On random ordering constraints – Revised 14/10/2008

Andreas Goerdt

Technische Universität Chemnitz, Fakultät für Informatik
Straße der Nationen 62, 09107 Chemnitz, Germany
goerdt@informatik.tu-chemnitz.de

Abstract. Ordering constraints are analogous to instances of the satisfiability problem in conjunctive normalform, but instead of a boolean assignment we consider a linear ordering of the variables in question. A clause becomes true given a linear ordering iff the relative ordering of its variables obeys the constraint considered.
The naturally arising satisfiability problems are **NP**-complete for many types of constraints. The present paper seems to be one of the first looking at random ordering constraints. Experimental evidence suggests threshold phenomena as in the case of random k-SAT instances and thus natural problems to be proved. We state some basic observations and prove two results:
First, random instances of the cyclic ordering and betweenness constraint (definition see Subsection 1.1) have a sharp threshold for unsatisfiability. The proof is an application of the threshold criterion due to Friedgut.
Second, random instances of the cyclic ordering constraint are satisfiable with high probability if the number of randomly picked clauses is $< 1 \cdot n$, where n is the number of variables considered.

Topics: Algorithms, logic, random structures, probabilistic analysis

1 Introduction

1.1 Results

Let V always be a set of n variables. A 3-clause over V is an *ordered* 3-tuple (x, y, z) consisting of three different variables. A formula, also called ordering constraint is a set of clauses.

In analogy with random 3-SAT formulas, a selection of the literature is [1] [3] [8] [20] we consider random ordering constraints. The random instance $F(V,p)$ or the corresponding probability space is obtained by picking each of the $n(n-1)(n-2)$ clauses independently with probability p. Thus $F(V,p)$ is analogous to the well known random graph $G(n,p)$. As common in the theory of random structures this paper deals with properties holding with high probability, that is $1-o(1)$ when n becomes large and $p = p(n)$ is a given function.

Given a linear ordering of all n variables a clause evaluates to true if its variables satisfy a given constraint with respect to the ordering. A formula becomes true when all its clauses are true. This is the satisfiability problem the present paper deals with.

The clause (x, y, z) interpreted as *cyclic ordering constraint* is true iff $x < y < z$ or $z < x < y$ or $y < z < x$ with respect to the ordering considered. This means there is a cyclic permutation of (x, y, z) which is monotonously increasing with respect to the ordering. Note that clauses which are cyclic permutations of each other are equivalent, syntactically we distinguish them nevertheless. (Our results do not depend on this.) The associated satisfiability problem is **NP**- complete as known for long [13].

In case of the *betweenness* problem the clause (x, y, z) is true iff y is between x and z, that is we have $x < y < z$ or $z < y < x$ with respect to the the ordering considered. The corresponding satisfiability problem is **NP**-complete, too [21]. Chor and Sudan [7] consider the optimization version of the betweenness constraint with methods of semidefinite programming.

Basic observations which follow readily from the literature as shown in Section 2 collects

Proposition 1. *For random instances $F(V,p)$ with $p = a/n^2$ the following events have high probability:*
(a) For $a < 0.8$ the random cyclic ordering and betweenness instance is satisfiable.
(b) For $a > 9 \cdot \ln 3 \approx 9.88$ the the cyclic ordering instance is unsatisfiable.
(c) For $a > 4 \cdot \ln 2 \approx 2.77$ the betweenness instance is unsatisfiable.

The expected number of clauses of $F(V,p)$ with $p = a/n^2$ is an. Moreover, the number of clauses is asymptotically equal to an with high probability. Techniques as detailed on pages 34/35 of [4] enable us to show that analogous results holds for the random instance obtained by picking a random set of exactly an clauses. This applies to all of our results.

The initial inspiration for this paper came from some experiments (performed only for $n \leq 40$ for running time reasons.) these experiments show that random cyclic ordering instances with up to $1.5n$ clauses tend to be satisfiable. For more than $1.6n$ many clauses we usually get unsatisfiability.

For the random betweennness constraint we experimentally observe the same phenomenon for an random clauses when a is between 1 and 1.2. The experimental observations seem to reflect the fact that a clause interpreted as a betweenness constraint is true for 2 orderings of its variables , whereas for the cyclic ordering constraint we get 3 (out of 6.)

As far as the author knows no theoretical results concerning the preceding observations or even random ordering constraints in general exist. We prove that the transition from satisfiability to unsatisfiability is swift, that is there is a sharp threshold.

Theorem 2. *There exists $C = C(n)$, $0.8 \leq C \leq 9.88(2.77)$ such that for each constant $\varepsilon > 0$ the cyclic ordering constraint (betweenness constraint) $F(V,p)$ is unsatisfiable with high probability if $p = (C + \varepsilon)/n^2$ and satisfiable if $p = (C - \varepsilon)/n^2$.*

Note that we have only bounds on C within which the threshold value C may vary depending on n. The situation is analogous to random 3-SAT or colourability of random graphs, see [11]. As in these cases the proof of Theorem 2 is an application of the threshold criterion of Friedgut [10] which gives no information of the value.

Given Proposition 1, to show that $F(V, a/n^2)$ is satisfiable with high probability for any a substantially larger than 0.8 seems to be non-trivial. Analyzing a heuristic algorithm which tries to find a satisfying ordering we make some progress.

Theorem 3. *For $p = a/n^2$ with $a \leq 1$ the random cyclic ordering instance $F(V, p)$ is satisfiable with probability $\geq \varepsilon$ for a constant $\varepsilon > 0$.*

Theorem 3 together with Theorem 2 implies a high probability result.

Corollary 4. *The random cyclic ordering constraint with $p = a/n^2$ and $a < 1$ is satisfiable with high probability.*

1.2 On the literature

Ordering constraints differ from traditional constraints like k-SAT or more general kinds of constraints in that the underlying assignment must be an *ordering* of all variables. This means on the one hand that each variable can receive one out of n values, its position in the ordering. On the other hand each of the n values can only be used once. Altogether we have $n!$ many assignments as opposed to only 2^n in the case of satisfiability.

Beyond random k-SAT there is a considerable body of work on random constraints with finite domain from which the values for each variable are taken. Only a small selection of the literature, in part due to Michael Molloy is [18], [19], [16] . The paper [18] points out that the investigation of thresholds is not only of structural interest, but has also algorithmic relevence: Random instances at thresholds often have some algorithmic hardness which makes them attractive as test cases for algorithms.

As far as we know systematic experimental studies of random ordering constraints have not been made by now. Our preliminary experiments indicate that instances closer to the threshold become harder.

Many real world notions like time and space involve some kind of ordering. Therefore it is not surprising that knowledge representation formalisms may contain ordering constraints. In [15] for example the cyclic ordering constraint occurs. In [6] a weighted version of an extended betweenness constraint is used to describe some biological situation. These applications provide additional motivation to investigate ordering constraints systematically. From the point of view of worst case complexity ordering constraints are investigated in [14]. As we have many different kinds of ordering constraints, it is natural to try to classify them according to their complexity. For some cases the authors can show a dichotomy either polynomial time solvable or NP-hard. Moreover, [14] has additional pointers concerning applications of ordering constraints.

More theoretically , besides the papers [7] and [14] mentioned above, we find [5] which deals with non-approximability results for the optimization version of the betweenness constraint.

2 Observations

We collect basic results which follow directly from the literature or are otherwise easy. The 2-core of a constraint F is the (unique) largest subformula of F in which each variable which occurs at all has degree ≥ 2 , that is occurs at least twice. The 2-core is non-trivial iff it is not empty.

Proposition 5. *Let F be a constraint which has only the trivial 2-core. F interpreted as a cyclic ordering (betweenness) constraint is satisfiable.*

Proof. The proof is based on the fact that the clauses of F can be satisfied one by one. An induction on the number of clauses of F formalizes this: F has a clause C which contains a variable x which occurs only once. F without C can be satisfied by induction hypothesis. Then C can be satisfied by putting the variable x in the right position. Note that we can always find a suitable position for x as we have a cyclic ordering (betweenness) constraint.

Satisfiable constraints with non-empty 2-core are easy to find:

Example 6. We consider the ordering $x < y < z < u$ and the constraint $(x, y, z),\ (x, z, u), (y, z, u)$. Each variable occurs twice. The constraint (interpreted as betweenness or cyclic ordering constraint) is satisfied by the ordering.

