Random MAX SAT, Random MAX CUT, and Their Phase Transitions

Don Coppersmith* David Gamarnik* Mohammad Hajiaghayi† Gregory B. Sorkin*

Abstract

With random inputs, certain decision problems undergo a "phase transition". We prove similar behavior in an optimization context.

Specifically, random 2-SAT formulas with clause/variable density less than 1 are almost always satisfiable, those with density greater than 1 are almost always unsatisfiable, and the "scaling window" is in the density range $1 \pm \Theta(n^{-1/3})$. We prove a similar phase structure for MAX 2-SAT: for density $c < 1$, the expected number of clauses satisfiable is $\lfloor cn \rfloor - \Theta(1/n)$; within the scaling window it is $\lfloor cn \rfloor - \Theta(1)$; and for $c > 1$, it is $\frac{3}{4}\lfloor cn \rfloor + \Theta(n)$. (Our results include further detail.)

For random graphs, a maximization version of the giant-component question behaves quite differently from 2-SAT, but MAX CUT behaves similarly.

For optimization problems, there is also a natural analog of the "satisfiability threshold conjecture". Although here too it remains just a conjecture, it is possible that optimization problems may prove easier to analyze than their decision analogs, and may help to elucidate them.

1 Introduction

In this paper, we consider random instances of MAX 2-SAT, MAX k-SAT, and MAX CUT. Just as random instances of the decision problem 2-SAT show a phase transition from almost-sure satisfiability to almost-sure unsatisfiability as the instance "density" increases above 1, so the maximization problem shows a transition at the same point, with the number of clauses *not* satisfied by an optimal solution suddenly changing from $\Theta(1/n)$ to $\Theta(n)$. MAX CUT experiences a similar phase transition: as a random graph's edge density crosses above $1/n$, the number of edges *not* cut in an optimal cut suddenly changes from $\Theta(1)$ to $\Theta(1/n)$.

Our methods are well established ones: the first-moment method for upper bounds; algorithmic analysis including the differential-equation method for lower bounds; and some more sophisticated arguments for the analysis of the scaling window. The interest of the work lies in the simplicity of the methods, and in the results. The question we ask seems very natural, and the answers obtained for MAX 2-SAT and MAX CUT are happily neat, and fairly comprehensive.

1.1 Outlook

Beyond our particular results for MAX 2-SAT and MAX CUT, we hope to spark further work on phase transitions in random instances of other optimization problems, in particular of CSPs (constraint satisfaction problems). Random instances of optimization problems have been studied extensively — some that come to mind are the travelling salesman problem, minimum spanning tree, minimum assignment, minimum bisection, minimum coloring, and maximum clique — but little has been said about *phase transitions* in such cases, and indeed many of the examples do not even have a natural parameter whose continuous variation could give rise to a phase transition.

Many problems, including all CSPs, have natural decision and optimization versions: one can ask whether a graph is k-colorable, or ask for the minimum number of colors it requires. We suggest that in a random setting, the optimization version is quite as interesting as the decision version. Furthermore, optimization problems may plausibly be easier to analyze than decision problems because the quantities of interest vary more smoothly. In fact, a recent triumph in the analysis of a decision problem, the characterization of the "scaling window" for 2-SAT, used as a smoothed quantity the size of the "spine" of a formula [BBC+01]. A way to view our MAX 2-SAT results is that instead of taking the size of the spine as our "order parameter", we take the size of a maximum satisfiable subformula. This seems comparably tractable (we reproduce the result of [BBC+01] incompletely, but more easily), and arguably more natural. Generally, when a decision problem has an optimization analog, the value of the optimum is both interesting in its own right, and an obvious candidate order parameter for studying the decision problem (in contrast to the cleverness of thinking of the spine).

*Department of Mathematical Sciences, IBM T.J. Watson Research Center, Yorktown Heights NY 10598, USA. e-mail {copper,gamarnik,sorkin}@watson.ibm.com

†Department of Mathematics, M.I.T., Cambridge MA 02139, USA. e-mail hajiagha@math.mit.edu

1.2 Problem and motivations Let F be a 2-SAT formula with n variables $X_1, \ldots, X_n$. An "assignment" of these variables consists of setting each X_i to either 1 (True) or 0 (False); we may write an assignment as a vector $\vec{X} \in \{0,1\}^n$. Let $F(\vec{X})$ be the number of clauses satisfied by $\vec{X}$. F is "satisfiable" if there exists an assignment satisfying all the clauses. The problem MAX 2-SAT asks for $\max F \doteq \max_{\vec{X}} F(\vec{X})$, i.e., the maximum, over all assignments $\vec{X}$, of the size (number of clauses) of a maximum satisfiable subformula of F.

We are concerned with this problem in a random setting. Let $\mathcal{F}(n,m)$ denote the set of all formulas with n variables and m clauses, where each clause is proper (consisting of two distinct variables, either of which may be complemented or not), and clauses may be repeated. Let $F \in \mathcal{F}$ be chosen uniformly at random; this is equivalent to choosing m clauses uniformly at random, with replacement, from the $2^2\binom{n}{2}$ possible clauses. Then $\max F$ is a random variable denoting the size of a largest satisfiable subformula of a random formula. We wish to find $f(n,m) \doteq \mathbb{E}(\max F)$, or bounds on this quantity, or other properties of the random variable $\max F$, as a function n and m. Typically, we consider the case $m/n \to c =$ constant as $n \to \infty$, and specifically we investigate the cases when $c < 1$, when c is large, and when c is close to 1. We also consider the "scaling window", which we will show to be the range $m = n \pm \Theta(n^{2/3})$.

Our investigation has two sources of motivation. First, *random* SAT and *max* SAT are both extensively studied, and so it is natural to look at *random max* SAT. On the *random* side, for 2-SAT formulas, it is known that for all $\varepsilon > 0$, as $n \to \infty$, random formulas $F \in \mathcal{F}(n,(1-\varepsilon)n)$ are satisfiable w.p. (with probability) approaching 1, while those in $\mathcal{F}(n,(1+\varepsilon)n)$ are satisfiable w.p. $\to 0$ [CR92a, FdlV92, Goe96]. When $\varepsilon = \Theta(n^{-1/3})$, the probability of satisfiability is bounded strictly between 0 and 1 [BBC+01]. Similarly, random 3-SAT formulas $F \in \mathcal{F}_3(n,cn)$ are known to be satisfiable w.p. $\to 1$ for $c < 3.26$ [AS00] and w.p. $\to 0$ for $c > 4.596$ [JSV00]; for any particular n there is known to be a sharp threshold $c(n)$ around which the transition occurs [Fri99], but it is only conjectured that $c(n)$ converges to a limit.

On the *max* side, it is known for example that for arbitrary 3-SAT formulas F, in polynomial time, $\max F$ can be approximated to within a factor of 7/8 [KZ97], but no better (unless P=NP) [Hås97]. Also, while 2-SAT is solvable in polynomial time, MAX 2-SAT is not: it can be approximated to within a factor 0.940 [LLZ02] but not better than 21/22 [Hås97].

