Some results for k-SAT on trees

Sumedha
National Institute of Science Education and Research, Institute of Physics Campus,
Bhubaneswar, Odisha- 751 005

Supriya Krishnamurthy
Department of Physics, Stockholm University, SE- 106 91, Stockholm, Sweden

Abstract. Phase transitions in random k-SAT problems are connected to their computational complexity. While polynomial time algorithms are known to solve the problem for \( k = 2 \), for \( k \geq 3 \) the problem is known to be NP-complete. Recently we have studied random k-SAT and many of its variants on regular infinite trees. We find that the solvability threshold for \( k = 2 \) matches the exact value of the threshold on regular random graphs. For higher \( k \), the values are very close to those predicted using other techniques like cavity method.

1. Introduction

K-satisfiability problems are an important example of constraint satisfaction problems, which have been fundamental in understanding computational complexity. The aim is to find a satisfying assignment for a randomly generated logical expression of \( M \) clauses(logical constraints). Each clause is an OR of \( k \) Boolean variables, which are chosen randomly from a set of \( N \) Boolean variables. For \( k \geq 3 \) variables, this problem is NP-complete [1], i.e., potential solutions can be verified easily for correctness, but finding a solution can take exponential time in the worst case. In addition, being NP-complete, should a polynomial-time algorithm be found for solving SAT, it is also possible to adapt it to solve any problem in NP in polynomial-time.

As the constraint density (\( \alpha = M/N \)) increases, the number of satisfying assignments decreases. In the limit of \( M \rightarrow \infty \) and \( N \rightarrow \infty \), the system is known to have a sharp threshold in constraint density \( \alpha_c \) below which the probability of finding satisfiable assignments approaches 1 and above which it vanishes [2, 3].

The problem is originally defined on a random graph, but because of the presence of loops, this is hard to solve exactly for arbitrary \( k \). Hence the location of a sharp threshold \( \alpha_c \) has been known rigorously only for \( k = 2 \) [4, 5]. For higher \( k \) only upper and lower bounds on this threshold are proven[6]. However using non-rigorous but powerful methods from statistical physics, namely the replica and cavity methods, estimates for the threshold are obtained which seem to be very close to the values obtained numerically [7, 8, 9].

The replica and cavity methods also predict that the solvability threshold is only one of many thresholds that exist in the problem, as the number of constraints is increased. Before the solvability transition occurs, it is conjectured that the set of solutions first breaks up into a large number of well separated clusters at the clustering transition \( \alpha_{d} \) [7, 10]. As the number of constraints further increases, it is argued that there is first a condensation transition [11] in which the number of clusters changes from being exponentially numerous to sub-exponential...
and a freezing transition beyond which some variables take the same value in all the solutions of a given cluster [12, 13]. Recently, both the existence of clusters [14] as well as the 1-RSB prediction for the satisfiability threshold has been established for random $k$-SAT [15] for large $k$. All the approaches mentioned above for $k \geq 3$ look at the properties of solution clusters.

Recently, we have studied the random $k$-SAT problem on regular $d$-ary rooted trees [16, 17, 18]. We study the probability of having a satisfiable assignment as a function of tree depth. In this paper we will review some of these results. We find that the probability of having a satisfiable assignment for an infinite tree gives a value for the solvability threshold that matches the value of the clustering (or dynamic) transition calculated using cavity method for random graphs. On a Bethe lattice we obtain a solvability threshold that matches the solvability threshold on the random graphs. We also present some preliminary results for a $(2+p)$-SAT problem. The plan of the paper is as follows: In Sec. 2, we define the model on a $d$-ary rooted tree. In Sec. 3 we present the solution of the random $k$-satisfiability problem and in Sec. 4, we compare our method with the tree reconstruction problem. In Sec. 5 we briefly discuss the $(2+p)$-SAT before concluding in Section 6.

