Discrete Mathematics and Theoretical Computer Science **AC**, 2003, 113–126

Linear Phase Transition in Random Linear Constraint Satisfaction Problems

David Gamarnik

Department of Mathematical Sciences, IBM T.J. Watson Research Center, Yorktown Heights NY 10598, USA.
gamarnik@watson.ibm.com

Our model is a generalized linear programming relaxation of a much studied random K-SAT problem. Specifically, a set of linear constraints $\mathcal{C}$ on K variables is fixed. From a pool of n variables, K variables are chosen uniformly at random and a constraint is chosen from $\mathcal{C}$ also uniformly at random. This procedure is repeated m times independently. We are interested in whether the resulting linear programming problem is feasible. We prove that the feasibility property experiences a linear phase transition, when $n \to \infty$ and $m = cn$ for a constant c. Namely, there exists a critical value c^* such that, when $c < c^*$, the problem is feasible or is asymptotically almost feasible, as $n \to \infty$, but, when $c > c^*$, the "distance" to feasibility is at least a positive constant independent of n. Our result is obtained using the combination of a powerful local weak convergence method developed in Aldous [Ald92], [Ald01], Aldous and Steele [AS03], Steele [Ste02] and martingale techniques. By exploiting a linear programming duality, our theorem implies the following result in the context of sparse random graphs $G(n, cn)$ on n nodes with cn edges, where edges are equipped with randomly generated weights. Let $\mathcal{M}(n, c)$ denote maximum weight matching in $G(n, cn)$. We prove that when c is a constant and $n \to \infty$, the limit $\lim_{n\to\infty} \mathcal{M}(n, c)/n$, exists, with high probability. We further extend this result to maximum weight b-matchings also in $G(n, cn)$.

Keywords: Random K-SAT, Satisfiability Threshold, Linear Programming, Sparse Random Graphs

1 Introduction

The primary objective of the present paper is studying randomly generated linear programming problems. We are interested in scaling behavior of the corresponding objective value and some phase transition properties, as the size of the problem diverges to infinity. Our random linear programming problems are generated in a specific way. In particular, our linear programs have a fixed number of variables per constraint and the number of variables and constraints diverges to infinity in such a way that their ratio stays a constant.

Our motivation to consider this specific class of random linear programs has several sources. The main motivation is recent explosion of interest in random instances of boolean satisfiability (K-SAT) problems and ensuing phase transition phenomenon. The main outstanding conjecture in this field states that the satisfiability property of random K-SAT problem experiences a linear phase transition as the function of the ratio of the number of clauses to the number of variables. Our linear programming problem can be viewed as a generalized linear programming relaxation of the integer programming formulation of such random K-SAT problem.

Tightly related to the K-SAT problem are problems of maximal cardinality cuts, independent sets, matchings and other objects, in sparse random graphs $G(n,cn)$, which are graphs on n nodes and cn edges selected uniformly at random from all the possible edges, and where $c>0$ is some fixed constant. It is easy to show that all of these objects scale linearly in n. It is conjectured that the size of each such object divided by n converges to a constant, independent of n. This convergence is established only for the case of maximal matchings using direct methods [KS81], where the limit can be computed explicitly, but is open in other cases.

The main result of this paper states that the objective value of the random linear programming problem we consider, when divided by the number of variables converges with high probability (w.h.p.) to a certain limit. As a corollary we prove that, suitably defined, distance to feasibility in the same random linear programming problem experiences a *linear* phase transition, just as conjectured for random K-SAT problem. Furthermore, we show that in a special case, the dual of this random linear programming problem turns out to be a linear programming relaxation of the maximum cardinality matching and more generally b-matching (defined later) problems in $G(n,cn)$. We show that these relaxations are asymptotically tight as the number of nodes n diverges to infinity. As a corollary of our main result, we prove that maximum cardinality b-matching when divided by n converge to a constant. These results hold even in the weighted version, where edges are equipped with randomly independently generated non-negative weights.

Our proof technique is a combination of a very powerful method of *local weak convergence* and martingale techniques. The local weak convergence method was developed in Aldous [Ald92], [Ald01], Aldous and Steele [AS03], Steele [Ste02]. The method was specifically used by Aldous for proving the $\zeta(2)$ conjecture for the random assignment problem. It was used in [Ald92] to prove that the expected minimum weight matching in a complete bipartite graph with randomly generated weights converges to a certain constant. Later Aldous proved [Ald01] that this limit is indeed $\zeta(2)$, as conjectured earlier by Mézard and Parisi [MP87]. Since then the local weak convergence method was used for other problems (see [AS03] for a survey), and seems to be a very useful method for proving existence of limits in problems like the ones we described, and in some instances also leads to actually computing the limits of interest. By an analogy with the percolation literature, we can call these problems existence and computation of *scaling limits* in large random combinatorial structures. Such questions, many of them open, abound in percolation literature [Sch01], [SW01], but are still open in the case of other lattices, like rectangular bond and site critical percolation, see Langlands [LPSA94]. Whether a local weak convergence is a useful technique for addressing these questions seems worth investigation.

To the extend that we know, our result is the first application of the local weak convergence method to establishing phase transitions in random combinatorial structures. In the following section we describe in details randomly generated combinatorial problems we mentioned above, describe the existing results in the literature and list some outstanding conjectures. In Section 3 we describe our model and state our main results. We also give an outline of the main proof steps. The full proof is too lengthy to fit this extended abstract. Section 4, is devoted to the application of our results to maximal matching and b-matching in sparse random graphs. Section 5 is devoted to conclusions and some open problems.

2 Background: random K-SAT, sparse random graphs and scaling limits

2.1 Random K-SAT problem

A satisfiability or K-SAT problem is a boolean constraint satisfaction problem with a special form. A collection of n variables $x_1, x_2, \ldots, x_n$ with values in $\{0,1\}$ is fixed. A boolean formulae of the form $C_1 \wedge C_2 \wedge \cdots \wedge C_m$ is constructed, where each C_i is a disjunctive clause of the form $x_{i_1} \vee \bar{x}_{i_2} \vee \bar{x}_{i_3} \vee \cdots \vee x_{i_K}$, where exactly K variables are taken from the pool $x_1, \ldots, x_n$, some with negation, some without. The formulae is defined to be satisfiable if an assignment of variables $x_i, i = 1, \ldots, n$ to 0 or 1 can be constructed such that all the clauses take value 1. The K-SAT problem is one of the most famous combinatorial optimization problem, see [PS98].

