RANDOM 2-SAT WITH PRESCRIBED LITERAL DEGREES

COLIN COOPER, ALAN FRIEZE[†], AND GREGORY B. SORKIN

ABSTRACT. Two classic "phase transitions" in discrete mathematics are the emergence of a giant component in a random graph as the density of edges increases, and the transition of a random 2-SAT formula from satisfiable to unsatisfiable as the density of clauses increases. The random-graph result has been extended to the case of prescribed degree sequences, where the almost-sure nonexistence or existence of a giant component is related to a simple property of the degree sequence. We similarly extend the satisfiability result, by relating the almost-sure satisfiability or unsatisfiability of a random 2-SAT formula to an analogous property of its prescribed literal-degree sequence.

The extension has proved useful in analyzing literal-degree-based algorithms for (uniform) random 3-SAT.

1. INTRODUCTION

There is considerable interest at present in displaying sharp transitions of probabilistic properties in combinatorial settings. One case of interest is that of random k-SAT formulae. In this note we discuss a model of random 2-SAT. In the standard model we have n variables and m random clauses. This model is well understood. Chvatál and Reed [CR92] showed that if $m = cn$, $c < 1$ constant then a random instance is satisfiable with high probability (**whp**) and that if $c > 1$ then a random instance is unsatisfiable **whp**. This result was sharpened by Goerdt [Goe96], Fernandez de la Vega [FdlV92] and Verhoeven [Ver99]. The tightest results are due to Bollobás, Borgs, Chayes, Kim and Wilson [BBC+01].

Just as in the case of the existence of a giant component in a random graph, Molloy and Reed [MR95], we obtain interesting results by considering models in which the number of occurrences of each literal is prescribed.

1.1. Notation.

Certain formalisms become confusing if not dealt with at once. An n-variable formula is built on a set of variables $V = \{x_1, \ldots, x_n\}$ and their complements $\{\bar{x}_1, \ldots, \bar{x}_n\}$, all $2n$ of which compose the set L of literals. Complementation pairs the literals, so complement$(x_i) = \bar{x}_i$ and complement$(\bar{x}_i) = \bar{x}_i$. For an arbitrary literal u we denote the complement by $\bar{u}$, so if $u = x_i$ then $\bar{u} = \bar{x}_i$, and if $u = \bar{x}_i$ then $\bar{u} = x_i$. We say that two literals u and v are distinct as variables if neither $u = v$ nor $\bar{u} = v$. A truth assignment $\sigma : V \to \{0, 1\}$ assigns each variable a value of 1 (true) or 0 (false), and extends naturally to the set of literals by $\sigma(\bar{x}_i) = 1 - \sigma(x_i)$. (We will often elide the function σ, simply writing for example $x_i = 0$; the complementarity condition then is that $x_i + \bar{x}_i = 1$.)

A clause C is an unordered pair of literals $\{u, v\}$, and C is satisfied by a truth assignment σ if $\sigma(u) + \sigma(v) \geq 1$. A 2-SAT formula F on n variables with m clauses consists of clauses $C_1, \ldots, C_m$ over the literals $\{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}$; it is satisfiable if there exists a truth assignment σ satisfying every clause. We say that F is *simple* if all the clauses are distinct and, in each clause $C_i = \{u_i, v_i\}$, u_i and v_i are distinct as variables.

Date: 2 May 2003.

[†] Supported by NSF grant CCR-0200945.

For $w \in L$ let $d_F(w)$ denote the *degree* of w, that is, the number of times w appears in the formula F. Suppose now that we fix a degree sequence $\mathbf{d} = d_1, \bar{d}_1, \ldots, d_n, \bar{d}_n$ and define the set of all formulas with degree sequence $\mathbf{d}$,

$$\Omega_{\mathbf{d}} = \left\{F : \; d_F(x_i) = d_i, d_F(\bar{x}_i) = \bar{d}_i, \; i = 1, 2, \ldots, n\right\}.$$

Denote the maximum degree by $\Delta_{\mathbf{d}} = \max\left\{d_1, \bar{d}_1, \ldots, d_n, \bar{d}_n\right\}$, and let

$$\begin{aligned} D_1 &= \sum_{i=1}^{n} (d_i + \bar{d}_i) = 2m \\ D_2 &= \sum_{i=1}^{n} d_i \bar{d}_i \end{aligned}$$

where m is the number of clauses in F.

We can assume that $d_i + \bar{d}_i \geq 1$ for all i. (Otherwise we can remove variable i from consideration.) Thus $D_1 \geq n$. Our random model is that

$$F \text{ is chosen uniformly at random from } \Omega_{\mathbf{d}}.$$

A degree sequence $\mathbf{d}$ is Δ-*proper* if

- $\Delta_{\mathbf{d}} \leq \Delta$.
- $D_1 = 2m$, i.e., D_1 is even.

We will write $f \ll g$ to mean $f = o(g)$ and $f \gg g$ for $g = o(f)$ whenever convenient.

1.2. Results and significance.

We prove the following criterion for satisfiability.

Theorem 1. *Let $0 < \epsilon < 1$ be constant and $n \to \infty$. Let $\mathbf{d}$ be any Δ-proper literal-degree sequence over n variables, with $\Delta = n^{1/11}$, and let F be a uniform random simple formula with degree sequence $\mathbf{d}$.*

(A): *If $2D_2 < (1 - \epsilon)D_1$ then*

$$\mathbf{P}(F \textit{ is satisfiable}) \to 1.$$

(B): *If $2D_2 > (1 + \epsilon)D_1$ then*

$$\mathbf{P}(F \textit{ is satisfiable}) \to 0.$$

Both limits are uniform in n (independent of $\mathbf{d}$).

For example in the case of the usual uniform model, with $m = cn$ randomly chosen clauses, we find that $D_1 = 2cn$ and **whp** $D_2 \approx c^2 n$, and we obtain the result of [CR92].

Constraint satisfaction problems are of interest in many fields, notably artificial intelligence, and there is a natural hope that understanding random formulas will help in understanding the instances that arise in those fields. By relaxing the assumption of uniform randomness, we increase the chance of making such a connection.

Theorem 1 has already been useful in a computer science-theoretic setting. In analogy to the 2-SAT phase transition result of [CR92], it is conjectured that there is a similar transition density for 3-SAT; while this remains unresolved, bounds are known: n-variable formulas with fewer than $3.4n$ clauses are almost surely satisfiable [KKL02], and those with greater than $4.596n$ clauses almost surely unsatisfiable [JSV00].