Our second motivation comes from a question of Achlioptas resolved by [BF01] and extended in [BFW02]. A classical question concerns when a "giant component" appears in a graph, as random edges are added one by one: there is a.s. no giant component when there are $(1/2-\varepsilon)n$ edges, and there a.s. is one when there are $(1/2+\varepsilon)n$ edges. Achlioptas asked how long a giant component can be avoided if random *pairs* of edges are given, and from each pair, just one is chosen. Bohman and Frieze showed that at least $0.55n$ edge pairs may be given while avoiding a giant component w.h.p. [BF01]. Bohman, Frieze and Wormald consider the same problem without the "pairing" aspect: how large c can be, such that from a random $2cn$ edges, it is typically possible to select cn edges avoiding a giant component [BFW02]. The "2" is then rather arbitrary; a more general question would be, given cn random edges, how many may be selected to avoid a giant component. We will call this problem "MAX giant-free spanning subgraph".

Because the "phase transition" from satisfiability to unsatisfiability in random 2-SAT (which occurs around clause density 1) is closely related to the appearance of a giant component in a random graph (which occurs around average degree 1), the results of [BFW02] for MAX giant-free spanning subgraph prompted us to look at MAX 2-SAT. As it turns out, these two problems are quite different, and a better graph analog of random MAX 2-SAT is random MAX CUT.

We consider several aspects of random MAX 2-SAT and random MAX CUT. We also extend the easiest results to arbitrary CSPs (constraint satisfaction problems).

2 Notation

We write $F(n,m)$ to denote a random 2-SAT formula. Typically we will fix a constant c and consider $F(n,\lfloor cn \rfloor)$; where it does not matter we will often write cn in lieu of $\lfloor cn \rfloor$. For any formula F, define $\max F$ to be the size of a largest satisfiable subformula of F. Our focus is the functional behavior of $\max F$.

Similarly, we write $G(n,m)$ for a random graph on n vertices with m edges. For any graph G, let $\vec{X}$ describe a partition of the vertices, and let $\operatorname{cut}(G,\vec{X})$ be the number of edges having one vertex in each part of the partition. Define $\max\operatorname{cut}(G) \doteq \max_{\vec{X}} \operatorname{cut}(G,\vec{X})$, and $f_{\text{cut}}(n,m) \doteq \mathbb{E}(\max\operatorname{cut}(G(n,m)))$.

We use standard asymptotic and "order" notation, so for example $f(n) \simeq g(n)$ means $f(n)/g(n) \to 1$ as $n \to \infty$, and $f(n) = o(n)$ means $f(n)/n \to 0$. We will also write $f(n) \lesssim g(n)$ to indicate that f is less than or equal to g *asymptotically* — $\limsup f(n)/g(n) \leqslant 1$ — though it may be that $f(n) > g(n)$ even for arbitrarily large values of n. Asymptotic results involving two variables, for example concerning 2-SAT formulae on n

variables with cn clauses, with c large (or $(1+\varepsilon)n$ clauses with ε small) should always be interpreted as taking the limit in n second; thus "for any desired error bound there exists a c_0, such that for all $c > c_0$ there exists an n_0, such that for all $n > n_0$," etcetera.

3 Summary of results

We establish several properties of random MAX 2-SAT, random MAX k-SAT, and random MAX CUT, focusing on 2-SAT. This section summarizes our main results and indicates the nature of the proofs; further results and proof sketches are given in subsequent sections.

Since for $c < 1$ a random formula $F(n, cn)$ is satisfiable w.h.p., we would expect $\max F$ to be close to cn in this case; the following theorem shows this to be true.

THEOREM 1. *For $c < 1$, $f(n, \lfloor cn \rfloor) = \lfloor cn \rfloor - \Theta(1/n)$.*

The proof comes from counting the expected number of the "bicycles" shown by [CR92b] to be necessary components of an unsatisfiable formula.

For any c, $f(n, cn) \geqslant \frac{3}{4}cn$, since a random assignment of the variables satisfies each clause with probability $\frac{3}{4}$. The next theorem shows that neither this bound nor the trivial upper bound cn is tight, although for large c, $\frac{3}{4}cn$ is close to correct.

THEOREM 2. *For c large, $(\frac{3}{4}c + \sqrt{c}\frac{\sqrt{8}-\sqrt{1}}{3\sqrt{\pi}} - O(1))n \lesssim f(n, cn) \lesssim (\frac{3}{4}c + \sqrt{c}\,\sqrt{3\ln(2)/8})n$.*

The values of $\frac{\sqrt{8}-\sqrt{1}}{3\sqrt{\pi}}$ and $\sqrt{3\ln(2)/8}$ are approximately 0.343859 and 0.509833, respectively. The upper bound is proved by a simple first-moment argument, and the lower bound by analyzing an algorithm; both techniques are exactly those demonstrated in [Spe94, Lecture 6] to analyze the Gale-Berlekamp switching game.

Our next results relate to the low-density case, when c is above but close to the critical value 1. How does $f(n, cn)$ depend on $c = 1 + \varepsilon$ for small ε?

THEOREM 3. *For any fixed $\varepsilon > 0$, $(1 + \varepsilon - \varepsilon^3/3)n \lesssim f(n, (1+\varepsilon)n)$; also, there exist absolute constants α_0 and ε_0, such that for any fixed $0 < \varepsilon < \varepsilon_0$, $f(n, (1+\varepsilon)n) \lesssim (1 + \varepsilon - \alpha_0\varepsilon^3/\ln(1/\varepsilon))n$.*

That is, a constant fraction of the clauses must remain unsatisfied, but this fraction — $\varepsilon^3/3$ at most — is surprisingly small. The lower bound is proved by using the "differential equation method" (see for example [Wor95]) to exactly analyze a version of the unit-clause heuristic. The upper bound's proof is a simple first-moment argument; however, for the probability that a sub-formula with density > 1 is satisfiable, it requires the exponentially small bound given by Bollobás et al. [BBC+01] (see Theorem 9 below). It is likely that, by replacing our use of [BBC+01] with structural properties of the kernel of a sparse random graph, the upper bound's $\varepsilon^3/\ln(1/\varepsilon)$ can be replaced by ε^3 to match the lower bound up to constants [JS02].

The major significance of [BBC+01] was to determine the "scaling window" for random 2-SAT. Without using their result, we prove an analogous result for MAX 2-SAT, and incidentally reproduce part of their 2-SAT result.

THEOREM 4. *Letting $c = 1 + \epsilon = 1 + \lambda n^{-1/3}$, we have $\lfloor cn \rfloor - f(n, \lfloor cn \rfloor) = O(\lambda^3)$ (if $\lambda > 1$); $= \Theta(1)$ (if $-1 \leqslant \lambda \leqslant 1$); $= \Theta(|\lambda|^{-3})$ (if $\lambda < -1$). Also, the probability that $F(n, cn)$ is satisfiable $= \exp(-O(\lambda^3))$ (if $\lambda > 1$); $= \Theta(1)$ (if $-1 \leqslant \lambda < 1$); $= 1 - \Theta(|\lambda|^{-3})$ (if $\lambda < -1$).*

In particular, in the scaling window $c = 1 \pm \lambda n^{-1/3}$, a random formula is satisfiable with probability strictly between 0 and 1 (the exact bounds depending on λ), and it can be made satisfiable by removing a constant-order number of clauses (the constant depending on λ).