Recently we have witnessed an explosion of interest in random instances of the K-SAT problem. This was motivated by computer science, artificial intelligence and statistical physics investigations, with phase transition phenomena becoming the focus of a particular attention. A random instance of a K-SAT problem with m clauses and n variables is obtained by selecting each clause uniformly at random from the entire collection of $2^K n^K$ ($2^K \binom{n}{K}$) possible clauses where repetition of variables is (is not) allowed. In particular, for each $j = 1, 2, \ldots, m$, K variables $x_{i_1}, x_{i_2}, \ldots, x_{i_K}$ or their negations are selected uniformly at random from the pool $x_i, 1 \le i \le n$ to form a clause $C_j = y_{i_1} \vee \ldots y_{i_k}$, where each $y_{i_r} = x_{i_r}$ or $\bar{x}_{i_r}$, equiprobably. This is done for all $j = 1, \ldots, m$ independently. Whether the resulting formulae has a satisfying assignment $\{x_1, \ldots, x_n\} \to \{0,1\}^n$ becomes a random event with respect to this random construction. The main outstanding conjecture for the random K-SAT problem is as follows.

Conjecture 1 *For every $K \ge 2$ there exists a constant c_K^* such that a random K-SAT formulae with n variables and $m = cn$ clauses is satisfiable when $c < c_K^*$ and is not satisfiable when $c > c_K^*$, w.h.p. as $n \to \infty$. In other words, the satisfiability experiences a* linear *sharp phase transition at $m = c_K^* n$.*

That the problem experiences a sharp phase transition is proven by Friedghut [Fri99] in a much more general context. It is the linearity which is the main outstanding feature of this conjecture. The conjecture can be rephrased as follows: there does not exist $c_1 > c_2$ and two infinite sequences $n_t^{(1)}, n_t^{(2)}, t = 1, 2, \ldots,$ such that instances of K-SAT problem with $n_t^{(1)}$ variables and $c_1 n_t^{(1)}$ clauses are satisfiable w.h.p., but instances with $n_t^{(2)}$ variables and $c_2 n_t^{(1)}$ clauses are not satisfiable w.h.p., as $t \to \infty$. The main result of our paper establishes an analogue of this conjecture for generalized linear programming relaxations of the integer programming formulation (to be described below) of the random K-SAT problem. Conjecture 1 is proven for the case $K = 2$. Specifically, $c_2^* = 1$ was established by Goerdt [Goe92], [Goe96], Chvatal and Reed [CR92], Fernandez de la Vega [dlV92]. For higher values of K many progressively sharper bounds on c_K^* (assuming it exists) are established by now. For $K = 3$ the best known upper and lower bounds are 4.506 and 3.42, obtained by Dubois, Boufkhad and Mandler [DBM00], and Kaporis, Kirousis and Lalas [KKL02], respectively. It is known that c_K^*, if exists, approaches asymptotically $2^K(\log 2 + o(1))$ when K is large, [APa]. See also [AM02], [FW02] for the related results.

2.2 Matching and b-matching in $G(n, cn)$

Let G be a simple undirected graph on n nodes $\{1, 2, \ldots, n\} \equiv [n]$ with an edge set E. A set of nodes $V \subset [n]$ in this graph is an independent set if no two nodes in V are connected by an edge. A matching is a collection of edges such that no two edges are incident to the same node. The size of the matching is the number of edges in it. Let $b \geq 1$ be a positive integer. A b-matching is a collection of edges $A \subset E$ such that every node is incident to at most b edges from A. Naturally, 1-matching is simply a matching. Note that 2-matching is collection of node disjoint paths and cycles. We will also call it path/cycle packing.

Fix a constant $c > 0$. Let $G(n, cn)$ denote a simple undirected sparse random graph on n nodes with $\lfloor cn \rfloor$ edges selected uniformly at random from all the possible $n(n-1)/2$ edges. This is a standard model of a sparse random graph. For future we drop the annoying notation $\lfloor \cdot \rfloor$, assuming that cn is always an integer. Denote by $IND(n,c)$ and $\mathcal{M}(n,c,b)$ the size of the maximum independent set and b-matching, respectively, in $G(n,cn)$. Suppose, in addition, the nodes and the edges of $G(n,cn)$ are equipped with random non-negative weights $W^{\text{node}}_{i,j}, W^{\text{edge}}_{i,j}$ drawn independently according to some common probability distributions $\mathbb{P}\{W^{\text{node}} \leq t\} \equiv w^{\text{node}}(t), \mathbb{P}\{W^{\text{edge}} \leq t\} \equiv w^{\text{edge}}(t)$. We assume throughout the paper that both W^{node} and W^{edge} have a bounded support $[0, B_w]$ (assumed the same for simplicity). Let $IND_w(n,c)$ abd $\mathcal{M}_w(n,c,b)$ denote the maximum weight independent set and b-matching, respectively.

It is well known and simple to prove that $IND(n,c)$ and $\mathcal{M}(n,c,1)$ are all $\Theta(n)$ w.h.p. as n diverges to infinity. It is natural to suspect then that the expected values of these objects divided by n converge to a constant, both in the unweighted and weighted cases. In other words, the scaling limits exist for these objects. The conjecture is stated in [Ald] and [AS03] for independent sets. The existence of these limits for expectation would also imply almost sure limits, by application of Azuma's inequality. It is known, though, that the limit $\lim_n \mathcal{M}(n,c)/n$ exists and it can be computed. This is obtained via Karp–Sipser algorithm [KS81]. The result was strengthened later by Aronson, Frieze and Pittel [APF98]. The Karp-Sipser algorithm is quite remarkable in its simplicity, however, it does not apply to the weighted case. Moreover, it is not clear how to extend the Karp-Sipser heuristic to b-matchings. In this paper we prove the existence of the limit $\lim_n \mathcal{M}_w(n,c,b)/n$ for the weighted case. The proof uses the main result of the paper and the linear programming duality, though we are not able to compute the limits. Naturally, our result applies to the non-weighted case – maximum cardinality b-matching. To the best of our knowledge this is a new result.

