

Ramsey Arrowing with k -SAT Solvers

Christopher A. Wood

Advisor: Professor Stanisław Radziszowski

May 8, 2013

1 Introduction

In his 1972 seminal paper entitled, “Reducibility Among Combinatorial Problems,” Richard Karp introduced a list of 21 NP-complete problems, including Boolean satisfiability, the maximum cut of a graph, and 0-1 integer programming [1]. The complexity of these problems was proven by deriving a polynomial-time reduction from CIRCUIT-SAT = $\{\langle C \rangle : C \text{ is a satisfiable Boolean combinational circuit } \}$, the first problem shown to be NP-complete by Cook in 1971 [2], starting the rush of complexity theory research.

The problem 3-SAT, or more formally, 3-CNFSAT, is a special case of satisfiability. It is a decision problem in which takes as input a 3-CNF Boolean formula and returns YES if the formula is satisfiable, and NO otherwise [3]. A 3-CNF formula, more formally known as a Boolean formula in 3-conjunctive normal form, is expressed as the Boolean AND of arbitrarily many clauses, where each clause is the Boolean OR of three literals, which is a Boolean variable or its negation. Such a Boolean formula is said to be satisfiable if and only if there exists an assignment of truth values to the variables such that substituting them into the literals of the formula will cause it to evaluate to true (or 1). Expressed as a formal language, we have that $3\text{-SAT} = \{\langle \phi \rangle : \phi \text{ is satisfiable } \}$.

In 2002 Hans van Maaren of and John Franco initiated the public SAT competition in search of optimal performing SAT solvers judged by a variety of criteria and specializations, including their ability to demonstrate satisfiability and exhaustively prove unsatisfiability. In addition, since SAT is a problem that often arises in academia and the industry, each of the candidate solvers are rigorously tested with massive application-

specific, crafted, and random Boolean formulas as input. With three different solver specializations tested against three different types of inputs, and a first, second, and third place awarded to the candidates, a total of 27 possible trophies are awarded each year. In the most recent competition held in 2011, solvers were tested using CPU time and world-clock time as a basis for their results, thus expanding the trophy space to 54 slots. The next competition is slated to take place in 2013.

The SAT problem is particularly interesting when applied to the classical Ramsey arrowing operator. In particular, it is possible to reduce the question $G \xrightarrow{?} (3, 3)^e$ to an equivalent 3-CNF formula ϕ_G such that $G \not\rightarrow (3, 3)^e \Leftrightarrow \phi_G$ is satisfiable. Intuitively, this is a very promising technique for determining if $G \rightarrow (3, 3)^e$ for K_4 -free graphs G . The result can also be applied to other arrowing parameters, as is discussed in Section 5.

The immediate application of this technique is to attack the upper bound of the Folkman number $F_e(3, 3; 4)$. In particular, to lower this bound, we will need to find a graph G on n vertices where $G \rightarrow (3, 3)^e$. It has been conjectured that $G_{127} = G(127, 3) = (\mathbb{Z}_{127}, E = \{(x, y) | x - y = \alpha^3 \pmod{127}\})$ is a prime candidate for witnessing an upper bound of 127 because of its denseness and large number of triangles. To determine whether G_{127} is indeed a witness we will decompose the problem of arrowing into problems on subgraphs H that witness $H \not\rightarrow (3, 3)^e$, and then carefully extend H to encompass all of G . For each subgraph H will generate a corresponding 3-CNF formula ϕ_G by mapping the edges in $E(H)$ to variables in $\phi_H \in 3\text{-SAT}$, and for edge adding the following clauses to ϕ_H :

$$(x + y + z) \wedge (\bar{x} + \bar{y} + \bar{z})$$

If H can be extended to encompass all of G and ϕ_G is shown to be unsatisfiable, then $G \rightarrow (3, 3)^e$, and so $F_e(3, 3; 4) = 127$. Otherwise, if ϕ_G is satisfiable, then $G \not\rightarrow (3, 3)^e$, disproving the conjecture made in by Radziszowski et al. [4].

Unfortunately, while this approach seems simple at first, the complexity of the formulas ϕ_H with $n \approx 85$ has proven to be very difficult for modern SAT solvers to handle. In this study, we will attempt to determine the structure of these formulas that makes them so difficult to solve, and as a result, discover the underlying cause for the phase transition that occurs at $n \approx 85$. We will also present a performance comparison for the popular SAT solvers entered into the SAT competition, including Minisat [5], zChaff [7, 6], and glucose [8] (which is based heavily on Minisat).