We introduce two online versions of MAX 2-SAT and propose an online algorithm suitable for both versions, and satisfying $\simeq (\frac{3}{4}c + \frac{3}{8})n$ clauses; we prove that this algorithm is optimal for one of the two versions.

For MAX k-SAT, including MAX 2-SAT we show that $f(n, m)/m$ is a non-increasing function of m. For the high-density (large-c) case of MAX k-SAT, Theorem 13 gives the analog of Theorem 2. Because little is known about the k-SAT threshold, we are not able to say anything about low-density MAX k-SAT.

Still more generally, Theorem 14 describes the high-density case for any MAX CSP.

Finally, we obtain corresponding results for the closely related MAX CUT problem for sparse random graphs. We previously defined $f_{\text{cut}}(n, cn) \doteq \mathbb{E}(\max\text{cut}\, G(n, cn))$ to be the expected size of a maximum cut in a random graph. We remark that this is the same as the size of a largest bipartite spanning subgraph G' of G, there being a natural correspondence between edges cut in a partition of G, and edges present in G'.

THEOREM 5. *For $c < 1/2$, $f_{\text{cut}}(n, cn) = \lfloor cn \rfloor - \Theta(1)$.*

THEOREM 6. *For c large, $\left(\frac{1}{2}c + \sqrt{c} \cdot \sqrt{8/(9\pi)}\right) n \lesssim f_{\text{cut}}(n, cn) \lesssim \left(\frac{1}{2}c + \sqrt{c}\sqrt{\ln(2)/2}\right) n$.*

The values of $\sqrt{8/(9\pi)}$ and $\sqrt{\ln(2)/2}$ are approximately 0.531922 and 0.588704, respectively. The upper bound was previously obtained in [BCP97].

THEOREM 7. *For any fixed* $\varepsilon > 0$, $(\frac{1}{2} + \varepsilon - \frac{16}{3}(\varepsilon^3))n \lesssim f_{\mathrm{cut}}(n, (1/2+\varepsilon)n) \lesssim (\frac{1}{2} + \varepsilon - \Omega(\varepsilon^3/\ln(1/\varepsilon)))n$.

The upper bound's $\varepsilon^3/\ln(1/\varepsilon)$ can probably be replaced by ε^3, just as for Theorem 3.

4 Random MAX 2-SAT

It is worth pointing out the following simple fact, upon which we will shortly improve.

REMARK 8. *For* $c > 1$, $f(n, cn) \gtrsim n(\frac{3}{4}c + \frac{1}{4})$.

Proof. It suffices to show that for any $\varepsilon > 0$, for all n sufficiently large, $f(n, cn) \geqslant (\frac{3}{4} + \frac{1}{4} - \varepsilon)n$. Select the first $(1-\varepsilon)n$ clauses, and let $\vec{X}$ be a best assignment for it. By Theorem 1, $\vec{X}$ satisfies an expected $(1-\varepsilon)n - o(1)$ of these first clauses, and an expected 3/4ths of the remaining $(c - 1 + \varepsilon)n$ clauses, yielding the claim. □

THEOREM 1: *Proof sketch.* See Summary of Results.

4.1 High-density random MAX 2-SAT

While it is well known that for $c > 1$, $F(n, cn)$ is a.s. unsatisfiable, is it possible that even for c large, *almost* all clauses are satisfiable? Theorem 2 rules this out by showing that a constant fraction of clauses must go unsatisfied; up to a constant, it also provides a matching lower bound.

THEOREM 2: *Proof of the upper bound.* The proof is by the first-moment method. If $\max F > (1-r)cn$ then there is a satisfying assignment of a subformula F' which omits rcn or fewer clauses, and where (taking F' to be maximal) all the omitted clauses are unsatisfied. Any fixed assignment satisfies each (random) clause of F' w.p. 3/4 and unsatisfies each omitted clause w.p. 1/4, so by linearity of expectations, the probability that there exists such an F' is

$$P = \Pr(\exists \text{ satisfiable } F') \leqslant 2^n \sum_{k=0}^{rcn} \binom{cn}{k} (\frac{3}{4})^{cn-k} (\frac{1}{4})^k.$$

For $r < \frac{1}{4}$ the sum is dominated by the last term. From Stirling's formula $n! \simeq \sqrt{2\pi n}\,(n/e)^n$,

$$\binom{cn}{rcn} \simeq 1/\sqrt{2\pi r(1-r)cn} \quad (r^{-r}(1-r)^{-(1-r)})^{cn}. \tag{4.1}$$

Substituting (4.1) into the previous expression,

$$P \lesssim 1/\sqrt{2\pi r(1-r)cn} \cdot \\ \cdot 2^n (r^{-r}(1-r)^{-(1-r)} (3/4)^{1-r} (1/4)^r)^{cn}.$$

Substituting $r = 1/4 - \varepsilon$,

$$\frac{1}{cn} \ln P \lesssim \ln(2)/c - (8/3)\varepsilon^2 + O(\varepsilon^3),$$

so that for $\varepsilon > \sqrt{(3/8)\ln 2/c}$, as $n \to \infty$, $P \to 0$. The conclusion follows.

Proof of the lower bound. The proof is algorithmic. When variables $X_1, \ldots, X_k$ have been set, define the reduced formula F_k in which any clause containing a True literal is removed and "scored", and False literals are removed from the remaining clauses. (Clauses with 0 variables remaining are permanently unsatisfied.) Define a potential function $q(F_k)$ to be the number of clauses already satisfied, plus 3/4 the number of 2-variable clauses ("2-clauses"), plus 1/2 the number of 1-variable clauses ("unit clauses"). Note that randomly assigning the remaining variables satisfies an expected total number of clauses precisely $q(F_k)$, so q is a lower bound on the number of clauses satisfiable.

After variables $X_1, \ldots, X_{k-1}$ have been set to define F_{k-1}, our algorithm sets X_k in whichever of the two ways gives an F_k with larger value $q(F_k)$. (Ties may be broken arbitrarily.) In F_{k-1}, let the number of appearances of X_k and $\bar{X}_k$ in unit clauses be denoted by A_1 and $\bar{A}_1$, and their number of appearances in 2-clauses by A_2 and $\bar{A}_2$. If X_k is set to True, then

$$q(F_k) - q(F_{k-1}) = \Delta_k \doteq \frac{1}{2}(A_1 - \bar{A}_1) + \frac{1}{4}(A_2 - \bar{A}_2),$$

and if X_k is set False, then $q(F_k) - q(F_{k-1}) = -\Delta_k$. Note that $q(F_k) = q(F_{k-1}) + |\Delta_k|$ is a lower bound on the number of satisfiable clauses, and $q(F_0) = \frac{3}{4}cn$.

