure: Begin with a list of the same 3 colours on every vertex of $G_{n,p=c/n}$. Repeatedly pick a random vertex v of minimum list-size, choose a random colour for v from its list, and delete that colour from the list of every neighbour of v. They show that this procedure a.s. fails for $c > 1.923...$ while for $c < 1.923...$ it does not a.s. fail. Corollary 5.5 (to follow) allows them to strengthen this latter fact to: for $c < 1.923..$, $G_{n.p=c/n}$ is a.s. 3-colourable. They also show that for large k, the analagous procedure succeeds for $c \leq k \ln k - \frac{3}{2}k$ but a.s. fails for $c > (1+\epsilon)k \ln k$ for any constant $\epsilon > 0$.

4.2 Satisfiability

Thus far, all lower bounds for d_3 have taken the following approach: (1) define an algorithm which looks for a satisfying solution in an instance of 3-SAT; (2) prove that for d less than a particular constant, this algorithm does not a.s. fail on $F^3_{n,M=cn}$. (The astute reader will notice that these two facts alone do not provide a bound on d_3; this point is addressed after Theorem 4.7 below.)

Of course, for this approach to work, there has to be a balance between the sophistication of the algorithm, and the ease with which this algorithm can be analyzed. Roughly speaking, increases in the lower bounds on d_3 over the years correspond to increases in the abilities of the researchers to analyse more complicated algorithms. So far, all algorithms studied have been fairly similar, they all involve setting variables one-at-a-time in a greedy manner. Achlioptas has recently written an excellent survey on this subject, so I will not go into many details here and instead refer the reader to [3].

All of the algorithms follow a similar strategy. Each time a variable is set, say $x_1 = T$, any clause that contains x_1 is removed from the formula since it is now satisfied. Any clause that contains $\bar{x}_1$ has $\bar{x}_1$ deleted from it, and it shrinks to a smaller clause. This is because there is no hope for this clause to be satisfied by setting $x_1 = F$, and so the $\bar{x}_1$ term is redundant. So neither x_1 nor $\bar{x}_1$ appears anywhere in the formula. This leaves us with a new formula on one fewer variables, with a few less clauses and a few clauses whose sizes have shrunk by one.

Initially, we have only 3-clauses, but as the procedure progresses we will also have 2-clauses and 1-clauses. 1-clauses are dealt with immediately since, for example, if we wish to satisfy a formula containing the clause $(\bar{x_7})$ then we have no choice but to set $x_7 = F$. So if there are any 1-clauses, then we choose one of them (usually at random) and set its variable appropriately. Where the various strategies differ is in the way they choose the variable to set when there are no 1-clauses.

The first result in this vein was due to Chao and Franco[24]:

Theorem 4.7 $d_3 \geq 8/3 = 2.666...$

To be precise, Chao and Franco proved something slightly weaker - that for $d < 8/3$, $F^3_{n,M=dn}$ is not a.s. unsatisfiable. Applying Corollary 5.5 from section 5 allows us to obtain Theorem 4.7 from their weaker result, i.e. to show that for the same range of d, $F^3_{n,M=dn}$ is a.s. satisfiable. Corollary 5.5 came many years after Chao and Franco's work.

The results in [24] come from analyzing what they call the Unit Clause algorithm. If there is a unit clause (a 1-clause) then they set that variable. Otherwise, they pick a variable x uniformly at random and choose its setting, T or F, uniformly at random.

They then improved this simple procedure by not picking the setting of x randomly. Instead, they count the number of 3-clauses that x appears in and the number of 3-clauses that $\bar{x}$ appears in, and set x according to which of these two counts is highest. Analyzing this procedure yields:

Theorem 4.8 $d_3 \geq 2.9$

Again, Chao and Franco only proved that for $d < 2.9$, $F^3_{n,M=dn}$ is not a.s. unsatisfiable; we require Corollary 5.5 to obtain this stronger result. A sharper analysis of the same procedure (see [3]) yields $d_3 \geq 3.001$.

The Unit Clause algorithm can obviously be run on k-SAT for any k. For general k, they proved

Theorem 4.9 *For $c < \frac{1}{2}\left(\frac{k-1}{k-2}\right)^{k-2}\frac{2^k}{k}$, the Unit Clause algorithm will find a satisfying assignment for $F^3_{n,M=cn}$ with probability at least $\zeta + o(1)$ for some $\zeta = \zeta(c) > 0$.*

Corollary 5.5 shows that this yields

Corollary 4.10 $d_k \geq \frac{1}{2}\left(\frac{k-1}{k-2}\right)^{k-2}\frac{2^k}{k}$.