2. The model

We define the $k$-SAT problem on a rooted tree as follows: Consider a regular $d$-ary tree $T$ in which every vertex has exactly $d$ descendents. The root of the tree $x_0$ has degree $d$ and its $d$ edges are connected to function nodes $\{c_1, c_2, ..., c_d\}$. Each function node has degree $k$, and each of its $k-1$ descendents $\{x_i = x_1, x_2, ..., x_{k-1}\}$ is the root of an independent tree (see Fig. 1). Hence the root has a degree $d$ while all the other vertices on the tree (except the leaves which have a degree = 1) have a degree $d+1$. Each vertex can take only two values: $-1$ or $1$. Each function node is associated independently with a clause $\phi(x_0, x_1, ..., x_{k-1}) = \ell_0 \lor \ell_1 \lor ... \lor \ell_{k-1}$. Here $\ell_i$ is one of the two literals $x_i$ or $\overline{x_i}$, determined by whether $x_i$ is joined to the function node by a dashed or a solid line (see Fig. 1). An assignment $\sigma$ of all the variables on the tree is a solution iff $\phi = 1$ for all the clauses on the tree. One configuration of dashed and solid lines on the tree defines a realization $R$.

We study the probability that a realization has no solution on this tree for a fixed boundary. This can happen if there is even a single variable on the graph, which, whether it takes the value $-1$ or $1$, causes at least one clause to be unsatisfied. Such a variable then is a variable that can take 0 values by our definition, and a realization that is not solvable has at least one variable of this type.

On the tree graph, we can define the probabilities of a variable taking 0, 1 or 2 values on the corresponding subtree. We define $P_i(0)$ as the conditional probability for a variable $x_i$ to take 0 values by our definition, and a realization that is not solvable has at least one variable of this type.
cause a contradiction, in the subtree of which it is the root, given that all the other variables in the subtree can take at least 1 value. Note that because of the tree structure and because of the definition of the specific quantity we are looking at, all variables $x_i$ at depth $n$ will have the same probability $P_n(0)$. The conditional probability that a variable at depth $n$, can take only one of the two values $-1$ or $1$ is defined to be $P_n(1)$ (the boundary nodes have $P_0(1) = 1$, for example). Similarly the conditional probability that a variable at depth $n$ can take both values is $P_n(2) = 1 - P_n(0) - P_n(1)$. The probability of a realization having a solution (or the fraction of realizations that have solutions) is then exactly equal to the product $\Pi_i(1 - P_i(0))$, where the product is over all the variables in the graph. The tree structure also gives us a way to calculate the $P_n$’s via recursions. For the problems we look at, we are interested in the recursions for these quantities deep within the tree, so that we can get rid of boundary effects.

3. Random $k$-SAT on a regular tree

Let us first calculate $P_{n+1}(0)$ for variable $x_0$ (assuming it is at depth $n+1$), given these quantities for its descendents. Assume variable $x_0$ has a degree $d$ and assume it is not negated on $d_1$ of these clauses. Variable $x_0$ will not be able to take the value $-1$ in the case when at least one of the $d_1$ clauses is not satisfied by the $k-1$ variables at the other end. In this case there will be at least one unsatisfied clause if $x_0$ takes the value $-1$. Similarly, if at least one of the $d-d_1$ clauses which are satisfied by $x_0$, are also not satisfied by the $k-1$ variables at the other end, then $x_0$ cannot take the value 1 either.

It is easy to see that averaging over all realizations at depth $n+1$ implies averaging over all values of $d_1$, as well as averaging over all realizations at depth $n$. It is important to note however that the realizations at depth $n+1$ are only built up from those realizations at depth $n$ that do have solutions. We define $Q_n$ as the conditional probability that a depth $n$ variable does not satisfy the clause above (to depth $n+1$), given that it has to be able to take at least one value (which satisfies the sub tree of which it is the root). In terms of $P_n(0)$ and $P_n(1)$, $Q_n = \frac{P_n(1)}{2(1-P_n(0))}$. The recursion for $P_{n+1}(0)$ is then:

$$P_{n+1}(0) = \frac{1}{2^d} \sum_{d_1=1}^{d_1=d-1} \binom{d}{d_1} [1 - (1 - Q_n^{-1})^{d_1}](1 - (1 - Q_n^{-1})^{d-d_1})$$

$$= 1 + (1 - Q_n^{-1})^d - 2(1 - 0.5Q_n^{-1})^d$$

(1)

Similarly one can work out the recursion for $P_{n+1}(1)$. This is:

$$P_{n+1}(1) = 2(1 - 0.5Q_n^{-1})^d - 2(1 - Q_n^{-1})^d$$

(2)