With $k-1$ variables already set, F_{k-1} a.s. has a.e. $(\frac{1}{2} \pm O(1/\sqrt{c}))2\frac{k-1}{n}\frac{n-k}{n} \cdot cn$ unit clauses, and $\left(\frac{n-k}{n}\right)^2 \cdot cn$ 2-clauses, on the remaining variables. (The reason for $\frac{1}{2} \pm O(1/\sqrt{c})$ instead of $\frac{1}{2}$ is that we set the previous variables in a biased manner.) Also, conditioned on the number of clauses, F_{k-1} is a uniformly random formula (each "slot" being equally likely to be filled by any of the remaining literals). For n large, A_1 and $\bar{A}_1$ are approximated by independent Poisson random variables with parameter $(\frac{1}{2} \pm O(1/\sqrt{c}))\frac{k-1}{n}c$, and A_2 and $\bar{A}_2$ by Poissons with parameter $\frac{n-k}{n}c$. By assumption, c is large, so each of these distributions is approximately Gaussian, and their sum Δ_k is also approximately Gaussian, with mean 0 (by symmetry) and variance

$$\begin{aligned} \sigma_k^2 &= 2 \cdot (\frac{1}{2})^2 \cdot \mathrm{Var}(A_1) + 2 \cdot (\frac{1}{4})^2 \cdot \mathrm{Var}(A_2) \\ &= c\left((\frac{1}{4} \pm O(1/\sqrt{c}))\frac{k-1}{n} + \frac{1}{8}\frac{n-k}{n} \right). \end{aligned}$$

For $Z \sim N(0,1)$, it is well known that $\mathbb{E}|Z| = \sqrt{2/\pi}$; thus $\mathbb{E}|\Delta_k| = \sqrt{2/\pi}\,\sigma_k = \sqrt{2/\pi}\sqrt{c(\frac{1}{4}\frac{k-1}{n} + \frac{1}{8}\frac{n-k}{n})} \pm O(1)$. Finally,

$$\begin{aligned}\mathbb{E}(q_n) &\geqslant \frac{3}{4}cn + \sum_{k=0}^{n-1}\mathbb{E}(|\Delta_k|)\\ &\approx \frac{3}{4}cn + \int_0^n \mathbb{E}(|\Delta_k|)dk\\ &\gtrsim \frac{3}{4}cn + \left(\sqrt{c}\frac{\sqrt{8}-\sqrt{1}}{3\sqrt{\pi}} - O(1)\right)n.\end{aligned}$$

□

4.2 Low-density random MAX 2-SAT For low-density formulas, with $c = 1+\varepsilon$ and $\varepsilon > 0$ a small constant, the bounds of Theorem 2 are inapplicable. It is still true (from Remark 8) that we expect to satisfy at least $(1+\frac{3}{4}\varepsilon)n$ clauses, but it is not obvious whether the best answer is this, or close to the full number of clauses $(1+\varepsilon)n$, or something in between. Theorem 3 shows that $(1+\varepsilon)n - f(n,cn)$, the number of clauses we must dissatisfy, lies between $\Theta(\varepsilon^3 n/\ln(1/\varepsilon))$ and $\Theta(\varepsilon^3 n)$. That is, a linear fraction of clauses must be rejected, but this fraction, at most $\Theta(\varepsilon^3)$, is surprisingly small. We will employ the following theorem of Bollobás et al. [BBC$^+$01] on random 2-SAT.

THEOREM 9. ([BBC$^+$01], Corollary 1.5) *There exist positive constants α_0 and ε_0 such that for any $0 < \varepsilon < \varepsilon_0$ and sufficiently large n,* $\Pr[F(n,(1+\varepsilon)n)$ *is satisfiable*$] \leqslant \exp(-\alpha_0\varepsilon^3 n)$.

(Here, α_0 is the $\liminf$ of the constant implicit in Θ in the theorem in [BBC$^+$01].) The $\exp(-\Theta(\varepsilon^3 n))$ probability of satisfiability in random 2-SAT translates into an expected $O(\varepsilon^3 n/ln(1/\varepsilon))$ unsatisfied clauses in random MAX 2-SAT.

THEOREM 3: *Proof of the upper bound.* The proof is by the first-moment method. Let $c = 1+\varepsilon$. Let F' range over subformulas of F which omit rcn or fewer clauses. Specifying $r < 1/4$, the conditions of Theorem 9 apply, so

(4.2)
$$P = \Pr(\exists \text{ satisfiable } F') \leqslant \sum_{k=0}^{rcn}\binom{cn}{k}e^{-\alpha_0(\varepsilon-\frac{k}{n})^3 n};$$

as $r < 1/4$, the sum is dominated by the last term. Using (4.1) to approximate $\binom{cn}{crn}$,

$$\frac{1}{cn}\ln P \lesssim -r\ln r - (1-r)\ln(1-r) - \alpha_0(\varepsilon - cr)^3/c.$$

First observe that as $\varepsilon \to 0$, for any $r = o(\varepsilon)$, this is

$$= -r\ln r(1+o(1)) - \alpha_0\varepsilon^3(1+o(1)).$$

For any constant $b < 1$, if $r = b\alpha_0\varepsilon^3/\ln(1/\varepsilon)$, this is

$$\begin{aligned}&= b\alpha_0\varepsilon^3(1+o(1)) - \alpha_0\varepsilon^3(1+o(1))\\ &< 0.\end{aligned}$$

That is, it is unlikely that asymptotically fewer than $\alpha_0\varepsilon^3/\ln(1/\varepsilon)$ clauses can go unsatisfied.

Proof of the lower bound (sketch). The proof is algorithmic, and of the sort familiar from [AS00] and previous works. It analyzes a version of the "unit-clause" heuristic. Initially, "seed" the algorithm by randomly deleting a variable from each of, say, $n^{1/10}$ random 2-clauses to convert them to unit clauses. While F has any unit clauses, select one at random and set its variable to satisfy the clause. The analysis consists of counting the clauses unsatisfied in these steps, and justifying the assertion that when there are no more unit clauses, $o(1)$ further clauses need be unsatisfied.

When k variables have been set, let the number of 2-clauses be denoted $m_2(k)$, the number of unit clauses $m_1(k)$, and the number of unset variables $m(k) = n-k$. In one step, the changes in these quantities are $\Delta m = -1$, $\mathbb{E}(\Delta m_2) = -\frac{2}{m}m_2$, and $\mathbb{E}(\Delta m_1) = -1 - \frac{1}{m}m_1 + \frac{1}{m}m_2$ (assuming that $m_1 > 0$ before the step). Over a large number of steps, the net changes will be a.s. a.e. equal to the expectations. Renormalizing with $\rho = m/n$, $\rho_1 = m_1/n$, and $\rho_2 = m_2/n$, the differential equation method (see for example [AS00, Wor95]) asserts that (ρ_1,ρ_2) a.s. a.e. obey the differential equations

$$d\rho_2/d\rho = \frac{2\rho_2}{\rho} \qquad d\rho_1/d\rho = 1 + \frac{\rho_1}{\rho} - \frac{\rho_2}{\rho}.$$

With boundary conditions that for $\rho = 1$ (i.e., initially), $\rho_2 = c$ and $\rho_1 = 0$, the unique solution is

$$\rho_2 = c\rho^2 \qquad \rho_1 = c\rho - c\rho^2 + \rho\ln\rho.$$

This results in $\rho_1 = 0$ at two times: initially, when $\rho = 1$, and also for $\rho = \rho^\star$ satisfying

