A Pseudo-Boolean Solution to the Maximum Quartet Consistency Problem [⋆]

António Morgado and João Marques-Silva

School of Electronics and Computer Science, University of Southampton, UK
ajrm@soton.ac.uk,jpms@ecs.soton.ac.uk

Abstract. Determining the evolutionary history of a given biological data is an important task in biological sciences. Given a set of quartet topologies over a set of taxa, the *Maximum Quartet Consistency* (MQC) problem consists of computing a global phylogeny that satisfies the maximum number of quartets. A number of solutions have been proposed for the MQC problem, including Dynamic Programming, Constraint Programming, and more recently Answer Set Programming (ASP). ASP is currently the most efficient approach for optimally solving the MQC problem. This paper proposes encoding the MQC problem with pseudo-Boolean (PB) constraints. The use of PB allows solving the MQC problem with efficient PB solvers, and also allows considering different modeling approaches for the MQC problem. Initial results are promising, and suggest that PB can be an effective alternative for solving the MQC problem.

1 Introduction

The amount of existing biological data (DNA and protein sequences) has increased the need for larger and faster determination of evolutionary history (or *phylogeny*) given a set of taxa (i.e. a set of related biological species [2]). Moreover, the availability of data is not always the same for different taxa. This is known as the data disparity problem [11, 12]. In recent years, quartet based methods have received greater attention from the computational biology community as a way to overcome the data disparity problem. Quartet-based methods are characterized by first inferring a set of evolutionary relationships between four taxa, and then from these relationships assemble a global evolutionary tree. Considering only four taxa in the first step to build the evolutionary relationships, leads to a greater confidence on the relationships produced. Nevertheless, the relationships obtained may be conflicting or even missing. The aim of this work is to obtain the evolutionary tree, under the parsimony assumption, that respects the maximum number of these relationships on four taxa.

Given a set of quartet topologies over a set of taxa, the *Maximum Quartet Consistency* (MQC) problem consists of computing a global phylogeny that satisfies the maximum number of quartets. A number of solutions have been proposed for the MQC problem, including Dynamic Programming, Constraint Programming, and more recently Answer Set Programming (ASP) [11, 9, 10]. ASP is currently the most efficient approach for optimally solving the MQC problem. This paper develops an encoding for the MQC problem with pseudo-Boolean (PB) constraints. Initial results are promising, and suggest that PB can be an effective alternative for solving the MQC problem.
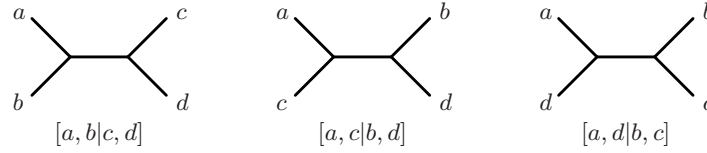
Fig. 1. Graphical representation of the quartet topologies $[a, b|c, d]$, $[a, c|b, d]$ and $[a, d|b, c]$.
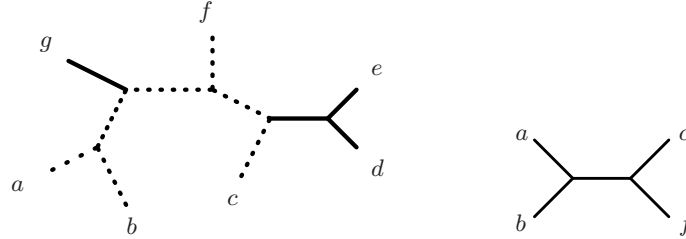

Fig. 2. Graphical representation of a phylogeny and of the quartet topology for the quartet $\{a, b, c, f\}$ derived from the phylogeny.

The paper is organized as follows. The first section introduces both the MQC problem and the MQI problem. The following section develops a Pseudo Boolean Optimization (PBO) model for the MQC problem and Section 4 proposes three optimizations to the PBO model. Section 5 shows the experimental results obtained and Section 6 presents some conclusions and points some directions for future research.