The case of maximum weight matching with random weights is also treated by Aldous and Steele [AS03] for the case of a randomly generated tree on n nodes. That is, consider a tree selected uniformly at random from the set of all possible n^{n-2} labelled trees. The limit $\lim_n \mathcal{M}_w(n,c)/n$ is proven and computed using the local weak convergence method, when the edges of this tree are equipped with exponentially distributed random weights. The tree structure of the underlying graph helps very much the analysis. In our case, however, the random graph $G(n,cn)$ contains a linear size non-tree "giant" component, [JLR00], when $c > 1/2$, and the results of ([AS03]) are not applicable.

3 Model and the main results

There is a natural way to describe a K-SAT problem as an integer programming problem. The variables are $x_i, i = 1,2,\dots,n$ which take values in $\{0,1\}$. Each clause C_j is replaced by a linear constraint of the form $x_{i_1} + (1 - x_{i_2}) + x_{i_3} + \ldots \geq 1$, where term $(1-x)$ replaces $\bar{x}$. For example a clause $C = x_3 \vee x_7 \vee \bar{x}_2 \vee \bar{x}_4$ in a 4-SAT problem is replaced by a constraint $x_3 + x_7 + (1 - x_2) + (1 - x_4) \geq 1$. It is easy to check that an assignment of x_2, x_3, x_4, x_7 to 0 and 1 gives C value 1 if and only if the corresponding constraint is satisfied. Clearly, these constraints can be created for all the possible clauses. In the present paper we study the linear programming (LP) relaxation of this integer programming problem, where the restriction $x_i \in \{0,1\}$ is replaced by a weaker restriction $x_i \in [0,1]$. Note, that this relaxation by itself is not interesting, as the assignment $x_i = 1/2$ for all $i = 1,2,\dots,n$ makes all of the linear constraints feasible. However, the problem becomes non-trivial when we generalize the types of constraints that can be generated on the variables x_i, and this is described in the following subsection.

3.1 Random K-LSAT problem

Our setting is as follows. Consider a fixed collection of K variables $y_1, y_2, \dots, y_K$ which take values in some bounded interval $B_x^1 \le y_i \le B_x^2$ and a fixed collection $\mathcal{C}$ of linear constraints on these variables: $\sum_{k=1}^{K} a_{rk} y_k \le b_r, r = 1,2,\dots,|\mathcal{C}|$, where the values a_{rk}, b_r are arbitrary fixed reals. The r-th constraint can also be written in a vector form $a_r y \le b_r$, where $a_r = (a_{r1},\dots,a_{rK})$ and $y = (y_1,\dots,y_K)$. We fix $c > 0$ and let $m = cn$, where n is a large integer. A random instance of a linear constraint satisfaction problem with $n+m$ variables $x_1,\dots,x_n,\psi_1,\dots,\psi_m$ and m constraints is constructed as follows. For each $j = 1,2,\dots,m$ we perform the following operation independently. We first select K variables $x_{i_1}, x_{i_2}, \dots, x_{i_K}$ uniformly at random from $x_i, i = 1,2,\dots,n$. Whether the variables are selected with or without replacement turns out to be irrelevant to the results of this paper, as it is the case for random K-SAT problem. However, the order with which the variables are selected is relevant, since the constraints are not necessarily symmetric. Then we select $1 \le r \le |\mathcal{C}|$ also uniformly at random. We then generate a constraint

$$C_j : \sum_{k=1}^{K} a_{rk} x_{i_k} \le b_r + \psi_j. \tag{1}$$

Here is an example of an instance with $K = 3, n = 10, m = 4, |\mathcal{C}| = 2$. Say the first constraint C_1 is $2y_1 + 3y_2 - y_3 \le 5$, and the second constraint C_2 is $-y_1 + y_2 + 4y_3 \le 2$. An example of an instance where first three constraints are type C_1 and the fourth is type C_2 is

$$(2x_5 + 3x_4 - x_9 \le 5 + \psi_1) \wedge (2x_1 + 3x_3 - x_4 \le 5 + \psi_2) \wedge$$
$$(2x_2 + 3x_1 - x_{10} \le 5 + \psi_3) \wedge (-x_5 + x_8 + 4x_7 \le 2 + \psi_4).$$

The central question is what are the optimal values of $B_x^1 \le x_i \le B_x^2, \psi_j \ge 0$, which minimize the sum $\sum \psi_j$ subject to the constraints C_j. That is, we consider the following linear programming problem:

$$\text{Minimize} \sum_{1 \le j \le m} \psi_j, \text{ subject to} : C_1, C_2, \dots, C_m,\ x_i \in [B_x^1, B_x^2], \psi_j \ge 0. \tag{2}$$

In words, we are seeking a solution x_j which is as close to satisfying the constraints $\sum_{k=1}^{K} a_{ri_k} x_{i_k} \le b_r$ as possible. If the optimal value of this linear programming problem is zero, that is $\psi_j = 0$ for all j, then all of these constraints can be satisfied. Naturally, the objective value of the linear program (2) is a random variable. We denote this random variable by $\mathcal{LP}(n,c)$. Note, that the linear program (2) is always feasible, by making ψ_j sufficiently large. In fact, clearly, in the optimal solution we must have $\psi_j = \max(0, \sum_{k=1}^{K} a_{ri_k} x_{i_k} - b_r)$. We refer to the linear program (2) as a *random linear constraint satisfaction (LSAT) problem*, or *random K-LSAT problem*.

The following conditions on the set of constraints $\mathcal{C}$ will be used below.

- **Condition $\mathcal{A}$.** For any constraint $a_r y \le b_r, 1 \le r \le |\mathcal{C}|$ for any $k \le K$ and for any value $z \in [B_x^1, B_x^2]$ there exist values $y_1,\dots,y_K \in [B_x^1, B_x^2]$ such that $y_k = z$ and the constraint is satisfied.