Molloy's Theorem 1.2, on page 666 of [20], for graphs and hypergraphs can be read as threshold theorem for the appearance of a non-trivial 2-core in $F(V,\ a/n^2)$:

Proposition 7. *Let*
$$t\ =\ \min_{0<f<1} -\ln(1-f)/(3\cdot f^2).$$

(a) If $a < t$ the 2-core of $F(V,\ a/n^2)$ is empty with high probability.

(b) If $a\ >\ t$ the random constraint $F(V\ ,a/n^2)$ has a 2-core of linear size with high probability.

First insight into the curve $-\ln(1-f)/(3\cdot f^2)$ can be obtained by looking at the logarithm series $-\ln(1-f) = f + f^2/2 + f^3/3 + \cdots$ which holds for $0 \leq f\ < 1$. The t in the preceding proposition is slightly larger than 0.8. Therefore the cyclic ordering (resp. betweenness) constraint $F(V, a/n^2)$ is satisfiable with high probability for $a\ < 0.8$ proving Proposition 1 (a) .

Concerning unsatisfiability we can use a standard first moment argument (as know from random k-SAT for example): We consider the cylic ordering constraint. Given a fixed linear ordering with the three variables x, y, z ordered as $x < y < z$. The clause (x, y, z) and its cyclic permutations become true under the ordering. The remaining 3 clauses over x, y, z are false. Altogether we have asymptotically $n^3/2$ clauses which are false under the ordering. As clauses are picked independently the probability that a random $F(V,\ a/n^2)$ is true under the ordering therefore is
$$(1-a/n^2)^{n^3/2} < \exp(-an/2).$$

The expected number of linear orderings satisfying a random $F(V, a/n^2)$ then is bounded above by $n! \times \exp(-an/2)$. This approaches 0 only when $a \geq \ln n$. Markov's inequality shows that the random formula is unsatisfiable with high probability in this case. Concerning betweenness constraints we get $2n^3/3$ instead of the $n^3/2$ before and thus essentially the same result. We can get unsatisfiability already for constant a.

Lemma 8. *(a) Let F be a satisfiable cyclic ordering constraint. There exists a partition of the variables into 3 sets K, L, M each with $n/3$ variables such that we have no clause from $K \times M \times L$ or its cyclic permutations in F.*

(b) Let F be a satisfiable betweenness constraint. There exists a partition of the variables into two sets K, L each with $n/2$ variables such that we have no clause from $K \times L \times K$ and no clause from $L \times K \times L$ in F.

Proof. (a) Let K be the first third of the variables of the satisfying ordering, L the second third , and M the last third of the variables. A clause from $K \times M \times L$ and its cyclic permutations is false under the ordering and thus must not belong to F.

(b) As (a) using the first and second half of the variables instead.

We come to the proof of Proposition 1 . Given a partition K, L, M as in the proof of Lemma 8. There are $3(n/3)^3$ clauses belonging to $K \times M \times L$ or its cyclic permutations. The probability that a random formula does not contain one of these clauses is

$$(1 - a/n^2)^{3(n/3)^3} \leq \exp(-an/9).$$

The expected number of partitions K, L, M such that we have no clause from $K \times M \times L$ or its cyclic permutations is bounded above by $3^n \cdot \exp(-an/9) = \exp((\ln 3 - a/9)n)$. This goes to 0 when $a/9 > \ln 3$ and the first moment argument implies (b) of Proposition 1. Concerning (c) we proceed analogously.

3 The threshold

We refer to [11] which we apply to our setting of random formulas. Random formulas can be seen as random 3-uniform directed hypergraphs. Unsatisfiability of a cyclic ordering (betweenness) constraint is a monotone property, as it is preserved under the additon of clauses. Morever, unsatisfiability is a property which is invariant under permutation of the variables. Therefore it satisfies the symmetry properties necessary for the application of the criterion from [11].

We need some abbreviations. For a formula F we abbreviate the property that F is unsatisfiable by UNSAT(F), in case of the random formula $F = F(V, p)$ we also write UNSAT(V, p). SAT(F) means F is satisfiable.

A variety of types of random formulas are needed: The random instance $F(V, p) \cup F(V, q)$ is constructed by first picking $F(V, p)$ and second and independently adding $F(V, q)$ to the formula picked . For a given formula M the random formula M^* is a random copy of M (formula isomorphic to M) on variables from V. A formula M is balanced iff its average degree (= $\sum_x$ degree of x in M / $\sharp$variables x of M) is not smaller that that of any subformula of M. The formula $F(V, p) \cup M^*$ is obtained by picking $F(V, p)$ and M^* independently and forming their union.

We say that UNSAT(V, p) has a coarse threshold iff we have a critical probability $p_c = p_c(n)$ and (small) constants $\varepsilon, \varepsilon'$ and δ such that $\varepsilon <$ Pr[UNSAT(V, p_c)] $< 1 - \varepsilon$ and $\varepsilon' <$ Pr[UNSAT$(V, (1 + \delta)p_c) < 1 - \varepsilon'$. In the case at hand $p_c = a/n^2$.

And a threshold is coarse iff adding an arbitrarily small, but linear number of clauses to $F(V, p_c)$ does *not* yield unsatisfiability with high probability (cf. the remark after Proposition 1.)

There exist several formulations of Friedgut's threshold criterion. In our case most convenient to apply is

Fact 9 (cf. Corollary 2.3 of [12]). *If UNSAT*(V, p) *has a coarse threshold then there exist*

- *a (critical) probability* $p_c = p_c(n)$,
- *a balanced formula* M *and a constant* $b > 0$ *, and*
- *a constant* $\varepsilon > 0$

such that for infinitely many n *we have*

- $\varepsilon <$ *Pr [UNSAT*(V, p_c) *]* $< 1 - \varepsilon$,
- *the expected number of copies of* M *in* $F(V, p_c)$ *is* $\geq b$ *, and*
- *Pr [UNSAT*$(F(V, p_c) \cup M^*)$ *]* $-$ *Pr [UNSAT*$(F(V, p_c) \cup F(V, \varepsilon p_c))$ *]* $\geq \varepsilon$.

Fact 9 says that in case of a coarse threshold there exists a formula M such that adding M^* to $F(V, p_c)$ is more likely to make $F(V, p_c)$ unsatisfiable than adding a (small) but linear number of random clauses. We will show that any M with the properties of Fact 9 such that Prob$[F(V, p_c \cup M^*)$ unsatisfiable $] \geq$ Prob$[F(V,\, p_c)$ unsatisfiable $] + \varepsilon$ necessary for the last inequality cannot satisfy the last inequality.

In the subsection to come we prove a crucial Lemma (from which the experienced reader already sees that we must have a sharp threshold.) The details for the contradiction are presented Subsection 3.2.

3.1 The Lemma

We split the of variables V into two disjoint subsets U and W. U is a set of k variables where k is a fixed constant independent of n, and W contains the remaining $n - k$ variables. The random instance $F(U, W, p)$ is obtained by picking each clause with at least one variable from U independently with probability p.

For arbitrary fixed formulas F over W and F^* over U (think of F^* as a fixed instance of the random formula M^* above) we consider the random formula over $V = W \cup U$ given by $F \cup F^* \cup F(U, W, p)$. We assume that $p = c/n^2$ where $c_1 < c < c_2$ for constants $c_1, c_2 > 0$ independent of n. Note that $c = c(n)$ itself may depend on n.

For the rest of this subsection F and F^* are given formulas and both satisfiable as a betweenness (cyclic ordering) constraint. Let $x_1 < x_2 < \cdots < x_{n-k}$ be an ordering satisfying F and $y_1 < y_2 < \cdots < y_k$ for F^* (where $\{x_1, \ldots, x_{n-k}\} = W$, $\{y_1, \ldots, y_k\} = U$.) We assume that the underlying n is sufficiently large and $\varepsilon > 0$ is a constant independent of n. Let UNSAT be the event that the random instance $F \cup F^* \cup F(U, W, p)$ is unsatisfiable as a betweenness (cyclic orering) constraint. Recall that $F(W, p)$ is the usual random instance with variables from W.