2   Preliminaries

A *phylogeny* is an unrooted tree whose leaves are bijectively mapped to a given set of taxa S, where each internal node has degree three. A *quartet* is a size four subset of S. For each quartet there exist three different possible phylogenies, called *quartet topologies*. Consider the quartet $\{a, b, c, d\}$, the three possible quartet topologies will be denoted by $[a, b|c, d]$, $[a, c|b, d]$ and $[a, d|b, c]$. Figure 1 gives a graphical representation of the three possible quartet topologies for the quartet $\{a, b, c, d\}$. For example, quartet topology $[a, b|c, d]$ means that the path that connects a and b does not intersect the path connecting c and d.

Given a phylogeny T on S and a quartet $q = \{a, b, c, d\}$, a quartet topology qt is said to be the quartet topology of q derived from T, if qt is the topology obtained from T, by removing all the edges and nodes not in the paths connecting the leaves that are mapped to taxa in q. Figure 2 represents a phylogeny, and the quartet topology derived from the phylogeny for the quartet $\{a, b, c, f\}$. The dotted branches show the path connecting the taxa in the quartet. Since the path that connects a and b does not intersect the path that connects c and f, then the derived quartet topology is $[a, b|c, f]$.

The set of quartet topologies derived from a phylogeny T is denoted by Q_T. If a quartet topology q is the same as the quartet topology derived from T, then T is said to

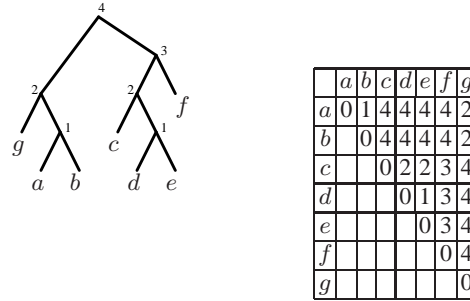|	a	b	c	d	e	f	g
a	0	1	4	4	4	4	2
b		0	4	4	4	4	2
c			0	2	2	3	4
d				0	1	3	4
e					0	3	4
f						0	4
g							0

Fig. 3. Graphical representation of a rooted phylogeny and the associated ultrametric matrix.

satisfy q and q is said to be *consistent* with T. In the example of Figure 2, $[a, b|c, f]$ is consistent with the phylogeny shown, but $[a, c|f, g]$ is not.

Given a set of quartet topologies Q on the set of taxa $S = \{s_1, \ldots, s_n\}$, if there exists a phylogeny T that satisfies all the quartet topologies in Q, then Q is said *compatible*. In practice the quartet topologies in Q may be inaccurate or even missing. If the set Q contains a quartet topology for each possible quartet of S, then Q is *complete* otherwise *incomplete*.

The problem of *Maximum Quartet Consistency (MQC)* is the problem where a set of quartet topologies Q on a set of taxa $S = \{s_1, \ldots, s_n\}$ is given, and returns a phylogeny T on S, that satisfies the maximum number of quartet topologies of Q.

The MQC problem is NP-hard [1] and if Q is complete, then MQC admits a polynomial-time approximation scheme [5]. If Q is incomplete, then MQC is MAX SNP-hard [5]. The dual problem to the MQC is the problem of *Minimum Quartet Inconsistency* (MQI). The MQI problem is the problem that given a set of quartet topologies Q (as in the MQC problem), returns a phylogeny that minimizes the number of quartet errors, where the set of quartet errors is the set $Q - Q_T$. The rest of the paper assumes that the set of quartet topologies Q is complete. In the recent past, different approaches have been reviewed in the literature for both the MQC and MQI problems. A detailed review is presented in [10].

3 Pseudo Boolean Model for the MQC Problem

This section develops a Pseudo Boolean Optimization(PBO) model for solving the MQC problem. The idea of the model is to obtain a rooted phylogeny, from which it is possible to construct an unrooted phylogeny [6]. Similarly to the existing ASP solution [10], the PBO model encodes the constraints of representing the rooted phylogeny tree as an ultrametric matrix. Moreover, an ultrametric phylogeny satisfies the maximum number of quartets topologies of a set Q if and only if the corresponding ultrametric matrix M satisfies the maximum number of quartets topologies in Q [10].