Eqs. 1 and 2 result in the following relation between $Q_{n+1}$ and $Q_n$:

$$Q_{n+1} = \frac{[1 - 0.5Q_n^{-1}]^d - [1 - Q_n^{-1}]^d}{2[1 - 0.5Q_n^{-1}]^d - [1 - Q_n^{-1}]^d}$$

(3)

From this equation the threshold at which the fraction of realizations goes to zero exponentially with the depth of the tree may be extracted. This is the solvability threshold for these models on an infinite tree. A fixed point analysis of Eq. 3 predicts a continuous transition for $k = 2$ and a first order transition for $k > 2$ (see Fig. 2 and 3). The value of $d$ at which the system undergoes a continuous transition for $k = 2$ can be extracted by expanding to order $Q^2$ in Eq. 3 at the fixed point. This gives, for $k = 2$:

$$Q_c = \frac{8(d/2 - 1)}{3(d-1)d}$$

(4)
**Figure 2.** Fixed points for $d < 2$ and $d > 2$ for 2-SAT

**Figure 3.** Fixed points for $d < 11.5$ and $d > 11.5$ for 3SAT
Table 1. We compare the values of \(d_c\) and \(d_s\) obtained from our tree calculations with the clustering (\(\alpha_d\)) and SAT-UNSAT(\(\alpha_s\)) threshold values known from the cavity calculations [8]. \(\alpha = (d + 1)/k\) for the tree. The difference in numbers for smaller values of \(k\) is due to the difference in the degree distributions. We study the problem on regular graphs, while [8] looks at the problem on poisson graphs.

| \(k\) | \((d_c + 1)/k\) | \((d_s + 1)/k\) | \(\alpha_d\) (from [8]) | \(\alpha_s\) (from [8]) |
|------|----------------|----------------|---------------------|---------------------|
| 2    | 1.5            | 1.5            | 1                   | 1                   |
| 3    | 4.166          | 4.55           | 3.927               | 4.267               |
| 4    | 8.4            | 10.15          | 8.297               | 9.93                |
| 5    | 16.2           | 21.26          | 16.12               | 21.12               |
| 6    | 30.5           | 43.41          | 30.5                | 43.37               |
| 7    | 57.28          | 87.84          | 57.22               | 87.79               |
| 8    | 107.13         | 176.57         | 107.24              | not available       |

which implies \(d_c = 2\). The value of \(\alpha\) corresponding to a give value of \(d\) should be \((d + 1)/k\) [19, 20]. Hence \(\alpha_c\) corresponding to \(d_c\) is 3/2, which matches the exact threshold for random 2-SAT on regular random graphs (these are random graphs in which each variable is connected with the same number of clauses)[17, 5].

For \(k > 3\) no exact results for random \(k\)-SAT on random graphs is known, but a lot of progress has been made using replica and cavity methods. Cavity calculations conjecture the breaking of solution space into many disconnected clusters before the SAT-UNSAT transition. This is the point where belief propagation (BP) iterations lead to contradiction. From the replica point of view this is also the point where the replica symmetry gets broken, and for systems with discontinuous transitions one has to look at 1-RSB solutions. For \(k = 3\), on regular random graphs, it has been found that BP iterations stop converging at \(d = 12\), which gives the value of clustering transition on a regular graph to be \(4 \leq \alpha_d < 4.33\) [21]. On the tree we find for \(k = 3\), Eq. 3 starts having non zero values of \(Q\) at around \(d_c = 11.5\), which gives \(\alpha_c = 4.166\) on an infinite tree. Given the fact that BP iterations were done for integer values of the degree on a regular random graph, the values match well within error bars.

In Table 1 we compare the solvability threshold on an infinite tree with the value of \(\alpha_d\) obtained via the cavity method. The later have been obtained for random graphs with poisson degree distribution. This also accounts for the difference in the value. As expected the effect of degree distribution goes down with increasing \(k\). We find for \(k \geq 6\), that the difference is less than 1% in the two values.

Unlike \(k = 2\), for \(k \geq 3\), the transition to the UNSAT phase does not happen at the point of first appearance of a new fixed point of Eq. 3. We find that if we study the problem on a graph which is only locally tree-like (Bethe lattice), with all nodes having the same degree \(d + 1\), the SAT-UNSAT threshold is different for the infinite tree and for the Bethe lattice, for \(k \geq 3\). We calculate the fraction of satisfied realizations per node on such a lattice [18, 22] and find that the calculation yields a value of threshold which is larger than \(\alpha_d\) for \(k > 2\) and is very close to the known 1-RSB static transition threshold (see Table 1).