We have

Lemma 10. *Let $\varepsilon > 0$ be a constant. If Pr[UNSAT] $> \varepsilon$ then for any constant $\delta > 0$ the random instance $F \cup F(W, \delta \cdot p)$ interpreted as a betweennness (cyclic ordering) constraint is unsatisfiable with high probability.*

Proof idea. Prob[UNSAT] $>$ ε means that Prob[$F \cup F^* \cup F(U, W, p)$ is unsatisfiable] $> \varepsilon$. Unsatisfiability is only caused by $F(U, W, p)$. If $F' = F(U, W, p)$ causes unsatisfiability then $F \cup F^* \cup F'$ is in particular false under all orderings with $W_1 < U < W_2$ for $W_1 \cup W_2 = W$. That is, we consider the orderings in which the variables of U are adjacent. When we substitute the variables from U in F' with an arbitrary variable x from W we get an unsatisfiable formula over W. As $F(U, W, p)$ causes unsatisfiabilty with probability $> \varepsilon$ we get that a linear number of random clauses over W makes F unsatisfiable even with high probability. This holds, as in this case *each* variable of W can serve as x.

Proof. As $F \cup F^*$ is satisfiable, unsatisfiability is due to the clauses from $F(U, W, p)$. The expected number of clauses of $F(U, W, p)$ is $\mathrm{E}[\sharp F(U, W, p)] = 3kc(1 + o(1))$. Let M be a sufficiently large constant such that $\sharp F(U, W, p) \le M$ with probability $\ge 1-(1/10)\varepsilon$. Such an $M = 10\,\mathrm{E}[\sharp F(U, W, p)]/\varepsilon$ can be found with Markov's inequality.

We have

$$\begin{aligned}
&\text{Pr[UNSAT } \mid \sharp F(U,W,p) \le M \text{]} \\
&\ge \text{Pr[UNSAT and } \sharp F(U,W,p) \le M \text{]} \\
&\ge \text{Pr[UNSAT]} - \Pr[\sharp F(U,W,p) > M] \\
&\ge \varepsilon \cdot (1 - 1/10) = (9/10)\varepsilon.
\end{aligned}$$

Let $F(U, W, M)$ be the random instance which is a uniform random set of exactly M clauses with at least one variable from U. The random instance $F \cup F^* \cup F(U, W, M)$ is unsatisfiable with probability $\ge 9/10\varepsilon$ by monotonicity, which is shown by the following (not unknown, cf. [4], page 34 for related arguments) general consideration:

We consider a general monotonously increasing event A over $\{0, 1\}^n$ endowed with the binomial distribution with p. Let $M \le n$ and a_m be the number of elements of A with exactly m 1's, let b_m be the binomial coefficient $\binom{n}{m}$.

We need to show that

$$\frac{\sum_{m=0}^{M} a_m p^m (1-p)^{n-m}}{\sum_{m=0}^{M} b_m p^m (1-p)^{n-m}} \le \frac{a_M}{b_M}. \tag{1}$$

Note that the left-hand side is the conditional probability and the right-hand side the probability for exactly M 1's under the uniform distribution.

We observe that

$$a_m \cdot \binom{n-m}{M-m} \le a_M \cdot \binom{M}{m}.$$

This because the number of ordered pairs of elements (a, b) with $a, b \in A$ such that a has exactly m 1's and b has M 1's and extends a is equal to the left hand side of the preceding inequality. Starting with b each a can be obtained by deleting a set of $M-m$ 1's out of M. This gives the upper bound.

Next we observe that

$$a_m \leq a_M \frac{\binom{M}{m}}{\binom{n-m}{M-m}} \text{ and } \frac{\binom{M}{m}}{\binom{n-m}{M-m}} b_M \;=\; b_m.$$

This implies the required result by substitution into inequality (1).

Let $F(W,x,M)$ be the random set of M clauses each of them containing the variable x. Next we show that $F \cup F(W,x,M)$ is unsatisfiable with probability $\geq (9/10)\varepsilon$.

Let $\{L_1, \dots L_M\}$ be an instance of $F(W,x,M)$. If $F \cup \{\, L_1, \dots,\, L_M \,\}$ is *satisfiable* then we have that $F \cup F^* \cup \{\, K_1, \dots,\, K_M \,\}$ is satisfiable where K_i is obtained by replacing the variable x in L_i with an arbitrary variable from U. For if

$$x'_1 < x'_2 \cdots < x'_i < \cdots < x'_{n-k} \text{ where } \; x'_i \;=\; x$$

makes $F \cup \{\, L_1, \dots,\, L_M \,\}$ true then $F \cup F^* \cup \{\, K_1, \dots,\, K_M \,\}$ is true for

$$x'_1 < x'_2 \dots x'_{i-1} < y_1 < y_2 < \; \dots \; < y_k < x'_i < x'_{i+1} < \cdots < x'_{n-k}.$$

This holds because $y_1 \;<\; y_2 \;<\; \dots \;<\; y_k$ is a satisfying ordering of F^* and the K_i's contain exactly one variable from U and do not contain the variable $x = x'_i$ any more.

Each satisfiable instance of $F \cup\; F(W,x,M)$ induces k^M distinct satisfiable instances of $F \cup F^* \cup F(U,W,M)$. Moreover, for different instances of $F(W,x,M)$ the corresponding sets of k^M instances of $F(U,W,M)$ are disjoint. Therefore the number of instances of $F(U,W,M)$ inducing satisfiabilty is at least as large as $k^M \times$ the number instances of $F(W,x,M)$ inducing satisfiability.

We come to the number of all instances. With high probability an instance of $F(U,W,M)$ does *not* contain:

- A clause with two or three variables from U.
- Two or more clauses with the same two variables from W.
- A clause with x.

This follows as k and M are both constant by elementary expectation calculations together with Markov's inequality. Therefore $F(U,W,M)$ has (asymptotically because of the three preceeding items)

$$\binom{3(n-k-1)(n-k-2)k}{M} \;=\; \frac{(3(n-k-1)(n-k-2)k)^M}{M!}(1+o(1))$$

many instances. For $F(W,x,M)$ we have

$$\binom{3(n-k-1)(n-k-2)}{M} \;=\; \frac{(3(n-k-1)(n-k-2))^M}{M!}(1+o(1))$$

instances. The quotient is k^M. As $F(U,W,M)$ and $F(W,x,p)$ are uniformly distributed we finally get

$$\text{Pr}\,[\,F \cup F(W,x,M)\ \text{sat.}]\ \leq\ \text{Pr}\,[\,F\ \cup F^* \cup\ F(U,W,M)\ \ \text{sat.}\ ] \leq (1/10)\varepsilon.$$

Let $\varepsilon' \geq (9/10)\varepsilon$ be the probability that $F \cup F(W,x,M)$ is unsatisfiable. An instance $\{L_1,\ldots,L_M\}$ of $F(W,x,M)$ is called *bad* iff $F \cup \{L_1,\ldots,L_M\}$ is unsatisfiable.

We come to $F(W,p')$ with $p' = \delta p$. Let BAD_x be the the event: There are *exactly* – only to simplifiy the subsequent calculation – M clauses containing x and this set of clauses is bad. The probability of BAD_x is (recall $p = c/n^2$)

$$\begin{aligned} &= \varepsilon' \cdot \binom{3(n-k-1)(n-k-2)}{M} p'^M \cdot (1-p')^{3(n-k-1)(n-k-2)-M} \\ &= \varepsilon' \cdot \frac{(3\delta c)^M}{M!} \cdot \exp(-3\delta c)(1+o(1)). \end{aligned}$$

We abbreviate the probability of BAD_x as

$$\varepsilon'' = \varepsilon' \cdot \frac{(3\delta c)^M}{M!} \cdot \exp(-3\delta c)\text{ , still a constant } > 0.$$

The expectation of the number of variables x such that BAD_x is $= \varepsilon'' n$. The subsequent second moment argument shows that we have at least one variable x with BAD_x with high probability, and the lemma holds.

The second moment argument rests on the observation that the events BAD_x and BAD_y are essentially independent for $x \neq y$: The number of clauses over W containing both x and y is $3 \cdot 2 \cdot (n-k-2)$ and we have asymptotically

$$\text{PROB [BAD}_x \text{ and BAD}_y\text{]}$$

$$\begin{aligned} &= \left(\varepsilon' \binom{3(n-k-1)(n-k-2) - 6(n-k-2)}{M} p'^M \cdot (1-p')^{3(n-k-1)(n-k-2)-M}\right)^2 \\ &= \varepsilon''^2 (1+o(1)), \end{aligned}$$

as the probability of clauses with both x and y is $O(1/n)$. The remaining calculation only applies Tchebyscheff's inequality, see [2] for the argument.