(4.3)
$$c = \ln(\rho^\star)/(\rho^\star - 1).$$

While $\rho > \rho^\star$, the only clauses ever unsatisfied are unit clauses which contain the negation of the variable being set, and the expected number of such rejected clauses per step is $\frac{1}{2m}m_1 = \frac{\rho_1}{2\rho}$. Integrating over the period $\rho^\star$ to 1,

$$\int_{\rho^\star}^{1} \frac{\rho_1}{2\rho} d\rho = \frac{1}{2}\int_{\rho^\star}^{1} (c - c\rho + \ln\rho)\, d\rho$$
$$= \frac{1}{2}\left(c\rho - c\rho^2/2 + \rho\ln\rho - \rho\right)\Big|_{\rho^\star}^{1}$$

which, substituting for c from (4.3)

$$= \frac{1}{2}(\rho^\star - 1) - \frac{1}{4}(\rho^\star + 1)\ln\rho^\star. \tag{4.4}$$

So from $\rho = 1$ to $\rho = \rho^\star$, the number of clauses dissatisfied by the algorithm is a.s. a.e. n times expression (4.4). After this time, the remaining (uniformly random) 2-SAT formula has density $\rho_2(\rho^\star) / \rho^\star = c\rho^\star = \ln(\rho^\star)/(\rho^\star(\rho^\star - 1)) < 1$ and thus (by Theorem 8) contributes $o(1)$ to the expected number of unsatisfied clauses. In short, the algorithm a.s. fails to satisfy a.e. $(\frac{1}{2}(\rho^\star - 1) - \frac{1}{4}(\rho^\star + 1)\ln\rho^\star)n$ clauses. For $\rho^\star$ (asymptotically) close to 1, the number of dissatisfied clauses is $\simeq n(1-\rho^\star)^3/24$. In particular, with $\varepsilon > 0$ asymptotically small and $c = 1+\varepsilon$, $\rho^\star \simeq 1 - 2\varepsilon$, and the number of dissatisfied clauses is $\simeq n\varepsilon^3/3$. □

Two remarks. First, in addition to the asymptote, the proof gives a precise parametric relationship (as functions of $\rho^\star$) between the clause density c (given by (4.3)) and the rejected-clause density (given by (4.4)). Solving numerically, for $c = 1.5$ we find rejected-clause density ≈ 0.0183275, and for $c = 2$ — where naively the rejected-clause density would be $\frac{1}{4}c = 0.5$ — we achieve rejected-clause density ≈ 0.0809517.

Second, with the solution in hand, the asymptotic behavior is easy to see without the need for differential equations. This alternate proof is less precise but more intuitive and more robust; it is the basis of the analysis within the scaling window (see Theorem 4.

THEOREM 3: *Alternate proof of lower bound.* Consider what happens when $m = (1 - \delta)n$ variables remain unset. The number of 2-clauses is a.s. $m_2 \simeq (1-\delta)^2(1+\varepsilon)n \simeq (1+\varepsilon-2\delta)n$. The expected increase in the number of unit clauses is then $\mathbb{E}(\Delta m_1) = -1-m_1/m+m_2/m \geqslant -1 + m_2/m$ (and the neglected m_1/m is not only conservative, but will also prove to be insignificantly small). Thus, $\mathbb{E}(\Delta m_1) \geqslant -1 + [(1 + \varepsilon - 2\delta)n]/[(1 - \delta)n] \simeq \varepsilon - \delta$. At $\delta = 0$, the number of unit clauses increases by ε per step, this increase linearly falls to 0 per step by $\delta = \varepsilon$, and further to $-\varepsilon$ by $\delta = 2\varepsilon$: the expected number of unit clauses is bounded by an inverted parabola, with base $2\varepsilon n$ and height $\frac{1}{2}\varepsilon^2 n$. At each step about $1/(2n)$th of the unit clauses get dissatisfied. The area under the parabola, times this $1/(2n)$ factor, is $\frac{2}{3} \cdot \text{base} \cdot \text{height} \cdot 1/(2n) = \frac{1}{3}\varepsilon^3 n$.

4.3 The scaling window For random MAX 2-SAT, we have seen that for fixed $c < 1$, $\lfloor cn\rfloor - f(n, \lfloor cn\rfloor) = \Theta(1/n)$, and for $c > 1$, $cn - f(n, cn) = \Theta(n)$. That is, random MAX 2-SAT experiences a phase transition around $c = 1$. It is natural to ask about the scaling window around the critical threshold: What is the interval around $c = 1$ within which $\lfloor cn\rfloor - f(n, \lfloor cn\rfloor) = \Theta(1)$? Theorem 4 shows that the scaling window is $c = 1 \pm \Theta(n^{-1/3})$.

THEOREM 4: *Proof sketch.* We apply a unit-clause heuristic like that in the lower-bound proof from Theorem 3. When there are unit clauses, satisfy a random one; otherwise, set a random variable to a random value. The algorithm proceeds in three phases: Phase 1 applies this unit-clause heuristic for $2\epsilon n$ steps; Phase 2 continues applying unit-clause until $m_1 = 0$; and Phase 3 treats the remaining formula by different means. Without loss of generality we may assume that $\lambda > 1$ or $\lambda < -1$ (results in the middle range follow from these), and if $\lambda < -1$ we go directly to Phase 3.

During Phases 1 and 2, when $t = \delta n$ variables have been set, the number of remaining variables is of course $n(t) = (1 - \delta)n$, and with overwhelming probability (exponentially close to 1), the number of remaining 2-clauses satisfies $m_2(t) = (1+o(1))\,(1+\varepsilon-2\delta)n$. Assuming this holds true, $m_1(t)$ is stochastically dominated by a Poisson-incremented random walk (r.w.) with drift $\varepsilon - \delta$, plus the number of restarts of that r.w. when it hits 0.

In Phase 1, the number of restarts is geometric with parameter $O(\epsilon)$ (mean $1/\varepsilon$), which is insignificantly small. The r.w. itself has drift $\leqslant \varepsilon$, and from Doob's inequality, Pr(r.w. ever exceeds $r\epsilon^2 n) = \exp(-\Omega(r))$.

In Phase 2, the number of restarts is by definition 0, and the r.w. has drift $\leqslant -\varepsilon$; a similar bound holds.

Phases 1 and 2 last for $O(\epsilon n)$ steps; $m_1(t)$ stays bounded by $O(\epsilon^2 n)$; and the expected number of unit clauses unsatisfied during each step is about $m_1(t)/(2n(t))$. Summing over t, the expected number of unsatisfied clauses is $O(\epsilon^3 n) = O(\lambda^3)$.

In Phase 3 the expected number of “bicycles” is $\Theta(\lambda^{-3})$, and this bounds the remaining number of unsatisfied clauses. When $\lambda < -1$, we only do Phase 3; in that case the lower bound follows from applying the second-moment method to the number of bicycles.

The probability results follow similarly. In Phases 1 and 2, the number of clauses unsatisfied at each step is $B(m_1(t), 1/(2n(t)))$, and, conditioned on $m_1(t)$, these values are independent from step to step, giving probability $\exp(-O(\epsilon^3 n))$ that no clauses are unsatisfied during these phases. Independence is preserved in Phase 3, where the probability of satisfiability is at least one minus the expected number of bicycles, i.e., $1 - O(|\lambda|^{-3})$. On the other hand, the probability of satisfiability is at most the probability that there is a "bad" bicycle, which from the second-moment method is $1-\Omega(|\lambda|^{-3})$. □

5 Random MAX k-SAT and MAX CSP

In this section we present some general facts and conjectures about MAX k-SAT and MAX CSP, and extend the previous high-density results.