2 SAT Solver Algorithms

The algorithms used in state-of-the-art SAT solvers tend to be based on the famous DPLL backtracking algorithm proposed by David, Putnam, Logemann, and Loveland in [9]. The DPLL algorithm works by recursively splitting a CNF formula into smaller formulas by assigning truth values to individual variables and simplifying the resulting formula at each level of iteration. Simplification works by removing all clauses in ϕ that become

Algorithm 1 DPLL Algorithm [9]

Require: ϕ **Ensure:** *true* or *false*

```
1: if  $\phi$  is a tautology with the current assignment then return true
2: end if
3: if  $\phi$  contains an empty clause or all literals in a clause are false (conflict) then return
   false
4: end if
5: for all unit clauses  $c \in \phi$  do
6:    $\phi \leftarrow \text{UnitPropagate}(c, \phi)$ 
7: end for
8: for all pure literal  $l \in \phi$  do
9:    $\phi \leftarrow \text{PureLiteralElimination}(l, \phi)$ 
10: end for
11:  $l \leftarrow \text{ChooseLiteral}(\phi)$ 
12: return  $\text{DPLL}(\phi \wedge l) \vee \text{DPLL}(\phi \wedge \neg l)$ 
```

true under the new assignment and all literals which are false. This process continues recursively until the formula is deemed satisfiable or a conflict emerges. Satisfiability occurs when all variables have been assigned truth values and the resulting formula has non-empty clauses remaining, or when all clauses can be evaluated to true. If a conflict arises, the algorithm backtracks to the last assigned truth value, attempts to substitute the negated truth value, and then proceeds as normal. A formula ϕ is deemed unsatisfiable if this algorithm exhaustively checks all possible truth value assignments without yielding a satisfiable formula.

The DPLL algorithm further enhances the backtracking search through the use of two additional procedures, unit propagation and pure literal elimination. Unit propagation works by assigning an appropriate truth value to all unit clauses (i.e. all clauses with only a single literal) and then propagating this truth value to the rest of the formula. For example, unit propagation of the formula $\phi = [(x_1) \wedge (x_1 \vee x_2 \vee x_3)]$ yields $\phi = [(x_2 \vee x_3)]$ by assigning a value of true to x_1 . Pure literal elimination works by finding variables which have uniform polarity (i.e. always positive or negative). For all such literals, there always exists a truth assignment such that the containing clauses will evaluate to true. Therefore, such clauses are deleted from ϕ . For example, given the formula $\phi = [(x_1 \vee x_2 \vee \neg x_3) \wedge (x_2 \vee \neg x_3 \vee x_4) \wedge (x_1 \vee x_2 \vee \neg x_4)]$, performing pure literal elimination yields $\phi = [(x_1 \vee x_2 \vee \neg x_4)]$ because we can assign $x_3 = \text{false}$ and eliminate the first two clauses. These two procedures are shown in the pseudocode description shown in Algorithm 1.

2.1 Advanced Heuristics

The Boolean Constraint Propagation (BCP) algorithm in [7] was a significant improvement on the traditional DPLL algorithm which seeks to minimize the number of times each clause is visited using *variable decisions* and *clause implications*. This procedure consists of modifying ϕ after assigning a variable, or deciding its truth value, by implying further assignments for unit clauses when detected. If a conflict emerges during propagation, then the algorithm must backtrack to the most recent literal decision, negate the value, and then continue down the search tree.

Generally speaking, most SAT solvers that employ this algorithm spend the majority of their time in this procedures. Chaff [7] introduced more advanced heuristics to optimize this part of the DPLL algorithm. Variable watching, which is an optimization that seeks to maintain some knowledge about which clauses recently become unit clauses after a variable assignment, helps reduce the number of clauses that are visited during the BCP procedure.

Another optimization is the Variable State Independent Decaying Sum (VSIDS) decision heuristic, which helps guide the solver when selecting a variable to assign at each layer in the search tree [7]. This procedure works by preprocessing ϕ at the beginning of the solve function to create a collection of counters for the number of times each literal appears in the formula. As the solver progresses, variables for decision are selected based on their count, and the solver will periodically reduce each count by a constant value. As a result, frequently occurring variables are selected first as they will propagate through ϕ more.