The $3.4n$ result follows from analyzing an algorithm which, essentially, chooses high-degree literals and sets them to 1. As with previous satisfiability bounds such as the $3.26n$ of [AS00], the algorithmic analysis relies on the "differential equation method" [Wor95], and it is in the nature of this method (which works in an open set) that the last part of the job generally must be done by some other technique. In the case of [AS00], the completion relies on the fact that a uniform random formula with fewer than n 2-clauses and $(2/3)n$ 3-clauses is almost surely satisfiable. For [KKL02], the completion comes from taking a now-sparse formula, treating its 3-clauses as if they were 2-clauses by ignoring a random literal, and applying our main theorem.

2. Random Formulas

This section is devoted to a statement and proof of Lemma 2, which parallels Theorem 1 but for formulas generated by a random configuration model to be described, rather than for simple random formulas. In Section 3 we will use Lemma 2 to prove Theorem 1.

Graphical Representation

A clause $\{u_j, v_j\}$ corresponds naturally to a pair of logical implications: for F to be satisfied, if u_j is 0 (false) then v_j must be 1 (true), and vice-versa; that is, $\bar{u}_j$ "implies" v_j, and $\bar{v}_j$ implies u_j.

Given a formula $F = \{\{u_j, v_j\} : j = 1, 2, \ldots, m\}$ we define a digraph $\Gamma = \Gamma(F) = (L, A)$ whose vertices are F's literals, and whose directed edges consist of the two implications derived from each of F's clauses: $A = \{(\bar{u}_j, v_j),\ (\bar{v}_j, u_j) : j = 1, 2, \ldots, m\}$.

It is well known (see for example Aspvall, Plass and Tarjan [APT79]) that F is unsatisfiable if and only if there is a variable x_j such that $\Gamma(F)$ contains a directed path from x_j to $\bar{x}_j$ *and* a directed path from $\bar{x}_j$ to x_j.

Configuration Model

Our model for generating a random $F \in \Omega_{\mathbf{d}}$ is based on the configuration model for graphs, Bollobás [Bol80]. We have a universe Z consisting of D_1 *points*, partitioned into subsets $Z(x), x \in L$, with $|Z(x_i)| = d_i$, $|Z(\bar{x}_i)| = \bar{d}_i$, $i = 1, 2, \ldots, n$; the points in $Z(x)$ are thought of as representatives, or occurrences, of literal x. "Inversely" to $Z(x)$, define $\phi : Z \to L$ by $\phi(w) = x$ iff $w \in Z(x)$; ϕ associates a point with the literal it represents. Let Ψ denote the set of *configurations*: partitions of Z into m disjoint 2-element sets. From a configuration $P \in \Psi$, we construct a formula F_P as follows: for each 2-element set $S = \{p, q\} \in P$ we create a clause $C_S = \{\phi(p), \phi(q)\}$.

An algorithmic description of the generation of a uniformly random $P \in \Psi$ can be useful:

Algorithm CONSTRUCT
begin
  paired-up points $P_0 := \emptyset$; free points $R_0 := Z$
  For $i = 1$ **to** m **do**
  begin
    Choose $u_i \in R_{i-1}$ *arbitrarily*
    Choose v_i *uniformly at random* from $R_{i-1} \setminus \{u_i\}$
    $P_i := P_{i-1} \cup \{\{u_i, v_i\}\}$; $R_i := R_{i-1} \setminus \{u_i, v_i\}$
  end
  Output $P := P_m$.
end

In the *random-configuration* model we choose P uniformly at random from Ψ and let F_P be the associated formula. F_P may not be *simple*, i.e., it may contain repeated clauses and/or clauses which contain 2 copies of the same literal. If however P is simple, then F_P is uniformly sampled from $\Omega_{\mathbf{d}}$: each simple formula is represented by exactly $\prod_{i=1}^{n} d_i!\, \bar{d}_i!$ distinct configurations. (By contrast, not-necessarily-simple formulas are *not* all represented by the same number of configurations, and here we do not consider the uniform distribution over such formulas.)

The distinction between random formulas and random simple formulas is typically unimportant in the usual (non-degree-sequence) model, because there, a constant fraction of formulas are simple. That is not the case in the degree-sequence model, and so the distinction is more important; we discuss this in section 3.

We will first study the likely satisfiability of F_P — a formula which is not necessarily simple — and later, in Section 3, show how to deal with the issue of simplicity.

Lemma 2. *Under the hypotheses of Theorem 1, but with $F = F_P$ a formula given by a random configuration, precisely the same conclusions (A) and (B) follow.*

The remainder of this section is devoted to a proof of Lemma 2; Part (A) is easy, Part (B) harder.

2.1. Case A: $2D_2 < (1-\epsilon)D_1$.

In this section we prove the result of part (A) of Lemma 2.

Chvátal and Reed [CR92] define a *bicycle* as a sequence of clauses $\{u, w_1\}, \{\bar{w}_1, w_2\}, \dots, \{\bar{w}_r, v\}$ where $w_1, w_2, \dots, w_r$ are distinct as *variables* and, for some $1 \le i, j \le r$, $u \in \{w_i, \bar{w}_i\}$ and $v \in \{w_j, \bar{w}_j\}$. That is, a bicycle is a chain of implications on distinct variables, except that some two middle literals, or their complements, also serve as the start and end literals. (It is the distinctness of the variables that makes bicycles a useful concept, for it translates into probabilistic independence.)

Chvátal and Reed argue that if an instance is infeasible then it contains a bicycle. We will show that **whp** $\Gamma(F_P)$ does not contain any bicycles. It is convenient first to show that **whp** $\Gamma(F_P)$ does not contain any long paths. Then we can restrict our attention to small bicycles.