3.2 The contradiction

Proof of Theorem 2. We show that the items as necessary for a coarse threshold as stated in Fact 9 together cannot exist. Thus the threshold cannot be coarse, but must be sharp and Theorem 2 holds.

We have from Proposition 1 that $p_c = a/n^2$ and $0.8 \leq a \leq 9.88$.

We show that the formula M must be satisfiable as cyclic ordering (betweenness) constraint. This follows because (as detailed next) M has no non-trivial 2-core, cf. Section 2.

Let k be the number of variables in M and t be the number of clauses. Then the expectation of the number of copies of M is $O\left(n^k \cdot \left(\frac{a}{n^2}\right)^t\right)$. We must have $t \leq k/2$ in order for the expectation to be $\geq b$. An analogous argument applies to any subformula of M as M is balanced. As $k/2$ clauses have $3k/2 < 2k$ slots for k variables there must always be a variable of degree < 2.

To visualize the calculation of some (conditional) probabilities we generate $F(V,p) \cup M^*$ with a different process from the standard one (first, generate $F(V,p)$ and second and independently add M^*.) The distribution however remains the same. Additional visualization (if necessary) can be gained by thinking in terms of a probability tree. Random steps occurring sequentially are independent.

1. Pick a random set $U \subseteq V$ of k variables (k is the number of variables of M.) Generate the random formula M^* isomorphic to M on U. We let $W = V \setminus U$.
2. Add $F(W, p_c)$ to M^*.
3. Add $F(U, W, p_c)$ (cf. the beginning of Subsection 3.1 for notation.)

We come to $F(V, p_c) \cup F(V, \varepsilon p_c)$. Given an arbitrary set U of k variables, $W = V \setminus U$, it can be generated as follows.

1. Generate $F(W, p_c)$.
2. Generate $F(W, \varepsilon p_c)$
3. Generate the rest necessary to get an instance of $F(V, p_c) \cup F(V, \varepsilon p_c)$. (Note $V \setminus W \neq \emptyset$.)

Let $\varepsilon > 0$ be fixed and assume that the last item of Fact 9 holds, that is

$$\text{Pr [UNSAT}(F(V,p_c) \cup M^*) \text{]} \;-\; \text{Pr [UNSAT}(F(V,p_c) \cup F(V,\varepsilon p_c)) \text{]} \;\geq\; \varepsilon.$$

This assumption is shown to contradict Lemma 10.

Concerning the process to generate $F(V,p) \cup M^*$, we observe that the probability spaces generated by the second and third step are isomorphic. Therefore we condition what comes on a fixed choice for the first step that is we fix U and M^* and start the proecess with the second step. Concerning the process for $F(V, p_c) \cup F(V, \varepsilon p_c)$, we can use the same U.

We have

$$\text{Pr [UNSAT}(F(V,p_c) \cup M^*) \text{]} \;=$$

$$= \sum_{F,\ \text{SAT}(F)} \text{Pr [} F = F(W,p_c) \text{]} \cdot \text{ Pr [UNSAT}(M^* \cup F \cup F(U,W,p_c)) \text{]}$$

$$+ \sum_{F,\ \text{UNSAT}(F)} \text{Pr [} F = F(W,p_c) \text{]}.$$

And

$$\Pr [\text{ UNSAT}(F(V,p_c) \cup F(V,\varepsilon p_c))]$$

$$= \sum_{F,\ \text{SAT}(F)} \Pr [\ F = F(W,p_c)\ ]\cdot \Pr [\text{ UNSAT}(F \cup F(W,\varepsilon p_c) \cup \text{“the rest, i. e. from Step 4”})]$$

$$+ \sum_{F,\ \text{UNSAT}(F)} \Pr [\ F = F(W,p_c)\ ].$$

Then

$$\Pr [\text{ UNSAT}(F(V,p_c) \cup M^*)] \ - \ \Pr [\text{ UNSAT}(F(V,p_c) \cup F(V,\varepsilon p_c))] \ =$$

$$= \sum_{F,\ \text{SAT}(F)} \Pr [\ F = F(W,p_c)\ ]\cdot$$

$$\cdot \Big(\Pr [\text{ UNSAT}(M^* \cup F \cup F(U,W,p_c))] - \Pr [\text{ UNSAT}(F \cup F(W,\varepsilon p_c) \cup \text{“ Step 4” }] \Big).$$

Our assumption above implies that we must have a satisfiable formula F over W such that

$$\Pr [\text{ UNSAT}(M^* \cup F \cup F(U,W,p_c))] - \Pr [\text{ UNSAT}(F \cup F(W,\varepsilon p_c) \cup \text{“ Step 4”})] \ > \varepsilon.$$

However, such an F cannot exist because by Lemma 10 any such F must already satisfy

$$\Pr [\text{ UNSAT}(F \cup F(W,\varepsilon p_c))] \ = 1 - o(1).$$

4 The cyclic ordering constraint

We come to the proof of Theorem 3.

4.1 The reduction

In addition to 3-clauses as above we need 2-clauses which are ordered pairs (x, y) consisting of two different variables. Their interpretation is $x < y$ throughout, 3-clauses are interpreted as cyclic ordering constraints from now on. The graph associated to the set of 2-clauses E is called G_E. It is obtained by viewing each 2-clause (x, y) as the directed edge $x \to y$. The random instance $F(V, p, q)$ is the union of two independent instances of $F(V, 3, p)$ and $F(V, 2, q)$ where $F(V, 3, p) = F(V, p)$ picks each 3-clause with p independently and $F(V, 2, q)$ each 2-clause with q. For subsets $A, B, C \subseteq V$ we use the notation $(A, B, C) = A \times B \times C$ and $(A, B) = A \times B$.

The following definition is the basis of our reductions.

Definition 11. *Let $F = D \cup E$ be a constraint with 2-clauses E and 3-clauses D over the set of variables V. Let A, B be any partition of V into two disjoint sets with $A \cup B = V$. If E contains clauses from both, (A, B) and (B, A) the constraints F_A and F_B are not defined. Otherwise the constraint F_A over A is defined as:*

The 2-clauses of E which belong to (A, A) are in F_A.

Let $(x, y, z) \in D$ with at least two variables belonging to A. It induces clauses as follows in F_A :

- $(x, y, z) \in (A, A, A)$ *implies* $(x, y, z) \in F_A$
- $(x, y, z) \in (A, A, B)$ *implies* $(x, y) \in F_A$
- $(x, y, z) \in (A, B, A)$ *implies* $(z, x) \in F_A$
- $(x, y, z) \in (B, A, A)$ *implies* $(y, z) \in F_A$.

Constraint F_B is defined in totally the same way exchanging the roles of A and B above.

Note that the 2-clauses from F belonging to (A, B) (resp. (B, A)) get lost when constructing F_A and F_B. When G_E has a cycle the whole cycle must either belong to A or to B in order that F_A and F_B are defined.

The next Lemma states the satisfiability properties which are preserved when comparing F with F_A and F_B.

Lemma 12. *F without 2-clauses from (B, A) is satisfiable by a linear ordering with $A < B$ iff F_A and F_B are satisfiable.*

Proof. Let F be without 2-clauses from (B, A). Then F_A and F_B are defined.

For "$\Rightarrow$", let F be satisfiable by an ordering with $A < B$. Then F_A and F_B are satisfied by the same ordering, or better its restrictions to A and B.

Look at F_A. The 3-clauses from F_A are 3-clauses from F and therefore are true under the ordering. Let $C = (x, y)$ be a 2-clause from F_A. Then C is true by a case distinction as to which clause from F induces C.

- C is a 2-clause from F. then C is true.
- C is induced by the 3-clause (x, y, z). As $z \in B$ in this case we have $x < y$ in the ordering making F true and C is true.
- C is induced by (y, z, x). Again $x < y$ as $z \in B$.
- C is induced by (z, x, y). Then $x < y$ as before.

By Definition 11 these are all possibilities which induce C in F_A and we are finished.

