SORBONNE UNIVERSITÉ

**ÉCOLE DOCTORALE EDITE DE PARIS (ED130)**
INFORMATIQUE, TÉLÉCOMMUNICATION ET ÉLECTRONIQUE

**THÈSE DE DOCTORAT DE L'UNIVERSITÉ SORBONNE UNIVERSITÉ**

SPÉCIALITÉ : **INGÉNIERIE / SYSTÈMES INFORMATIQUES**

PRÉSENTÉE PAR : **HAKAN METIN**

POUR OBTENIR LE GRADE DE :

**DOCTEUR DE L'UNIVERSITÉ SORBONNE UNIVERSITÉ**

**SUJET DE LA THÈSE :**

# EXPLOITATION DES SYMÉTRIES DYNAMIQUES POUR LA RÉSOLUTION DES PROBLÈMES SAT

SOUTENUE LE :

DEVANT LE JURY COMPOSÉ DE :

| | | |
|---|---|---|
| *Rapporteurs :* | Y. XXX | XXX |
| | Y. XXX | XXX |
| *Examinateurs :* | Y. XXX | XXX |
| | Y. XXX | XXX |
| | Y. XXX | XXX |
| | Y. XXX | XXX |

LIP6

*Ah la these*

# CONTENTS

SAT Theory NP complete Cite [COOK 71]

Used In
- Formal methods Hardware model checking;Software model checking; Termination analysis of term-rewrite systems ; Test pattern generation (testing of software hardware) ; etc
- AI planning, game n queen sudoku
- Bio info Haplotype inference ; Pedigree checking ; Analysis of Genetic; Regulatory Networks; etc.
- Design Automation: Equivalence checking; Delay computation; Fault diagnosis ; etc
- Security: Cryptanalysis ; Inversion attacks on hash functions; etc

Where can found
- Computationally hard problems: Graph coloring, Traveling salesperson
- Mathematical: van der Waerden numbers ; Quasigroup open problems
- Core engine for other solvers :0-1 ILP/Pseudo Boolean ;QBF ; #SAT ; SMT ; MAXSAT;
- Integrated into theorem provers :HOL ; Isabelle ;

- Integrated into widely used software: Eclipse provisioning system based on a Pseudo Boolean solver Suse 10.1 dependency manager based on a custom SAT solver

Interest in BMC [ ]Biere 99]

SAT solver Def

Input: Can encode any Boolean formula into Normal Form

Classical simplification: - Resolution - Subsumption

More complicated Hidden tautology BVA ... Loo Heule simplification slides

Algo CDCL

CHAPTER 1

# INTRODUCTION

~~Nowadays, computers are powerfull and used in many applications in different domains. On of this domain is critical application, these applications are running in planes, cars and some software must be secure. Proving the correctness of these software leads to combinatorial explosion.~~

Over the years, computers scientist have developed many techniques to solve this kind of problems like *constraint programming* (CP) [7], *Propositional Satisfiability* (SAT) [2], *Solver Modulo Theory* (SMT) [1].

In this thesis, we focus on *propositional satisfiability* that is used in many applications in different domains: *formal methods*: hardware model checking, software model checking, etc; *artificial inteligence*: planning; *games resolution*: sudoku, n-queens, *Bioinformatics* : Haplotype inference, *design automation* : equivalence checking

At its most basic, symmetry is some transformations of an object that leaves this object unchanged. In the case of satisfiability problems it maps a solution of a problem to another solution of the problem.

CHAPTER 2

# PRELIMINARIES

## 2.1 SAT basics

### Satisfiability problem

A *Boolean variable*, or *propositional variable*, is a variable that has two possible values : true or false (noted $\top$ or $\bot$, respectively). A *literal* $l$ is a propositional variable or its negation. For a given variable $x$, the positive literal is represented by $x$ and the negative one by $\neg x$. A *clause* $\omega$ is a finite disjunction of literals represented equivalently by $\omega = \bigvee_{i=1}^{k} l_i$ or the set of its literals $\omega = \{l_i\}_{i \in [\![1,k]\!]}$. A clause with a single literal is called *unit clause*. A *conjunctive normal form (CNF) formula* $\varphi$ is a finite conjunction of clauses. A CNF can be either noted $\varphi = \bigwedge_{i=1}^{k} \omega_i$ or $\varphi = \{\omega_i\}_{i \in [\![1,k]\!]}$. We denote $\mathcal{V}_\varphi$ ($\mathcal{L}_\varphi$) the set of variables (literals) used in $\varphi$ (the index in $\mathcal{V}_\varphi$ and $\mathcal{L}_\varphi$ is usually omitted when clear from context).