Claim 3. *$\Gamma(F_P)$ has no long directed paths, **whp**.*

Proof. Let $k_0 = \lceil 3\epsilon^{-1}\log n \rceil$ and let X_0 be the number of directed paths of length $k_0 - 1$ in $\Gamma(F_P)$. In the estimation of $\mathbf{P}(w_1 \to w_2 \to \cdots \to w_{k_0} \in \Gamma)$ below, we do not use the procedure CONSTRUCT. Rather, first, $\mathbf{P}(w_1 \to w_2) \le \mathbf{E}(\#\{w_1 \to w_2\}) = d(\bar{w}_1) \cdot d(w_2)/(D_1 - 1)$. Then, $\mathbf{P}(w_2 \to w_3 \mid w_1 \to w_2)$ has the same form, but in a random configuration where $d(\bar{w}_1)$ and $d(w_2)$ have both been reduced by 1; thus, $\mathbf{P}(w_2 \to w_3 \mid w_1 \to w_2) \le d(\bar{w}_2)d(w_3)/(D_1 - 3)$, and so forth.

$$
\begin{aligned}
\mathbf{E}(X_0) &\le \sum_{w_1,\dots,w_{k_0} \in L} \mathbf{P}(w_1 \to w_2 \to \cdots \to w_{k_0} \in \Gamma) \\
&\le \sum_{w_1,\dots,w_{k_0} \in L} \frac{d(\bar{w}_1)d(w_2)}{D_1 - 1} \times \frac{d(\bar{w}_2)d(w_3)}{D_1 - 3} \times \cdots \\
&\qquad \times \frac{d(\bar{w}_{k_0-1})d(w_{k_0})}{D_1 - 2k_0 + 3} \\
&\le \sum_{w_1, w_{k_0}} \frac{\Delta^2}{D_1 - 2k_0} \sum_{w_2,\dots,w_{k_0-1}} \prod_{i=2}^{k_0-1} \frac{d(w_i)d(\bar{w}_i)}{D_1 - 2k_0} \\
&\le \frac{n^2\Delta^2}{D_1 - 2k_0} \left(\frac{2D_2}{D_1 - 2k_0} \right)^{k_0-2} \\
&\le (1+o(1))\, n\Delta^2 \left((1-\epsilon)(1+o(1)\right)^{k_0-2} \\
&= O(n\Delta^2 n^{-2.9}) \\
&= o(1).
\end{aligned}
\tag{1}
$$

So, **whp**, Γ has no directed path of length $\ge k_0$. □

From Claim 3 it follows immediately that **whp**, Γ contains no long bicycles, of length $> k_0$.

Claim 4. *$\Gamma(F_P)$ has no short bicycles, **whp**.*

Proof. Let Y_r be the number of bicycles of length r, and $Y = \sum_{r=2}^{k_0} Y_r$. Then the probability that Γ contains a short bicycle is

$$\begin{aligned}
\mathbf{P}(Y>0) &\le \mathbf{E}(Y)\\
&\le \sum_{r=2}^{k_0} \sum_{\substack{w_1,\dots,w_r\in L\\ \pm i,\pm j}} \frac{d(\bar{w}_1)d(w_2)}{D_1-1} \times \frac{d(\bar{w}_2)d(w_3)}{D_1-3} \times \\
&\quad \cdots \times \frac{d(\bar{w}_{r-1})d(w_r)}{D_1-2r+5} \qquad (2)\\
&\quad \times \frac{\Delta d(w_1)}{D_1-2r+3}\frac{\Delta d(\bar{w}_r)}{D_1-2r+1}\\
&\le (1+o(1))\frac{4\Delta^2}{D_1}\times\\
&\quad \times \sum_{r=2}^{k_0}\sum_{w_1,\dots,w_r\in L} r^2 \prod_{i=1}^{r} \frac{d(w_i)d(\bar{w}_i)}{D_1}\\
&= (1+o(1))\frac{4\Delta^2}{D_1}\sum_{r=2}^{k_0} r^2\left(\frac{2D_2}{D_1}\right)^r\\
&\le (1+o(1))\frac{4\Delta^2}{D_1}\frac{\epsilon(1+\epsilon)}{(1-\epsilon)^3} \qquad (3)\\
&= o(1).
\end{aligned}$$

Here we have used that $\Delta = o(n^{1/2})$, much weaker than the hypothesis $\Delta = n^{1/11}$. (Throughout this article, we shall frame the constraints on various parameters. These parameters weave through the propositions leading up to the main results in Theorem 1 and Lemma 2, and optimizing the parameters subject to the framed constraints leads to the parameter values in the main results.) □

The preceding pair of claims shows that **whp**, Γ contains no bicycles and thus is satisfiable; this verifies Part (A) of Lemma 2.

2.2. **Case B:** $2D_2 > (1+\epsilon)D_1$.
For $w \in L$ we let

$$\mathrm{span}(w) = \{v : \ \Gamma(F_P) \text{ contains a directed path from } w \text{ to } v\}.$$

We show that **whp** there exists a literal w and variables x, y such that $x, \bar{x} \in span(w)$ and $y, \bar{y} \in span(\bar{w})$. This forces the formula to be unsatisfiable since

$$w \Longrightarrow x \wedge \bar{x} \text{ and } \bar{w} \Longrightarrow y \wedge \bar{y}. \qquad (4)$$

We do this by arguing that we can **whp** find a pair $w, \bar{w}$ such that both $\mathrm{span}(w)$ and $\mathrm{span}(\bar{w})$ are "large" and that **whp** large spans contain complementary pairs.

Generating spans

In generating a configuration P, that is a pairing of the configuration points Z, we will consider the set Z to be partitioned into three sets: a set D of "dead", paired-up points (which only grows with time); a set L of "live" points actively seeking pairs; and a set U of "untouched" points. We will call $L \cup U$ the set of "unpaired" points.

We work in terms of the configuration model and make a series of passes of the following "truncated span" algorithm, TSPAN. (If we were to repeat until all points are paired up, this would be a particular version of CONSTRUCT; in fact we stop earlier.) TSPAN takes as arguments initial values of D, L, and U, an iteration threshold s, and a live-size threshold ℓ. The algorithm causes the sets D, L, U to evolve, preserving the property that they are a partition of Z.

TSPAN(D, L, U, s, ℓ)

> **While** $0 < |L| \le \ell$ (while there are live points but no more than ℓ of them), and for at most s iterations:
> > **Choose** a live point $p \in L$.
> > **Randomly choose** an unpaired point $p' \neq p$ ($p' \in U \cup L \setminus p$).
> > **Pair** p with p' and move them to the paired-up set. (Set $D \leftarrow D \cup \{p, p'\}$, $L \leftarrow L \setminus \{p, p'\}$, $U \leftarrow U \setminus \{p, p'\}$.)
> > **Identify** the literal v represented by p'. (Set $v = \phi(p')$.)
> > **Make live** the free points associated with $\bar{v}$. (Set $L \leftarrow L \cup (U \cap \phi^{-1}(\bar{v}))$, $U \leftarrow U \cap \phi^{-1}(v)$).

Note that in a TSPAN pass starting with $L = \{p\}$, with $u = \phi(p)$, all the literals v identified in the pass are implied by u: if u is 1 then v must be 1 to satisfy the formula F_P (in particular, to satisfy the subformula generated in the pass).

The Plan

Let us now outline our grand scheme. We will run a sequence of executions of TSPAN, doing it few enough times and each with a short enough time bound that we can ensure that the number of points we ever touch (pair up) is a small fraction of the total. More specifically, we will perform a sequence of "pair iterations" as follows.