Consider the set of taxa $S = \{s_1, \ldots, s_n\}$ and a set of quartets Q. An *ultrametric* matrix M is a symmetric square matrix $n \times n$, where for each i such that $1 \leq i \leq n$ then $M(i, i) = 0$, for each i, j such that $1 \leq i < j \leq n$ then $1 \leq M(i, j) = M(j, i) \leq n$,

and for each triple of indices i, j, k such that $1 \leq i, j, l \leq n$, there is a tie between the maximum value of $M(i, j)$, $M(i, l)$ and $M(j, l)$.

The values in the ultrametric matrix M, represent the lowest common ancestor in the rooted phylogeny, that is the value of $M(i, j)$ corresponds to the internal node of the phylogeny that is the lowest common ancestor between taxa i and j. Figure 3 presents a rooted phylogeny, where the internal nodes have been labeled. The labels correspond to integers in decreasing order from the root to the leaves. On the right side of the figure is represented half of the associated ultrametric matrix. In [4] it is explored the relationship between rooted phylogenies and ultrametric matrixes and presents an algorithm to obtain a rooted phylogeny from the associated ultrametric matrix in polynomial time.

It was proven in [10] that in order to obtain an optimal phylogeny, the values of the entries of M can be restricted to $1 \leq M(i, j) \leq \lceil \frac{n}{2} \rceil$. To encode the values of $M(i, j)$ the PBO model introduces a set of Boolean variables $M_{i,j,k}$ where $1 \leq i < j \leq n$ and $1 \leq k \leq \lceil \frac{n}{2} \rceil$. $M_{i,j,k}$ has value 1 iff $M(i, j) = k$, otherwise $M_{i,j,k}$ is 0. To ensure that, for each pair (i, j), one and only one of the variables $M_{i,j,k}$ is selected to be true, the model introduces the following constraint:

$$\sum_{k=1}^{\lceil \frac{n}{2} \rceil} M_{i,j,k} = 1 \tag{1}$$

The value of each $M(i, j)$ variable is given by $M(i, j) = \sum_{k=1}^{\lceil \frac{n}{2} \rceil} k \times M_{i,j,k}$.

To ensure that the resulting matrix M is ultrametric, one of the following three conditions must be satisfied, for each $1 \leq i < j < l \leq n$:

$$M(i, j) = M(i, l) \wedge M(i, l) > M(j, l), \text{ or} \tag{2}$$
$$M(i, j) = M(j, l) \wedge M(j, l) > M(i, l), \text{ or} \tag{3}$$
$$M(j, l) = M(i, l) \wedge M(i, l) > M(i, j) \tag{4}$$

The PBO model associates three new Boolean variables $c1_{i,j,l}$, $c2_{i,j,l}$, $c3_{i,j,l}$ with constraints (2), (3) and (4), respectively. Each of the variables $cx_{i,j,l}$ is true iff the associated constraint is satisfied.

Constraint (2) is the logical AND of an equality constraint and a greater than constraint. In the PBO model each of these constraints is associated with additional Boolean variables, respectively, $c1_{i,j,l}^1$ and $c1_{i,j,l}^2$. $c1_{i,j,l}^1 = 1$ iff $M(i, j) = M(i, l)$, and can be implemented with a comparator circuit on the unary representation of $M(i, j)$ and $M(i, l)$, using variables $M_{i,j,k}$ and $M_{i,l,k}$. $c1_{i,j,l}^2 = 1$ iff $M(i, l) = M(j, l)$, and can also be implemented with a comparator circuit on the unary representation of $M(i, l)$ and $M(j, l)$, using variables $M_{i,l,k}$ and $M_{j,l,k}$. As a result, $c1_{i,j,l}$ is defined as:

$$c1_{i,j,l} = AND(c1_{i,j,l}^1, c1_{i,j,l}^2) \tag{5}$$

Variables $c2_{i,j,l}$ and $c3_{i,j,l}$ are encoded similarly. Finally to guarantee that one of the conditions (2), (3) or (4) is satisfied, the PBO model uses the following constraint:

$$c1_{i,j} + c2_{i,j} + c3_{i,j} \geq 1 \tag{6}$$

As the objective is to compute the phylogeny that maximizes the number of quartets that can be satisfied, then with each quartet is associated with a Boolean variable q_t,

where $1 \leq t \leq |Q|$. q_t will be true if quartet number t is consistent, otherwise q_t is false. A quartet $[i,j|l,m]$ is consistent if and only if one of the following conditions is satisfied [10]:

$$M(i,l) > M(i,j) \wedge M(j,m) > M(i,j) \text{, or} \tag{7}$$
$$M(i,l) > M(l,m) \wedge M(j,m) > M(l,m) \tag{8}$$

Suppose that quartet number t is the quartet $[i,j|l,m]$. The model associates two new variables to each of the conditions (7) and (8). Let $d1_{i,j,l,m}$ be associated with condition (7) and $d2_{i,j,l,m}$ be associated with condition (8). The associated variable q_t is encoded as a gate OR:

$$q_t = OR(d1_{i,j,l,m}, d2_{i,j,l,m}) \tag{9}$$

Both the conditions (7), (8) consist of logical ANDs of two greater than conditions. Thus variable $d1_{i,j,l,m}$ and $d2_{i,j,l,m}$ are encoded as gates AND in a analogous way to variables $c1_{i,j,l}$.

The cost function of the PBO model is then to maximize the number of quartets that are consistent, that is:

$$\max : \sum_{t=1}^{|Q|} q_t \tag{10}$$

4   Optimizations to the PBO Model

This section describes three optimizations to the basic PBO model. The first optimization aims reusing auxiliary variables that serve for encoding of some of the circuits associated with the PBO model. The second optimization is related with the Boolean variables used for representing the value of each entry in the ultrametric matrix. The third optimization sets the values for some of $M(i,j)$ variables when it is known that s_i and s_j are siblings.

4.1   First Optimization

The objective of the first optimization is to reduce the number of variables used in the encoding. The reduction is achieved by exploiting the information provided by the auxiliary variables used for encoding cardinality constraints. In order to implement this optimization, sequential counters [8] are used. The uniqueness constraint (1) of the PBO model in Section 3 is split into two constraints. The first constraint deals with the need to have one at least one variable selected by adding the constraint:

$$\sum_{k=1}^{\lceil \frac{n}{2} \rceil} M_{i,j,k} \geq 1 \tag{11}$$

The second constraint is:

$$\sum_{k=1}^{\lceil \frac{n}{2} \rceil} M_{i,j,k} \leq 1 \tag{12}$$

and is encoded in CNF with a sequencial counter [8]. This sequential counter introduces variables $s_{k,1}$. These variables have the property that if $M_{i,j,a} = 1$ then for $1 \leq k < a$ all variables have $s_{k,1} = 0$ and for $a \leq k \leq \lceil \frac{n}{2} \rceil$ then $s_{k,1} = 1$. The property enables the encoding of $M(i,j) < M(l,m)$ by considering the associated variables $s_{k,1}$ of $M(i,j)$ and of $M(l,m)$. In order to better understand, let the variables $s_{k,1}$ associated to the sequential counter of $M(i,j)$ be denoted by $s_k^{i,j}$. The objective is to encode that $M(i,j) < M(l,m)$ by re-using the variables $s_k^{i,j}$ and $s_k^{l,m}$. Using the above property, this can be done by searching for the k where $s_k^{i,j} = 1$ and $s_k^{l,m} = 0$, which can be encoded in a variable $e_k^{(i,j)(l,m)}$ as a gate AND:

$$e_k^{(i,j)(l,m)} = AND(s_k^{i,j}, NOT(s_k^{l,m})) \tag{13}$$

Then variable $LT_{i,j,l,m}$ encodes that $M(i,j) < M(l,m)$ by a gate OR:

$$LT_{i,j,l,m} = OR(e_k^{(i,j)(l,m)} : 1 \leq k \leq \lceil \frac{n}{2} \rceil) \tag{14}$$

For this optimization, all the other constraints of the PBO model of Section 3 are maintained, but making use of the variables $LT_{i,j,l,m}$ as appropriate.