It turns out that, in the interesting range of d, there tend to be many 2-clauses throughout most of the procedure. Chvátal and Reed[27] analysed an algorithm which takes advantage of this fact. If there is no 1-clause then, instead of choosing a random variable, we choose a random 2-clause and set one of its variables to satisfy the clause (we choose which of the two variables to set uniformly at random). If there is no 2-clause then we choose a literal to set uniformly at random. This guarantees that we will satisfy at least one of the 2-clauses. They proved that for general k, this algorithm will a.s. produce a satisfying solution to $F^k_{n,M=dn}$ if $c < \frac{1}{8}\left(\frac{k-1}{k-3}\right)^{k-3}\frac{k-1}{k-2} \times \frac{2^k}{k}$. At first glance, this seems worse than Theorem 4.9, but one should note that here the algorithm suceeds a.s., rather than simply with probability bounded away from 0.

Frieze and Suen[42] studied this algorithm more closely for 3-SAT and used it to show that:

Theorem 4.11 $d_3 \geq 3.003$.

Frieze and Suen actually obtained this bound without appealing to Corollary 5.5 (which did not exist at the time). To do this, they introduced a complicated backtracking component to the algorithm. Thus, unlike the versions of Theorems 4.7 and 4.8 found in [24], they showed that their algorithm a.s. succeeds, rather than the weaker fact that it does not a.s. fail.

This remained the best bound on d_3 for several years. Recently, Achlioptas[2] suggested a similar procedure where, instead of setting one of the variables in the chosen 2-clause, we set both variables. The main advantage of this approach is the following: Recall that Chao and Franco increased the performance of the Unit Clause algorithm significantly by choosing between the two possible settings of the selected variable. In Chvátal and Reed's variant, the variable is chosen specifically to satisfy a particular 2-clause and so there is no choice for its setting. However, when setting both variables in a 2-clause, Achlioptas could choose between 3 different settings which satisfy that clause. Choosing the best of these 3 settings leads to improved performance, and Achlioptas proved:

Theorem 4.12 *$d_3 \geq 3.165$.*

Besides setting two variables at once, Achlioptas' procedure involved a second innovation: He did not always set 1-clauses as soon as they appeared. Rather, at each time step he made a random choice as to whether to set a 1-clause or a 2-clause. The probability of choosing to set a 1-clause is chosen so that, in the long haul, he is setting 1-clauses at roughly the same pace as they are formed. Certainly this does not make his algorithm run more efficiently since, as we noted earlier, when a 1-clause appears we have no choice but to set it eventually, so we might as well set it right away. Nevertheless, this innovation makes the algorithm dramatically easier to analyse, since it completely dispenses with analyzing the timing of the 1-clause appearances, which can be very difficult. For a more detailed description of the advantages, see [2, 3].

Most recently, Achlioptas and Sorkin[7] determined what they proved to be the most efficient possible algorithm from amongst all algorithms which fit the general framework described in this section. (See their paper for a formal definition of the class of algorithms amongst which theirs is the optimum.) The analysis of their algorithm yielded:

Theorem 4.13 *$d_3 \geq 3.26$.*

At the time this article was written, this stood as the best bound thus far.

In the setting of random 3-SAT, the analogue of the 3-core comes from applying what is known as the *pure literal rule*. A *pure literal* is one whose negation does not appear in the formula; a *pure variable* is a variable at least one of whose literals is pure. In this procedure, one repeatedly looks for a pure variable v and sets v appropriately. If we consider the degree of a variable v to be an ordered pair containing the number of clauses that v and $\bar{v}$ each appear in, then the pure literal rule will continue until we find what can be thought of as a (1,1)-core, and so we define it to be an *impure core.*

For $k \geq 3$, we define

$$\gamma_k = \gamma_k^- = \sup\{\gamma : \text{a.s. } F^k_{n,M=\gamma n} \text{ has no impure core}\}.$$

We define γ_k^+ in the obvious way, analagous to the definition of c_k^+, d_k^+ and λ_k^+. Clearly, $d_k \geq \gamma_k$.

The first analysis of γ_3 was due to Broder et al[23] who determined:

Theorem 4.14 $1.63 \leq \gamma_k \leq 1.7$

The paper of Pittel et al[67] in determining the threshold for the appearance of a 3-core provided a short and elegant, but non-rigorous, heuristic for predicting the value of λ_k. (Their long rigorous proof used a different approach.) A number of different people independently noted that this heuristic also yields a good prediction for the value of γ_k, as do other non-rigorous approaches (see, eg. [62]). In [65], Molloy and Pittel found a simple technique for making this heuristic rigorous. In [66], Molloy and Wormald use it to provide the first rigorous proof of the value of γ_k along with the thresholds for many similar cores.

Theorem 4.15

$$\gamma_k = \gamma_k^- = \gamma_k^+ = \min_{\beta>0} \frac{2\beta}{k(1-e^{-\beta})^{k-1}}.$$

Thus, for example, $\gamma_3 = 1.6369..., \gamma_4 = 1.5445..., \gamma_5 = 1.4035....$ See [44] for another application of the same technique.

5 Sharp Thresholds

As described in the introduction, we still do not know whether there exist absolute constants $c_3^*, c_4^*, ...$ such that for any $\epsilon > 0$, $G_{n,M=(c_k^*-\epsilon)n}$ is a.s. k-colourable and $G_{n,M=(c_k^*+\epsilon)n}$ is a.s. not k-colourable. However, we know that something very close to this is true. Before describing that result more precisely, we will discuss some work which led up to it.