For a given formula $\varphi$, an *assignment* of the variables of $\varphi$ is a function $\alpha : \mathcal{V} \mapsto \{\top, \bot\}$. As usual, $\alpha$ is *total*, or *complete*, when all elements of $\mathcal{V}$ have an image by $\alpha$, otherwise it is *partial*. By abuse of notation, an assignment is often represented by the set of its true literals. The set of all (possibly partial) assignments of $\mathcal{V}$ is noted $Ass(\mathcal{V})$.

The assignment $\alpha$ *satisfies* the clause $\omega$, denoted $\alpha \models \omega$, if $\alpha \cap \omega \neq \emptyset$. Similarly, the assignment $\alpha$ satisfies the propositional formula $\varphi$, denoted $\alpha \models \varphi$, if $\alpha$ satisfies all the clauses of $\varphi$. Note that a formula may be satisfied by a partial assignment. A formula is said to be *satisfiable* (SAT) if there is at least one assignment that satisfies it; otherwise the formula is *unsatisfiable* (UNSAT).

## Algorithm

The state of the art sound and complete algorithm to resolve a SAT problem is Conflict-Driven Clause learning (CDCL) algorithm 1. This algorithm is inspired by Davis Putnam Logemann Loveland [3].

The CDCL algorithm walks a binary search tree. It first applies unit propagation to the formula $\varphi$ for the current assignment $\alpha$ (line 5). A conflict at level 0 indicates that the formula is not satisfiable, and the algorithm reports it (lines 7-8). If a conflict is detected, it is analyzed, which provides a *conflict clause* explaining the reason for the conflict (line 9). The analysis is completed by the computation of a backjump point to which the algorithm backtracks (line 10). This clause is learnt (line 11), as it does not change the satisfiability of $\varphi$, and avoids encountering a conflict with the same causes in the future. Finally, if no conflict appears, the algorithm chooses a new decision literal (line 13-14). The above steps are repeated until the satisfiability status of the formula is determined.

```
1  function CDCL(φ: CNF formula)
2      returns ⊤ if φ is SAT and ⊥ otherwise
3      dl ← 0;                                          // Current decision level
4      while not all variables are assigned do
5          isConflict ← unitPropagation();
6          if isConflict then
7              if dl = 0 then
8                  return ⊥;                                          // φ is UNSAT
9              ω ← analyzeConflict();
10             dl ← backjumpAndRestartPolicies();
11             φ ← φ ∪ {ω};
12         else
13             assignDecisionLiteral();
14             dl ← dl + 1;
15     return ⊤;                                                      // φ is SAT
```

**Algorithm 1:** The CDCL algorithm.

## 2.2 Groups basics

Symmetries is related to a branch of mathematics called group theory. This section give us an overview of group theory.

### Groups

A *group* is a structure $\langle G, * \rangle$, where $G$ is a non empty set and $*$ a binary operation such the following axioms are satisfied:

- *associativity*: $\forall a, b, c \in G, (a * b) * c = a * (b * c)$
- *closure*: $\forall a, b \in G, a * b \in G$.
- *identity*: $\forall a \in G, \exists e$ such that $a * e = e * a = a$
- *inverse*: $\forall a \in G, \exists b \in G$, commonly denoted $a^{-1}$ such that $a * a^{-1} = a^{-1} * a = e$

Note that *commutativity* is not required i.e $a * b = b * a$, for $a, b \in G$. The group is *abelian* if it satisfies the commutativity rule. Moreover, the last definition leads to important properties which are: i) uniqueness of the identity element. To prove this property, assume $\langle G, * \rangle$ a group with two identity elements $e$ and $f$ then $e = e * f = f$. ii) uniqueness of the inverse element. To prove this property, suppose that an element $x_1$ has two inverses, denoted $b$ and $c$ in group $\langle G, * \rangle$, then

$$
\begin{array}{rcll}
b & = & b * e & \\
 & = & b * (a * c) & c \text{ is an inverse of } a, \text{so } e = a * c \\
 & = & (b * a) * c & \textit{associativity} \text{ rule} \\
 & = & e * c & b \text{ is an inverse of } a, \text{so } e = a * b \\
 & = & c & \textit{identity} \text{ rule}
\end{array}
$$

The structure $\langle G, * \rangle$ is denoted as G when clear from context that G is a group with a binary operation. In this thesis, we interested only with the *finite* groups i.e with a finite number of elements.