4.2   Second Optimization

For the PBO model described in Section 3, for each pair of taxa (i,j), the values of the variables $M(i,j)$ are encoded through selection variables $M_{i,j,k}$ where $1 \leq k \leq \lceil \frac{n}{2} \rceil$.

The first optimization described here replaces the encoding of the selection variables. Variables $M_{i,j,k}$ are still going to be used to encode $M(i,j)$, but here $M_{i,j,k}$ represents the $k-$th bit of the binary representation of $M(i,j)$. Now k is limited by $0 \leq k \leq \lfloor \log_2(\lceil \frac{n}{2} \rceil) \rfloor$. With this encoding $M(i,j)$ can be obtained by $M(i,j) = \sum_{k=0}^{\lfloor \log_2(\lceil \frac{n}{2} \rceil) \rfloor} 2^k \times M_{i,j,k}$. Moreover, the constraints used in the encoding need to be modified. The constraints in Equation (1) that encode the uniqueness of the selection variables are no longer used. All the other constraints are maintained, but with the new limit for variable k. Instead of the uniqueness constraints, this optimization requires that the encoded variables $M(i,j)$ are restricted to $\{1, \ldots, \lceil \frac{n}{2} \rceil\}$, that is $1 \leq M(i,j)$ and $M(i,j) \leq \lceil \frac{n}{2} \rceil$. The first part is obtained by adding the constraint:

$$\sum_{k=0}^{\lfloor \log_2(\lceil \frac{n}{2} \rceil) \rfloor} M_{i,j,k} \geq 1 \tag{15}$$

For the second part, a new Boolean variable $ltb_{i,j}$ is introduced, that captures the condition that $M(i,j)$ is not larger than $\lceil \frac{n}{2} \rceil$. The variables $M_{i,j,k}$ are used to representing this constraint as a comparator circuit.

In order to ensure that $ltb_{i,j}$ is true, the following constraint is added to the model:

$$ltb_{i,j} \geq 1 \tag{16}$$

4.3   Third Optimization

The optimization described in this section follows [11, 9, 10]. The objective of this optimization is to previously determine the value of some variables, namely when a pair of taxa is know to be siblings. The optimization can be used independently of the model (or optimization) used.

Let $S = \{s_1, \ldots, s_n\}$ be a set of taxa and Q be a complete set of quartets . A *Bipartition* of S is a pair (X, Y) of nonempty subsets of S, such that $S = X \cup Y$ and $X \cap Y = \emptyset$. Consider a bipartition (X, Y) of S, such that $|X| \geq 2$ and $|Y| \geq 2$, let $Q_{(X,Y)}$ be defined as $Q_{(X,Y)} = \{[x_1, x_2 | y_1, y_2] : x_i \in X \wedge y_i \in Y$ for $i \in \{1, 2\}\}$. Suppose that three taxa from Y are fixed and also that $|X| = l$. An *l-subset* with respect to (X, Y) is the set of l quartets from Q that contain the three fixed taxa from Y and one taxa from X. There are a total of $\binom{n-l}{3}$ of l-subsets.