3 3-SAT Phase Transitions

The phase transition for the SAT decision problem, or any constraint satisfaction problem, is defined as the *sharp* change in probability that the input problem instance has a satisfying solution [14]. In the case of 3-SAT, the phase transition, which is proportional to the ratio of clauses to variables, has been empirically determined to be approximately 4.2. In the case of G_{127} , the ratio of clauses to variables is roughly 7.483, which is well beyond this threshold. However, it is estimated that the difficulty of deciding $\phi_{G_{127}}$ reaches a point of singularity on the phase transition boundary. Therefore, analyzing the difficulty of solving reduced formulas beyond this threshold may offer insight into the overall computational difficulty of deciding $\phi_{G_{127}}$. Estimating the hardness of Ramsey arrowing problems converted to SAT formulas is the primary interest of this report.

3.1 Probabilistic Estimation of the Hardness

Table 1: Performance of Minisat solver by removing N_R vertices from G_{127}

m_r	CPU Time	Restarts	Conflicts	Decisions	Propagations	Conflict Literals
46	1228.97s	28668	18968226	41661327	1999748077	930184526
45	4362.29s	66558	52952393	106164381	5779728314	2755169058
44	5576.51s	81915	62420966	126120884	6748259547	3254804365
43	21681.5s	245755	199022470	384078647	21687371442	10581492993

3.2 Empirical Tests for Hardness

In our experiments we denote the number of vertices removed from the graph G_{127} as m . The CNF formula ϕ_G for G_{127} experiences a peculiar phase transition when $m \approx 41$, making it seemingly intractable for modern SAT solvers to halt with an answer in a realistic amount of time. Table 1 shows how the Minisat solver time increases as m decreases from 46 to 43. The CPU time continues to increase as m decreases.

To understand the nature of the phase transition for G_{127} we conducted the following experiments: random subgraph selection and variable fixation, induced subgraph modifications by adding edges without creating cliques of size 4, and induced subgraph modification by removing edges. Before discussing the results of these experiments, we first describe the software tools that were used to aid in these experiments.

Breaking the problem of $G_{127} \rightarrow (3, 3)^e$ into small graphs H is a natural way to approach this problem. In addition, clause reduction and unbalancing through variable assignment may provide further simplifications to ϕ_H . Therefore, for this experiment, we have the following parameters:

- m_s - the number of maximal independent sets $V \subset V(G)$ removed from G . This is a structured removal of vertices that doesn't impact the symmetry of the graph in a significant way.
- m_r - the number of vertices in $V \subset V(G)$ such that $H = G[V]$. The selection of these vertices will be both structured and unstructured. That is, we will experiment with removing structured groups of vertices, such as those contained within independent sets, as well as unstructured groups composed of randomly selected vertices.
- r - the number of variables $x_i \in \phi_H$ assigned truth values that are propagated through the rest of the formula.
- p - the number of levels of recursive propagation that occur after assigning r truth values. By default, $p = 1$, indicating a single pass of unit propagation.

Table 2: SAT solver guidance experimental results. Values of NA indicate that the computation could not be completed (i.e. the SAT solver could not decide a formula instance in a reasonable amount of time). NA

m_s	m_r	r	Average time [s]	Minimum time [s]	Estimated CPU years
0	37	18	NA	15402s	512.1 (based on minimum time)
0	30	20	NA	8334.8	276.9 (based on minimum time)
0	37	22	NA	1858.1	246.9 (based on minimum time)
0	37	23	NA	NA	NA
0	37	24	NA	2361.6s	1256 (based on minimum time)
0	37	25	NA	NA	NA
0	37	26	NA	NA	NA
0	37	27	NA	NA	NA
0	37	28	NA	NA	NA
0	37	29	NA	NA	NA
0	37	30	NA	NA	NA
0	37	31	NA	NA	NA
0	37	32	NA	NA	NA

- n_e - the number of edges randomly added to H such that no K_4 is formed. Since G_{127} is edge critical, it is not possible to add any edge to the graph without forming a K_4 .

Clause reduction will use the same unit propagation and pure literal elimination techniques presented in the DPLL algorithm as a preprocessing step for ϕ_H before it is run with a solver.

3.3 Experiment 1: Induced subgraph generation with SAT guided variable fixation

The hardness of solving ϕ_H can be estimated by summing together the time for all individual 2^r truth value assignments. In particular, if ϕ_H^i is the resulting formula after assigning the i th variable configuration, and $T(\phi_H^i)$ is the time required to decide ϕ_H^i , then the hardness (estimated decision time) for ϕ_H is $\sum_{i=1}^{2^r} T(\phi_H^i)$. Table 2 lists some of the results from these experiments and the estimated computation time required to decide ϕ_H for various $H = G[n]$.