5.1 Concentration on a few points

If the constants c_k^* existed, then for every $c \geq \frac{1}{2}$ which is not equal to one of them, we would know, with high probability, the chromatic number of $G_{n,M=cn}$. Thus, for each $\frac{1}{2} \leq c \neq c_k^*$, the random variable $\chi(G_{n,M=cn})$ a.s. takes on one particular value; i.e. it is concentrated on one point. In [68], Shamir and Spencer proved that for every constant c, $\chi(G_{n,M=cn})$ is concentrated on a small range of at most 4 points.

Theorem 5.1 *For every constant c, there is an integer $u(n)$ such that a.s. $u(n) \le \chi(G_{n,M=cn}) \le u(n) + 3$.*

Unfortunately, their proof technique does not help us to compute $\lim_{n\to\infty} u(n)$ in terms of c. In fact, for most constants c, we don't even know whether this limit converges, although it is hard to imagine a situation where it doesn't.

Proof It is more convenient to work in the $G_{n,p}$ model. We begin with the following fact:

Fact 1: *Define $u = u(n)$ to be the largest integer such that $\mathbf{Pr}(\chi(G_{n,p=2c/n}) \ge u) > \frac{1}{\ln\ln n}$. Then a.s. $G_{n,p=2c/n}$ has a partial u-colouring which colours all but at most $\ln^2 n\sqrt{n}$ vertices.*

The theorem follows easily from Fact 1. A.s. $G = G_{n,p=2c/n}$ has a partial u-colouring which leaves only $Y \subset V(G)$ uncoloured where $|Y| \le \ln^2 n\sqrt{n}$. The same calculations as in the proof of Theorem 4.2 show that a.s. G has no subgraph on $o(n)$ vertices with average degree at least 2.99. Therefore, a.s. the vertices of Y do not contain a 3-core and so they can be coloured using at most 3 new colours.

To prove Fact 1, we let X denote the minimum over all partial u-colourings of the number of uncoloured vertices. Thus G is u-colourable iff $X = 0$. We will apply the Simple Concentration Bound to prove that X is concentrated. Our vertex set is $v_1, ..., v_n$. For each i, we let T_i be the random trial which chooses all the edges from v_i to $\{v_1, ..., v_{i-1}\}$. Note that the outcome of each T_i can only affect X by at most 1. Therefore,

$$\mathbf{Pr}(|X - \mathbf{Exp}(X)| > \frac{1}{2}\ln^2 n\sqrt{n}) \le 2\mathrm{e}^{(\frac{1}{2}\ln^2 n\sqrt{n})^2/2n} = 2e^{-\frac{1}{8}\ln^2 n}.$$

Since $\mathbf{Pr}(X = 0) \ge \frac{1}{\ln\ln n}$, this implies $\mathbf{Exp}(X) \le \frac{1}{2}\ln^2 n\sqrt{n}$ and so $\mathbf{Pr}(X > \ln^2 n\sqrt{n}) = o(1)$ as required. □

Remarks Actually, Theorem 5.1 was originally stated with "$u(n) + 4$" rather than "$u(n) + 3$", and it was more general in that it allowed c to be as high as $n^{1/6-\epsilon}$ for any constant $\epsilon > 0$. Also, Fact 1, as stated, does not appear in their paper; instead they use an analagous but more complicated result. Frieze (see [54]) noted that this slightly simplified version follows easily from the ideas contained in their original proof.

To a reader who is very familiar with Azuma's Inequality and its ilk, this seems like a fairly simple proof. However, it should be noted that this was the paper which introduced Azuma's Inequality to the combinatorics community and so, at its time, the proof was quite substantial, not to mention influential.

Shortly thereafter, Luczak [54] showed that a simple modification of their proof reduces the length of the range from 4 to 2.

Theorem 5.2 *For every constant c, there is an integer $u(n)$ such that a.s. $u \le \chi(G_{n,M=cn}) \le u + 1$.*

Proof Outline Again, we use the $G_{n,p}$ model. By Theorem 4.3 we can assume that $c > 1.0001$ and so $u \geq 3$. Consider $G = G_{n,p=2c/n}$ and the set Y as in the proof of Theorem 5.1. Łuczak showed that a.s. we can find a superset $Y' \supseteq Y$ such that (i) $|Y'| < 3\ln^3 n\sqrt{n}$ and (ii) $N(Y')$ is a stable set. ($N(Y')$ is the neighbourhood of Y', i.e. the set of vertices in $G - Y'$ which are each adjacent to at least one vertex in Y'.) Thus, we can obtain a $(u+1)$-colouring of G as follows. Take the u-colouring of $V(G) - Y' - N(Y')$ obtained by uncolouring some of the vertices in the partial colouring of $V(G) - Y$. Use a new colour on $N(Y')$. Then colour Y' using 3 of the colours already used in the u-colouring of $V(G) - Y' - N(Y')$.