In the same way we check that F_B is true under the ordering of F. For completeness the details: The 3-clauses are true. Again let $C = (x,\, y)$ be a 2-clause from F_B. We look at the cases which induce C.

- C is a 2-clause from F. Then C is true.
- C is induced by the 3-clause (x, y, z). As $z \in A$ and $A < B$ we have $x < y$ in the ordering making F true and C is true.

- C is induced by (y, z, x). Again $x < y$ as $z \in A$.
- C is induced by (z, x, y). Then $x < y$ as before.

These are all possibilities to induce C and F_B becomes true.

Now "$\Leftarrow$". Assume that F_A is true under an ordering of A and F_B under one of B. We show that F becomes true under the ordering in which the one of B is appended to that A so that in particular $A < B$.

The 2-clauses from F belong to $(A\,,\,A)$, $(B\,,\,B)$, or $(A\,,B)$ and they are all true under the ordering because F_A and F_B are true.

Concerning the 3-clauses of F we have some cases. 3-clauses from (A, A, A) and (B, B, B) are true as F_A and F_B are true.

We look at 3-clauses from F with exactly two variables from B. Let $C = (x, y, z)$ be such a clause.

- $x,\ y \in B$. Then $x < y$ as we have the 2-clause (x, y) in F_B and C becomes true as $A < B$.
- $x, z \in B$. Then $z < x$ and C is true as $A < B$.
- $y, z \in B$. Then $y < z$ and we are done.

We omit the analogous argument for 3-clauses with two variables from A.

We consider a set of 2-clauses E over the set of variables V. The following notions are the natural graph theoretic ones as induced by G_E the graph associated to E. The outdegree of a variable x, denoted as $\mathrm{Odeg}_E(x)$, is the number of clauses $(x, -) \in E$. The set of neighbors of x is $N_E(x) = \{y|(x, y) \in E\}$. A neighbor of x is reached by *one* edge from x in G_E.

Definition 13. *(a) The boundary $B_0 = B_{0,E}$ is defined by*

$$B_0 = \{x|\,\textit{Odeg}(x) = 0\}.$$

(b) $B_1 = B_{1,E}$ is

$$B_1 = \{x|\,N_E(x) \subseteq B_0 \textit{ and } N_E(x) \textit{ is not empty}\}.$$

(c) $B = B_E = B_0 \cup B_1$ is the boundary.
(d) The interior is

$$\textit{Int} = \textit{Int}_E = V \setminus B.$$

(e) When F consists of 3-clauses D and 2-clauses E we let $B_F = B_E$ .... for all notions introduced.

The proof of Theorem 3 is an analysis of the reduction as visualized in Figure 1. F is a formula over V consisting only of 3-clauses. A is a fixed set of one half of the variables from V and B (not to be confused with a boundary) is the other half.

Note that all formulas of the tree comply with Definition 11 and thus are defined, as any formula F has no 2-clauses from (B_F, Int_F) by Definiton 13. With Lemma

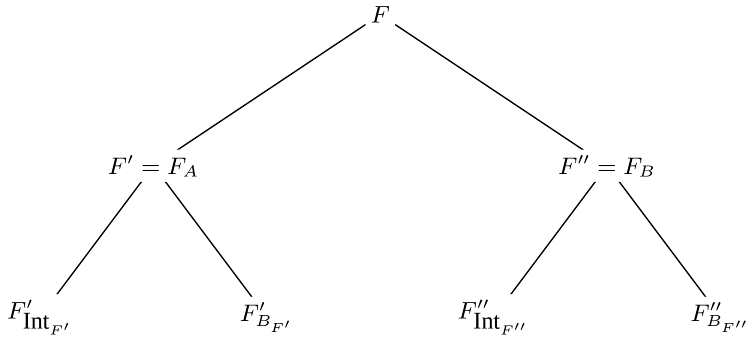

Fig. 1.

12 the root F is satisfiable if all formulas at the leaves are satisfiable. We prove that for $F = F(V, 1/n^2)$ the probability of the event that *all* formulas at the leaves $F'_{\mathrm{Int}_{F'}}, F'_{B_{F'}}, F''_{\mathrm{Int}_{F''}}, F''_{B_{F''}}$ are simultaneously satisfiable does not go to 0.

Theorem 14 concerns the first step of the reduction and shows that F' and F'' can be treated independently. Note that to us only the case $a = 1$ of the Theorem is of interest.

Theorem 14. *Let a be a constant. $F(V, 3, a/n^2)$ is satisfiable with probability $> \varepsilon' > 0$ if $F(V, b/n^2, c/n)$ with $b = (1/4)a$ and $c = (3/4)a$ is satisfiable with probability $> \varepsilon > 0$.*

Proof. We consider an arbitrary but fixed partition A, B of the n variables in V into two halves, for example $A = \{x_1, \ldots, x_{n/2}\}$. When $F = F(V, 3, a/n^2)$ the 3-clauses of F_A according to Definition 11 are distributed as $F(A, 3, a/n^2)$.

A given 2-clause (x, y) is induced by $3n/2$ many 3-clauses as can easily checked in Definition 11. It is present in F_A with probability $1 - (1 - a/n^2)^{3n/2} = 3a/(2n)$, asymptotically, independently. Therefore the 2-clauses of F_A are distributed as $F(A, 2, 3a/(2n)$ and independent from the 3-clauses. Substituting $m = n/2$ we get that F_A is distributed as $F(A, (1/4)a/m^2, (3/4)a/m)$.

The same applies to F_B and the claim follows with Lemma 12 with $\varepsilon' = \varepsilon \cdot \varepsilon$. Note that F_A and F_B are independent as A, B are sets fixed beforehand.

4.2 The second step of the reduction

We analyze the probability that the formulas at the leaves of Figure 1 are satisfiable by investigating (the absence of) cycles in graphs associated to these formulas.

Definition 15. *Let F be a constraint of 2- and 3-clauses.*

(a) A trivial ordering of F is an ordering which has $x < y$ for each 2-clause (x, y) and $x < y$ and $y < z$ for each 3-clause (x, y, z).

(b) The directed graph associated to F is denoted by G_F. It has as vertices the variables of F. Its edges are the 2-clauses of F and for each 3-clause (x, y, z) of F the three edges (x, y), (y, z) and (x, z).

(c) By a cycle of length $s \geq 2$ in G_F resp. F we mean a set of edges $(x_1, x_2),\ (x_2, x_3), \ldots\ (x_s\, x_1)$ in G_F with $x_1, \ldots x_s$ all different.

Obvious consequences of the previous definition collects

Lemma 16. *Let F be a formula.*

(a) F is satisfiable iff there exists a formula H which contains the 2-clauses of F and for each 3-clause of F a cyclic permutation of this clause such that the graph G_H has no cycle.

(b) F is satisfiable by a (or any) trivial ordering iff G_F is cycle free.

(c) A trivial ordering of F can be found efficiently by sorting G_F topologically.

The proof of Theorem 3 proceeds by showing the random graphs associated to the formulas at the leaves of Figure 1 are cycle free with probability not going to 0.

We take a look at cycles in standard random graphs. Following [9] who treat the undirected case, [22] seems to be the first work on the directed case. The classical directed random graph $G(n,p)$ is obtained by picking each of the n^2 directed edges (including loops) with probability p independently. The basic result concerning cycles is from [22] and [17], the giant component threshold.

Fact 17. *Let $G = G(n, c/n)$.*
(a) For constant $0 \leq c \leq 1$ Prob$[G$ has a directed cycle $] = c(1 + o(1))$.
(b) If $c > 1$ the G has a strongly connected component of linear size with high probability. For $c < 1$ we have no strongly connected component of linear size.

Note that the our graphs neither are standard random graphs, in particular edges may be dependent, nor are they independent of each other, $F'_{\mathrm{Int}_{F'}}$ and $F'_{B_{F'}}$ are dependent as are the F'''s.

To provide some perspective, we sketch the proof in [22] of (a) (it should be well known to the random graph expert). Let $p = c/n$ with $c < 1$.

For $k \geq 1$ we have $\binom{n}{k} \cdot (k-1)!$ possible cycles of length k over n vertices. The probability of a fixed cycle is p^k and the expected number of cycles of length k is asymptotically

$$\sum_C \text{Prob } [C] \;=\; \frac{c^k}{k}$$

where C ranges over all possible cycles of length k. Disregarding any asymptotic detail we have that the expectation of the number of all cycles is

$$\sum_D \text{Prob } [D] \;=\; \sum_{k \geq 1} \sum_C \text{Prob } [C] \;=\; \sum_k \frac{c^k}{k} \;=\; -\ln(1-c)$$

by the logarithm series. Here D ranges over all possible cycles and given k, C over all cycles of length k.