During this experiment we intuitively assumed that those formulas with a higher ratio of clauses of length two as compared to clauses of length three will be decided faster. Therefore, to aid the SAT solver in its decision for all $1 \leq i \leq 2^r$ possible ϕ_H^i formulas we have sorted list of ϕ_H^i in descending order based on this ratio. These results have aided in the evaluation of these small experiments.

Table 3: Structural induced subgraph generation and variable fixation experimental results. Values of NA indicate that the computation could not be completed (i.e. the SAT solver could not decide a formula instance in a reasonable amount of time).

m_s	m_r	r	Average time [s]	Minimum time [s]	Estimated CPU years
3	0	28	NA	98334.8	837022
3	0	29	NA	NA	NA
3	0	30	NA	NA	NA

Table 4: Induced subgraph modification experimental results. Values of NA indicate that the computation could not be completed (i.e. the SAT solver could not decide a formula instance in a reasonable amount of time).

m_s	m_r	r	n_e	Average time [s]	Minimum time [s]	Estimated CPU years
0	27	28	60	NA	NA	NA
0	27	29	60	NA	81923.4	NA
0	27	30	60	NA	67219.6	NA

3.4 Experiment 2: Structural induced subgraph generation and variable fixation

In this set of experiments we combine the structural removal of vertices from G_{127} with the random removal of vertices. In particular, we specify the minimum size of the induced subgraph H and a number of maximal independent sets to remove from G_{127} . If the removal of these independent sets does not meet the designated order of H , then the remaining vertices that need to be removed are done so at random. Table 3 shows the results that have been compiled from this experiment so far.

3.5 Experiment 3: Induced subgraph modification.

This experiment consists of adding edges to induced subgraphs of G_{127} (formed from the vertex set $V \subseteq V(G_{127})$) such that the resulting graph $G[V] + e$ is K_4 -free. The intuition for this experiment is that adding additional edges will increase the effectiveness of variable fixation and unit propagation in the SAT formula preprocessing stages of the experiments. Algorithm 2 shows the general procedure that is followed for creating subgraphs $H = G[V]$. The preliminary experimental results for this experiment are shown in Table 4

4 Formula Balancedness

An interesting result from these hardness estimation experiments was that the balancedness of the formula ϕ_H had a surprising impact on the ease with which solvers can

Algorithm 2 Algorithm for constructing candidate graphs for G_{127} SAT hardness experiments

Require: Non-negative integers m_s , m_r , and n_e

Ensure: $H \subset G_{127} \cup \{e_1, \dots, a_{n_e}\}$

```

1:  $H \leftarrow G_{127}$ 
2: for  $i = 0 \rightarrow m_s$  do
3:    $H \leftarrow H - V$ , where  $V \subset V(H)$  is a maximal independent set in  $H$ 
4: end for
5: for  $i = 0 \rightarrow m_r$  do
6:    $H \leftarrow H - v$ , where  $v$  is a randomly selected vertex  $v \in V(H)$ 
7: end for
8: for  $i = 0 \rightarrow n_e$  do
9:    $e \leftarrow (u, v)$ , where  $u, v \in V(H)$  are two randomly selected vertices
10:   $cc \leftarrow 0$ 
11:  while  $e \cup E(H)$  forms a  $K_4$  and  $cc < \text{THRESHOLD}$  do
12:     $e \leftarrow (u, v)$ , where  $u, v \in V(H)$  are two randomly selected vertices
13:     $cc \leftarrow cc + 1$ 
14:  end while
15:  if  $cc = \text{THRESHOLD}$  then return  $H$ 
16:  else
17:     $H \leftarrow (V(H), E(H) \cup e)$ 
18:  end if
19: end for

```

determine the formula's satisfiability. If we force a certain coloring of the edges of H , which equates to assigning truth values to the corresponding variables in ϕ_H , it is clear that the number of clauses will either decrease or the number of literals in each clause will decrease, or both, after propagating these truth value assignments throughout the formula. In this experiment, we are after the phase shift that occurs when comparing the ratio of clauses of size 2 to clauses of size 3. We note that if a clause has size 1 the literal's truth value can be fixed and then propagated throughout the formula again (this process can recurse indefinitely).