To prove that Y' exists, we build it greedily, repeating the following step until $N(Y')$ is a stable set: Choose some edge uv such that $u, v \in N(Y')$ and add both u and v to Y'. Clearly, when this stops, $N(Y')$ is a stable set. Each time we add a pair $\{u, v\}$, we add 2 vertices and at least 3 edges to the subgraph induced by Y'. Thus, adding more than $\ln^3 n\sqrt{n}$ pairs would produce a subgraph on fewer than $3\ln^3 n\sqrt{n}$ vertices with average degree at least $3 - o(1) > 2.99$. But a.s. no such subgraph exists. Therefore, a.s. we add fewer than $\ln^3 n\sqrt{n}$ pairs and so $|Y'| < 3\ln^3 n\sqrt{n}$, as required. $\square$

Łuczak proved that, as with Theorem 5.1, Theorem 5.2 holds for p as high as $n^{-5/6}$. More recently, Alon and Krivelevich[10] proved that it holds for p as high as $n^{-1/2-\epsilon}$ for any constant $\epsilon > 0$.

5.2 Friedgut's Theorem

In [39], Friedgut developed an approach which enables us to show that certain properties exhibit a threshold which is almost as sharp as was predicted by Erdős in Question 1.1(a). For simplicity, we will present his theorem in a somewhat weakened form. For one thing, we will only be concerned with $G_{n,p}$ for p of order $\Theta(1/n)$. We begin with a few definitions.

We say that a graph property, P, is *monotonically increasing* if (i) P is invariant under graph automorphisms, and (ii) if H is a subgraph of G on the same vertex set, and H has P, then G must have P.

A monotonically increasing property P exhibits a *transition* if for some $0 < c_1 < c_2$, (i) $G_{n,p=c_1/n}$ a.s. does not have P and (ii) $G_{n,p=c_2/n}$ a.s. has P. Insisting on a property being monotonically increasing rather than monotonically decreasing is a convenience which is easily satisfied by negating the property if necessary. For example, we can let P be the property of non-3-colourability.

A monotonically increasing property P which exhibits a transition is said to have a *sharp threshold* if there exists a function $c(n)$ such that for any $\epsilon > 0$, $G_{n,p=(c(n)-\epsilon)/n}$ a.s. has P and $G_{n,p=(c(n)+\epsilon)/n}$ a.s. does not have P. Note that we must have $c_1 - o(1) \leq c(n) \leq c_2 + o(1)$.

Note that the only difference between this and the stronger property in Question 1.1(a) is that here we allow c to vary with n, just as in Theorems

5.1, 5.2. All natural properties which are known to have sharp thresholds are conjectured to also satisfy the stronger property where for all n, $c(n)$ can be taken to be some constant c. However, we do not have any general tools for proving this fact.

Given two properties P_1, P_2, their *symmetric difference*, $P_1 \triangle P_2$ is the property of satisfying one but not both of P_1, P_2. Given a set of graphs $\mathcal{H} = \{H_1, ..., H_k\}$, we define $Q(\mathcal{H})$ to be the property of having a subgraph which is isomorphic to one of the members of $\mathcal{H}$.

The following comes from Theorem 1.1 of [39]:

Friedgut's Theorem *Let P be any monotonically increasing graph property which exhibits a transition but does not have a sharp threshold. Then there exists some $c(n)$ such that for every $\delta > 0$ there is an infinite sequence of values of n for which the following holds:*

There is a set of unicyclic graphs $\mathcal{H} = \{H_1, ..., H_k\}$ where each H_i has size at most $K = K(\delta, P)$ (i.e. the sizes of the H_i cannot grow with n) such that the probability of $G_{n,p=c(n)/n}$ having $P \triangle Q(\mathcal{H})$ is less than δ.

In other words, when $p = c(n)/n$, P can be arbitrarily closely approximated in probability by the property of having a subgraph from a fixed list of unicyclic graphs.

For example, suppose that P is the property of being non-bipartite. Then P is equivalent to the property of having a subgraph from a list of unicyclic graphs, namely the list of odd cycles, $\{C_3, C_5, ...\}$. But there is a subtlety here: the sizes of the graphs in this list grow with n, since we include all cycles of odd length up to n. However, it is not important that P be equivalent to the property of containing a subgraph from the list; P need only be closely approximated by that property. And by truncating the list, we can obtain lists of graphs with bounded size which approximate P arbitrarily well, so long as we are in the range $c < 1$. More specifically, it turns out that for every $\delta > 0$, there exists $t = t(\delta)$ such that if $\mathcal{H} = \{C_3, C_5, ..., C_{2t+1}\}$, then the probability of $G_{n,p=1/2n}$ having $P \triangle Q(\mathcal{H})$ is less than δ.

Theorem 1.1 of [39] is in fact much more general than the version presented above, in that (i) this is essentially an "if and only if" statement, and (ii) it allows for more general sharp thresholds which occur when M is not linear in n.

