A note on random 2-SAT with prescribed literal degrees

Colin Cooper* Alan Frieze† 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 a prescribed literal sequence.

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 $x_1, x_2, \ldots, x_n$ and m random clauses. This model is quite well understood. Chvatál and Reed [4] 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 [8], Fernandez de la Vega [7] and Verhoeven [12]. The tightest results are due to Bollobás, Borgs, Chayes, Kim and Wilson [3].

Just as in the case of the existence of a giant component in a random graph, Molloy and Reed [9], we can obtain interesting results by considering models in which the number of occurrences of each literal is prescribed.

Let the set of literals be $L = \{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}$. A *2-SAT formula* F is then a set of m *distinct* clauses $C_1, C_2, \ldots, C_m$ where each C_i is a 2-element subset of L (we exclude clauses in which the 2 literals are identical, i.e., *loops*). A *truth assignment* σ is a mapping $\sigma : L \to \{0, 1\}$ which satisfies $\sigma(x_j) + \sigma(\bar{x}_j) = 1$ for $j = 1, 2, \ldots, n$. σ satisfies F if $\sigma(C_i) \geq 1$ for $i = 1, 2, \ldots, m$. (Here $\sigma(C) = \sigma(w_1) + \sigma(w_2)$ if $C = \{w_1, w_2\}$).

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

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

*School of Mathematical and Computing Sciences, Goldsmiths College, University of London, London SE14 6NW, UK. e-mail: c.cooper@gold.ac.uk.

†Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA 15213, USA. Supported in part by NSF grant CCR-9818411. e-mail alan@random.math.cmu.edu

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

Let $\Delta_{\mathbf{d}} = \max\left\{d_1, \bar{d}_1, \ldots, d_n, \bar{d}_n\right\}$ and

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

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 that variable from consideration. Thus $D_1 \geq n$. Our random model is that

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

The degree sequence $\mathbf{d}$ is *proper* if

- $\Delta_{\mathbf{d}} \leq n^{\alpha}$ where $\alpha < 1/13$ is a constant.
- $D_1 = 2m$, i.e., D_1 is even.

We prove the following criterion for satisfiability:

Theorem 1. *Let* $\mathbf{d}$ *be proper and let* $0 < \epsilon < 1$ *be constant. Then*

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

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

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

$$\mathbf{P}(F \text{ is unsatisfiable}) \to 1.$$

For example in the case of $m = cn$ randomly chosen clauses we find that $D_1 = 2cn$ and **whp** $D_2 \approx c^2 n$, and we obtain the result of [4].

2 Proof of the theorem

Graphical Representation

Given a formula $F = \{\{u_j, v_j\} : j = 1, 2, \ldots, m\}$ we define a digraph $\Gamma = \Gamma(F) = (L, A)$ where $A = \{(\bar{u}_j, v_j), (\bar{v}_j, u_j) : j = 1, 2, \ldots, m\}$. (If $w \in L$ then $\bar{w}$ is defined as follows: if $w = x_j$ then $\bar{w} = \bar{x}_j$ and if $w = \bar{x}_j$ then $\bar{w} = x_j$.)

It is well known (see for example Aspvall, Plass and Tarjan [1]) that F is unsatisfiable if and only if there is a variable x_j such that Γ_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 [2]. We have a universe Z consisting of D_1 *points*, partitioned into subsets $Z(x), x \in L$, with $|Z(x_i)| = d_i$, $|Z(\bar{x}_i)| = \bar{d}_i$, $i = 1, 2, \ldots, n$. "Inversely" to $Z(x)$, define $\phi : Z \to L$ by $\phi(w) = x$ iff $w \in Z(x)$. 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 $C_S = \{\phi(p), \phi(q)\}$. In the configuration model we choose P uniformly at random from Ψ and let F_P be our random 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. We will first study the likely satisfiability of F_P, and later, in Section 3, show how to deal with the issue of simplicity.