Select a pair of "complementary points" $p, q \in Z$, with $\phi(p) = \overline{\phi(q)}$. Define a run of TSPAN to be a "success" if it terminates because $|L| = \ell$, rather than because $|L| = 0$ or because of the iteration bound. Run TSPAN with the current set of dead points D, with $L = \{p\}$ (and with U the rest). If the run fails, perform a new pair iteration. Also, if q got paired up, perform a new pair iteration. Claim 6 will show that the run succeeds with decent probability, and Claim 5 shows that it is unlikely to pair up q in the process. Given this, run TSPAN with $L = \{q\}$. If the run on q fails, perform a new pair iteration. Claim 7 shows that any pair p, q succeeds with decent probability. Once we find a pair p, q both of which succeeded, Claims 8 and 10 (with Remark 9) show that with high probability both p and q imply contradictions, and thus F is unsatisfiable.

Let $t = \frac{1}{2}|D|$ denote the number of point-pairings, a natural index of "time" as points are repeatedly paired in a single or multiple executions of TSPAN, and let $|L(t)|$ be the number of live points at this time. We shall throughout respect the condition that

$$|D| + |L| = o(D_2/\Delta^2), \tag{5}$$

which implies that $t = o(D_1)$.

Claim 5. *Given condition* (5), *each pairing ("while") step in* TSPAN *gives an expected increase in the live points,*

$$\mathbf{E}(|L(t+1)| - |L(t)|) \ge \frac{\epsilon}{2}. \tag{6}$$

Also, for any untouched point, $q \in U(t)$, the probability q gets paired with p or made live is

$$\mathbf{P}(q \notin U(t+1)) \le (1 + \Delta)/(D_1 - 2t - 1). \tag{7}$$

Proof. Start with the first property. Let d'_j be the number of unpaired points representing literal x_j. The point p' paired with p is equally likely to be any of the $D_1 - 2t - 1$ other unpaired points. Call a literal x_j untouched if none of the points representing x_j or $\bar{x}_j$ is in $D \cup L$, *i.e.*, j is untouched if $(Z(x_j) \cup Z(\bar{x}_j)) \cap (D \cup L) = \emptyset$. Note that the number of touched literals is $\le |D \cup L|$. If $x_j = \phi(p')$ is untouched, then all $\bar{d}_j$ representatives of x_j become new live points. At the same time, at least one live point (p) gets paired up, and possibly a second (if $p' \in L(t)$). Thus the expected increase in the number of live points is

$$\begin{aligned}
&\mathbf{E}(|L(t+1)| - |L(t)|) \\
&\geq -1 - \frac{|L|}{D_1 - 2t - 1} + \frac{1}{D_1 - 2t - 1} \sum_{j \text{ untouched}} (d_j \bar{d}_j + \bar{d}_j d_j) \\
&\geq -1 + \frac{2}{D_1 - 2t - 1} \left(\sum_{j=1}^{n} d_j \bar{d}_j - \sum_{j \text{ touched}} d_j \bar{d}_j - \frac{1}{2}|L| \right) \\
&\geq -1 + \frac{2}{D_1 - 2t - 1} \left(D_2 - |L \cup D| \cdot \Delta^2 - \frac{1}{2}|L| \right) \\
&= -1 + \frac{2}{D_1 - o(D_1)} (D_2 - o(D_2)) \\
&\geq -1 + \frac{2D_2}{D_1}(1 - o(1)) \\
&\geq -1 + (1+\epsilon)(1 - o(1)) \\
&\geq \epsilon/2.
\end{aligned}$$

For the second property, the probability that q is paired with p is $1/(D_1 - 2t - 1)$. For q to become live, p has to be paired with one of at most Δ points q' with $\phi(q') = \overline{\phi(q)}$. □

Claim 6. *For $s \gg \Delta^2 \epsilon^{-2}$ and $s = o(D_2/\Delta)$, if* TSPAN *is run starting with no more than $t = o(D_2/\Delta^2)$ paired-up points D, with any single live point $L = \{p\}$, with a time bound s and size bound $s\epsilon/4$, then with probability $P \geq 1/(2s)$ it terminates with live size $s\epsilon/4$.*

Proof. Fix an arbitrary linear order on the configuration points Z. Let $\sigma_1, \sigma_2, \ldots \in [0,1)$ be independent uniform random reals. Implement TSPAN by, in the ith iteration, choosing the "first" point $p \in L$ (in the fixed ordering), and pairing it with $p' \in L \cup U \setminus p$, where p' is the $\lceil \sigma_i(|L| + |U| - 1) \rceil$th element of $L \cup U \setminus p$.

The preceding process may terminate at $i < s$ steps, if the live size $|L|$ hits 0 or ℓ; we now define another process which always continues for s steps. If $|L(i-1)| > 0$ (even if $|L(i-1)| \geq s\epsilon/4$), just proceed with TSPAN as above, using σ_i. If $|L(i-1)| = 0$, then restore D, L, and U to their initial values before making a single step of TSPAN as above, using σ_i. We call the latter case a "restart", and define the jth "start time" by $I_j = i - 1$ (with $I_1 = 0$).

Let $\xi_j = 1$ if during the jth start we ever achieve $|L| \geq s\epsilon/4$ ("success"), and $\xi_j = 0$ if there is no jth start or it fails to achieve this live-size. The probability of interest is $P = \mathbf{P}(\xi_1 = 1)$, since this is the probability that the original process TSPAN reaches the size bound $s\epsilon/4$. Note that the jth start is (stochastically) just like the first, except with time bound $s - I_j$ in lieu of s, and therefore ξ_j is stochastically dominated by ξ_1. In particular, if $|L(s)| \geq s\epsilon/4$ then one of the starts (the last one) was successful, and there are no more than s starts, which is to say that

$$\mathbf{P}(|L(s)| \geq s\epsilon/4) \leq \mathbf{P}(\sum_{j=1}^{s} \xi_j \geq 1) \leq \mathbf{E} \sum_{j=1}^{s} \xi_j \leq s\mathbf{E}\xi_1 = s\mathbf{P}(\xi_1 = 1),$$

and thus

$$P = \mathbf{P}(\xi_1 = 1) \geq \mathbf{P}(|L(s)| \geq s\epsilon/4)/s. \tag{8}$$

It remains only to find a good bound on the latter quantity, which we do in a manner patterned on the Azuma-Hoeffding inequality.