- **Condition $\mathcal{B}$.** There exist a positive integer l and a constant $\nu > 0$ such that for any K-dimensional cube I of the form $\prod_{1\le k\le K}[\frac{i_k}{l},\frac{i_k+1}{l}], B_x^1 \le \frac{i_k}{l} < B_x^2$, i_k integer, there exists at least one constraint $\sum a_{rk}y_k \le b_r$ from $\mathcal{C}$ such that for every $y \in I$, $\sum a_{rk}y_k - b_r \ge \nu$. That is, every point of the cube I deviates from satisfying this constraint by at least ν.

The analogue of the Condition $\mathcal{A}$ clearly holds for random K-SAT problem. Given any clause $y_1 \vee y_2 \vee \cdots \vee y_K$ and $k \le K$, if y_k is set to be 0 or 1, we still can satisfy the clause, by satisfying any other variable. The following is an example of an LSAT problem where Conditions $\mathcal{A}$ and $\mathcal{B}$ are satisfied. Fix $K = 3$. Let $B_x^1 = 0, B_x^2 = 1$, and let $\mathcal{C}$ be a collection of all eight constraints of the type $-y_1 - y_2 - y_3 \le -7/4, -(1-y_1) - y_2 - y_3 \le -7/4, \dots, -(1-y_1)-(1-y_2)-(1-y_3) \le -7/4$. Condition $\mathcal{A}$ is checked trivially. We claim that Condition $\mathcal{B}$ holds for $l = 2$ and $\nu = 1/4$. Select any cube I with side-length $1/l = 1/2$. For example $I = [0,1/2]\times[1/2,1]\times[1/2,1]$. Consider constraint $-y_1-(1-y_2)-(1-y_3) \le -7/4$. For any $y \in I$ we have $-y_1-(1-y_2)-(1-y_3) \ge -7/4+1/4 = -7/4+\nu$. Other cases are analyzed similarly.

Consider now the following generalization of the linear program (2). For each $j = 1,2,\dots,m$ generate a random variable W_j, independently from some common distribution $\mathbb{P}\{W_j \le t\}$ with a bounded support $[-B_w, B_w]$. Let $w_x \ge 0$ and $w_\psi > 0$ be fixed non-negative constants. Our random linear program in variables x_i, ψ_j is constructed exactly as above except each constraint $C_j : \sum_{1\le r\le K} a_{rk}x_{i_k} \le b_r + \psi_j$ is replaced by

$$C_j : \sum_{1\le r\le K} a_{rk}x_{i_k} \le b_r + W_j + \psi_j, \tag{3}$$

and the objective function is replaced by

$$\text{Minimize } w_x \textstyle\sum_{1\le i\le n} x_i + w_\psi \sum_{1\le j\le m} \psi_j, \tag{4}$$

$$\text{subject to :} \qquad C_1, C_2, \dots, C_m,\ x_i \in [B_x^1, B_x^2],\ \psi_j \ge 0.$$

This particular form of the linear program might look unnatural at first. But note that setting $B_w = w_x = 0, w_\psi = 1$, turns this into exactly linear program (2). We will show later that this general format is useful when we study b-matchings in sparse random graphs $G(n, cn)$. We denote the optimal value of the linear program (4) by $\mathcal{GLP}(n,c)$. As before, this linear program is always feasible, by making ψ_j sufficiently large. Since we assumed $w_\psi > 0$, then in the optimal solution

$$\psi_j = \max(0, a_{rk}x_{i_k} - b_r - W_j). \tag{5}$$

We now state the main result of this paper. In words, our result asserts that the scaling limit of $\mathcal{GLP}(n,c)/n$ exists.

Theorem 1 *For every $c \ge 0$, the limit*

$$\lim_{n\to\infty} \frac{\mathcal{GLP}(n,c)}{n} \equiv f(c) \tag{6}$$

exists w.h.p. That is, there exists $f(c) \ge 0$ such that for every $\varepsilon > 0$,

$$\mathbb{P}\{|\frac{\mathcal{GLP}(n,c)}{n} - f(c)| > \varepsilon\} \to 0$$

as $n \to \infty$.

Our first application of Theorem 1 is the following result. It establishes a linear phase transition property for the random K-LSAT problem. Recall that $\mathcal{LP}(n,c)$ is the optimal value of the linear programming problem (2).

Theorem 2 *There exists a constant $c_K^* > 0$ such that, w.h.p. as $n \to \infty$,*

$$\lim_{n\to\infty} \frac{\mathcal{LP}(n,c)}{n} = 0, \tag{7}$$

for all $c < c_K^*$, and

$$\liminf_{n\to\infty} \frac{\mathcal{LP}(n,c)}{n} > 0, \tag{8}$$

for all $c > c_K^*$. Moreover, if Condition $\mathcal{A}$ holds, then $c_K^* > 0$, and if Condition $\mathcal{B}$ holds, then $c_K^* < +\infty$.

In what sense does the theorem above establish a *linear* phase transition? It is conceivable that for a collection of constraints $\mathcal{C}$, the following situation occurs: there exist two constants $c_1 > c_2$ and two sequences $n_t^{(1)}, n_t^{(2)}, t = 0,1,2,\ldots$, such that for $c = c_1$ the corresponding optimal values (2) of the random K-LSAT problem satisfy w.h.p. $\lim_t \mathcal{LP}(n_t^{(1)}, c)/n_t^{(1)} = 0$, but for $c = c_2$, $\liminf_t \mathcal{LP}(n_t^{(2)}, c)/n_t^{(2)} \geq \delta(c) > 0$. In other words, the critical density c oscillates between different values. This is precisely the behavior that Conjecture 1 rules out for random K-SAT problem. Our theorem states that such a thing is impossible for the random K-LSAT problem. There exists a linear function $c_K^* n$ such that, w.h.p., below this function the instance is very close to being feasible, but above this function the scaled "distance" $\min(1/n)\sum \psi_j$ to feasibility is at least a positive constant.

In this paper we use local weak convergence method to prove Theorem 1. While our approach is very much similar to the one used in [Ald92], there are several distinctive features of our problem. In particular, we do not use an infinite tree construction and instead consider a sequence of finite depth trees with some corresponding sequence of probability measures. Then we use a Martingale Convergence Theorem for the "projection" step. This simplifies the proofs significantly.