As part of this experiment we looked at fixing 50 and 60 variables in ϕ_H , where H is an induced subgraph on 85, 90, and 95 vertices. To further this analysis, we also experimented with disjoint edge colorings (i.e. choosing edges that don't belong to the same triangle) and random edge colorings. The phase shift is visualized in Figures 1, 2, 3, 4, and 5 for H graphs of varying size using both the disjoint and random edge selection strategies.

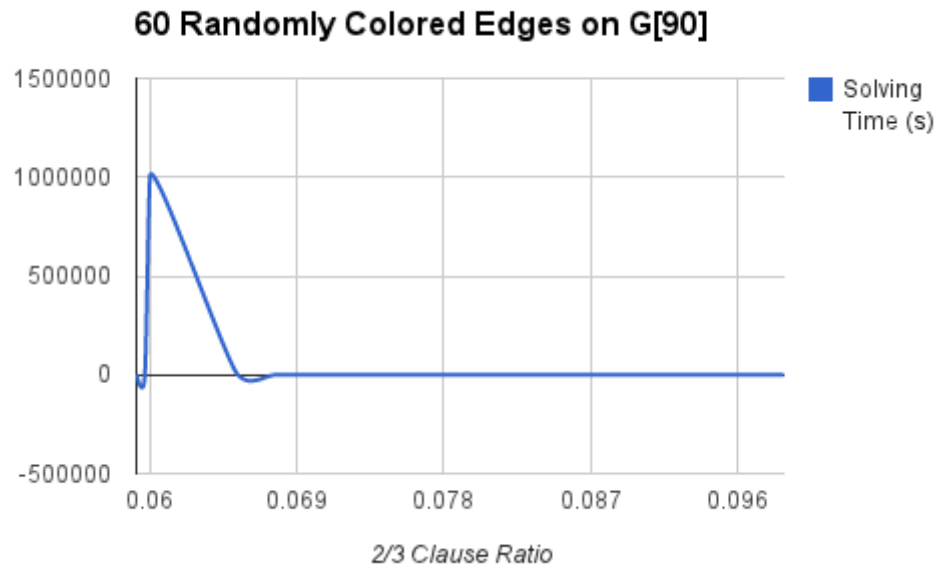


Figure 1: The global maximum should lie at infinity (the solver did not finish within a time frame of 5 hours).

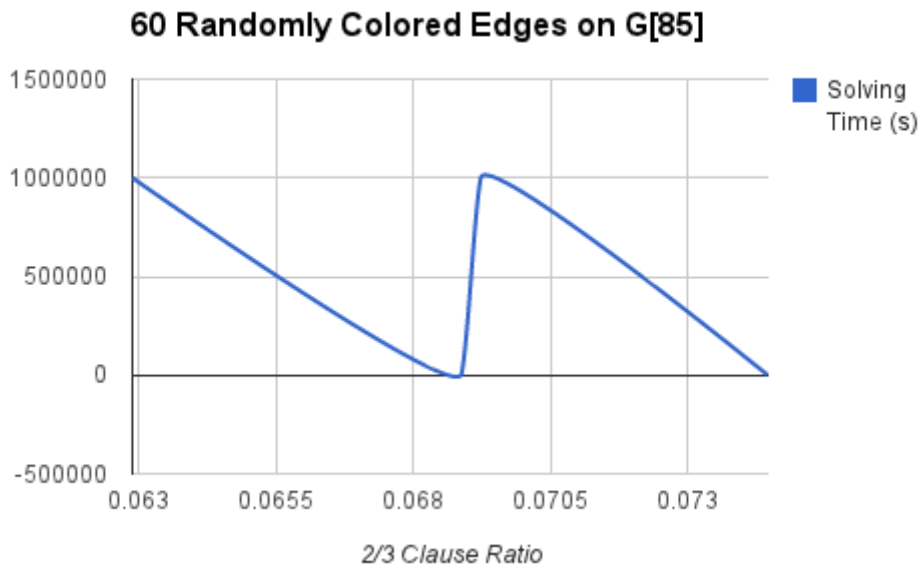


Figure 2: The global maximum should lie at infinity (the solver did not finish within a time frame of 5 hours).