An l-subset is said to be *exchangeable* on X, if by ignoring the difference of the taxa from X on the quartets in the l-subset, it produces a unique quartet topology, otherwise the l-subset is said to be *nonexchangeable*. In the case where $l = 2$, then both taxa in X are said to be siblings and the following corollary holds:

Proposition 1 (Corollary 2.5 from [10]). *Let* $S = \{s_1, \ldots, s_n\}$ *be a set of taxa, Q be a complete set of quartets on taxa S. For the pair of taxa (s_i, s_j) from S, let $p_1 = |Q_{(\{s_i, s_j\}, Y)} - Q|$, p_2 be the number of nonexchangeable pairs on $\{s_i, s_j\}$. If $2p_1 + p_2 \leq n - 3$ then s_i, s_J are siblings in an optimal phylogeny.*

In the optimization described in this section, for every pair of taxa, the condition of the corollary is tested. When the condition is true, for example for taxa i and j, then the PBO model is augmented with the following constraints:

$$M_{i,j,1} \geq 1 \tag{17}$$

$$-1 \times M_{i,j,k} \geq 0 \quad , k \in \{2, \ldots, upperLimit\} \tag{18}$$

The $upperLimit$ in Equation (18) is dependent on the encoding of variable $M_{i,j,k}$ (either as described in Section 3 or as described in Section 4.2).

5   Experimental Results

This section presents experimental results comparing the PBO model proposed in Section 3 and the ASP model described in [10]. The instances considered were obtained from [10]. These instances correspond to quartet topologies derived from random generated trees with a percentage of quartet topologies randomly altered. The percentage of altered quartet topologies introduces errors in the quartet topologies. Higher percentage of altered quartet topologies means a higher possibility of errors in the quartet topologies of the instance.

In the experiments four models were considered, three obtained from the PBO formulation and one from the ASP formulation. The first PBO model considers the first optimization described in Section 4.1 and will be referred as *PBO+fst*. The second PBO model includes both the optimizations of Section 4.2 and Section 4.3. This second model will be referred as *PBO+(scd+trd)*. The last PBO model, called *PBO+trd*,

% Altered	N. Variables			N. Constraints		
	PBO+fst	PBO+(scd+trd)	PBO+trd	PBO+fst	PBO+(scd+trd)	PBO+trd
01	5760	4514.4	6276.6	19890	16238.8	24464
05	5760	4537.2	6310.8	19890	16301.5	24568.5
10	5760	4566.4	6354.4	19890	16385.2	24708
15	5760	4587.6	6386.4	19890	16448.8	24814
20	5760	4611.2	6421.8	19890	16519.6	24932
25	5760	4628.4	6447.6	19890	16571.2	25018
30	5760	4648.4	6477.6	19890	16631.2	25118

Table 1. Average number of variables and number of constraints for instances with 10 taxa.

% Altered	CPU Time			
	phy+SModels	PBO+fst	PBO+(scd+trd)	PBO+trd
01	0.0464	0.7696	0.4704	0.7316
05	0.3048	2.2673	1.686	7.0885
10	1.3264	5.7819	5.8872	28.8291
15	2.4324	12.7119	11.78235	52.6487
20	9.0915	32.2536	17.78277	68.77968
25	28.4901	60.7041	28.0254	117.6832
30	65.4176	121.3564	52.75086	239.2057

Table 2. Average CPU time in seconds for instances with 10 taxa.

includes only the third proposed optimization (Section 4.3). In all the PBO models an encoder was implemented that receives as input the quartet topologies and returns as output a file in PB format. The generated file was then given as input to the PBO solver. For all experiments the PBO solver used was *minisat+* [3].

The fourth model is the ASP model described in [10]. The phy program, that encodes the quartet topologies into answer set programming, was obtained from [10]. The instances were given to phy, and for each, the parameters given were the number of taxa involved and the maximum number of quartet errors known in the instance. This last parameter was set as the number of quartet topologies in the instance. After obtaining the encoded instance, the encoded file was given to the ASP-solver SModels [7] SModels was configured to obtain all the stable models in order to maximize the number of quartets satisfied.

The results were obtained on an Intel Xeon 5160, 3GHz server, with 4 GB of RAM. The results comparing the average number of variables and number of constraints between the three PBO models is shown in Table 1. As can be seen from the table the model that requires more variables and more constraints is the PBO+trd model, whereas, the model that requires less variables and less constraints is the PBO+(scd+trd).