5.1 Concentration and limits It is known that random k-SAT has a sharp threshold: that is, there exists a threshold function $c(n)$ such that for any $\varepsilon > 0$, as $n \to \infty$, a random formula on n variables with $(c(n) - \varepsilon)n$ clauses is a.s. satisfiable, while one with $(c(n)+\varepsilon)n$ clauses is a.s. unsatisfiable [Fri99]. To prove an analogous result for random MAX k-SAT is much easier; this was first done by [BFU93].

Let $F_k(n,m)$ be a rnadom k-SAT formula on n variables with m clauses, and let $f_k(n,m) = \mathbb{E}(\max F_k)$; we may omit the subscripts k.

THEOREM 10. ([BFU93]) *For all k, n, c, and λ,* $\Pr(|\max F_k(n,cn) - f_k(n,cn)| > \lambda) < 2\exp(-2\lambda^2/(cn))$.

Proof. Let X_i represent the ith clause in F. Replacing X_i with an arbitrary clause cannot change $\max F$ by more than 1. The result follows from Azuma's inequality. □

Since for any fixed k and c we know that $f_k(n,cn) = \Theta(n)$, letting $f'(n,cn) \doteq f(n,cn)/n$, the theorem may be interpreted as saying that for any $\varepsilon > 0$, w.h.p., $(f'(c,n) - \varepsilon)n < \max F(n,cn) < (f'(c,n)+\varepsilon)n$. To conjecture that $f'(c,n)$ is asymptotically independent of n is analogous to the "satisfiability threshold conjecture".

CONJECTURE 11. (MAX k-SAT limit conjecture) *For every k, for almost every constant $c > 0$, as $n \to \infty$, $f_k(n,cn)/n$ converges to a limit.*

Since we can prove asymptotic upper and lower bounds on $f_k(n,cn)/(cn)$, the conjecture's truth would follow if $f_k(n,cn)/(cn)$ were monotone in n, but we do not know if this is so. However, monotonicity of $f_k(n,cn)/(cn)$ in c (with n held fixed) is implied by the next theorem; it shows that as the number of clauses increases, the expected fraction of clauses that can be satisfied can only decrease.

THEOREM 12. *For any k and n, $f_k(n,m)/m$ is a non-increasing function in m.*

Proof. In a uniform random instance of $F_k(n,m)$, let the maximum number of satisfiable clauses be J, so that $\mathbb{E}(J) = f(n,m)$. By deleting single clauses, we obtain m uniform random instances F of $F(n,m-1)$. Of these, $m-J$ each have $\max F = J$, while the remaining J each have $\max F \in \{J-1, J\}$. The average of these m values is at least $\frac{(m-J)(J)+(J)(J-1)}{m} = \frac{J(m-1)}{m}$. Taking expectations, we find $\frac{f(n,m-1)}{m-1} \geqslant \frac{1}{m-1} \times \mathbb{E}(\frac{J(m-1)}{m}) = \mathbb{E}(\frac{J}{m}) = \frac{f(n,m)}{m}$, as desired. □

5.2 Bounds for MAX k-SAT and MAX CSP In this section we extend Theorem 2.

THEOREM 13. *For all k, for all c sufficiently large,* $(\frac{2^k-1}{2^k}c + \frac{2}{k+1}\sqrt{\frac{ck}{\pi 2^k}} - O(1))n \lesssim f_k(n,cn) \lesssim (\frac{2^k-1}{2^k}c + \sqrt{c}\sqrt{\frac{(2^k-1)\ln 2}{2^{2k-1}}})n$.

Note that the leading terms are equal, and the second-order terms equal to within constant $\cdot\sqrt{k}$.

Proof. The proof is similar to that of Theorem 2, and is omitted in the interest of space. □

Still more generally, we may consider a CSP (constraint satisfaction problem). Let g be a k-ary "constraint" function, $g : \{0,1\}^k \to \{0,1\}$. A random formula $F_g(n,m)$ over g is defined by m clauses, each chosen uniformly at random (with replacement) from the $2^k n(n-1)\cdots(n-k+1)$ possible clauses defined by an ordered k-tuple of distinct variables each appearing positively or negated. (Formally, a clause consists of a k-tuple $(i_1,\ldots,i_k)$ of distinct values in $[n]$, specifying the variables, and a binary k-vector $(\sigma_1,\ldots,\sigma_k)$, specifying their signs.) A clause with literals (signed variables) $X_1,\ldots,X_k$ is satisfied if $g(X_1,\ldots,X_k) = 1$. (Formally, an assignment $x_1,\ldots,x_n$ of the full set of variables $X_1,\ldots,X_n$ satisfies a clause as above if $g(x_{i_1} \oplus \sigma_1, \ldots, x_{i_n} \oplus \sigma_n) = 1$, where "$\oplus$" denotes XOR, or addition modulo 2.) As ever, such a formula F is satisfiable if there exists an assignment of the variables satisfying all the clauses; and $\max F$ is the maximum, over all assignments, of the number of clauses satisfied.

Generally a CSP may be based on a finite family of constraint functions, of "arities" bounded by k, but for notational convenience we limit ourselves to a single function.

Let a k-ary clause function g be given, with $\mathbb{E}(g(X)) = p$ over random inputs. Define $P = \min\{p, 1-p\}$ and $Q = 1-P$. Let $F_g(n,m)$ be a random formula over g on n variables, with m clauses, and let $f_g(n,m) = \mathbb{E}(\max F)$.

THEOREM 14. *Given an arity k and a constraint function g, for all c sufficiently large, $(p\,c+\sqrt{PQ^2c/k}\,)n \lesssim f_g(n,m) \lesssim (p\,c+\sqrt{2PQ\ln(2)c}\,)n$.*

6 Online random MAX 2-SAT

In this section, we discuss online versions of the MAX 2-SAT problem. [BF01, BFW02] consider an online version of MAX giant-free spanning subgraph, in which random edges e_i are given one by one, and we must accept or reject e_i based on the previous edges $e_1,\ldots,e_{i-1}$, with the goal of accepting as many edges as possible without creating a giant component.

There are two natural online interpretations of random MAX 2-SAT. In both, we are told in advance the total number of variables n and clauses m; also, in both, clauses c_i are presented one by one, and we must choose "on line" whether to accept or reject c_i based on the previously seen clauses $c_1,\ldots,c_{i-1}$. When we accept a clause we are guaranteeing to satisfy it; when we reject a clause we are free to satisfy or dissatisfy it. Our goal is to maximize the number of clauses accepted.

In our first interpretation of online MAX 2-SAT, ONLINE I, when we accept a clause, we are also required to satisfy it immediately, by setting at least one of its literals True; once a variable is set, it may never be changed. The second interpretation, ONLINE II, is more generous: the variables' assignments may be decided after the last clause is presented. Let $f_{\text{O-I}}(n,m)$ be the expected number of clauses accepted by an optimal algorithm for ONLINE I, and $f_{\text{O-II}}(n,m)$ that for ONLINE II. Clearly, $\frac{3}{4}m \leqslant f_{\text{O-I}}(n,m) \leqslant f_{\text{O-II}}(n,m) \leqslant f(n,m)$. Here we present a "lazy" algorithm applicable to both $f_{\text{O-I}}(n,cn)$ and $f_{\text{O-II}}(n,cn)$. ONLINE-LAZY begins with no variables "set". On presentation of a clause, ONLINE-LAZY rejects it only if it must, and otherwise does the least it can to accept it. Specifically, on presentation of clause c_i, which without loss of generality we may consider to be $(X\vee Y)$, it takes the following action. If $X=$ True or $Y=$ True, accept c_i. If $X=$ False and $Y=$ False, reject c_i. If $X=$ False and Y is unset (or vice-versa), set $Y=$ True (resp. $X=$ True) and accept c_i. If X and Y are both unset, arbitrarily choose one, set it True, and accept c_i.

