Optimisation Assignment

Matt Ingram

February 21, 2019

1 Installation

A few small prerequisites need to be installed before running my program. These are: Python 3, GNU Multiple Precision Arithmetic Library (GMP), and the SoPlex solver. I shall include the GMP and SoPlex archived files (for installation on a Unix-based system) in my final submission. Alternatively, follow the blue links to go to their respective download pages. To install these dependencies, please follow the instructions below:

• GNU Multiple Precision Arithmetic Library

Please change directory to where the *qmp-6.1.2.tar.lz* file is located and run the following commands:

- lzip -d gmp-6.1.2.tar.lz
- 2 tar -xf gmp-6.1.2.tar
- cd gmp-6.1.2
- ./configure
- 5 make
- 6 make check
- 7 make install

• SoPlex Solver

Please change directory to where the soplex-4.0.1.tgz file is located and run the following commands:

- tar -xf soplex-4.0.1.tgz
- 2 cd soplex-4.0.1
- 3 mkdir build
- d cd build
- 5 cmake <path/to/SoPlex> -DGMP=true
- 6 make
- 7 make test
- 8 make install

• Pyomo

The easiest way to install Pyomo is to use Python's built-in package manager, *pip*. Once *pip* has been installed, please run the following command:

```
1 pip install pyomo
```

Should you run into any issues installing any of these dependencies, please consult the relevant INSTALL files and documentation.

2 Usage

2.1 Directory Structure

Once the dependencies have been installed, the directory structure should look something like Figure 1 (varying slightly if you installed the dependencies in a different location):

Note: The program will **not** work unless the four folders (*graphs*, *models*, *settings*, *solutions*) are present. This is what is contained in each folder:

• /graphs: This folder contains the input graph files. Graph files should be saved with a .txt extension, and follow the following format:

[V] [E] [v11] [v12] [v21] [v22] ...

[vE1] [vE2]

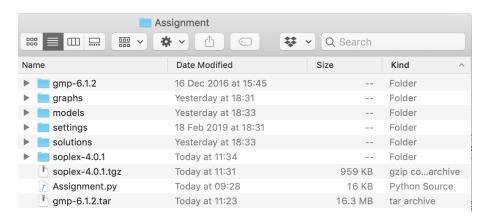


Figure 1: Directory Structure

where V is the number of vertices, E is the number of edges, and each [vi1] [vi2] pair (for $i \in \{1, 2, ..., E\}$ represents the edge between vertex [vi1] and [vi2]. For example, the complete four-vertex graph would be represented as follows:

Note: Vertex numbering *must* start at 0, not 1, for the program to work.

- /models: This folder contains the .mps linear program modelling files generated by my program, which are then solved by the SoPlex solver.
- /settings: This folder contains the settings files for the SoPlex solver.
- /solutions: This folder contains the detailed solutions outputted by the SoPlex solver. Each time the program is executed, it will output two files one for the Fractional Clique Cover Number, and one for the Shannon Entropy.

2.2 Running

To run the program, please run the following command (from the base directory):

The program will output the Fractional Clique Cover Number, the Shannon Entropy, and the solutions to both of the linear programs concisely, and in rational format. As mentioned above, more detailed output for both linear programs is saved in the /solutions folder.

3 Linear Program Reformulation

3.1 Fractional Clique Cover Number

- In the specification, variables are introduced for all the subsets S of V, the set of all vertices. However, as the only non-zero variables are those corresponding to the cliques of V, I introduced variables only for the cliques. To find the cliques of V efficiently, I modified the well-known $Bron\ Kerbosch$ algorithm, so that it recursively finds all the cliques of V, not only the maximal ones.
- Whilst searching for how to further simplify the linear program, I found an article by Peter Cameron[1] which states that "It can be shown that the same minimum value [the Fractional Clique Cover Number] is obtained if we restrict to regular fractional clique covers.". A regular fractional clique cover is one

where, for every vertex, the sum of the values given to the cliques that contain this vertex **equals** one, rather than being bounded above by 1. The proof of this claim can be found in the paper referenced in the article[2], and so I implemented this simplification in my program.

3.2 Shannon Entropy

For the Shannon Entropy linear program, variables are required for all of the subsets of the set of vertices V. However, many of the constraints 'overlap' and so are redundant.