Next an "obvious sieve" (citation from [9]). The probability that a cycle exists is

$$\sum_{D} \text{Prob}\,[D] \;-\; \sum_{D_1,D_2} \text{Prob}\,[D_1 \text{ and } D_2] + \cdots \tag{2}$$

where D ranges over all cycles, D_1, D_2 ranges over all *sets* of two different cycles. $\ldots D_1, \ldots D_s$ over all sets of exactly s cycles. Some more detailed analysis of the probability of cycles with common edges shows that sufficiently many sets of different cycles are independent and

$$\sum_{D_1,\ldots,D_m} \text{Prob}\,[D_1 \text{ and } \ldots \text{ and } D_k] \;=\; \frac{1}{m!}\cdot\left(\sum_{D} \text{Prob}\,[D]\right)^m \;=\; \frac{1}{m!}(-\ln(1-c))^m$$

The definiton of the exponential function yields the result by plugging this into equation (2). In our case the detailed analysis of the dependencies between different cycles required here may well be possible, but should be lengthy. We use a different approach (Lovasz Local Lemma) to bound the probability of the absence of cycles from below.

The subsequent Theorem 18 together with Lemma 12 and Theorem 14 implies Theorem 3.

Theorem 18. *For $F = F(V, b/n^2, c/n)$ with $b = 1/4$ and $c = 3/4$ we have Prob$[F_{\text{Int}}$ and F_B are satisfiable $]$ is bounded strictly above 0.*

To prove Theorem 18 we show that both F_{Int} and F_B are cycle free in the sense of Definition 15 (c). Note that F_{Int} and F_B are not any more independent as F_A and F_B are in Theorem 14.

The stochastic dependency problem requires some preparation. Usually one thinks of $F = F(V, p, q)$ as being generated first by picking each 3-clause with its probability p and then each 2-clause with q. We use a different generation process which in the end yields the right distribution. In its first step the process determines the boundaries $B_{0,E}$, $B_{1,E}$ where E is the set of 2-clauses of a random formula F. This allows to calculate probabilities conditional on given boundaries B_0 and B_1 in a transparent way.

The process is motivated by the following formula. Given disjoint sets B_0 and B_1 and we observe that Prob$[B_{0,E} = B_0$ and $B_{1,E} = B_1]$ is equal to the following big product

$$\begin{aligned}\prod_{x\in B_0}(1-q)^{n-1} \;\times\; &\prod_{x\in B_1} \text{Prob}[N_E(x) \subseteq B_0 \text{ and } |N_E(x)| \geq 1] \times \\ &\times \prod_{x\in \text{Int}} \text{Prob}[N_E(x) \cap V \setminus B_0 \;\text{ is not empty}].\end{aligned} \tag{3}$$

This is the case as edges starting at different vertices are independent as G_E, the graph associated to E is a directed graph.

For any x we have $x \in B_{0,E}$ with probability $(1-q)^{n-1}$ independent of anything else. Let B_0 be a fixed set, $b_0 = |B_0|$. Condition on $B_{0,E} = B_0$ and assume that the 2-clauses E of a random formula come. For any $x \notin B_0$ we get that $x \in B_{1,E}$ with probability
$(1-(1-q)^{b_0})(1-q)^{n-1-b_0}$ independent of anything else. . In the same way we get for $x \notin B_0$ that $x \in \mathrm{Int}_E$ with probability $1-(1-q)^{n-1-b_0}$.

This suggests the following generation process of $F = F(V,p,q)$. The process has five independent steps, that is the probabilities multiply. Recall the notation of Definition 13 (e).

1. For $x \in V$ decide independently $x \in B_{0,F}$ with probability $(1-q)^{n-1}$ and $x \notin B_{0,F}$ with probability $1-(1-q)^{n-1}$. We abbreviate $b_0 = |B_{0,F}|$.

2. For $x \notin B_0$ decide independently $x \in B_{1,F}$ with

$$\frac{(1-q)^{n-1-b_0} \cdot (1-(1-q)^{b_0})}{1-(1-q)^{n-1}},$$

decide $x \notin B_{1,E}$ with

$$\frac{1-(1-q)^{n-1-b_0}}{1-(1-q)^{n-1}}$$

(Note that the sum of these probabilities is 1.)

3. Now we start to generate the 2-clauses in the boundary $B = B_0 \cup B_1$. For each $x \in B_1$ we consider all clauses in (x, B_0). For $b_0 \geq k \geq 1$ every set with k such clauses has probability

$$\frac{q^k(1-q)^{b_0-k}}{1-(1-q)^{b_0}}.$$

Add such a random set. (Note that for each vertex the sum of the probabilites is 1.)

4. We start to generate the 2-clauses with at least one variable from the interior Int $= V \setminus B$. For $x \notin B$ we consider all clauses in $(x, V \setminus B_0)$. Each set of $n-1-b_0 \geq k \geq 1$ of these clauses has probability

$$\frac{q^k(1-q)^{n-1-b_0-k}}{1-(1-q)^{n-1-b_0}}$$

We add such a random set with its probabilty.

5. We add each clause from (Int, B_0) with probability q independently.

With the big product above it is easy to show that this process generates the 2-clauses from $F(V,p,q)$ The 3-clauses are added and cause no problem as they are independent of anything done by now.

The next lemma is easily proved based on the process above. The conditional probabilities are simply calculated by starting the process with Step 3.

Lemma 19. *Let $F = F(V,p,q)$ and let B_0, B_1 be given disjoint sets of variables. Conditional on the event $B_{0,F} = B_0$ and $B_{1,F} = B_1$ the constraints F_{Int_F} and F_{B_F} are independent of each other.*

We fix some notation.

$$\begin{aligned}\beta_0 &= \exp(-c) + o(1),\\ \beta\ &= \exp(-c(1-\beta_0)) + o(1)\\ &= \exp\left(-c(1-\exp(-c))\right) + o(1),\\ \beta_1\ &=\ \beta\ -\ \beta_0 \text{ and } \gamma\ =\ 1-\beta.\end{aligned}$$

The $o(1)$-terms and in particular their quantification depend on the context. In the assumption of a theorem we have universal quantification, in the conclusion existential one.

We get some concentration results.

Lemma 20. *For $F = F(V, b/n^2, c/n)$ we have with high probability the following equalities:*

$$\begin{aligned}|B_{0,F}|\ &=\ \beta_0 n,\\ |B_F|\ &=\ \beta n,\\ |B_{1,F}|\ &=\ \beta_1 n\\ |\mathit{Int}| &= \gamma n.\end{aligned}$$

Proof. (a) follows from independence. (b) from independence, conditioning on the fact that $B_{0,F}$ is a set of variables satisfying (a). Note that we only need a standard second moment argument for the concentration (no stronger bounds like Chernoff.)

Lemma 21. *We consider $F(V, b/n^2, c/n)$ with $b = 1/4, c = 3/4$. Let B_0 be a set consisting of $\beta_0 n$ variables. Let B_1 be disjoint from B_0 with $\beta_1 n$ variables. Let F be a random instance conditional on the event that $B_{0,F} = B_0$ and $B_{1,F} = B_1$. For the conditional probabilities holds:*
(a) Prob$[F_{\mathit{Int}}$ has no cycle $] > \varepsilon > 0$
(b) Prob$[F_B$ has no cycle$] > \varepsilon > 0$

Given the preceding three lemmas, Theorem 18 follows by combining them. Just for orientation (by pocket calculator) some values:

$$\begin{aligned}\beta_0 = \exp(-c) \approx 0.4723,\ \ c(1-\beta_0) \approx 0.3957,\\ \beta\ =\ \exp(-c(1-\beta_0))\ \approx\ 0.67319\\ \beta_1 \approx 0.2,\quad \gamma \approx 0.3268\end{aligned}$$

Proof of Lemma 21 (a). We abbreviate

$$d\ =\ 3b + \frac{c}{\gamma}\ \ =\ \ \frac{3}{4}\cdot\frac{\gamma+1}{\gamma}\ \ =\ \frac{3}{4}\cdot\frac{2-\beta}{\gamma}. \tag{4}$$

In estimates we sometimes and without explicitly mentioning it enlarge d by a sufficiently small but constant amount. This in order to bound $d + o(1)$ above by (the enlarged) d. Not that this is particularly simplifying when bounding terms like $(d+o(1))^n$.