3.2 Maximum weight b-matching

We return to the setting of Subsection 2.2. We have a sparse random graph $G(n,cn)$, where c is a positive constant. The edges of these graph are equipped with random weights $W_{i,j}$ which are selected independently from a common distribution $\mathbb{P}\{W_{i,j} \leq t\} = w^{\text{edge}}(t), 0 \leq t \leq B_w < \infty$, where $[0, B_w]$ is the support of this distribution. Again let $\mathcal{M}_w(n,c,b)$ denote the maximum weight b-matching in $G(n,cn)$, where $b \geq 1$ is an integer.

Theorem 3 *For every $c > 0$ the limit*

$$\lim_{n\to\infty} \frac{\mathcal{M}_w(n,c,b)}{n} \equiv g(c) \tag{9}$$

exists w.h.p.

The probability in the statement of the theorem is both with respect to the randomness of $G(n,cn)$ and with respect to the random weights. This theorem is proven in Section 4. We use linear programming duality and certain linear programming formulation of the maximum weight b-matching problem in order to related it to our main result, Theorem 1.

3.3 Proof plan

The proof of our main result, Theorem 1 is too lengthy to fit this extended abstract, so below we just outline the main proof steps. The proof of Theorem 2 follows quickly from Theorem 1 by taking $c_K^* = \sup\{c : f(c) = 0\}$. Both proofs are in the full version of the paper, which is available upon request.

Let $\mathbb{E}[\cdot]$ denote the expectation operator. The general scheme of the proof is similar to the one of Aldous' [Ald92].

1. We first observe that, as in the case of a random K-SAT problem, in the limit as $n \to \infty$, the (random) number of constraints containing a fixed variable x from the pool $x_1, \dots, x_n$ is distributed as a Poisson random variable with parameter cK, denoted henceforth as $\mathrm{Pois}(cK)$.

2. For every $c > 0$ we introduce
$$\lambda(c) \equiv \liminf_{n\to\infty} \frac{\mathbb{E}[\mathcal{G}LP(n,c)]}{n}. \tag{10}$$
Our goal is to show that in fact convergence $\lim_n \frac{\mathbb{E}[\mathcal{G}LP(n,c)]}{n}$ holds, and therefore we can set $f(c) = \lambda(c)$. The convergence w.h.p. will be a simple consequence of Azuma's inequality. Then, in order to prove Theorem 2, we prove that $c_K^* \equiv \sup\{c : f(c) = 0\}$ satisfies the properties required by the theorem.

3. We consider a subsequence $n_1, n_2, \dots, n_i, \dots$ along which $\frac{\mathbb{E}[\mathcal{G}LP(n_i,c)]}{n_i}$ converges to $\lambda(c)$. Let $X_1, \dots, X_{n_i}$, $\Psi_1, \dots, \Psi_{cn_i} \in [B_x^1, B_x^2]^{n_i} \times [0,\infty)^{cn_i}$ denote a (random) optimal assignment which achieves the optimal value $\mathcal{G}LP(n_i)$. For each n_i we pick a variable x_1 from the pool $x_1, \dots, x_{n_i}$ (the actual index is irrelevant) and consider its d-neighborhood appropriately defined, where d is some fixed constant. We then consider the optimal solution $(X(n_i,d), \Psi(n_i,d))$ restricted to this d-neighborhood. We consider the probability distribution $\mathcal{P}(d,n_i)$ which describes the joint probability distribution for the values of (X_i, Ψ_j, W_j) for X_i, Ψ_j in the d-neighborhood as well as the graph-theoretic structure of this neighborhood.

   We show that for each fixed d, the sequence of probability measures $\mathcal{P}(d,n_i)$ is tight in its corresponding probability space. As a result, there exists subsequence of n_i along which the probability distribution $\mathcal{P}(d,n_i)$ converges to a limiting probability distribution $\mathcal{P}(d)$ for every fixed d. Moreover, we show that the subsequence can be selected in such a way that the resulting probability distributions are consistent. Namely, for every $d < d'$, the marginal distribution of $\mathcal{P}(d')$ in d-neighborhood is exactly $\mathcal{P}(d)$. We will show that, since the sequence n_i was selected to achieve the optimal value $\mathbb{E}[\mathcal{G}LP(n_i,c)] \approx \lambda(c)n$, then
$$\mathbb{E}[w_x X_1 + \frac{w_\Psi}{K} \sum_j \Psi_j] = \lambda(c), \tag{11}$$
   where the expectation is with respect to $\mathcal{P}(d)$ and the summation is over all the constraints C_j containing X_1.

   The sequence of probability distributions $\mathcal{P}(d), d = 1, 2, \dots$ was used by Aldous in [Ald92] to obtain an invariant (with respect to a certain pivot operator) probability distribution on some infinite tree. Our proof does not require the analysis of such a tree, although similar invariant measure can be constructed.

4. We consider a random sequence $\mathbb{E}[X_1|\Im_d], d = 1,2,\dots$, where $X_1 \in [B_x^1, B_x^2]$ is, as above, the value that is assigned to the variable x_1 by an optimal solution, and $\Im_d$ is the filtration corresponding to the sequence of probability measures $\mathcal{P}(d), d = 1,2,\dots$. We prove that the sequence $\mathbb{E}[X_1|\Im_d], d = 1,2,\dots$ is a martingale.

