Independent Study Proposal

Ramsey and Folkman Numbers
And the potential SAT Solvers and Basis Reduction have in computing them

Professor Stanislaw Radziszowski
Winter 2011/2012

Alexander Lange
Graduate Student
Department of Computer Science
Rochester Institute of Technology
arl9577@rit.edu

Goals of Proposed Study

- Research current methods in computing bounds on Ramsey and Folkman Numbers
- Develop a better understanding of the satisfiability problem, leading SAT solvers, basis reduction and the LLL algorithm
- Explore the potential these methods have in computing new better bounds on such problems

Introduction

Combinatorial computing is a branch of computer science which involves developing and implementing algorithms to generate, enumerate and search through combinatorial structures [1]. These combinatorial structures are found in a variety of both theoretical and applied mathematical fields, but because of their size, pure mathematical analysis becomes impractical. Computers are needed in order to study and understand them further.

During this fall quarter at RIT, I have studied under Professor Radziszowski in the course VCSG-801: Combinatorial Computing. In this course, I was introduced to a number of these structures, the interesting problems that are carried with them, and the tools that have been used in order to solve such problems. The area I found most interesting was the search for Ramsey and Folkman numbers. Ramsey and Folkman theory studies the properties combinatorial structures need in order to guarantee desired structures contained within them [3][4]. Specifically, Ramsey and Folkman numbers are the smallest size of graphs needed to necessarily contain particular subgraphs.

The ability to find Ramsey and Folkman numbers depends directly on the ability to find the maximum clique or some similar structure within a graph. The maximum clique problem is NP-complete and therefore all tools that are used in solving the problem become more and more difficult to use as the sizes of the questioned graphs increase. However, because it is NP-complete, it can be reduced to other NP-complete problems [2]. The satisfiability problem, which is also NP-complete, is of particular interest. This problem asks whether or not a Boolean formula, consisting of Boolean variables and operations, can be evaluated to true by any assignment of values to the given variables. This problem has been studied extensively, and a number of “SAT solving” algorithms have been created that make use of heuristics and other tricks to solve specific examples of the problem.
Since the max clique problem can be reduced to a Boolean expression, it can also be transformed into a system of linear equations. In class, the LLL algorithm was introduced, which reduces the basis of lattices formed from integer linear equations.

The purpose of this independent study will be to further study the satisfiability problem, the current leading SAT solving algorithms, the LLL algorithm, and the potential that these tools have in finding new better bounds on the currently unknown Ramsey and Folkman numbers. The following sections will discuss in more detail the four areas which will be studied.

Techniques

Satisfiability and SAT Solvers

The satisfiability problems of particular interest are those that are in conjunctive normal form (CNF). A Boolean expression is in conjunctive normal form if it contains sets of Boolean variables connected with the Boolean OR operator, which are all connected with the Boolean AND operator. For example, the expression

$$(x_1 \lor x_2 \lor x_3) \land (x_2 \lor x_3)$$

is in CNF, and is clearly satisfiable – the assignment $x_1 = 1, x_2 = 1$ and $x_3 = 0$ evaluates the formula to 1.

It can be shown that the maximum clique problem can be reduced to Boolean expressions of this form [2]. The advantage of reducing the max clique problem is that many algorithms have been developed for solving specific types of satisfiability problems. Although not guaranteed, these algorithms can find satisfying solutions in time that is much less than the expected exponential time of backtracking algorithms used in finding the maximum clique of graphs.

In addition to studying the polynomial reductions of graphs to CNF-SAT, a variety of SAT-solving software will be surveyed and experimented with. Each year, competitions are held which rank the leading SAT solvers on their ability to solve certain types of problems. These solvers (found at [8]) will be a good starting point for study.

The LLL Algorithm

In addition to being reduced to CNF-SAT, graph searching problems can also be reduced to solving matrix equations of the form $AU = B$, where $U$ is a $(0,1)$ vector [1]. However, since the graphs in question will presumably be quite large, using similar backtracking techniques on the matrices will most likely be just as difficult. We therefore consider a separate, but equivalent equation:

$$\begin{bmatrix} I & 0 \\ A & -B \end{bmatrix} \begin{bmatrix} U \\ 1 \end{bmatrix} = \begin{bmatrix} U \\ 0 \end{bmatrix}$$

By picking a lattice with basis equal to

$$\begin{bmatrix} I & 0 \\ A & -B \end{bmatrix}$$

we can try to reduce this basis to obtain smaller structures to work with. We do this by obtaining as close to an orthogonal basis as possible (since computations are with integers, a true orthogonal basis is unlikely). The LLL algorithm, invented by Lenstra, Lenstra and Lovász in 1982, takes advantage of the Gram-Schmidt Process to perform such a reduction in polynomial time [1].