• The first constraint that I simplified was the:

$$x_T - x_S \ge 0 \quad \forall S \subseteq T \subseteq V$$

constraint. For very subset T of V, we do not need to create a constraint for every subset S of T; it is sufficient to only create constraints for subsets S that are of size one less than T. This is because if $R \subseteq S \subseteq T$, then $x_R \le x_S$ and $x_S \le x_T \implies x_R \le x_T$. This is more easily shown diagramatically:

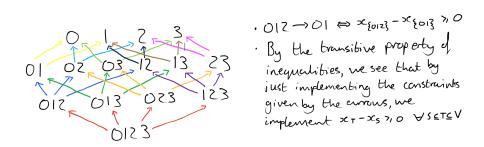


Figure 2: Constraint Reformulation

For a graph of N vertices, the total number of constraints without the reformulation would be $\sum_{n=2}^{N} \binom{N}{n}(2^n-1)$ (where constraints involving the empty set have been removed as the variable corresponding to the empty set is 0), whereas the the number of constraints with this reformulation is $\sum_{n=2}^{N} \binom{N}{n} n$. For the N = 8 case, this is a reduction of 6297 – 1016 = 5281 constraints.

- To increase the efficiency of the constraint generation, I calculated the above constraint and the $X_{N(v)\cup\{v\}}-x_{N(v)}=0$ constraint in one traversal of the subsets pairs, as the neighbourhood constraint also involves two sets differing in size by only one variable. I don't think it is possible to further simplify the neighbourhood constraint, and as there is only one constraint per vertex, the number of these constraints is negligible with respect to the total number of constraints.
- The next constraint that I simplified significantly was the:

$$x_S + x_T - x_{S \cup T} - x_{S \cap T} \ge 0 \quad \forall S, T \subseteq V$$

constraint. Instead of creating a constraint for every pair of vertices S and T, of which there would be $2^N \times 2^N = 2^{2N}$, constraints are only required for pairs of vertices where $S \not\subseteq T$ and $T \not\subseteq S$. If $S \subseteq T$, then:

$$x_S + x_T - x_{S \cup T} - x_{S \cap T} = x_S + x_T - x_T - x_S = 0$$

and so the constraint becomes trivial. The same argument holds for $T \subseteq S$ by simply switching the variable labels.

• The final constraint I attempted to simplify was the:

$$x_{\{v\}} \leq 1 \quad \forall v \in V$$

constraint. After lots of experimentation, I believe that the inequality in this constraint can be replaced with an equality. This intuition is backed up by the proof that this modification holds in the Fractional Clique Cover Number linear program. Furthermore, when testing my program with the constraint as an inequality, SoPlex reported that there were a small number of redundant constraints, which I believe to relate to this constraint. However, as I could not come up with a formal proof to back up this intuition, I decided to leave the constraint as an inequality in my final implementation.

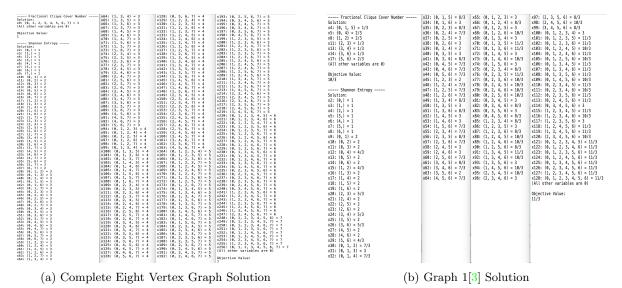


Figure 3: Example Solutions

4 Examples

In Figure 3, the outputs for the complete eight vertex graph and Graph 1 are shown. As expected, the Fractional Clique Cover Number of the complete graph is 1, the Shannon Entropy of the complete graph is 8-1=7, the Fractional Clique Cover Number of Graph 1 is 10/3, and finally the Shannon Entropy of Graph 1 is 11/3. Please note that multiple other example graphs are included in the /graphs folder, including all 6 graphs from the paper[3].

5 Additional Information

- If you would like the program to print out the linear program (the objective function and the constraints) before it is solved, please uncomment out line 435 of Assignment.py.
- I inverted all of the greater than constraints so that the model only consists of equality and less than constraints, in order to simplify the implementation.
- The vertex numbering in the input files must begin at vertex **0**, and continue sequentially.
- In my initial implementation, I used the linear program solver included in the SciPy package, and then used the Fraction module to convert the floating point output to a rational format. However, this seemed unsatisfactory and susceptible to floating-point arithmetic nuances, and so I reimplemented my code using Pyomo and SoPlex. My program first generates all the variables and constraints, which are then passed to Pyomo. Pyomo builds the LP model and outputs it in the standard .mps format, and then my program spawns a new process and executes the SoPlex program, which solves the model. My program then reads the exact solution outputted by SoPlex, parses the rational numbers as strings, creates exact fractions from these strings using the Fraction module, and then finds the exact optimal value using the exact optimal solution and the objective function.

References

- [1] Peter Cameron: Guessing Numbers of Graphs, https://cameroncounts.wordpress.com/2016/03/20/guessing-numbers-of-graphs/
- [2] Peter Cameron, Anh N. Dang, Søren Riis: Guessing Games on Triangle-free Graphs, The Electronic Journal of Combinatorics
- [3] Maximilien Gadouleau: On the possible values of the entropy of undirected graphs, Journal of Graph Theory