There is an algorithmic description of the generation of P which can be useful:

Algorithm CONSTRUCT
begin
  $P_0 := \emptyset$; $R_0 := Z$
  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

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

A *bicycle* is a sequence of clauses $\{u, w_1\}$, $\{\bar{w}_1, w_2\}$, ..., $\{\bar{w}_r, v\}$ where $w_1, w_2, \ldots, w_r$ are distinct literals and $u \in \{w_i, \bar{w}_i\}$, $v \in \{w_j, \bar{w}_j\}$ for some $1 \leq i, j \leq r$.

Chvátal and Reed [4] 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 show that **whp** $\Gamma(F_P)$ does not contain any long paths. Then we can restrict our attention to small bicycles.

Claim 2. *$\Gamma(F_P)$ has no long directed paths* **whp**.

Proof of Claim 2
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 are implicitly using CONSTRUCT with the initial sequence $u_1, u_2, \ldots$, taken from $Z(\bar{w}_1)$, then $Z(\bar{w}_2)$ and so on.

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

So, **whp**, Γ has no directed path of length $\geq k_0$.
End of proof of Claim 2.

Using Claim 2 we see that we need only consider the existence of bicycles of length $r \leq 2k_0$.

So, if Y_r is the number of bicycles of length r and $Y=\sum_{r=2}^{2k_0} Y_r$ then

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

since $\Delta_{\mathbf{d}}=o(n^{1/2})$.

This verifies Part (a) of Theorem 1 in respect of the random formula F_P. We translate this result to uniformly chosen formulae in Section 3.

2.2 Case 2: $2D_2>(1+\epsilon)D_1$

For $w\in L$ we let $span(w)=\{v:\ \Gamma(F_P)$ contains a directed path from w 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

$$w\Longrightarrow x\wedge\bar{x} \text{ and } \bar{w}\Longrightarrow y\wedge\bar{y}. \qquad (3)$$

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

We work in terms of the configuration model and consider the following algorithm. We have $z\in Z$ and generate points reachable from z at the same time as we generate P via CONSTRUCT.

In the execution of the algorithm SPAN(z,Z) the elements of Z are partitioned into

- A_P: paired-up points — $A_P=\emptyset$ initially.
- A_L: live points — $A_L=\{z\}$ initially.
- A_U: untouched points — $A_U=Z\setminus\{z\}$ initially.

At a *general step* we arbitrarily choose $z'\in A_L$, move it to A_P and randomly pair it with an element z'' of $A_L\cup A_U$. We place z'' into A_P. Suppose now that $z'\in Z(u)$ and $z''\in Z(v)$. We consider that we have created a clause $\{u,v\}$ and if any points of $Z(\bar{v})\setminus A_P$ are in A_U, we move them to A_L. We repeat such steps until $A_L=\emptyset$, and we denote the final value of A_P by $R_{z,Z}$.

The span of literal w can be computed as follows. First, as a minor detail, we generalize SPAN so that for $z\notin Z$ we define $R_{z,Z}=\emptyset$. Let $Z(\bar{w})=\{z_1,z_2,\dots,z_d\}$. Then we run SPAN$(z_1,Z)$, let $Z_1=Z\setminus R_{z_1,Z}$, run SPAN(z_2,Z_1), let $Z_2=Z_1\setminus R_{z_2,Z_1}$, run SPAN$(z_3,Z_2)$ and so on. $span(w)$ is w together with the set of literals λ for which λ appears as v in a general step.

We have to consider a sequence of *truncated* executions of SPAN, denoted by TSPAN. We add the extra stopping condition:

$$|A_L|\geq b=\Delta^2\log n.$$

If this occurs we say that z is *large*. We let $R_{z,Z}$ be as in SPAN. We run this sequence searching for a pair of large points z, z' where $z \in Z(w)$ and $z' \in Z(\bar{w})$ for some literal w.