It is also more precise about the manner in which K depends on P. For any fixed n, we can consider $\mathbf{Pr}(P)$ as a continuous function in the variable p, the edge-probability. We denote this function by $\mu(p)$. Friedgut observes that if P exhibits a transition, but does not have a sharp threshold, then there exist absolute constants A and $\epsilon > 0$ such that for arbitrarily large values of n, there will be a point $p^* = p^*(n)$ for which (i) $\zeta < \mu(p^*) < 1 - \zeta$, and (ii) $p^* \times \mu'(p^*) < A$. $K(P, \delta)$ is in fact a function only of A and δ.

Friedgut also showed that an analogue to his theorem applies to random instances of k-SAT, and his first application was to almost answer Question 1.2(a).

Theorem 5.3 *For $k \geq 2$, non-satisfiability of $F^k_{n,M}$ has a sharp threshold.*

Theorem 1.2 of [39] is a more technical variation of Theorem 1.1 of [39], and can be easier to apply. (For simplicity of presentation, we won't state it here.) Achlioptas and Friedgut [4] used Theorem 1.2 to prove

Theorem 5.4 *For $k \geq 3$, non-k-colourability of $G_{n,M}$ has a sharp threshold.*

Thus, they came very close to answering Question 1.1(a). These two theorems have the following corollaries, which were helpful towards establishing some of the bounds in sections 4.1, 4.2.

Corollary 5.5 *(a) If $G_{n,M=cn}$ is k-colourable with probability at least $\zeta + o(1)$ for some constant $\zeta = \zeta(c) > 0$, then for each $c' < c$, $G_{n,M=c'n}$ is a.s. k-colourable.*

(b) If $F^k_{n,M=dn}$ is satisfiable with probability at least $\zeta + o(1)$ for some constant $\zeta = \zeta(c) > 0$, then for each $d' < d$, $F^k_{n,M=d'n}$ is a.s. satisfiable.

Of course, the analagous corollaries also hold for the $G_{n,p}$, $F^k_{n,p}$ models and when we replace "satisfiable" by "not satisfiable", or "k-colourable" by "not k-colourable"

At first glance, it would seem that the sharpness of 3-colourability might follow quite easily from Friedgut's Theorem, since it is well known that non-3-colourability is not equivalent to containing one of a list of small subgraphs. However, to apply Friedgut's Theorem, one needs to show that it cannot even be closely approximated with such a list, and proving this is far from simple. In fact, to date, Friedgut's Theorem has only been successfully applied to a handful of properties [4, 39, 40, 41]. For example, we don't know how to apply his theorem to k-list-colourability.

Question 5.6 *For $k \geq 3$, does non-k-list-colourability have a sharp threshold?*

We don't even know the answer to the following, for any small value of $k \geq 3$:

Question 5.7 *Consider $k \geq 3$. Run the random graph process until the first k-core appears. Is the resulting graph a.s. not k-list-colourable?*

I think that the answer to this question is almost certainly NO for every $k \geq 3$. In fact, the answer to the following question is probably YES:

Question 5.8 *Consider $k \geq 3$. Run the random graph process until the graph is not k-list-colourable. Is the resulting graph a.s. not k-colourable?*

So, I think that the thresholds for k-list-colourability and k-colourability are equal.

We know that the answer to Question 5.7 is NO for large values of k, since Alon and Krivelevich[11] proved that for large k and $c < \frac{1}{100} k \ln k$, $G_{n,M=cn}$ is a.s. k-list colourable.

6 Asymptotics

As we saw in sections 3.2, 4.2, we know that $(\frac{1}{2}+o(1))\frac{2^k}{k} < d_k \leq \ln 2/\ln(2^k/(2^k-1)) < \ln 2 \times 2^k$. Subsequent efforts have improved the constant term in this lower bound, but no one has yet been able to improve the asymptotic order of either bound. It would be very interesting to get asymptotic upper and lower bounds with the same order of magnitude.

Question 6.1 *What is the asymptotic order of d_k?*

We understand the asymptotics of c_k much better than those of d_k, thanks to the following theorem of Łuczak[55]:

Theorem 6.2 $c_k = (1 + \mathrm{o}(1))k \ln k$.

Recall that every lower bound on d_3 established in section 4.2 was obtained by analyzing an algorithm which will a.s. produce a satisfying assignment for $F^3_{n,M=cn}$. Each of those algorithms, when extended to general k, will a.s. fail for $c > O(2^k/k)$. It is entirely plausible that $d_k = 2^k(\ln 2 + \mathrm{o}(1))$ but that there is no efficient algorithm which will find a satisfying solution for $c > O(2^k/k)$.

Question 6.3 *Is there a constant $\epsilon > 0$ and a polytime algorithm P such that for all $k \geq 3$, a.s. P will find a satisfying solution for $F^k_{n,M=\epsilon\times d_k n}$?*

A similar situation is true for colourability. We know of no algorithm which can guarantee a k-colouring of $G_{n,M=cn}$ for $c > (\frac{1}{2}+\epsilon)k\ln k$. In fact, it may be the case that no such algorithm exists.