Table 2 compares the average CPU times on the instances considered for all the PBO models and the phy+Smodes model.

A few conclusions can be drawn from the results. First comparing the *PBO+fst* and the basic *PBO+trd* model. The sharing of auxiliary variables introduced by the first optimization is an important aspect in this problem. This optimization reduces the number of variables used by the encoding as well as the number of constraints. This reduction

leads to lower CPU time spent by the PBO-solver. Nevertheless, model *PBO+(scd+trd)* reduces even further the model by considering the selection variables as bits of the binary representation of values in M. Again, it can be seen from Table 2, that the reduction on the number of variables and constraints used by the encoding resulted in lower CPU times spent by the PBO-solver, where the model *PBO+(scd+trd)* is on average approximately 4 times faster than the *PBO+trd* and 1.6 times faster than *PBO+fst*.

Comparing the best of our PBO models (*PBO+(scd+trd)*) with the ASP model, the ASP model is more effective when the percentage of modified quartets is small, but the *PBO+(scd+trd)* model becomes more when the percentage of modified quartets increases.

6   Conclusions

This paper proposes a first attempt at solving the MQC problem with PBO. The new PBO model is compared with a recent solution based on ASP [10], which is currently the most efficient for the MQC problem. Despite the number of the taxa considered being modest, the results show that the PBO model can be beneficial when the number of expected quartet errors is high. The PBO model is still recent, and additional modeling insights and corresponding performance improvements are to be expected in the near future.

Future research will involve developing optimizations to the PBO model. For example, by encoding with PB constraints some of the optimizations proposed in the literature for the MQC problem. Furthermore, experiments will consider larger sets of taxa as well as real world data.

References

1. V. Berry, T. Jiang, P. Kearney, M. Li, and T. Wareham. Quartet cleaning: Improved algorithms and simulations. *Proceedings of the Seventh European Symposium on Algorithms (ESA99), Lecture Notes in Computer Science*, 1643:313–324, 1999.
2. B. Chor. From quartets to phylogenetic trees. In *Conference on Current Trends in Theory and Practice of Informatics*, pages 36–53, 1998.
3. N. Een and N. Sorensson. Translating pseudo-boolean constraints into SAT. *Journal on Satisfiability, Boolean Modeling and Computation*, 2:1–26, 2006.
4. D. Gusfield. *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. Cambridge University Press, January 1997.
5. T. Jiang, P. Kearney, and M. Li. Orchestrating quartets: approximation and data correction. *Proceedings of the 39th IEEE Symposium on Foundations of Computer Science*, pages 416–425, 1998.
6. D. Pelleg. *Algorithms for Constructing Phylogenies from Quartets*. PhD thesis, Masters thesis, Israel Institute of Technology, 1998.
7. P. Simons. Computing the stable model semantics.
8. C. Sinz. Towards an optimal cnf encoding of boolean cardinality constraints. In *Principles and Practice of Constraint Programming - CP 2005, 11th International Conference*, volume 3709, pages 827–831, 2005.
9. G. Wu, G. Lin, J.-H. You, and X. Wu. Faster solution to the maximum quartet consistency problem with constraint programming. *Proceedings of 3rd Asia-Pacific Bioinformatics Conference*, pages 329–338, 2005.

10. G. Wu, J. You, and G. Lin. Quartet-based phylogeny reconstruction with answer set programming. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, 4(1):139–152, 2007.
11. G. Wu, J.-H. You, and G. Lin. A lookahead branch-and-bound algorithm for the maximum quartet consistency problem. *Algorithms in Bioinformatics: 5th International Workshop, WABI 2005, Mallorca, Spain, October 3-6, 2005: Proceedings*, 2005.
12. G. Wu, J.-H. You, and G. Lin. A polynomial time algorithm for the minimum quartet inconsistency problem with O(n) quartet errors. *Information Processing Letters*, 100(Issue 4):167 –171, 2006.