THEOREM 15. *For any fixed c,* ONLINE-LAZY *is the unique (up to its arbitrary choice) optimal algorithm for* ONLINE I*, and* $f_{\text{O-I}}(n,cn) \simeq (\frac{3}{4}c+(1-e^{-c})/4+(1-e^{-c})^2/8)n \geqslant (\frac{3}{4}c+\frac{3}{8})n$.

We note that for $c=1$, $f_{\text{O-I}}(n,n)\approx 0.957997n$, and for c asymptotically large, $f_{\text{O-I}}(n,cn)\simeq(\frac{3}{4}c+\frac{3}{8})n$.

Proof of optimality (sketch). There are two central ideas. First, the "future" performance of an optimal algorithm is solely a (random) function of the number of unset variables and the number of clauses remaining. Second, we may compare ONLINE-LAZY with a putative optimal algorithm. The only case of interest is when ONLINE-LAZY sets X_k, while the optimal algorithm sets nothing (rejecting a clause instead). We will show that the optimal algorithm then isn't optimal at all. Once the two algorithms part ways, modify ONLINE-LAZY to do exactly what the optimal algorithm does until such time (if any) as the latter sets X_k. If the optimal algorithm never sets X_k, or sets it to the same value as ONLINE-LAZY did, then ONLINE-LAZY (modified) wins by at least one clause. If the optimal algorithm sets X_k oppositely, then the current numbers of clauses satisfied are equal, and the number of unset variables and future clauses are also equal, so the remaining (optimal, expected) performances are equal. That is, in some cases (of probability >0), ONLINE-LAZY (modified) beats the "optimal" algorithm, and in all cases it is at least as good, so the "optimal" algorithm isn't optimal. It follows that an optimal algorithm can never reject a clause having an unset variable.

Proof of performance (sketch). Full proof in Proceedings version. In "round" k, when k variables are yet to be set, the expected number of clauses satisfied is easily calculated (arithmetic omitted!); so is the expected duration of the round. From the expected durations we compute a "nominal" number of rounds before all cn clauses are exhausted; we then show that the true (random) number of rounds is likely to be close to the nominal value. Summing the expected numbers of clauses satisfied in round k over the (random) number of rounds yields the claimed result.

□

Note that ONLINE-LAZY does not, in fact, need to know the number of clauses in advance!

A variant of ONLINE I is that if we accept a clause we must set *both* its variables. In this case, similar arguments show that an optimal algorithm simply sets each new literal True.

We know essentially nothing about ONLINE II.

7 Random MAX CUT

In this section we will show that random MAX CUT behaves like random MAX 2-SAT in nearly every respect. Since the transition from satisfiability to unsatisfiability for a random formula $F(n,m)$ is generally related to the birth of a giant component in a random graph $G(n,m)$, let us first explain why the obvious analogy fails: why random MAX 2-SAT is very different from "random MAX giant-free".

Random MAX giant-free (without the dreadful name) was recently studied by [BF01] and [BFW02], in answer to, and then extending, a question of Achlioptas. For definiteness, they say that $G(n, m)$ has a giant component if it has a component with more than $\ln^2(n)$ vertices. Let MAX giant-free(G) be the size of a largest spanning subgraph G' of G such that G' has no giant component. One dramatic difference between MAX 2-SAT and MAX giant-free is that for arbitrarily high density c, at least 3/4 of the clauses in a random formula $F(n, cn)$ can be satisfied, while for MAX giant-free, at most $1/c$th of the edges in a random graph $G(n, cn)$ can be accepted. (Lemma 1 from [BFW02] can be used to show that for any $\varepsilon > 0$, w.h.p. every subgraph of size $(1 + \varepsilon)n$ of a random graph $G(n, cn)$ contains a linear-size component.)

Intuitively, 2-SAT and giant components in a graph are related because as a graph becomes dense enough to have a giant component, and long cycles, a 2-SAT formula becomes dense enough to have "cycles" of implications including unsatisfiable ones $A \Rightarrow \bar{A} \Rightarrow A$. However, a 2-SAT formula containing cycles of implications can remain satisfiable if no cycle contains two complementary literals, as a graph with cycles can be bipartite if all cycles are even. While these conditions do not hold for a random formula/graph, they do hold for the pruned formulas/graphs that are the solutions to MAX problems. In short, the proper analogy proves to be between satisfiable 2-SAT formulas and bipartite graphs; it just happens that a random graph is bipartite (except for a few small cycles) exactly when it has no giant component. We will now give some more tangible connections between random MAX 2-SAT and random MAX CUT.

To begin with, thinking of a graph edge (u, v) as imposing a "cut constraint" $(u \oplus v)$ on boolean variables u and v, MAX CUT is, like MAX 2-SAT, a MAX CSP with XOR constraints replacing 2-SAT's disjunctions. 2-SAT is polynomial-time solvable, and MAX 2-SAT is 0.940-approximable [LLZ02] but not better than 21/22-approximable [Hås97] in polynomial time, unless P=NP. Similarly, graph bipartiteness is (trivially) polynomial-time solvable, and MAX CUT is 0.878-approximable [GW95] but not better than 16/17-approximable [TSSW00] in polynomial time, unless P=NP.

The methods we have applied to random MAX 2-SAT are equally applicable to MAX CUT, and yield analogous results. In analogy with MAX 2-SAT, for MAX CUT we have the trivial bounds $\frac{1}{2}cn \leqslant f_{\text{cut}}(n, cn) \leqslant cn$.

A difference between random MAX 2-SAT and random MAX CUT is that while below the 2-SAT threshold, $c < 1$, we had $cn - f(n, \lfloor cn \rfloor) = O(1/n)$, below the giant-component threshold, $c < \frac{1}{2}$, the gap does not tend to 0, per the Theorem 5.

THEOREM 5: *Proof.* It is well known that for $c < 1/2$, w.h.p. a random graph $G(n, cn)$ consists of small trees and an expected number $\Theta(1)$ of small unicyclic components, including $\Theta(1)$ with odd cycles. A bipartite spanning subgraph of G must lack at least one edge from each odd cycle, and deleting only these edges gives a bipartite subgraph. □

Despite this difference, like random MAX 2-SAT, random MAX CUT does exhibit a phase transition: while Theorem 5 shows that for $c < 1/2$, the gap $\lfloor cn \rfloor - f(n, cn) = \Theta(1)$, Theorem 6 shows that for $c > 1/2$, the gap jumps to $\Theta(n)$.

7.1 High-density random MAX CUT

THEOREM 6: *Proof.* The proof is similar to that of Theorem 2, and is omitted in the interest of space. The upper bound was shown in [BCP97].