Define the differences

$$X_i = |L(i)| - |L(i-1)| \tag{9}$$

for the process above. Note that $-2 \le X_i$ (at worst 2 live points get paired up and no new ones created); $X_i \le \Delta$ (at least 0 live points get paired up — 0 rather than 1 in the case of a restart where the single live point p is reintroduced and immediately paired off again — and at most Δ new ones created); and (whether or not we are making a restart), by (6),

$$\mathbf{E}(X_i \mid \sigma_1, \ldots, \sigma_{i-1}) \ge \epsilon/2. \tag{10}$$

In particular, if we set

$$\lambda = \epsilon/(6\Delta^2) \tag{11}$$

then for any values $\sigma_1, \ldots, \sigma_{i-1}$,

$$\begin{aligned}
\mathbf{E}\left(e^{-\lambda X_i \mid \sigma_1,\ldots,\sigma_{i-1}}\right) &\le 1 - \lambda \mathbf{E}(X_i \mid \sigma_1, \ldots, \sigma_{i-1}) + \lambda^2\Delta^2 \\
&\le e^{-\lambda\epsilon/2+\lambda^2\Delta^2} \\
&\le e^{-\lambda\epsilon/3}. && (12)
\end{aligned}$$

Note that $X = \sum_{i=1}^{s} X_i = |L(s)|$, and we are interested in $\mathbf{P}(X \le s\epsilon/4)$. For any bound w,

$$\begin{aligned}
\mathbf{P}(X \le w) &= \mathbf{P}(e^{\lambda(w-X)} \ge 1) \\
&\le \mathbf{E}(e^{\lambda(w-X)}) = e^{\lambda w}\mathbf{E}e^{-\lambda X} \\
&= e^{\lambda w}\mathbf{E}\left(\prod_{i=1}^{s} e^{-\lambda X_i}\right) \\
&= e^{\lambda w}\mathbf{E}_{\sigma_1,\ldots,\sigma_{s-1}}\mathbf{E}_{\sigma_s}\left(\prod_{i=1}^{s-1} e^{-\lambda X_i} \cdot e^{-\lambda X_s}\right) \\
&= e^{\lambda w}\mathbf{E}_{\sigma_1,\ldots,\sigma_{s-1}}\left(\prod_{i=1}^{s-1} e^{-\lambda X_i} \cdot \mathbf{E}_{\sigma_s}\left(e^{-\lambda(X_s \mid \sigma_1,\ldots,\sigma_{s-1})}\right)\right).
\end{aligned}$$

Applying (12), this is

$$\le e^{\lambda w}\mathbf{E}_{\sigma_1,\ldots,\sigma_{s-1}}\left(\prod_{i=1}^{s-1} e^{-\lambda X_i} \cdot e^{-\lambda\epsilon/3}\right)$$

which, inductively, is

$$\le e^{\lambda w - \lambda\epsilon s/3}. \tag{13}$$

Taking $w = s\epsilon/4$ as desired for (8), this is

$$\begin{aligned}
&\le e^{-\lambda\epsilon s/12} \\
&\le e^{-s\epsilon^2/(72\Delta^2)} && (14)
\end{aligned}$$

for λ as in (11). With $s \ge 72\Delta^2/\epsilon^2$ as hypothesized, this is

$$\le 1/2. \tag{15}$$

Together with (8) this proves the claim. □

Claim 7. *Given a set D of paired-up points with $|D| = o(D_2/\Delta)$, choose a pair of complementary points $p, q \in Z \setminus D$, so $\phi(p) = \overline{\phi(q)}$. Run* TSPAN *with paired-up points D, a single live point $L = \{p\}$, and $U = Z \setminus D \cup \{p\}$, with iteration bound s as prescribed in Claim 6 and also satisfying $s^2 = o(D_1/\Delta)$, and live-size bound $s\epsilon/4$, to produce a larger set D' of paired-up points.* If *the live-size bound was reached* and $q \notin D'$, *similarly run* TSPAN *with paired-up points D', a single live point $L = \{q\}$, and $U = Z \setminus D' \cup \{q\}$.*

Then, with probability $\geq 1/(5s^2)$, both runs occur, and both reach live-size $s\epsilon/4$.

Proof. By (7), the probability that the TSPAN run starting from p "kills" q is $\leq s(1+\Delta)/(D_1 - |D| - 1) = o(1/s)$ by hypothesis. Let P be the event that TSPAN starting from p grows to size $s\epsilon/4$, Q the same event for q, and K the event that p does not kill q. Then,

$$\mathbf{P}(P \wedge \neg K) \geq \mathbf{P}(P) - \mathbf{P}(K) \geq 1/(2s) - o(1/s).$$

Then

$$\begin{aligned}\mathbf{P}(Q \wedge (P \wedge \neg K)) &= \mathbf{P}(P \wedge \neg K) \cdot \mathbf{P}(Q \mid (P \wedge \neg K)) \\ &\geq [1/(2s) - o(1/s)] \cdot 1/(2s) \geq 1/(5s).\end{aligned}$$

□

Claim 8. *Starting with $t = o(D_2/\Delta^2)$ paired-up points and $s\epsilon/4$ live points, run* TSPAN *with time bound $l - s$ and live-size bound $l\epsilon/4$. If $s > 12 \ln n\, \Delta^4/\epsilon^2$ and $l \geq 2s$ and $l = o(D_2/\Delta^2)$ then with probability $\geq 1 - \exp(-\Omega(\Delta^2))$,* TSPAN *terminates with a live set of size $l\epsilon/4$.*

Proof. Starting from live-size $\epsilon s/4$, and growing for $l - s$ additional steps, we expect to reach size roughly $\epsilon(l/2 - s/4)$, and ask that it reach the lesser size $\epsilon l/4$. This can fail in either of two ways: first, by failing to have live-size $\geq \epsilon l/4$ at step $l - s$, or second, by having live-size 0 at any of the $l - s$ steps.