To this end we let $Z = \{z_1, z_2, \dots, z_{2m}\}$ where if $\delta_j = \min\{d_j, \bar{d}_j\}, j = 1, 2, \dots, n$, the first $2\delta_1$ points are from $Z(x_1) \cup Z(\bar{x}_1)$, the next $2\delta_2$ points are from $Z(x_2) \cup Z(\bar{x}_2)$ and so on. Furthermore, the points corresponding to a particular variable alternate between the variable and its complement. For example, $Z(x_1) = \{z_1, z_3, \dots, z_{2\delta_1 - 1}\}, Z(\bar{x}_1) = \{z_2, z_4, \dots, z_{2\delta_1}\}$. The ordering of the points $Z_j, j > 2(\delta_1 + \cdots + \delta_n)$ is arbitrary. We run TSPAN(z_1, Z), let $Z_1 = Z \setminus R_{z_1,Z}$, run TSPAN$(z_2, Z_1)$, let $Z_2 = Z_1 \setminus R_{z_2,Z_1}$, run TSPAN$(z_3, Z_2)$ and so on. When we run TSPAN(z_2, Z_1) for example, we re-set $A_L \leftarrow \{z_2\}$ and $A_U \leftarrow Z_1 \setminus \{z_2\}$. We let A_P grow naturally.

Note that our assumptions imply that there are at least $n^{1-2\alpha}$ values of i for which $\delta_i \geq 1$ ($d_j \bar{d}_j \leq n^{2\alpha}$ for all j and $D_2 > n$).

Now consider the change Θ_t in the size of A_L after t general steps.

$$\mathbf{E}(\Theta_t) \geq -1 + \frac{1}{D_1}(2D_2 - 2t\Delta^2) \geq \frac{\epsilon}{2} \tag{4}$$

provided $t = o(n/\Delta^2)$.

Explanation: -1 due to z' being removed from A_L. Let $d'_j = |Z(x_j) \setminus A_P|$, $\bar{d}'_j = |Z(x_j) \setminus A_P|$ for $j = 1, 2, \dots, n$. The expected number of new members of A_L is then $\frac{1}{D_1 - 2t} \sum_{j=1}^{n} d'_j \bar{d}'_j$.

Now consider the execution of TSPAN(z_k, Z_{k-1}), for some k such that $kb^2 = o(n/\Delta^2)$. Let the sequence of sizes of A_L be $Y_0 = 1, Y_1, \dots,$. Then in general we have

$$Y_l - 2 \leq Y_{l+1} \leq Y_l + \Delta \tag{5}$$

and

$$\mathbf{E}(Y_{l+1} - Y_l \mid \text{previous history}) \geq \frac{\epsilon}{2}. \tag{6}$$

Suppose now that we consider a modified process which proceeds as follows: If Y_t reaches zero before reaching b then we undo all the pairings and start again with new random pairings at each step. This constitutes a sequence of Bernouilli trials (= executions of TSPAN(z_k, Z_{k-1})) whose probability of success p is to be estimated. This is to be considered as a thought experiment used to estimate p and not a way of generating a favourable formula.

It follows from (5), (6) and Chernoff bounds that

$$\mathbf{P}\left(Y_l \leq \frac{\epsilon}{4} l\right) \leq \exp\left\{-\frac{\epsilon^2 l^2}{8l\Delta^2}\right\} \tag{7}$$

Putting $s = 4\epsilon^{-1}b$ we see that

$$\mathbf{P}(Y_s \leq b) \leq e^{-2(\log n)^2}.$$

This implies that **whp** there is a successful trial within the first s trials. (At this point we should deal with the events $z_k \notin Z_{k-1}$. The size of A_P will be $O(s^2 b \log n)$ and the probability that $z_k \in A_P$ is $O(|A_P|\Delta/n) = o(1/s)$ with our assumptions.) This then implies that if $k = o(n/(b^2\Delta^2))$ then

$$\begin{aligned} \mathbf{P}(z_k \text{ is large} \mid \text{the outcomes of TSPAN}(z_i, Z_{i-1}), \\ 1 \leq i < k) \quad \geq \quad \frac{1}{2s}. \end{aligned} \tag{8}$$