4. Connection with tree reconstruction problem

The reconstruction problem, as originally defined, is a broadcast model on a tree, such that information is sent from the root to the leaves, across edges which act as noisy channels. The problem then is to understand whether we can recover information about the root from a knowledge of the configuration of the leaves. It has been shown that the recursions developed in the reconstruction context are the same as obtained by other means (such as the replica or
cavity methods) for the dynamical glass transition on a random graph [23] or the clustering transition for k-SAT.

In terms of reconstruction, these fixed point recursions are developed for the unconditional probability distribution at the root of the tree to have a certain 'bias' ; namely, the fraction of boundary conditions (out of all boundary conditions that have a non-zero solution set for a fixed instance), weighted by the total number of solutions each of these boundary conditions possesses, that leads to the root taking the value \( -1 \) a certain number of times and the value \( 1 \) a certain number of times.

Our approach apriori looks very different from the reconstruction problem mainly because the quantities \( P_n(1) \) and \( P_n(2) \) are the fraction of realizations that have a non zero solution set (for any fixed boundary condition) and not the fraction of boundaries that have a non-zero solution set for any fixed instance. However, we show in [17] that if we define the probability space over boundary conditions instead of realizations, then we can derive the recursion for fraction of boundary conditions that fix the value unambiguously at the root at level \( n \), given this quantity at level \( n - 1 \). These equations turn out to be the same as Eqs. 1 and 2 in Sec. 3. The fact that for k-SAT, cavity recursions for the clustering transition matches the tree reconstruction recursions [23], gives a connection between the solvability transition on infinite trees as calculated in Sec. 3 and clustering transition on random graphs.

5. Variants of Random k-SAT

The method outlined in Sec. 3 can straightforwardly be used to study many variants of random k-SAT. We have used it to successfully predict transitions for biased k-SAT, balanced k-SAT [17], k-NAE SAT and \((2 + p)\)-SAT [22]. As a simple illustration of the wide applicability of the method we briefly mention the results for \((2 + p)\)-SAT here.

For \( k = 2 \) the random k-SAT is known to have a unique continuous transition, while for \( k = 3 \) it exhibits multiple transitions. \((2 + p)\)-SAT interpolates between these two systems as every clause has a probability \( p \) of being a 3-clause and \((1 - p)\) of being a 2-clause. The problem is NP-complete for any \( p > 0 \) as it contains a subformula of 3-clauses. However using the replica method and numerical simulations [24, 25], it was found that random \((2 + p)\)-SAT with poisson degree distribution continues to have a continuous transition for \( p < 0.41 \). Also the computational cost of proving a formula SAT or UNSAT stays linear in this regime for all values of \( \alpha \).

We studied the problem on a rooted tree. As before we define \( Q_n \) as the conditional probability that a depth \( n \) variable does not satisfy the clause above (to depth \( n + 1 \)). The recursions for \( P_{n+1}(0) \) and \( P_{n+1}(1) \) come out to be

\[
P_{n+1}(0) = \frac{1}{2^d} \sum_{d_1=1}^{d-1} \left[ 1 - \{p(1 - Q_n^2) + (1 - p)(1 - Q_n)\}^{d_1} \right] \left[ 1 - \{p(1 - Q_n^2) + (1 - p)(1 - Q_n)\}^{d-d_1} \right]
\]

\[
= 1 + \left[ p(1 - Q_n^2) + (1 - p)(1 - Q_n) \right]^d - 2 \left[ \frac{1 + p(1 - Q_n^2) + (1 - p)(1 - Q_n)}{2} \right]^d
\]  

(5)

Similarly for

\[
P_{n+1}(1) = 2 \left[ \frac{1 + p(1 - Q_n^2) + (1 - p)(1 - Q_n)}{2} \right]^d - 2 \left[ p(1 - Q_n^2) + (1 - p)(1 - Q_n) \right]^d
\]  

(6)

\[
Q_{n+1} = \frac{\left[ 1 + p(1 - Q_n^2) + (1 - p)(1 - Q_n) \right]^d - \left[ 2p(1 - Q_n^2) + 2(1 - p)(1 - Q_n) \right]^d}{2 \left[ 1 + p(1 - Q_n^2) + (1 - p)(1 - Q_n) \right]^d - \left[ 2p(1 - Q_n^2) + 2(1 - p)(1 - Q_n) \right]^d}
\]  

(7)
6. Conclusions
In conclusion, we have looked at the $k$-satisfiability class of constraint satisfaction problems on a regular tree as well as a Bethe lattice. Our approach directly looks at all realizations giving them weights 0 and 1 if they are unsuccessful or successful respectively. We do not weight them by the number of solutions. Interestingly, even though we do not look at the solution space, the solvability threshold on the infinite tree seems similar to the clustering (or dynamic) transition in solution space obtained via BP iterations.