5. This is the "projection" step in which for any $\varepsilon > 0$ and an arbitrary large n we construct a feasible solution to the system of constraints (4) which achieves the expected objective value at most $(\lambda(c)+\varepsilon)n$. Given any large n and an instance of a random linear program (4) with variables $x_1, x_2, \dots, x_n, \psi_1, \dots, \psi_{cn}$ and constraints $C_j, 1 \le j \le cn$, for each variable $x_i, 1 \le i \le n$ we consider its d-neighborhood, where d is a very large constant. We let the value of x_i be $\mathbb{E}[X_i|\Im_d]$ where the expectation is conditioned on the observed d-neighborhood of the variable x_i and this information is incorporated by filtration $\Im_d$. By construction, this value is in $[B_x^1, B_x^2]$. Then we set Ψ_j to the minimal value which satisfies the constraint C_j, for the selected values of x_i, for all $j = 1,2,\dots,cn$. Using a martingale convergence theorem and property (11) we show that for a randomly chosen variable x_i, the corresponding value of $\mathbb{E}[w_x X_i + \frac{w_\psi}{K}\sum_j \Psi_j]$ is smaller than $\lambda(c)+\varepsilon$, when n and d are sufficiently large. We sum the expectation above over all x_i and observe that each constraint belongs in the sum $\sum_j$ of exactly K variables x_i. Then the sum of these expectations is $\mathbb{E}[w_x \sum_{1\le i\le n} X_i + w_\psi \sum_{1\le j\le cn} \Psi_j]$ which is exactly the objective function. We use this to conclude that the expected value of the objective function is at most $(\lambda(c)+\varepsilon)n$.

4 Applications to maximum weight b-matching in sparse random graphs

The main goal of this section is proving Theorem 3. We begin with a linear programming formulation of the maximum weight matching problem. Suppose we have (a non-random) graph with n nodes and m undirected edges represented as pairs (i,j) of nodes. Denote by E the edge set of the graph. The edges are equipped with (non-random) weights $0 \le w_{i,j} \le w_{\max}$. Given $V \subset [n]$, let $\delta(V)$ denote the set of edges with exactly one end point in V. A classical result from the theory of combinatorial optimization ([Sch03], Theorem 32.2) states that the following linear programming problem provides an exact solution (namely, it is a tight relaxation) of the maximum weight b-matching problem:

$$\text{Maximize} \sum_{i,j} w_{i,j} x_{i,j} \tag{12}$$

subject to :

$$\sum_j x_{i,j} \le b, \qquad \forall i = 1,2,\dots,n, \tag{13}$$

$$\sum_{i,j\in V} x_{i,j} + \sum_{(i,j)\in A} x_{i,j} \le \frac{b|V|+|A|-1}{2}, \qquad \forall\, V \subset [n], A \subset \delta(V) \text{ such that } b|V|+|A| \text{ is odd}, \tag{14}$$

$$0 \le x_{i,j} \le 1. \tag{15}$$

Specifically, the optimal basic solution of this linear programming problem is always integral, it corresponds to a b-matching, and and the optimal value is equal to maximum weight b-matching. We denote the optimal value of the linear program above by $\mathcal{LPM}(G)$.

Our plan for proving Theorem 3 is as follows. We first show that when the graph has very few small cycles (and this will turn out to be the case for $G(n,cn)$), the optimal value $\mathcal{L}PM^0(G)$ of the modified linear program, obtained from (12)-(15) by dropping the constraints (14), is very close to $\mathcal{L}PM(G)$. In the context of random graphs $G(n,cn)$, this will imply that the difference $\Big|\mathcal{L}PM(G(n,cn)) - \mathcal{L}PM^0(G(n,cn))\Big| = o(n)$, w.h.p. We then take the dual of the modified linear program (12), (13), (15) and show that it has the form (4). Applying Theorem 1 we will obtain that the limit $\lim_n \mathcal{L}PM^0(G(n,cn)/n$ exists w.h.p. This will imply the existence of the limit $\lim_n \mathcal{L}PM(G(n,cn)/n$ w.h.p. Finally, since $\mathcal{L}PM(G(n,cn)$ is the maximum weight b-matching in $G(n,cn)$, that is $\mathcal{M}_w(n,c,b)$, and Theorem 3 will follow.

Naturally, $\mathcal{L}PM^0(G) \geq \mathcal{L}PM(G)$.

Proposition 1 *Given a weighted graph G and $d \geq 3$ let $L(d)$ denote the set of cycles of length $< d$ in G and let $M(d)$ denote the total number of edges in $L(d)$. Then*

$$\frac{d-1}{d}(\mathcal{L}PM^0(G) - M(d)w_{\max}) \leq \mathcal{L}PM(G) \leq \mathcal{L}PM^0(G). \tag{16}$$

Proof : We already observed that the right-hand side of (16) holds. We concentrate on the left-hand side bound.

Let $x^0 = (x^0)_{i,j}$ denote an optimal solution of the linear program (12), (13), (15) with its optimal value $\mathcal{L}PM^0(G)$. In the graph G we delete all the $M(d)$ edges which contribute to $L(d)$ together with the corresponding values of $x^0_{i,j}$. Consider the resulting solution $x^1 = (x^1_{i,j})$ to the linear program corresponding to the reduced weighted graph G^1, which now does not contain any cycles of length less than d. The objective value of the linear program (12), (13), (15) corresponding to the solution x^1 is at least $\mathcal{L}PM^0(G) - M(d)w_{\max}$, since, by constraint (15), $x^0_{i,j} \leq 1$ for all the edges (i,j). We further modify the solution x^1 to x^2 by letting $x^2_{i,j} = (1-1/d)x^1_{i,j}$ for every edge (i,j) in the graph G^1. The objective value of the linear program (12), (13), (15) corresponding to x^2 is then at least $\frac{d-1}{d}(\mathcal{L}PM^0(G) - M(d)w_{\max})$. We claim that in fact x^2 is a feasible solution to the linear program (12)-(15), implying (16) and completing the proof of the proposition.

Clearly, constraints (13) and (15) still hold. We concentrate on (13). Consider any set $V \subset [n]$ and $A \subset \delta(V)$ such that $b|V| + |A|$ is odd. Assume first $|V| + |A| < d$. Let $\hat{V}$ denote the union of V and the end points of edges in A. Then $|\hat{V}| < d$. Since G^1 does not contain any cycles of size $< d$, then $\hat{V}$ does not contain any cycles at all, and therefore is a forest. In particular it is a bipartite graph. Let $\hat{x}^1$ denote the sub-vector of the vector x^1 corresponding to edges with both ends $\hat{V}$. Since (13) holds for x^1, then it also holds for the vector $\hat{x}^1$ for all nodes in $\hat{V}$. A classical result from a combinatorial optimization theory [Sch03] states that for every bipartite graph, the polytope corresponding to the degree constraints (13) has only integral extreme points, which are b-matchings (in the reduced graph $\hat{V}$). This follows from the fact that this polytop when described in matrix form $Bx \leq b$ corresponds to the case when B is totally unimodular. We refer the reader to [Sch03] for the details. But since any integral solution $\hat{x}^1$ corresponding to the b-matching must satisfy the constraints (14) by Theorem 32.2 in [Sch03], then these constraints are automatically satisfied by x^1. Moreover then their are satisfied by x^2. We proved that (14) holds whenever $|V| + |A| < d$.