Crucial is $d\gamma = 3/4 \cdot (2-\beta) < 1$. The value β above yields $d\gamma \approx 0.9951$. More exactly: $\ln(15/7) > 0.76 > 3/4 = c$ (pocket calculator) then $\exp(-c) < 7/15$ then $c(1-\exp(-c)) < 2/5$. As $\exp(2/5) < 3/2$ (pocket calculator) we have $\beta = \exp(-c(1-\exp(-c))) > 2/3$ and $3/4(2-\beta) < 1$.

We denote $G_{\text{Int}} = (V, E_{\text{Int}})$. Consider two fixed variables $x, y \in$Int, then $x \rightarrow y \in E_{\text{Int}}$ with probability

$$1 - \left(1 - \frac{b}{n^2}\right)^{3(n-2)} \left(1 - \frac{c}{n} \cdot \frac{1}{1-(1-c/n)^{n-1-\beta_0 n}}\right)$$
$$= \left(\frac{1}{n}\left(3b + \frac{c}{\gamma}\right)\right)(1+o(1)) = \frac{d}{n}(1+o(1)).$$

This is so because $x \rightarrow y$ can be induced by one of $3(n-2)$ many 3-clauses or by the 2-clause (x, y). The probability that the 2-clause (x, y) is present is $c/n \cdot 1/(1-(1-c/n)^{n-1-\beta_0 n}) = (1/n) \cdot (c/\gamma)(1+o(1))$. Note that two 2-clauses like (x, y) and (x, y') are not stochastically independent. As $|$ Int $| = \gamma n$ and $d\gamma < 1$ it turns out that we are in a situation analogous to Lemma 17 (a), at least as far as cycles are concerned.

W.l.o.g. we can restrict attention to cycles which do not contain two edges which are induced by one 3-clause in F. If this is the case we would have a piece like $\dots \rightarrow x \rightarrow y \rightarrow z \dots$ with $x, y, z \in$Int on the cycle and a 3-clause $(x, y, z) \in F$. We substitute $\dots \rightarrow x \rightarrow y \rightarrow z \dots$ with $\dots x \rightarrow z \dots$ to get a shorter cycle. Then we proceed inductively.

Given a possible cycle $x_1 \rightarrow x_2 \rightarrow x_3 \dots x_s \rightarrow x_1$ with $x_i \in$Int the edges which induce this cycle are stochastically independent. The probability of the cycle $\leq (d/n)^s$ Here the preceding restriction is used. The expected number of cycles of length $n \geq s \geq 2$ is

$$\leq \binom{\gamma n}{s} \cdot (s-1)! \cdot \left(\frac{d}{n}\right)^s \leq \frac{(d\gamma)^s}{s}. \tag{5}$$

The expected number of cycles is asymptotically a constant, $-\ln(1-d\gamma) - d\gamma$ by the logarithm series. As the present situation seems to have (slightly) more stochastic dependencies than the $G_{n,p}$-case, the argument to come is not based on a direct (and tedious) analysis of the dependencies between different cycles. We use the Lovasz Local Lemma instead. Our formulation is from page 53/54 of [2].

Given a constant $\varepsilon > 0$ we have a constant S such that the expected number of cycles of length $> S$ is $< \varepsilon$ (by (5) and $d\gamma < 1$.) And the probability to have a cycle of length $> S$ is $< \varepsilon$. We fix ε sufficiently small and S accordingly. Then

$$\text{Prob}[\text{No cycle}] \geq \text{Prob}[\text{No cycle of length} \leq S] - \varepsilon.$$

To apply the Lovasz Local Lemma we need some notational preparation. For $2 \leq s \leq S$ let

$$c_s = \binom{\gamma n}{s} \cdot (s-1)! = (\gamma n)_s / s \tag{6}$$

be the number of all cycles of length s possible in G_{Int}. Recall the standard notation $(m)_s = m\cdot(m-1)\cdot(m-2)\cdots(m-s+1)$. We number all possible cycles of length s with $1,\dots c_s$ and consider events $C_{s,j}$, $1\le j\le c_s$. Event $C_{s,j}$ says that cycle j of length s is present. The $C_{s,j}$ correspond to the events A_i in the Lovasz Local Lemma.

We set

$$x_s \;=\; (d/n)^s \tag{7}$$

where the present d is larger by an arbitrarily small but constant amount than the d introduced in (4). This because we need Prob$[C_{s,-}]\le x_s(1-o(1))$.

The event $C_{s,i}$ has stochastic dependencies only with those events $C_{t,-}$ whose cycle has variables in common with the cycle of $C_{s,i}$. There are $O((\gamma n)^{t-1})$ such events. Note that $s,t\le S$ a constant. We have

$$x_s\cdot\prod_{t=2}^{S}(1-x_t)^{O(n^{t-1})} = x_s(1-o(1)). \tag{8}$$

As Prob$[C_{s,-}]\le x_s(1-o(1))$ for all $S\ge s\ge 2$ the assumptions of the Local Lemma hold. We conclude using $c_s \;\le\; (\gamma n)^s/s$ and the logarithm series in the subsequent calculation

$$\begin{aligned}
\text{Prob}\,[\bigwedge_{i,s}\neg\, C_{s,i}] &\ge \prod_{s=2}^{S}(1-x_s)^{c_s} \;\ge\; \prod_{s=2}^{S}(1-x_s)^{(\gamma n)^s/s}\\
&= \prod_{s=2}^{S}\exp\left(-(d\gamma)^s/s\right) + o(1) \;=\; \exp\left(-\sum_{s=2}^{S}(d\gamma)^s/s\right) \;+\; o(1)\\
&= \exp\left(\ln(1-d\gamma) + d\gamma + \sum_{s\ge S+1}(d\gamma)^s/s\right) + o(1)\\
&> (1-d\gamma)\cdot\exp(d\gamma)+o(1)
\end{aligned} \tag{9}$$

As the final term is a constant $>\;0$ independent of the S picked above the proof is finished.

Proof of Lemma 21 (b). Denote $G_B \;=\; (V,E_B)$. As in the proof of (a) w.l.o.g. we restrict attention to those cycles in G_B for which F does not have any 3-clause which induces two edges belonging to the cycle.

Let x,y be two variables from B. For $(x,y)\notin(B_1,B_0)$ the edge $x\to y$ can only be induced by one of $3(n-2)$ 3-clauses and

$$\Pr[\,x\to y\;\in E_B\,] \;=\; 3b\cdot(1/n)+O(1/n^2) = (3/4)(1/n)(1+o(1))$$

For $(x,y)\in(B_1,B_0)$ the edge $x\to y$ can be induced by the 2-clause (x,y) or by one of the 3-clauses. We get

$$
\begin{aligned}
&\Pr[\, x \to y \,\in E_B \,] \\
&= \; 1 \,-\, \left(1 - \frac{b}{n^2}\right)^{3(n-2)} \cdot \left(1 \,-\, \frac{c}{n} \cdot \frac{1}{1-(1-c/n)^{\beta_0 n}}\right) \\
&= \; \frac{1}{n}\left(3b + \frac{c}{1-\exp(-c\beta_0)}\right)(1+o(1)) \\
&= \; \frac{1}{n} \cdot \frac{3}{4} \cdot \left(1 + \frac{1}{1-\exp(-c\beta_0)}\right)(1+o(1)).
\end{aligned}
$$

We abbreviate

$$
d \,=\, 1 + \frac{1}{1-\exp(-c\beta_0)} \;=\; \frac{2 \,-\, \exp(-c\beta_0)}{1 -\, \exp(-c\beta_0)}.
$$

We have $\exp(-c\beta_0) \approx 0.7$ and $d \approx 4.34$ and $(3/4)d \approx 3.3255$ which is relatively large as we need constants < 1. But (B_1, B_0) has only $\approx 0.1n^2$ many candidate edges. This is important for our argument.