We reason about the increments $X_i = |L(i)| - |L(i-1)|$ just as in (9) and (10). For any step i, the probability of a second-type failure — of hitting $|L(i)| = 0$ — is

$$\mathbf{P}(\text{type-II failure at } i) \leq \mathbf{P}(X_1 + \cdots + X_i + s\epsilon/4 \leq 0)$$

which, as in (13) but with i in lieu of s, and bound $w = -s\epsilon/4$, is

$$\leq \exp(-\lambda s\epsilon/4 - \lambda i\epsilon/3).$$

Then the probability of any second-type failure is

$$\mathbf{P}(\text{any type-II failure}) \leq l e^{-\lambda s\epsilon/4}$$

which, substituting $\lambda = \epsilon/(6\Delta^2)$ from (11), and the bounds on s and l from the hypothesis, is

$$\begin{aligned}&\leq \frac{2n\Delta}{\Delta^2}\exp(-2\ln n\, \Delta^2) \\ &\leq n^{1-2\Delta^2} \leq n^{-\Delta^2}. \qquad (16)\end{aligned}$$

The probability of a first-type failure after l steps is

$$\mathbf{P}(\text{type-I failure}) = \mathbf{P}\left(\sum_{i=1}^{l-s} X_i < (l-s)\epsilon/4\right)$$

which by (14), with $l - s$ substituted for s, is

$$\leq e^{-(l-s)\epsilon^2/(72\Delta^2)}$$

which under the hypothesis $l \geq 2s$ and the hypothesized bound on s is

$$\leq n^{-\Delta^2/36}. \tag{17}$$

Summing (16) and (17) proves the claim. □

Remark 9. Starting from any point p and running TSPAN, if two live points are ever paired with one another, then in any satisfying assignment of the formula F, $\phi(p)$ must be true.

Proof. For any live point q (including p itself), when q is paired with q', this means F includes a clause $(\phi(q), \phi(q'))$, so that if $\phi(q)$ is false then $\phi(q')$ must be true. The points made live when q is paired to q' are those q'' with $\phi(q'') = \neg\phi(q')$, and so $\phi(q'')$ must be false. Thus (inductively), *if* the original live point p corresponds to a false literal, then every subsequent live point must also be false. But then if two live points r_1, r_2 are paired, the clause $(\phi(r_1), \rho(r_2))$ would be unsatisfied. So if two live points are paired, F can only be satisfiable if $\phi(p)$ is true. □

Claim 10. *Under condition* (5)*, and for* $|L| \geq 5$*, if* TSPAN *is run with no time nor live-size bound, the probability that no two points in* L *get paired with one another is* $\leq \exp(-|L|^2/(6D_1))$.

Proof. Let D, L, and U denote the initial sets of paired, live, and untouched points (rather than the corresponding sets as TSPAN progresses). Perform TSPAN by successively pairing off points in (the original set) L. The first point from L is paired "successfully" (to a point in U rather than to another point in L) with probability $|U|/(|L| + |U| - 1)$. Conditioned upon this, the second point from L is paired successfully with probability $(|U| - 1)/(|L| + |U| - 3)$, and similarly for the following points. Thus the probability that all points in L are paired to points in U is

$$\mathbf{P}(\text{success}) = \prod_{i=0}^{|L|-1} \frac{|U| - i}{|L| + |U| - 2i - 1}.$$

The terms are increasing with i: it is easily checked that this is so if $|U| - |L| + 1 > 0$, and this follows from condition (5) which assures that $|L| = o(|U|)$. Since each term is also ≤ 1, the product is dominated by the ith power of the ith term, for any i. Choosing $i = (|L| - 1)/2$ gives

$$\begin{aligned}
\mathbf{P}(\text{success}) &\leq \left(\frac{|U| - (|L| - 1)/2}{|U|}\right)^{(|L|-1)/2} \\
&= \left(1 - \frac{|L| - 1}{2|U|}\right)^{(|L|-1)/2} \\
&\leq \exp(-(|L| - 1)^2/(4|U|)) \\
&\leq \exp(-|L|^2/(5|U|)).
\end{aligned}$$

Since by definition $D_2 \leq \frac{1}{2}\Delta D_1$, $D_2/\Delta^2 \leq D_1/\Delta \leq D_1$, and so from (5) it also follows that $|D| + |L| = o(D_1)$. Then, as $|D| + |L| + |U| = D_1$, $|U| = D_1 - o(D_1)$, and $5|U| \leq 6D_1$, giving $\mathbf{P}(\text{success}) \leq \exp(-|L|^2/(6D_1))$. □

We are now ready to prove the main result for random configurations, part B of Lemma 2.

Proof of Lemma 2 part B. Algorithmically, we will make k trials as in Claim 7, where we use TSPAN to try to grow both of a complementary pair of points p and q to live-size $s\epsilon/4$, for appropriate values of the parameters k and s. By Claim 7, each of the k trials succeeds with probability at least $1/(5s^2)$. We then reason probabilistically by Claims 8 and 9 that for any trial that succeeds, p and q *would* each continue growing to live size $l\epsilon/4$ (for an appropriate value of l), and two points in each live set would pair with one another, certifying the unsatisfiability of F. In particular we will show that each of the k trials would certify the unsatisfiability of F, with probability at least $1/(6s^2)$. Thus, if k is much larger than s^2, with overwhelming probability, F must be unsatisfiable.

(It is an important subtlety that while we algorithmically grow the spans to live-size $s\epsilon/4$, we do *not* actually continue growing them and look for a certificate, but just reason that we could do so with high probability: if we actually did it we could touch on the order of kl points, while this way we touch only on the order of $ks+l$ points.)

We will first derive appropriate parameter settings, and then fill in the few logical gaps. We will use the notation $f \ll g$ to indicate that $f/g \to 0$ as $n \to \infty$, and in particular (sacrificing an arbitrarily small amount in making Δ as large as possible) we will impose the additional assumption that $f/g = o(n^{-\delta})$ for some $\delta > 0$. Also, $f \gg g$ is defined similarly, and is equivalent to $g \ll f$.

Since each trial will certify F unsatisfiable with probability $1/(6s^2)$ and we wish k trials to give an overwhelming probability of supporting such a certificate (probability $\leq \exp(-\Delta^2/4)$ will be needed after (20) to prove Part B of Theorem 1) we will need $k \gg s^2\Delta^2$, so (for some tiny constant $\delta > 0$) we may as well fix $\boxed{k = n^\delta s^2 \Delta^2}$. To satisfy (5), we need $ks = n^\delta s^3 \Delta^2 \ll D_2/\Delta^2$. In choosing values to ensure this, we know little about D_2 beyond $D_2 \geq D_1 \geq n$, so we will ensure $n^\delta s^3 \Delta^2 \ll n/\Delta^2$ by setting $\boxed{s^3 = n^{1-2\delta}\Delta^{-4}}$ thereby constraining $\Delta \ll n^{1/4}$.

With s as above, $s^2 \ll n/\Delta$ and so Claim 7's additional constraint $s^2 \ll D_1/\Delta$ is satisfied. Claim 6 requires $s \gg \Delta^2\epsilon^{-2}$. Since ϵ is fixed it is irrelevant asymptotically, so we need $s^3 \gg \Delta^6$ which is to say $n^{1-2\delta}\Delta^{-4} \gg \Delta^6$, thereby giving the still stronger constraint $\boxed{\Delta \ll n^{1/10}}$.