The LLL algorithm and the linear algebraic theory involved with it will be studied [6]. Strategies in reducing Ramsey and Folkman type problems to systems of integer linear equations will also
be examined, with the hopes that using this approach will increase the possibility of solving such problems.

Problems

Ramsey Numbers

The Ramsey Number $R(s, t)$ is the least positive integer $n$ such that for any graph $G$ with $|G| = n$, $G$ contains a clique of order $s$ or independent set of order $t$. This is equivalent to stating that any edge two-coloring of $K_n$ contains either a monochromatic coloring of size $s$ in the first color or $t$ in the second [5].

Although Ramsey theory has been under significant study, many of the questions in the field remain open [3]. Large computations are performed in order to find solutions to specific instances of the problem. These instances usually involve modifications to the original stated problem, where more than two colors are used and/or subgraphs other than cliques are searched for. For example, $R(3, 3, 3) = 17$ means that $K_{17}$ is the smallest complete graph such that any three-coloring of the edges of this graph results in a monochromatic triangle in one of the colors. $R(W_5, C_5) = 9$ states that 9 is the smallest size possible such that for any graph with such size, there is always either a $W_5$ subgraph or $C_5$ independent set found within the graph.

The naive approach to computing Ramsey numbers involves generating all possible non-isomorphic graphs of a certain size, and searching them for the desired substructures. Unfortunately, generating these graphs becomes impossible at relatively low numbers. Likewise, searching through them is very hard to do, even with various pruning methods applied to backtracking algorithms. Therefore, mathematicians and computer scientists are posed with an interestingly difficult problem, where creative and innovative methods are needed in order to find the solution.

Folkman Numbers

A separate branch of Ramsey theory involves computing vertex and edge Folkman numbers. These problems are similar, but tend to be more difficult because an additional variable is introduced. For the edges, the Folkman Number $F_e(s, t; k)$ is the smallest possible size of a graph such that $K_k$ is not a subgraph, and any two-coloring yields either $K_s$ in the first color or $K_t$ in the second [4].

As an example, it can be shown that $F_e(3, 3; 6) = 8$. This is because the graph $C_5 + K_3$ (which has 8 vertices) does not contain $K_6$, but at the same time, produces a monochromatic triangle with any two-coloring. Because no graph with 7 vertices has this property, the answer is 8.

Normally, when computing Ramsey numbers, massive computations are helpful in finding counter examples for the lower bound. However, with Folkman Numbers, this strategy is reversed, and the upper bound is where witnesses are needed. This property makes them harder to find – the larger the graph, the harder it is to generate graphs which lead to witnesses. As $k$ decreases in size, there is a larger chance that a clique of size $k$ will exist in a graph. Therefore, Folkman numbers with smaller parameters are both the most sought after and the most difficult to compute. For example, $F_e(3, 3; 4)$ is considered to be “most wanted”, while at the same time, is not very well understood.

Summary of Proposed Study

I plan on approaching this independent study with equal parts theory and practice. With my background in mathematics, I will be able to develop a strong understanding of the mathematical concepts behind these different, yet connected areas of study. Although I recognize the importance
in theory, it is also quite evident to me that large computations are necessary in order to obtain results. Educated and creative experimentation is needed to further the understanding of these problems. The software package nauty, developed by Brendan McKay, will be studied and used extensively to generate and manipulate graphs [7]. Other software packages, such as SAT-solvers, will be used, and multiple algorithms, including the LLL algorithm, will be implemented. An obvious goal of this independent study will be to obtain results – finding new bounds and/or values for Ramsey and Folkman numbers would be wonderful. But at the very least, I will gain a significantly deeper understanding and appreciation for these interesting and inherently difficult fields of study.

Preliminary Bibliography