First, we count the number of possible paths through B of length s with exactly $k \leq s/2$ edges from (B_1, B_0) starting in k fixed slots, the first slot following the last one. We have

$$
\leq (\beta_1 n \beta_0 n)^k \cdot (\beta n)^{s-2k} \tag{10}
$$

possibilities. The probability that the cycle as induced by such a path is present is

$$
\leq \left(\frac{1}{n} \cdot \frac{3}{4} \cdot d\right)^k \cdot \left(\frac{3}{4} \cdot \frac{1}{n}\right)^{s-k} \;=\; \left(\frac{1}{n} \cdot \frac{3}{4}\right)^s d^k \tag{11}
$$

We multiply both preceding upper bounds with the (generous) bound $\binom{s}{k}$ for the number of positions where the edges from (B_1, B_0) start and with $1/s$ because of cyclic permutations. This yields that the expected number of cycles of length s with exactly k edges from (B_1, B_0) is

$$
\leq \frac{1}{s}\left(\frac{3}{4}\right)^s \binom{s}{k} \cdot (\beta_1 \beta_0 d)^k \cdot \beta^{s-2k} \;=\; \frac{1}{s}\left(\frac{3}{4} \cdot \beta\right)^s \cdot \binom{s}{k}\left(\frac{\beta_1 \beta_0 d}{\beta^2}\right)^k.
$$

The expected number of cycles of length s is

$$
\begin{aligned}
&\leq \; \frac{1}{s}\left(\frac{3}{4} \cdot \beta\right)^s \cdot \sum_{k=0}^{s/2} \binom{s}{k}\left(\frac{\beta_1 \beta_0 d}{\beta^2}\right)^k \\
&\leq \; \frac{1}{s}\left(\frac{3}{4} \cdot \beta\right)^s \left(1 \,+\, \frac{\beta_1 \beta_0 d}{\beta^2}\right)^s \\
&= \; \frac{1}{s}\left(\frac{3}{4}\right)^s \cdot \left(\beta \,+\, \frac{\beta_1 \beta_0 d}{\beta}\right)^s.
\end{aligned} \tag{12}
$$

We need to show that the base under the exponent s is strictly less than 1. We can write $\beta = \exp(-c(1-\beta_0)) = \beta_0 \cdot \exp(c\beta_0)$. We recall $\beta_1 = \beta - \beta_0$ and calculate

$$\begin{aligned}
&\beta + \frac{\beta_1\beta_0}{\beta} \cdot d \\
&= \beta + \frac{(\beta_0 \cdot \exp(c\beta_0) - \beta_0) \cdot \beta_0}{\beta_0 \exp(c\beta_0)} \cdot \frac{2 - \exp(-c\beta_0)}{1 - \exp(-c\beta_0)} \\
&= \beta + \frac{\beta_0(\exp(c\beta_0) - 1)}{1} \cdot \frac{2 - \exp(-c\beta_0)}{\exp(c\beta_0) - 1} \\
&= \beta + \beta_0 \cdot (2 - \exp(-c\beta_0)) \\
&= \beta_0 \exp(c\beta_0) + \beta_0 \cdot (2 - \exp(-c\beta_0)) \\
&= \beta_0 \cdot (2 - \exp(-c\beta_0) + \exp(c\beta_0)) \,.
\end{aligned}$$

We bound (by calculator) $\beta_0 \leq 0.48$ then $c \cdot \beta_0 \leq 0.36$ and $\exp(c\beta_0) \leq 144/100$ and $-\exp(-c\beta_0) \leq -100/144$. Using these bounds we get

$$\beta_0 \cdot (2 - \exp(-c\beta_0) + \exp(c\beta_0)) \leq 1897728/1440000 < 4/3$$

as $3 \cdot 1897727 = 5693184 < 5760000 = 1440000 \cdot 4$ and the base of the exponentiation in (12) is bounded above by a constant < 1. Let const be this constant. We continue as in the proof of (a) only with two parameters, s, k instead of s. The probability that long cycles exist can be made arbitrarily small as const < 1.

For a candidate cycle of length s with exactly k edges from (B_1, B_0) we introduce the event $C_{s,k,i}$ where $1 \leq i \leq$ the number of all such cycles. We let, compare (6) and (10)

$$c_{s,k} = \frac{1}{s} \cdot \binom{s}{k} \cdot (\beta_1 n \beta_0 n)^k \cdot (\beta n)^{s-2k}$$

be an upper bound to the number of all such cycles.

Next we set, compare (7) and (11)

$$x_{s,k} = \left(\frac{1}{n} \cdot \frac{3}{4}\right)^s \cdot d^k.$$

Any cycle of length s has dependencies only with $O((\beta n)^{t-1})$ cycles of length t whereas $x_{t,-} = O((1/n)^t)$. The assumptions of the Local Lemma hold, compare (8), and we finally get that we have no cycle in G_B with probability

$$\geq (1 - \text{const}) \exp(\text{const}) + o(1)$$

compare (9), which is a constant > 0.

Acknowledgement

The author thanks Andre Lanka for helpful hints and thoughtful questions.

References

1. D. Achlioptas, , A. Naor, A.,Y. Peres. The fraction of satisfiable clauses in a typical formula. In Procceedings 44th FOCS ,2003, 362–370.
2. Noga Alon, Joel H. Spencer. The Probabilistic Method. Wiley-Interscience Publications 1992.
3. Paul Beame ,R. Karp ,T. Pitassi, M. Saks., The efficiency of resolution and Davis-Putnam procedures. SIAM Journal on Computing 31(4) ,2002, 1048 - 1075
4. Bela Bollobas. Random Graphs. Academic Press 1985.
5. Nir Ailon, Noga Alon. Hardness of fully dense problems. Information and Computation. To appear.
6. Thomas Christof, Michael Junger, John Kececioglu, Petra Mutzel, Gerhard Reinelt. A branch-and-cut approach to physical mapping with end-probes.
7. B. Chor, M. Sudan. A geometric approach to betweenness, SIAM Journal on Disctete Mathematics, 11(4), 1998, 511 - 523.
8. O. Dubois, Y. Boufkhad , J. Mandler. Typical random 3-SAT formulae and the satisfiability threshold. In Proceedings 11th SODA , 2000, 126-127.
9. Pal Erdös, A. Renyi. On the evolution of random graphs. Publications of the Mathematical Institute of the Hungarian Academy of Science 5, 1960, 17 - 61.
10. Ehud Friedgut. Necessary and sufficient conditions for sharp thresholds of graph properties and the k-SAT problem. Journal of the American Mathematical Society 12, (1999, 1017-1054.
11. Ehud Friedgut. Hunting for sharp thresholds. Random Structures and Algorithms 26(1-2),2005, 37-51 .
12. Ehud Friedgut, Michael Krivelevich. Sharp Thresholds for Ramsey Properties of Random Graphs. Random Structures and Algorithms 17 (1), 2000, 1- 19.
13. Zvi Galil, N. Meggido. Cyclic ordering is **NP**-complete. Theoretical Computer Science 5(2),1977, 179 - 182.
14. Walter Guttmann,Markus Maucher. Variations on an ordering theme with constraints. In Proceedings Fourth IFIP International Conference on Theoretical Computer Science, TCS 2006, IFIP International Federation for Information Processing 209, Springer, 77-90.
15. Amar Isli, Anthony G. Cohen. An Algebra for cyclic ordering of 2D orientation. In Proceedings of the 15th American Conference on Artificial Intelligence (AAAI), AAAI/MIT press, 1998, 643 - 649.
16. Hamed Hatami, Michael Molloy. Sharp Thresholds for Constraint Satisfaction Problems and Homomorphisms. Manuscript.
17. Richard M. Karp. The transitive closure of a random digraph. Random Structures and Algorithms 1(1), 73 - 93.
18. Michael Molloy. Models and thresholds for random constraint satisfaction problems. In Proceedings STOC 2002, ACM, 209-217.
19. Michael Molloy. When does the giant component bring unsatisfiability ? Combinatorica. To appear.
20. Michael Molloy. The pure literal rule threshold and cores in random hypergraphs. In Proceedings SODA 2004, 665 - 674.
21. J. Opatrny. Total ordering problem. SIAM Journal on Computing. 8(1), 1979, 111- 114.
22. I. Palasti. On the threshold distribution function of cycles in a directed random graph. Studia Scientiarium Mathematicarum Hungarica 6, 1971, 67 - 73.