Suppose now $|V| + |A| \geq d$. For the solution x^1, let us sum the constraints (13) corresponding to all the nodes in V and sum the right-hand side constraints (15) corresponding to all the edges in A. Each value $x_{i,j}$ for $i,j \in V$ is counted twice, once for node i and once for node j. Each value $x_{i,j}$ for $(i,j) \in A$

is also counted twice, once for constraint (13) for the node i and once for constraint (15) for the edge (i,j). Then we obtain $2(\sum_{i,j\in V} x^1_{i,j} + \sum_{(i,j)\in A} x^1_{i,j}) \leq b|V| + |A|$. This implies $\sum_{i,j\in V} x^2_{i,j} + \sum_{(i,j)\in A} x^2_{i,j} \leq (1-1/d)\frac{b|V|+|A|}{2} \leq \frac{b|V|+|A|-1}{2}$, since by assumption, $b|V|+|A| \geq |V|+|A| \geq d$. Again we showed that the constraint (13) holds for the solution x^2. □

We return to our main setting – sparse random graph $G(n,cn)$, with edges equipped with randomly generated weights $W_{i,j}$, drawn according to a distribution function $w^{\text{edge}}(t), t \geq 0$ with support $[0,B_w]$. Let $E = E(G(n,cn))$ denote the edge set of this graph. We denote the value of $\mathcal{LPM}^0(G(n,cn))$ by $\mathcal{LPM}^0(n,c)$ for simplicity. That is, $\mathcal{LPM}^0(n,c)$ the optimal (random) value of the linear program (12), (13), (15) on the graph $G(n,cn)$.

Proposition 2 *W.h.p. as $n \to \infty$*

$$\mathcal{LPM}^0(n,c) - o(n) \leq \mathcal{M}_w(n,c,b) \leq \mathcal{LPM}^0(n,c). \tag{17}$$

Proof : Let $d(n)$ be a very slowly growing function of n. It is well known that in $G(n,cn)$ w.h.p. the total number of edges which belong to at least one cycle with size $< d(n)$ is $o(n)$ (far more accurate bounds can be obtained [JLR00]). Thus $M(d(n)) = o(n)$ w.h.p. Note also that from (15), we have $\mathcal{LPM}^0(n,c) \leq w_{\max} bn = O(n)$. Applying Proposition 1 with $d = d(n)$, we obtain the result. □

Our final goal is proving the convergence w.h.p. of $\mathcal{LPM}^0(n,c)/n$. We use linear programming duality for this purpose. Consider the dual of the linear program (12), (13), (15), generated on the weighted graph $G(n,cn)$. It involves variables $y_1,\ldots,y_n$ and has the following form.

$$\text{Minimize } b \sum_{1\leq i\leq n} y_i + \sum_{(i,j)\in E} \psi_{i,j} \tag{18}$$

subject to :

$$y_i + y_j + \psi_{i,j} \geq W_{i,j}, \qquad \forall\ (i,j) \in E \tag{19}$$

$$y_i, \psi_{i,j} \geq 0. \tag{20}$$

The objective value of this linear program is also $\mathcal{LPM}^0(n,c)$, thanks to the strong duality of linear programming. The linear program above is almost of the form (4) that we need in order to apply Theorem 1. Let us rewrite the linear program above in the following equivalent form

$$\text{Minimize } b \sum_{1\leq i\leq n} y_i + \sum_{(i,j)\in E} \psi_{i,j} \tag{21}$$

subject to :

$$(-1)y_i + (-1)y_j \leq -W_{i,j} + \psi_{i,j}, \qquad \forall\ (i,j) \in E \tag{22}$$

$$y_i, \psi_{i,j} \geq 0, \tag{23}$$

Let the set of constraints $\mathcal{C}$ of the linear program (4) contain only one element $(-1)y_1 + (-1)y_2 \leq 0$. We set $B^1_x = 0, B^2_x = B_w$, where, as we recall, $[0,B_w]$ is the support of the distribution of $W_{i,j}$. We also set $w_x = b$ and $w_\psi = 1$. Our linear program has now the form (4) except for we need to consider in addition the constraints $y_i \leq B_w \geq W_{i,j}$. We claim that in fact these constraints are redundant. In fact any value of y_i which exceeds B_w can be decreased to B_w, resulting in smaller value of the objective function and still honoring all the constraints. Thus, we may replace $0 \le y_i \le B_w$ simply by $y_i \ge 0$. We conclude that the linear program (21), (22), (23) has form (4). Applying Theorem 1, there exists a function $g(c) \ge 0$ such that w.h.p. $\mathcal{LPM}^0(n,c)/n \to g(c)$ as $n \to \infty$. Finally, applying Proposition 2, we obtain (9). This concludes the proof of Theorem 3. □

5 Discussion

The results of the present paper lead to several interesting open questions.

An interesting group of questions relates to the behavior of the function $f(c) = \lim_n \mathbb{E}[\mathcal{GLP}(n)]/n$, which, by results of this paper is equal to zero for $c < c^*$, in the specific context of linear program (2). What can be said about $f(c)$ near c^*_K, when $c^*_K < \infty$? It is not hard to show that $f(c)$ is continuous and non-decreasing function of c. Is it differentiable in c^*_K? Is it convex, concave or neither? Similar questions arise in connection with percolation probability $\theta(p)$ which is a function of a bond or site probability p. It is known that for every dimension d there exists a critical value p^*_d, for which the probability $\theta(p)$ of existence of an infinite component containing the origin in d-dimensional bond/site percolation model, is equal to zero for $p < p^*_d$ and is positive for $p > p^*_d$, (whether $\theta(p) = 0$ for $p = p^*_d$ is a major outstanding problem in percolation theory), see [Gri99]. The behavior of $\theta(p)$ near p^*_d is a well-known open problem. We also refer the reader to [APb] for the related discussion on scaling limits and universality of random combinatorial problems.