Question 6.4 *Is there a constant $\epsilon > 0$ and a polytime algorithm P such that for all $k \geq 3$, a.s. P will find a k-colouring of $G_{n,M=(\frac{1}{2}+\epsilon)k\ln k\times n}$?*

We don't even know of any efficient algorithm which will find a stable set large enough to form one of the colour classes in such a colouring.

Question 6.5 *Is there a constant $\epsilon > 0$ and a polytime algorithm P such that for all $k \geq 3$, a.s. P will find a stable set of size n/k in $G_{n,M=(\frac{1}{2}+\epsilon)k\ln k\times n}$?*

An analogous question can be asked for $G_{n,M}$ for much larger values of M. See [12] for some recent work on this problem.

7 A new question

We close this survey with a new question, posed by Achlioptas and Molloy, which is of the same flavour as Bollobás' Question 4.1.

Consider a graph G and any collection of k disjoint stable sets $V_1, ..., V_k$. We say that this collection is *bilinked* if for every pair V_i, V_j, the bipartite graph induced by $V_i \cup V_j$ is connected.

Question 7.1 *Consider a random graph process on n vertices and $k \geq 3$. Does the chromatic number of the graph jump from k to $k+1$ within $o(n)$ steps of the appearance of the first bilinked collection of k stable sets, other than a triangle?*

The motivation for this question is as follows.

Consider any bilinked collection, set $a_i = |V_i|$, and let B be the subgraph induced by $V_1 \cup ... \cup V_k$. There are at least $a_i + a_j - 1$ edges between V_i, V_j and so the total number of edges in B is at least $(k-1)(a_1 + ... + a_k) - \binom{k}{2}$. Since $|B| \geq k$, this is at least $\frac{k-1}{2}|B|$ and so B has average degree at least $k-1$. Furthermore, for $k = 3$ if B is not a triangle then $|B| \geq 4$ and this shows that the average degree in B is at least 2.5. Therefore, a straightforward argument along the lines of Theorem 4.2 shows that a.s. any bilinked collection of at least 3 stable sets in $G_{n,p=c/n}$ which is not a triangle, must contain at least ϵn vertices for some $\epsilon = \epsilon(c) > 0$.

Let B be the first non-triangle bilinked collection to appear. We conjecture that B will not a.s. induce a uniquely k-colourable subgraph, but that B will a.s. contain a uniquely k-colourable subgraph B' of size $|B| - o(n)$. If this is the case, then with high probability, within a small number of steps, an edge will be added between two vertices in the same colour class of B'. The subgraph induced by B' plus that edge will have chromatic number $k+1$.

To prove the other side of our question, that a.s the chromatic number doesn't jump before the appearance of such a bilinked collection, it suffices to answer the following question in the affirmative:

Question 7.2 *Consider the random graph process. Is it true that a.s. by the time the chromatic number jumps to $k+1$, a uniquely k-colourable subgraph which is not a triangle has appeared?*

Note that the colour classes of any uniquely k-colourable graph must be bilinked, as otherwise we could find a different colouring by switching on a Kempe chain.

It is easy to see that there are many graphs of chromatic number $k+1$ which do not contain any uniquely k-colourable subgraphs. Nevertheless, we think that such graphs do not arise in the random graph process. If one could prove that the answer to Question 7.2 is YES, then Corollary 1 of Aldous[9] would immediately yield an alternate proof of the sharp threshold for k-colourability. We can ask the same question for satisfiability:

Question 7.3 *Consider the random k-SAT process. Is it true that a.s. by the time the formula is unsatisfiable, a uniquely satisfiable subformula has appeared?*

Similarly, if one could prove that the answer to this question is YES, then we would have an alternate proof of the sharp threshold for satisfiability.

If Question 7.1 is answered in the affirmative, then it would reduce the quest for the threshold for k-colourability to finding the threshold for the appearance of a bilinked collection of k stable sets. This latter threshold seems a little less daunting because this structure is somewhat reminiscent of a core. However, none of the now standard techniques for finding the threshold for the appearance of core-like structures seem to apply here, and so even this latter threshold may be very hard to determine.

Acknowledgements

I would like to thank Dimitris Achlioptas, Ehud Friedgut, Tomasz Łuczak, Colin McDiarmid, Joel Spencer and an anonymous referee for much help during the preparation of this paper.

References

[1] D. Achlioptas, Threshold Phenomena in Random Graph Coloring and Satisfiability, Ph.D. Thesis, University of Toronto, 1999.

[2] D. Achlioptas, Setting two variables at a time yields a new lower bound for random 3-SAT, in *32nd Annual Symposium on Foundations of Computer Science (STOC)* (2000), pp. 28 - 37.

[3] D. Achlioptas, A survey of lower bounds for random 3-SAT via differential equations, *Theoretical Computer Science* (to appear).

[4] D. Achlioptas and E. Friedgut, A sharp threshold for k-colorability, *Random Structures & Algorithms* **14** (1999), 63 - 70.

[5] D. Achlioptas and M. Molloy, Analysis of a list-colouring algorithm on a random graph, in *38th Annual Symposium on Foundations of Computer Science (FOCS)* (1997), pp. 204 - 212.