In terms of s, for Claim 8 it suffices (and is more or less necessary) to have $s \gg \Delta^4$, i.e., $s^3 = n^{1-2\delta}\Delta^{-4} \gg \Delta^4$ or $n^{1-2\delta} \gg \Delta^8$, which is weaker than the preceding constraint on Δ.

We now consider a suitable value of the parameter l. To obtain a useful result from Claim 10 requires $|L|^2 \gg D_1$. Assuming a successful outcome per Claim 8, the live set's size is $|L| = l\epsilon/4$, and again we know little about D_1 beyond $D_1 \leq 2n\Delta$, so we are more or less forced to require $l\epsilon/4 \gg n^{1/2}\Delta^{1/2}$ which is equivalent to the simpler $l \gg n^{1/2}\Delta^{1/2}$. This is stronger than the constraint coming from Claim 8's hypothesis $l \geq 2s$, since s is only about $n^{1/3}\Delta^{-4/3}$. We must however respect Claim 8's other constraint, that $l \ll D_2/\Delta^2$, which we ensure by requiring $l \ll n/\Delta^2$. Thus for Claims 8 and 10 it suffices (and is more or less necessary) to have $l^2 \gg n\Delta$ and $l \ll n\Delta^{-2}$ (or $l^2 \ll n^2\Delta^{-4}$), giving the constraint $n \gg \Delta^5$; again this is strictly weaker than a previous constraint. So setting $\boxed{l = n^{1-\delta}\Delta^{-2}}$ satisfies the hypothesis of Claims 8 and 10, and for Claim 10 gives an exponentially small probability that no two live points are paired with one another.

After these calculations there is one small mathematical technicality to see to. Claim 7 shows that complementary points p and q are both likely to grow to live-size $s\epsilon/4$ (call these events P and Q); and Claims 8 and 10 show that, given event P, p would be likely to grow to live-size $l\epsilon/4$ and then imply that $\phi(p)$ must be true (call this event P'), and similarly for Q and q (event Q'). That is, we know that $P \wedge Q$ is likely, and we know that P' is likely given P and that Q' is likely given Q, but we need to know that $P \wedge Q \wedge P' \wedge Q'$ is likely.

For arbitrary events,

$$\begin{aligned}
\mathbf{P}(P \wedge Q \wedge P' \wedge Q') &= \mathbf{P}(P \wedge Q) - \mathbf{P}(P \wedge Q \wedge ((\neg P') \vee (\neg Q'))) \\
&\geq \mathbf{P}(P \wedge Q) - \mathbf{P}(P \wedge Q \wedge \neg P') - \mathbf{P}(P \wedge Q \wedge \neg Q') \\
&\geq \mathbf{P}(P \wedge Q) - \mathbf{P}(P \wedge \neg P') - \mathbf{P}(Q \wedge \neg Q') \\
&\geq \mathbf{P}(P \wedge Q) - \mathbf{P}(\neg P' \mid P) - \mathbf{P}(\neg Q' \mid Q).
\end{aligned}$$

For the particular events in question, and with the parameters chosen as above, Claim 7 shows that $\mathbf{P}(P \wedge Q) \geq 1/(5s^2)$, while Claims 8 and 10 show that $\mathbf{P}(\neg P' \mid P) = \exp(-\Omega(n^\delta)) = o(1/5s^2)$ and likewise for Q' and Q, yielding

$$\mathbf{P}(P \wedge Q \wedge P' \wedge Q') \geq 1/(6s^2).$$

That is, each complementary pair p and q would, with probability at least $1/(6s^2)$, prove F unsatisfiable, and so over our $k = n^\delta s^2 \Delta^2$ choices of such pairs, the chance no pair certifies F

unsatisfiable is at most $\exp(-n^\delta\Delta^2)$. Thus,

$$\mathbf{P}(F \text{ is satisfiable}) \leq \exp(-n^\delta\Delta^2). \tag{18}$$

□

3. Uniform Simple Formulas

We have now proved Theorem 1, but for random formulas F_P generated according to the configuration model, rather than for *simple* random formulas F chosen uniformly from $\Omega_{\mathbf{d}}$.

If $\mathbf{d}$ satisfies $\sum_{i=1}^n(d_i^2+\bar{d}_i^2)=O(m)$, then the expected number of repeated clauses, and clauses with a repeated literal, is $O(1)$, and there is a positive probability that there are none and the formula is simple. In that case, the high-probability results for the configuration model imply high-probability results for the uniform model $F \in \Omega_{\mathbf{d}}$.

To obtain the same conclusion with a weaker constraint on the degree sequence, namely for all proper degree sequences with $2D_2 < (1-\epsilon)D_1$, we use the idea of switchings; see [McK85, MW91, CFRR02]. In a pairing $P \in \Psi$, a pair (u,v) is a *loop* if $\phi(u)=\phi(v)$, it is a *tautology* if $\phi(u)=\overline{\phi(v)}$, and if it is neither a loop nor a tautology then it is *redundant* if P contains another pair $\{u',v'\}$, $u'<v'$ with $\phi(u')=\phi(u)$, $\phi(v')=\phi(v)$, and $u<u'$. The following algorithm will remove loops, tautological edges and repeated clauses; it assumes some total ordering on the points Z such that each $Z(x)$ forms an interval.

The study of random graphs with a fixed degree sequence rests on the analysis of algorithms similar to the one given next (at least for "small" maximum degree), based on an easy observation. F_P is simple iff the following multi-graph $G=G(P)$ is simple and has no *tautological* edges of the form $\{u,\bar{u}\}$, $u \in L$. G has vertex set L, and contains an edge $\{\phi(x),\phi(y)\}$ for every pair $\{x,y\} \in P$. Our algorithm below differs from the graphical versions in that it removes tautological clauses as well. These can be handled in the same way as loops and so we will not provide a proof of our claims. It suffices to refer the reader to the proofs of the graphical case.

Algorithm SIMPLIFY
begin
  Construct P using CONSTRUCT.
  Let the a loops, b tautological edges and c redundant clauses be
   enumerated as $\{u_i,v_i\} \subseteq Z, i=1,2,\dots,a+b+c$.
  If $a+b+c \geq n^{1/5}$ then terminate — **FAILURE**.
  For $i=1$ **to** $a+b+c$ **do**
  begin
   Choose $\{x,y\}$ randomly from P – **Step A**.
   Replace the two pairs $\{u_i,v_i\},\{x,y\}$ by
    $\{u_i,x\},\{v_i,y\}$, where $u_i<v_i$ and we choose
    randomly the order $x<y$ or $x>y$.
  end
  If F_P is not simple then terminate — **FAILURE**.
end

Let Q denote the output of SIMPLIFY.