Computing the values of c^*_K is another question which seems so far beyond the techniques applied in this paper. The cases where local weak convergence methods lead to computation of limiting values seem to be related to special structures of the corresponding problems.

A separate course of investigation is to formulate and study randomly generated integer programming problems as a unifying framework for studying random k-SAT, coloring, maximum k-cuts , maximum independent sets, and other problems. For example, under which conditions on the set of prototype constraints does the feasibility problem experience a sharp transition?

Finally, it should be clear what goes wrong when one tries to use local weak convergence approach for random K-SAT problem, for example along the lines of Theorem 1. Our approach is built on using the values like $\mathbb{E}[X_k\,|\,\cdot]$ to construct a feasible solution, but these expectations are not necessarily integers. Digging somewhat deeper into the issue, it seems that local weak convergence method in general is not very hopeful for resolving Conjecture 1, since it looks only into constant size neighborhoods of nodes. To elaborate somewhat this point consider maximal independent set problem in r-regular random graphs, discussed in [AS03]. For almost any node in such a graph, its constant size neighborhood is a r-regular tree, and, as such, the neighborhoods are indistinguishable. In such circumstances it seems hard to try to concoct a solution which is based only on neighborhoods of nodes. Some long-range structural properties of these graphs like structure of cycles have to be considered. We refer the reader to [AS03] for a further discussion of this issue.

Acknowledgements

The author gratefully acknowledges interesting and fruitful discussions with David Aldous, Michael Steele and Alan Frieze. The author is also grateful to Jonathan Lee and Maxim Sviridenko for references to the results in polyhedral combinatorics.

References

[Ald] D. Aldous. My favorite 6 open problems in mathematical probability. *http://stat-www.berkeley.edu/users/aldous/Problems.ps*.

[Ald92] D. Aldous. Asymptotics in the random assignment problem. *Probab.Th.Rel.Fields*, (93):507–534, 1992.

[Ald01] D. Aldous. The $\zeta(2)$ limit in the random assignment problem. *Random Structures and Algorithms*, (18):381–418, 2001.

[AM02] D. Achlioptas and C. Moore. The asymptotic order of the random k-SAT threshold. *Proc. 43d IEEE Symposium on Foundations of Computer Science*, 2002.

[APa] D. Achlioptas and Y. Peres. The threshold for random k-SAT is $2^k(\ln 2 + o(1))$. *Submitted*.

[APb] D. Aldous and A. Percus. Scaling and universality in continuous length combinatorial optimization. *Preprint*.

[APF98] J. Aronson, B. Pittel, and A. Frieze. Maximum matchings in sparse random graphs: Karp-Sipser revisited. *Random Structures and Algorithms*, 12:11–178, 1998.

[AS03] D. Aldous and J. M. Steele. The objective method: Probabilistic combinatorial optimization and local weak convergence. In *Discrete Combinatorial Probability*. H. Kesten Ed., Springer-Verlag, 2003.

[CR92] V. Chvatal and B. Reed. Mick gets some (the odds are on his side). *Proc. 33d Symposium on Foundations of Computer Science*, 1992.

[DBM00] O. Dubois, Y. Boufkhad, and J. Mandler. Typical random 3-SAT formulae and the satisfiability threshold. *Proc. 11th ACM-SIAM Symposium on Discrete Algorithms*, 2000.

[dlV92] W. Fernandez de la Vega. On random 2-SAT. *Unpublished manuscript*, 1992.

[Fri99] E. Friedghut. Sharp thresholds of graph proprties, and the *k*-SAT problem. *J. Amer. Math. Soc.*, 4:1017–1054, 1999.

[FW02] A. Frieze and N. Wormald. Random k-SAT: A tight threshold for moderately growing k. In *Proceedings of the Fifth International Symposium on Theory and Applications of Satisfiability Testing*, pages 1–6, 2002.

[Goe92] A. Goerdt. A threshold for unsatisfiability. *Mathematical Foundations of Computer Science, 17th International Symposium.* I. M. Havel and V. Koubek, Eds. Lecture Notes on Computer Science, No. 629, Springer Verlag, pages 264–274, 1992.

[Goe96] A. Goerdt. A threshold for unsatisfiability. *J. Computer and System Sciences*, 53:469–486, 1996.

[Gri99] G. Grimmett. *Percolation*. Springer, 1999.

[JLR00] S. Janson, T. Luczak, and A. Rucinski. *Random Graphs*. John Wiley and Sons, Inc., 2000.

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

[KS81] R. Karp and M. Sipser. Maximum matchings in sparse random graphs. In *22nd Annual Symposium on Foundations of Computer Science*, pages 364–375, 1981.

[LPSA94] R. Langlands, P. Pouliot, and Y. Saint-Aubin. Conformal invariance for two-dimensional percolation. *Bull. Amer. Math Soc. (N.S.)*, 30(1):1–61, 1994.

[MP87] M. Mézard and G. Parisi. On the solution of the random link matching problem. *J. Physique*, 48:1451–1459, 1987.

[PS98] C. Papadimitriou and K. Steiglitz. *Combinatorial optimization: algorithms and complexity*. Dover, 1998.

[Sch01] O. Schramm. A percolation formula. *Electronic Communications in Probability*, 6:115–120, 2001.

[Sch03] A. Schrijver. *Combinatorial Optimization. Volume A*. Springer, 2003.

[Ste02] J. M. Steele. Minimal spanning trees for graphs with random edge lenghts. *Mathematics and Computer Science II. Algorithms, Trees, Combinatorics and Probabilities. Editors: B. Chauvin, P. Flajolet, D. Gardy, A. Mokkadem.*, pages 223–246, 2002.

[SW01] S. Smirnov and W. Werner. Critical exponents for two-dimensional percolation. *Math. Research Letters*, 8(5-6):729–744, 2001.