7.2 Low-density random MAX CUT The following fact follows from small-ε asymptotics of classical random graph results; see, e.g., [Bol98, VII.5, Theorem 17].

CLAIM 16. *For $\varepsilon > 0$, a random graph $G(n, (1/2+\varepsilon)n)$ a.s. has a giant component of size $(4\varepsilon + o(\varepsilon))n$.*

CLAIM 17. *The probability that a random graph $G(n, (1/2 + \varepsilon)n)$ is bipartite, conditioned on the existence of a component of size $\Theta(\varepsilon n)$ created by the "first" $(1/2 + \varepsilon/2)n$ edges, is $\exp(-\Omega(\varepsilon^3 n))$.*

Proof. If the presumed giant component is not bipartite, we are done. If it is, by connectivity, it has a unique bipartition; let the sizes of the parts be n_1 and n_2. Each of the remaining $\varepsilon n/2$ edges has both endpoints in the giant component w.p. $\Theta(\varepsilon^2)$, so there are $\Theta(\varepsilon^3 n)$ of these, w.p. $1 - \exp(-\Omega(\varepsilon^3 n))$. The probability that each such edge preserves bipartiteness is $(2n_1 n_2)/(n_1 + n_2)^2 \leqslant 1/2$; over the $\Theta(\varepsilon^3 n)$ independent edges it is $\exp(-\Omega(\varepsilon^3 n))$. □

THEOREM 7: *Proof sketch.* The proof of the upper bound uses the first-moment method and Claim 17. The proof of the lower bound is algorithmic and similar to that of Theorem 3.

8 Conclusions and open problems

We have presented a road map for MAX 2-SAT and MAX CUT in a random setting, establishing that there is a phase transition, and deriving asymptotics below the critical value, for constants slightly above the critical value and in the scaling window around it, and for larger constants. For constant densities slightly above threshold there is a logarithmic gap between our lower and upper bounds; we need to confirm that the $\ln(1/\varepsilon)$ factors are extraneous. In the other cases, our bounds are only separated by a constant. However — especially in light of the exact result of [BFW02] for random MAX giant-free spanning subgraph— it would be wonderful to get the *exact* asymptotics.

If exact asymptotics could be found, that would automatically imply the truth of the MAX k-SAT limit conjecture, Conjecture 11. It is also interesting to speculate even further, on a MAX CSP limit conjecture.

Another issue entirely concerns online random MAX 2-SAT. For ONLINE I, where when accepting a clause we also commit to a variable assignment satisfying it, we argued that ONLINE-LAZY is an optimal algorithm, and we analyzed it exactly. For ONLINE II, however, where we commit to clauses right away but do not commit to variable assignments until the end, the only bounds we have are the weak ones inherited from ONLINE I and the offline problem, $f_{\text{O-I}}(n,m) \leqslant f_{\text{O-II}}(n,m) \leqslant f(n,m)$. It would be interesting to understand this problem better and obtain improved bounds, or, ideally, again to find a provably optimal algorithm and analyze it exactly.

References

[AS00] Dimitris Achlioptas and Gregory B. Sorkin, *Optimal myopic algorithms for random 3-SAT*, 41st Annual Symposium on Foundations of Computer Science, IEEE Comput. Soc. Press, Los Alamitos, CA, 2000, pp. 590–600.

[BBC+01] Béla Bollobás, Christian Borgs, Jennifer T. Chayes, Jeong Han Kim, and David Bruce Wilson, *The scaling window of the 2-SAT transition*, Random Structures and Algorithms **18** (2001), no. 3, 201–256.

[BCP97] Alberto Bertoni, Paola Campadelli, and Roberto Posenato, *An upper bound for the maximum cut mean value*, Graph-theoretic concepts in computer science (Berlin, 1997), Lecture Notes in Comput. Sci., vol. 1335, Springer, Berlin, 1997, pp. 78–84. MR 99d:68185

[BF01] Tom Bohman and Alan M. Frieze, *Avoiding a giant component*, Random Structures and Algorithms **19** (2001), no. 1, 75–85.

[BFU93] Andrei Z. Broder, Alan M. Frieze, and Eli Upfal, *On the satisfiability and maximum satisfiability of random 3-CNF formulas*, Proceedings of the Fourth Annual ACM-SIAM Symposium on Discrete Algorithms (Austin, TX, 1993) (New York), ACM, 1993, pp. 322–330. MR 94b:03023

[BFW02] Tom Bohman, Alan M. Frieze, and Nicholas C. Wormald, *Avoiding a giant component II*, Unpublished manuscript, February 2002.

[Bol98] Béla Bollobás, *Modern graph theory*, Springer, New York, 1998.

[CR92a] Vasěk Chvátal and Bruce Reed, *Mick gets some (the odds are on his side)*, 33th Annual Symposium on Foundations of Computer Science (Pittsburgh, PA, 1992), IEEE Comput. Soc. Press, Los Alamitos, CA, 1992, pp. 620–627.

[CR92b] ______, *Mick gets some (the odds are on his side)*, 33th Annual Symposium on Foundations of Computer Science (Pittsburgh, PA, 1992), IEEE Comput. Soc. Press, Los Alamitos, CA, 1992, pp. 620–627.

[FdlV92] Wenceslas Fernandez de la Vega, *On random 2-SAT*, Manuscript, 1992.

[Fri99] Ehud Friedgut, *Necessary and sufficient conditions for sharp thresholds of graph properties, and the k-SAT problem*, J. Amer. Math. Soc. **12** (1999), 1017–1054.

[Goe96] Andreas Goerdt, *A threshold for unsatisfiability*, J. Comput. System Sci. **53** (1996), no. 3, 469–486.

[GW95] Michel X. Goemans and David P. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, JACM **42** (1995), 1115–1145.

[Hås97] Johan Håstad, *Some optimal inapproximability results*, STOC '97 (El Paso, TX), ACM, New York, 1997, pp. 1–10 (electronic). MR 1 715 618

[JS02] Svante Janson and Gregory B. Sorkin, Personal communication, 2002.

[JSV00] Svante Janson, Yiannis C. Stamatiou, and Malvina Vamvakari, *Bounding the unsatisfiability threshold of random 3-SAT*, Random Structures Algorithms **17** (2000), no. 2, 103–116.

[KZ97] Howard Karloff and Uri Zwick, *A 7/8-approximation algorithm for MAX 3SAT?*, Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science, Miami Beach, FL, USA, IEEE Press, 1997.

[LLZ02] Michael Lewin, Dror Livnat, and Uri Zwick, *Improved rounding techniques for the MAX 2-SAT and MAX DI-CUT problems*, Proc. of IPCO, 2002, pp. 67–82.

[Spe94] Joel H. Spencer, *Ten lectures on the probabilistic method*, second ed., SIAM, 1994.

[TSSW00] Luca Trevisan, Gregory B. Sorkin, Madhu Sudan, and David P. Williamson, *Gadgets, approximation, and linear programming*, SIAM J. Comput. **29** (2000), no. 6, 2074–2097. MR 2001j:68046

[Wor95] Nicholas C. Wormald, *Differential equations for random processes and random graphs*, Ann. Appl. Probab. **5** (1995), no. 4, 1217–1235.