So for a successive pair z_{k-1}, z_k with $k = o(n/(b^2\Delta^2))$,

$$\begin{aligned} \mathbf{P}(z_{k-1}, z_k \text{ are both large} \mid \text{the outcomes of} \\ \text{TSPAN}(z_i, Z_{i-1}),\ 1 \leq i < k-1) \quad \geq \quad \frac{1}{4s^2}. \end{aligned} \tag{9}$$

It follows that

$$\begin{aligned}
&\mathbf{P}(\nexists k \leq s^2\Delta^2 \log n: \ z_{2k-1}, z_{2k} \text{ are both large}) \\
&\qquad \leq \ n^{-\Delta^2/4}. \qquad (10)
\end{aligned}$$

(Note that our assumptions imply $s^2\Delta^2 \log n = o(n/(b^2\Delta^2))$.)

We now show that if during the execution of SPAN(z, Z) the size of A_L reaches b then **whp** it reaches $\ell_0 = n/(\Delta^2 \log n)$. Indeed,

$$\begin{aligned}
&\mathbf{P}(|A_L| \text{ fails to reach } \ell_0 \ \big| \ |A_L| \text{ reaches } b) \\
&\qquad \leq \ e^{-2\ell_0/\Delta^2}. \qquad (11)
\end{aligned}$$

This follows directly from (4), (7) by taking $t = 4\epsilon^{-1}\ell_0$.

Now assume that $|X_L|$ reaches ℓ_0 and consider the set V_0 of variables v for which $Z(v), Z(\bar{v}) \subseteq A_U$ at the stage when $|X_L|$ first reaches ℓ_0. We choose $V_1 \subseteq V_0$ such that $|V_1| = n^{.65}$ and $d(v), d(\bar{v}) \neq 0$ for $v \in V_1$. Then

$$\begin{aligned}
&\mathbf{P}(\nexists v \in V_1 : \ R_{z,Z} \cap Z(v) \neq \emptyset \text{ and} \\
&\qquad R_{z,Z} \cap Z(\bar{v}) \neq \emptyset) \\
&\leq \prod_{v \in V_0} \left(1 - \frac{(\ell_0 - 2n^{.65}\Delta)^2 d(v) d(\bar{v})}{D_1^2}\right) \\
&\leq \exp\left\{-(1 - o(1)) \frac{n^{.65} n^2}{\Delta^4 (\log n)^2 (n\Delta)^2}\right\} \\
&\leq e^{-n^{.04}}. \qquad (12)
\end{aligned}$$

Explanation: The probability that $R_{z,Z} \cap Z(v) \neq \emptyset$ is at least $\frac{d(v)\ell_0}{D_1}$ and conditional on this, the probability that $R_{z,Z} \cap Z(\bar{v}) \neq \emptyset$ is at least $\frac{d(\bar{v})\ell_0}{D_1}$. As we run through V_1 the factor ℓ_0 in the numerator decreases by at most $2n^{.65}\Delta$.

In summary, (10) shows that we will **whp** find a pair of complementary literals, both having a large span and then (11), (12) imply that both of these spans contain a complementary pair, verifying the existence of w, x, y such that (3) holds.

The next section requires us to give an estimate of the probability that F_P is satisfiable in part (b). Adding the failure probabilities from (10), (11) and (12) we get a failure probability of order

$$n^{-\Delta^2/4} + e^{-2n/(\Delta^4 \ln n)} + e^{-n^{.65}/(\Delta^6 (\ln n)^2)} \leq n^{-\Delta^2/5}. \qquad (13)$$

3 Uniform Sampling

We have now proved Theorem 1, but for random formulas F 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 [10, 11, 5]. Observe that F_P is simple iff the following multi-graph $G = G(P)$ is simple. The vertex set of G is L. It contains an edge $\{\phi(x), \phi(y)\}$ for every pair $\{x, y\} \in P$.