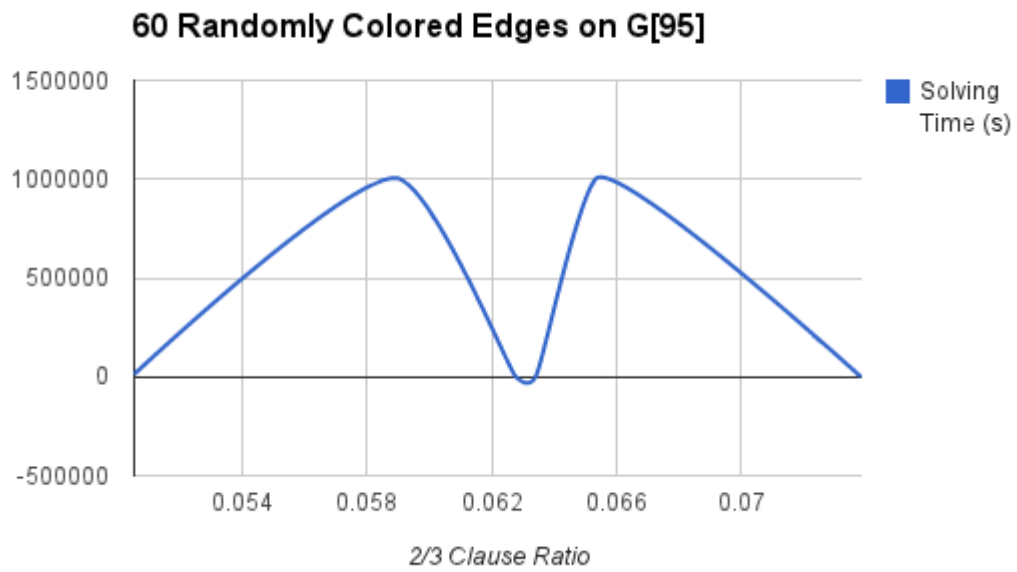


Figure 3: The global maximum should lie at infinity (the solver did not finish within a time frame of 5 hours).

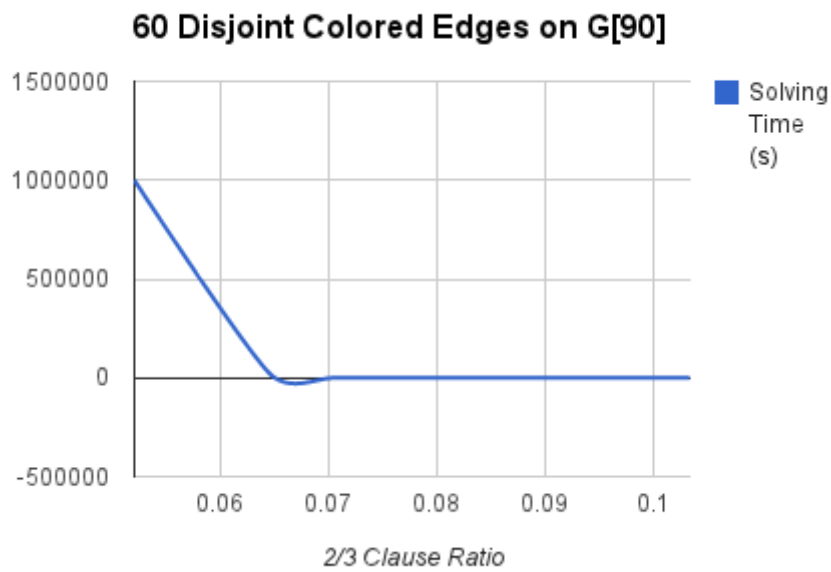


Figure 4: The global maximum should lie at infinity (the solver did not finish within a time frame of 5 hours).