[6] D. Achlioptas and M. Molloy, Almost all graphs with $2.522n$ edges are not 3-colourable, *Electronic J. Comb.* **6** R29. (1999),

[7] D. Achlioptas and G. Sorkin, Optimal myopic algorithms for random 3-SAT, in *Proceedings of the 41st Annual Symposium on Foundations of Computer Science (FOCS)* (2000), pp. 590-600.

[8] M. Ajtai, J. Komlós and E. Szemerédi, Topological complete subgraphs in random graphs, *Studia Sci. Math. Hungar.* **14** (1979), 293 - 297.

[9] D. Aldous, Threshold limits for cover times, *J. Theoretical Prob.* **4** (1991), 197 - 211.

[10] N. Alon and M. Krivelevich, The concentration of the chromatic number of random graphs, *Combinatorica* **17** (1997), 303 - 313.

[11] N. Alon and M. Krivelevich, List colouring of random and pseudo-random graphs, *Combinatorica* **19** (1999), 453 - 472.

[12] N. Alon, M. Krivelevich and B. Sudakov, Finding a large hidden clique in a random graph, *Random Structures Algorithms* **13** (1998), 457 - 466.

[13] N. Alon and J. Spencer, *The Probabilistic Method*, Wiley, New York (1992).

[14] K. Azuma, Weighted sums of certain dependent random variables, *Tokuku Math. Journal* **19** (1967), 357 - 367.

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

[16] B. Bollobás, The chromatic number of random graphs, *Combinatorica* **7** (1988), 49-55.

[17] B. Bollobás, The evolution of random graphs, *Trans. Amer. Math. Soc.* **286** (1984), 257 - 274.

[18] B. Bollobás, The evolution of sparse graphs, in *Graph Theory and Combinatorics* (ed. B. Bollobás), (1984), 35-57.

[19] B. Bollobás, C. Borgs, J. Chayes, J.H. Kim, and D. Wilson, The scaling window of the 2-SAT transition, *Random Structures and Algorithms* (to appear).

[20] B. Bollobás and P. Erdős, Cliques in random graphs, *Math. Proc. Camb. Phil. Soc.* **80** (1976), 419-427.

[21] B. Bollobás, T. Fenner and A. Frieze, Hamilton cycles in random graphs of minimal degree at least k, in *A tribute to Paul Erdős* (eds. A. Baker, B. Bollobás, A. Hajnal), (1990). 59-95.

[22] B. Bollobás and A. Thomason, Random graphs of small order, in *Random Graphs '83* (eds. M. Karoński and A. Ruciński), *Annals of Discrete Math*, 28, (1985), pp. 47-97.

[23] A. Broder, A. Frieze and E. Upfal, On the satisfiability and maximum satisfiability of random 3-CNF formulas, in *41st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)* (1993), pp. 322-330.

[24] M. Chao and J. Franco, Probabilistic analysis of two heuristics for the 3-satisfiability problem, *SIAM J. Comput.* **15** (1986), 1106 - 1118.

[25] M. Chao and J. Franco, Probabilistic analysis of a generalization of the unit-clause literal selection heuristic for the k-satisfiability problem, *Inform. Sci.* **51** (1990), 289 - 314.

[26] V. Chvátal, Almost all graphs with $1.44n$ edges are 3-colorable, *Random Structures & Algorithms* **2** (1991), 11 - 28.

[27] V. Chvátal and B. Reed, Mick gets some (the odds are on his side), in *33rd Annual Symposium on Foundations of Computer Science* (1992), pp. 620 - 627.

[28] O. Dubois and Y. Boufkhad, A general upper bound for the satisfiability threshold of random r-SAT formulas, *J. Algorithms* **24** (1997), 395 - 420.

[29] O. Dubois, Y. Boufkhad and J. Mandler, Typical random 3-SAT formulae and the satisfiability threshold, in *48th Annual ACM-SIAM Symposium on Discrete Algorithms* (2000), pp. 126-127.

[30] P. Dunne and M. Zito, An improved upper bound on the non 3-colourability threshold, *Information Processing Letters* **65** (1998), 17-23.

[31] J. Edmonds, Paths, trees and flowers, *Can. J. Math.* **17** (1965), 449-467.

[32] P. Erdős, On circuits and subgraphs of chromatic graphs, *Mathematika* **9** (1962), 170-175.

[33] P. Erdős and A. Rényi, On random graphs I, *Publ. Math. Debrecen* **6** (1959), 290-297.

[34] P. Erdős and A. Rényi, On the evolution of random graphs, *Pub. of the Math. Inst. of the Hungarian Academy of Sciences* **5** (1960), 17-61.

[35] P. Erdős and A. Rényi, On the existence of a factor of degree one of a connected random graph, *Acta Math. Acad. Sci. Hungar.* **17** (1966), 359-368.

[36] W. Fernandez de la Vega, On random 2-SAT, manuscript. 1992.

[37] N. Fountoulakis, M.Sc. dissertation, Oxford University, 2000.