The following algorithm removes loops and repeated clauses: assume some total ordering on the points Z such that each $Z(x)$ forms an interval. A non-loop pair $\{u, v\}$, $u < v$ is redundant in $P \in \Psi$ if P contains another pair $\{u', v'\}$, $u' < v'$ with $\phi(u') = \phi(u), \phi(v') = \phi(v)$ and $u < u'$.

Algorithm SIMPLIFY
begin
Construct P using CONSTRUCT.
Let the a loops and b redundant clauses be enumerated as $\{u_i, v_i\} \subseteq Z, i = 1, 2, \ldots, a + b$.
If $a + b \geq 2n^{2\alpha}$ then terminate — **FAILURE**.
For $i = 1$ **to** $a + b$ **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.

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 (graph version of the) following lemma see e.g. McKay [10] or Cooper, Frieze, Reed and Riordan [6].

Lemma 3. *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 3 that we need only prove the equivalent of Theorem 1 with Q in place of F.

Consider the proof of Claim 2. We argue that in (1), we can replace the terms

$$\frac{d(\bar{w}_i)d(w_{i+1})}{D_1 - 2i + 1} \text{ by } \frac{d(\bar{w}_i)d(w_{i+1})}{D_1 - 2i + 1} + O\left(\left(\frac{\Delta n^{2\alpha}}{n}\right)^2\right). \tag{14}$$

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 $O\left(\frac{\Delta n^{2\alpha}}{n}\right)$, and are approximately independent of one another. The bound on the extra term applies in the context of Claim 2, 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 (14) does not affect the conclusion of Claim 2.

A similar correction can be applied in the rest of the proof of Theorem 1(a). In this case the last two terms in (2) should be given a slightly larger correction, $+O\left(\frac{\Delta n^{2\alpha}}{n}\right)$: 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.

For Theorem 1(b) we need (13) and

$$\frac{|\Psi^\star|}{|\Psi|} \geq e^{-O(\Delta^2)}. \tag{15}$$

Indeed, (13) and (15) 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) (15), see [6].

References

[1] B. Aspvall, M.F. Plass and R.E. Tarjan, A linear time algorithm for testing the truth of certain quantified boolean formulas, *Information Processing Letters* 8 (1979) 121–123.

[2] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *European Journal of Combinatorics* 1 (1980) 311–316.

[3] B. Bollobás, Christian Borgs, Jennifer Chayes, J.H. Kim, and D.B. Wilson, The scaling window of the 2-SAT transition, *Random Structures and Algorithms* (2001) 201–256.

[4] V. Chvatál and B. Reed, Mick gets some (the odds are on his side), *Proceedings of the 33rd Annual IEEE Symposium on the Foundations of Computer Science* (1992) 620–627.

[5] C. Cooper, A.M. Frieze and B. Reed, Random regular graphs of non-constant degree: connectivity and Hamilton cycles, to appear.

[6] C. Cooper, A.M. Frieze, B. Reed and O. Riordan, Random regular graphs of non-constant degree: independence and chromatic number, to appear.

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

[8] A. Goerdt, A threshold for unsatisfiability, *Journal of Computer and System Sciences* 33 (1996) 469–486.

[9] M. Molloy and B. Reed, The size of the giant component of a random graph with a given degree sequence, *Combinatorics, Probability and Computing* 7 (1998) 295–305.

[10] B.D. McKay, Asymptotics for symmetric 0–1 matrices with prescribed row sums, *Ars Combinatoria* 19A (1985) 15–26.

[11] B.D. McKay and N.C. Wormald, Asymptotic enumeration by degree sequence of graphs with degree $o(n^{1/2})$, *Combinatorica* 11 (1991) 369–382.

[12] Y. Verhoeven, Random 2-SAT and unsatisfiability, *Information Processing Letters* 72 (1999) 119–123.

This research was sponsored in part by National Science Foundation (NSF) grant no. CCR-0122581.