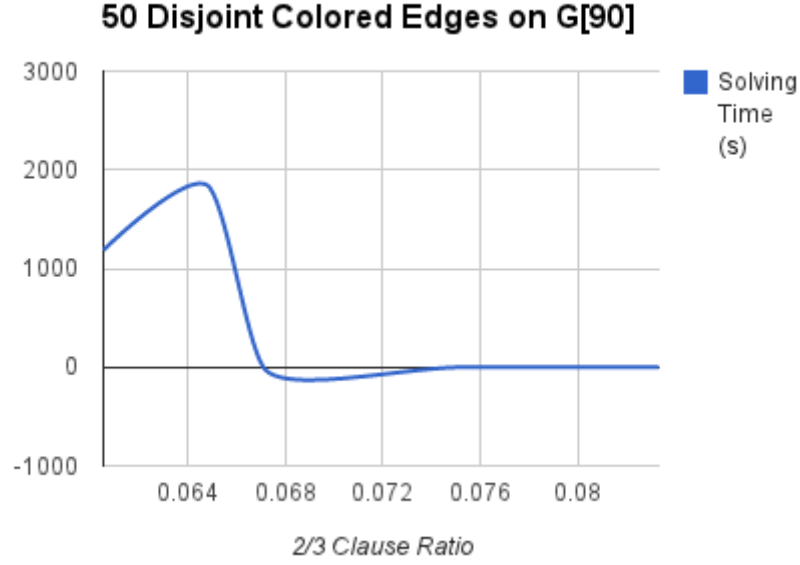


Figure 5: The global maximum should lie at infinity (the solver did not finish within a time frame of 5 hours).

5 Arrowing Problems

In this section we present some new results for edge Folkman numbers, some of which are improvements upon results found in the literature.

5.1 $F_e(3, 5; 13) = 18$

The problem of determining if $G = K_8 + C_5 + C_5 \rightarrow (3, 5)^e$ was proved by Nenov [12] to determine equality in the bound $F_e(3, 5; 13) \geq 18$. According to Kolev in 2008 [13], no one has been able to solve this arrowing problem. Using SAT reduction techniques, we have reduced this problem to an equivalent SAT formula as follows. For each triangle xyz in G , add $(xy + xz + yz)$ to ϕ_G . Then, search for all other pairs of vertices u, v such that $\{x, y, z, u, v\}$ is a clique of size 5. For all such sets of vertices, add $(\bar{x}y + \bar{x}z + \bar{x}u + \bar{x}v + \bar{y}z + \bar{y}u + \bar{y}v + \bar{z}u + \bar{z}v + \bar{u}v)$ to ϕ_G . The resulting formula has 143 variables, each one corresponding to an edge in G , and 4982 clauses. Minisat was able to prove that this formula is unsatisfiable in 0.83s, thus showing that $G \rightarrow (3, 5)^e$, and therefore $F_e(3, 5; 13) = 18$.

References

- [1] R. Karp, Reducibility among combinatorial problems, *Complexity of Computer Computations*, (RE Miller and JM Thatcher, eds.) (1972), 85-103.
- [2] S. A. Cook, The Complexity of Theorem-Proving Procedures, *In Proceedings of the third annual ACM symposium on Theory of computing (STOC '71)*. ACM, New York, NY, USA (1971), 151-158. DOI=10.1145/800157.805047. <http://doi.acm.org/10.1145/800157.805047>
- [3] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, Introduction to Algorithms, *MIT Press* **44** (1990), 97-138.
- [4] S. P. Radziszowski, X. Xu, On the Most Wanted Folkman Graph, *Geombinatorics* **16** (4) (2007), 367-381.
- [5] N. Sörensson, N. Eén, Minisat v1.13 - A SAT Solver with Conflict-Clause Minimization, *SAT 2005* (2005) 53.
- [6] Y. Mahajan, Z. Fu, S. Malik, Zchaff2004: An Efficient SAT Solver, *Theory and Applications of Satisfiability Testing*. Springer Berlin/Heidelberg (2005).
- [7] M. W. Moskewicz, C. F. Madigan, Y. Zhao, L. Zhang, S. Malik, Chaff: Engineering an efficient SAT solver, *Proceedings of the 38th annual Design Automation Conference* ACM (2001).
- [8] G. Audemard, L. Simon, GLUCOSE: a solver that predicts learnt clauses quality, *SAT Competition* (2009), 7-8.
- [9] M. Davis, G. Logemann, and D. W. Loveland, A machine program for theorem-proving, *Communications of the ACM* **5(7)** (1962), 394-397.
- [10] Y. Boufkhad, O. Dubois, Y. Interian, B. Selman. Regular Random k-SAT: Properties of Balanced Formulas, *Journal of Automated Reasoning* **35.1-3** (2005), 181-200.
- [11] Random k-SAT: Two moments suffice to cross a sharp threshold
- [12] N. Nenov, On the Zykov numbers and some of their applications in Ramsey theory (in Russian), *Serdica* **9(2)** (1983), 161-167.
- [13] N. Kolev, New Upper Bound for the Edge Folkman Number $F_e(3, 5; 13)$, arXiv:0806.1403 (2008).

- [14] D. M. Pennock, Q. F. Stout, Exploiting a theory of phase transitions in three-satisfiability problems, *Ann Arbor* **1001** (1996), 48109.