We remark that $O(\Delta^2)$ is a high-probability upper bound on $a+b+c$, and the algorithm's cap of $n^{1/5}$ on $a+b+c$ is chosen simply to satisfy $n^{1/5} \gg \Delta^2 = n^{2/11}$. It follows by routine calculation that the probability the algorithm terminates in failure is $o(1)$. Let $\Psi^\star$ denote the set of configurations $P \in \Psi$ for which F_P is simple.

For a proof of the following lemma (in the case of graphs with a fixed degree sequence) see e.g. McKay [McK85] or Cooper, Frieze, Reed and Riordan [CFRR02].

Lemma 11. *There exists $\tilde{\Psi} \subseteq \Psi^\star$ such that*

(a):

$$\frac{|\tilde{\Psi}|}{|\Psi^\star|} = 1 - o(1).$$

(b):

$$\mathbf{P}(Q \in \tilde{\Psi}) = 1 - o(1).$$

(c): *For all* $P_1, P_2 \in \tilde{\Psi}$,

$$\frac{\mathbf{P}(Q = P_1)}{\mathbf{P}(Q = P_2)} = 1 \pm o(1).$$

It follows from Lemma 11 that we need only prove the equivalent of Theorem 1 with Q in place of F.

3.1. Proof of part A of main Theorem.

Consider the proof of Claim 3. We argue that in (1), we can replace the terms

$$\frac{d(\bar{w}_i)d(w_{i+1})}{D_1 - 2i + 1} \text{ by } \frac{d(\bar{w}_i)d(w_{i+1})}{D_1 - 2i + 1} + O\left(\left(\frac{\Delta n^{1/5}}{n}\right)^2\right). \tag{19}$$

The extra term comes from considering the chance that the arc (w_i, w_{i+1}) is created by SIMPLIFY. For this to happen, (i) one of $\bar{w}_i$ or w_{i+1} must be incident with a redundant pair or a loop, and (ii) the other one must be incident with a pair $\{x, y\}$ chosen in Step A. (We say that $\{a, b\}$ is *incident with* $\{c, d\}$ if the corresponding edges are incident in the graph $G(P)$, i.e., if $\{\phi(a), \phi(b)\} \cap \{\phi(c), \phi(d)\} \neq \emptyset$.) Events (i) and (ii) each occur with probability $O\left(\frac{\Delta n^{1/5}}{n}\right)$, and are approximately independent of one another. The bound on the extra term applies in the context of Claim 3, where the relevant probabilities are conditioned upon the existence of previous arcs in a path under consideration: there are only $O(\log n)$ arcs in each path considered, and the new arc is by definition disjoint from the old ones. The correction in (19) does not affect the conclusion of Claim 3.

A similar correction can be applied in the rest of the proof of part (A) of Theorem 1. In this case the last two terms in (2) should be given a slightly larger correction, $+O\left(\frac{\Delta n^{1/5}}{n}\right)$: condition (i) may be implied by the existence of a previous arc, so we simply bound its probability by 1, while the probability of condition (ii) is as in the preceding paragraph.

3.2. Proof of part B of main Theorem.

For part B of Theorem 1 we need (18) and

$$\frac{|\Psi^\star|}{|\Psi|} \geq e^{-O(\Delta^2)}. \tag{20}$$

Indeed, (18) and (20) imply that

$$\begin{aligned}
&\mathbf{P}(F \text{ is satisfiable}) \\
&\quad = \mathbf{P}(F_P \text{ is satisfiable} \mid P \text{ is simple}) \\
&\quad \leq \mathbf{P}(F_P \text{ is satisfiable}) \,/\, \mathbf{P}(P \text{ is simple}) \\
&\quad \leq e^{O(\Delta^2)} n^{-\Delta^2/5} \\
&\quad = o(1).
\end{aligned}$$

For a proof of (a graph version of) (20), see [CFRR02].

References

[APT79] Bengt Aspvall, Michael F. Plass, and Robert Endre Tarjan, *A linear-time algorithm for testing the truth of certain quantified Boolean formulas*, Inform. Process. Lett. **8** (1979), no. 3, 121–123. MR **80b:**68050

[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.

[Bol80] Béla Bollobás, *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs*, European J. Combin. **1** (1980), no. 4, 311–316. MR **82i:**05045

[CFR] Colin Cooper, Alan M. Frieze, and Bruce Reed, *Random regular graphs of non-constant degree: connectivity and hamilton cycles*, to appear.

[CFRR02] Colin Cooper, Alan Frieze, Bruce Reed, and Oliver Riordan, *Random regular graphs of non-constant degree: independence and chromatic number*, Combin. Probab. Comput. **11** (2002), no. 4, 323–341. MR 1 918 719

[CR92] 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.

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

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

[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.

[KKL02] A. C. Kaporis, L. M. Kirousis, and E. Lalas, *The probabilistic analysis of a greedy satisfiability algorithm*, 5-th International Symposium on the Theory and Applications of Satisfiability Testing, 2002, pp. 362–376.

[McK85] Brendan D. McKay, *Asymptotics for symmetric* 0-1 *matrices with prescribed row sums*, Ars Combin. **19** (1985), no. A, 15–25. MR **87e:**05081

[MR95] Michael Molloy and Bruce Reed, *A critical point for random graphs with a given degree sequence*, Proceedings of the Sixth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science, "Random Graphs '93" (Poznań, 1993), vol. 6, 1995, pp. 161–179. MR **97a:**05191

[MW91] Brendan D. McKay and Nicholas C. Wormald, *Asymptotic enumeration by degree sequence of graphs with degrees* $o(n^{1/2})$, Combinatorica **11** (1991), no. 4, 369–382. MR **93e:**05002

[Ver99] Yann Verhoeven, *Random 2-SAT and unsatisfiability*, Inform. Process. Lett. **72** (1999), no. 3-4, 119–123. MR **2000k:**68074

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

(Colin Cooper) School of Mathematical and Computing Sciences, Goldsmiths College, University of London, London SE14 6NW, UK.

E-mail address: `c.cooper@gold.ac.uk.`

(Alan Frieze) Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA 15213, USA.

E-mail address: `alan@random.math.cmu.edu`

(Gregory B. Sorkin) Department of Mathematical Sciences, IBM T.J. Watson Research Center, Yorktown Heights NY 10598, USA.

E-mail address: `sorkin@watson.ibm.com`