For $k > 2$ almost for all variants of the satisfiability problem, the clustering transition is known to be well below the solvability transition of random graphs. Solvability transition for random graphs via the cavity approach is estimated by looking at a reduced entropy called complexity, which is defined as the log of the number of clusters in the solution space for a typical realization. Complexity is conjectured to be equivalent to counting different backbones of frozen variables, evaluated using populations of typical realizations [26]. We find [18, 22] that the value of the solvability transition obtained via complexity matches very well with the value of solvability threshold on Bethe lattices. Our method does not make any assumption about the structure of solution space and assigns the same weight to all realizations with solutions.

Inspite of these differences in the two approaches, the expression and threshold obtained by us matches the corresponding expressions and thresholds on regular random graphs, obtained using the cavity method. This intriguing connection need to be explored further to get a better understanding of the problem. Our method can easily be extended to $k$-SAT problems with non-uniform degree distributions, making it a useful tool even to study the real-world SAT applications.

References
[1] Cook S, 1971 Proc. 3rd Annual ACM Symp. on Theory of Computing (Shaker Heights, Oh) (New York: ACM press) p151
[2] Mitchell D, Selman B and Levesque H 1992, Proc. 10th Nat. Conf. Artif. Intel., 459
[3] Kirkpatrick S and Selman B 1994 Science 264 1297
[4] Chvatal V and Reed B, 1992 33rd FOCS, 620; Goerdt A, 1996 J. Comput. System. Sci. 53 469
[5] Cooper C, Frieze A and Sorkin G B, 2002 Proceedings of the 13th annual ACM-SIAM symposium on Discrete Algorithms 316-320; Fernandez de la vega W, 2001 Theoret. Comput. Sci. 265 131
[6] Achlioptas D 2001 Theoret. Compt. Sci. 265 159
[7] Mézard M, Parisi G and Zecchina R 2002 Science 297 812
[8] Mertens S, Mézard M and Zecchina R 2006 Random Structures and Algorithms 28 340
[9] Monasson R and Zecchina R 1996 Phys. Rev. Lett. 76 3881
[10] Biroli G, Monasson R and Weigt M 2000 Eur. Phys. J. B 14 551
[11] Krzakala F, Montanari A, Ricci-Tersenghi F, Semerjian G and Zdeborová L 2007 Proc. Natl. Acad. Sci. U.S.A. 104 10318
[12] Semerjian G 2008 J. Stat. Phys 130 251
[13] Zdeborová L and Krzakala F 2007 Phys. Rev. E 76 031131
[14] Achlioptas D and Coja-Oghlan A 2008 Proc. 49th FOCS 793
[15] Ding J, Sly A and Sun N 2014 arXiv:1411.0650
[16] Krishnamurthy S and Sumedha 2012 J. Stat. Mech P05009.
[17] Sumedha, Krishnamurthy S and Sahoo S 2013 Phys. Rev. E 87 042130.
[18] Krishnamurthy S and Sumedha 2014 arXiv:1412.2460
[19] Baxter R J 1982 Exactly solvable models in statistical mechanics (London: Academic Press)
[20] Gujrati P D 1995 Phys. Rev. Lett. 74 809
[21] Castellana M, Zdeborová L 2011 J. Stat. Mech. P03023
[22] Krishnamurthy S and Sumedha, in preparation
[23] Mézard M and Montanari A 2006 J. Stat. Phys. 124 1317
[24] Monasson R, Zecchina R, Kirkpatrick S, Selman B and Troyansky L 1999 Random Structure and Algorithms 15 414
[25] Monasson R and Zecchina R 1988 J. Phys. A 31 9209
[26] Parisi G 2004 Lecture Notes in Computer Science 2919 Springer 203