[38] J. Franco and M. Paull, Probabilistic analysis of the Davis-Putnam procedure for solving the satisfiability problem, *Disc. Appl. Math.* **5** (1983), 77 - 87.

[39] E. Friedgut, Necessary and sufficient conditions for sharp thresholds of graph properties, and the k-SAT problem, *J. Amer. Math. Soc.* **12** (1999), 1017 - 1054.

[40] E. Friedgut and M. Krivelevich, Sharp thresholds for certain Ramsey properties of random graphs, *Random Structures & Algorithms* **17** (2000), 1 - 19.

[41] E. Friedgut, V. Rődl, A. Rucinski and P. Tetali, in preparation.

[42] A. Frieze and S. Suen, Analysis of two simple heuristics on a random instance of k-SAT, *J. Algorithms* (1996), **20** 312 - 355.

[43] A. Goerdt, A threshold for unsatisfiability, *J. Comput. System Sci.* **53** (1996), 469 - 486.

[44] A. Goerdt and M. Molloy, Analysis of edge deletion processes on random regular graphs, in *Proceedings of Latin American Theoretical Informatics* (2000), pp. 38-47.

[45] S. Janson, D. Knuth, T. Łuczak and B. Pittel, The birth of the giant component, *Random Structures & Algorithms* **3** (1993), 233 - 358.

[46] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley, New York (2000).

[47] S. Janson, Y. Stamatiou and M. Vamvakari, Bounding the unsatisfiability threshold of random 3-SAT, *Random Structures & Algorithms* **17** (2000), 103 - 116.

[48] A. Kamath, R. Motwani, K. Palem and P. Spirakis, Tails bounds for occupancy and the satisfiability threshold conjecture, in *Proceedings of the 35th Annual Symposium on Foundations of Computer Science (FOCS)* (1994), pp. 592-603.

[49] A. Kaporis, L. Kirousis and Y. Stamatiou, A note on the non-colorability threshold of a random graph, *Elec. J. Comb.* **7** (2000), R29.

[50] A. Kaporis, L. Kirousis, Y. Stamatiou, M. Vamvakari and M. Zito, The unsatisfiability threshold revisted, manuscript.

[51] L. M. Kirousis, E. Kranakis, D. Krizanc, and Y. Stamatiou, Approximating the unsatisfiability threshold of random formulas, *Random Structures & Algorithms* **12** (1998), 253 - 269.

[52] J. Komlós and E. Szemerédi, Limit distributions for the existence of Hamilton cycles in a random graph, *Disc. Math.* **43** (1984), 55-63.

[53] T. Łuczak, Size and connectivity of the k-core of a random graph, *Discrete Math.* **91** (1991), 61 - 68.

[54] T. Łuczak, A note on the sharp concentration of the chromatic number on random graphs, *Combinatorica* **11** (1991), 295 - 297.

[55] T. Łuczak, The chromatic number of random graphs, *Combinatorica* **11** (1991), 45 - 54.

[56] T. Łuczak and J. Wierman, The chromatic number of random graphs at the double jump threshold, *Combinatorica* **9** (1989), 39 - 49.

[57] T. Łuczak, Cycles in a random graph near the critical point, *Random Structures & Algorithms* **2** (1991), 421 - 439.

[58] C. McDiarmid, Concentration, in *Probabilistic Methods for Algorithmic Discrete Mathematics* (eds. M. Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed), Springer, Berlin (1998), 195 - 248.

[59] D. Matula, The largest clique size in a random graph, Tech. Rep., Dept. Comp. Sci., Southern Methodist University, 1976.

[60] D. Matula, Expose-and-merge exploration and the chromatic number of a random graph, *Combinatorica* **7** (1987), 275 - 294.

[61] D. Matula and L. Kucera, An expose-and-merge algorithm and the chromatic number of a random graph, in *Random Graphs '87* (eds. M. Karoński, J. Jaworski and A. Ruciński), (1990), pp. 175-188.

[62] M. Mitzenmacher, Tight thresholds for the pure literal rule, Technical Note 1997-011, Digital Systems Research Center, Palo Alto, CA, 1997.

[63] M. Molloy, The Chromatic Number of Sparse Random Graphs., M.Math. Thesis, University of Waterloo, 1992.

[64] M. Molloy, A gap between the appearance of a k-core and a k-chromatic graph, *Random Structures & Algorithms* **8** 159 - 160. (1996),

[65] M. Molloy and B. Pittel, Subgraphs with average degree at least 3 in a random graph, in preparation.

[66] M. Molloy and N. Wormald, in preparation.

[67] B. Pittel, J. Spencer and N. Wormald, Sudden emergence of a giant k-core in a random graph, *J. Comb. Th. B* **67** (1996), 111 - 151.

[68] E. Shamir and J. Spencer, Sharp concentration of the chromatic number on random graphs $G_{n,p}$, *Combinatorica* **7** (1987), 121 - 129.

[69] D. West, *Introduction to Graph Theory*, Prentice Hall, Upper Saddle River, NJ (1996).

Department of Computer Science
University of Toronto
Toronto, Canada
molloy@cs.toronto.edu