Given a group $G$, a *subgroup* is a non empty subset of $G$ which is also a group with the same binary operation. If $H$ is a subgroup of $G$, we denote as $H \leq G$. A group has at least two subgroups: i) the subgroup composed by identity element $\{e\}$, denoted *trivial* subgroup. All other subgroups are *nontrivial*; ii) the subgroup composed by itself, denoted *improper* subgroup. All other subgroups are *proper*.

#### Generators of a group

If every elements in a group G can be expressed as a linear combination of a set of group of elements S = $\{g_1, g_2, ..., g_n\}$ then we say G is generated by the S. we denote this as G = $\langle S \rangle$ = $\langle \{g_1, g_2, ..., g_n\} \rangle$

## Permutation groups

A *permutation* is a bijection from a set $X$ to itself.

Example: given a set $X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$, $g = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ x_2 & x_3 & x_1 & x_4 & x_6 & x_5 \end{pmatrix}$

$g$ is a permutation that maps $x_1$ to $x_2$, $x_2$ to $x_3$, $x_3$ to $x_1$, $x_4$ to $x_4$, $x_5$ to $x_6$ and $x_6$ to $x_5$.

Permutations are generally written in *cycle notation* and the self mapped elements are omitted. So the permutation in cycle notation will be : $g = (x_1\ x_2\ x_3)\ (x_5\ x_6)$. We say *support* of the permutation $g$ noted $supp(g)$ the elements that not mapped to themselves, $supp(g) = \{x \in X \mid g(x) \neq x\}$.

The set of permutations of a given set $X$ form a group $G$, with the composition operation $(\circ)$ and called *permutation group*. The *symmetric group* id the set of all possible permutations of a set $X$ and noted $\mathfrak{S}(X)$. So, a *permutation group* is a subgroup of $\mathfrak{S}(X)$.

A permutation group $G$ induces a *equivalence relation* on the set of element $X$ being permuted. Two elements $x_1, x_2 \in X$ are equivalent if there exists a permutation $g \in G$ such that $gx_1 = x_2$. Then equivalence relation partitions $X$ into *equivalence classes* referred to as the *orbits* of $X$ under $G$. The orbit of an element $x$ under group $G$ (or simply orbit of $x$ when clear from the context) is the set $[x]_G = \{g.x \mid g \in G\}$

CHAPTER 3

# SYMMETRY AND SAT

The group of permutations of $\mathcal{V}$ (i.e. bijections from $\mathcal{V}$ to $\mathcal{V}$) is noted $\mathfrak{S}(\mathcal{V})$. The group $\mathfrak{S}(\mathcal{V})$ naturally acts on the set of literals: for $g \in \mathfrak{S}(\mathcal{V})$ and a literal $\ell \in \mathscr{L}$, $g.\ell = g(\ell)$ if $\ell$ is a positive literal, $g.\ell = \neg g(\neg\ell)$ if $\ell$ is a negative literal. The group $\mathfrak{S}(\mathcal{V})$ also acts on (partial) assignments of $\mathcal{V}$ as follows: for $g \in \mathfrak{S}(\mathcal{V})$, $\alpha \in Ass(\mathcal{V})$, $g.\alpha = \{g.\ell \mid \ell \in \alpha\}$. Let $\varphi$ be a formula, and $g \in \mathfrak{S}(\mathcal{V})$. We say that $g \in \mathfrak{S}(\mathcal{V})$ is a symmetry of $\varphi$ if for every *complete* assignment $\alpha$, $\alpha \models \varphi$ if and only if $g.\alpha \models \varphi$. The set of symmetries of $\varphi$ is noted $S(\varphi) \subseteq \mathfrak{S}(\mathcal{V})$.

The previous mathematical definitions of group theory is applied to the CNF formula. So, the group of permutations of $\mathcal{V}$ (i.e. bijections from $\mathcal{V}$ to $\mathcal{V}$) is noted $\mathfrak{S}(\mathcal{V})$. We say that $g \in \mathfrak{S}(\mathcal{V})$ is a symmetry of $\varphi$ if following conditions holds:

- permutation fixes the formula, $g(\varphi) = \varphi$
- $g$ commutes with the negation: $g(\neg l) = \neg g(l)$

The set of symmetries of $\varphi$ is noted $S(\varphi) \subseteq \mathfrak{S}(\mathcal{V})$. The sets of symmetries of a formula $\varphi$ preserves the satisfaction, for every *complete* assignment $\alpha$, $\alpha \models \varphi \leftrightarrow g(\alpha) \models \varphi$ for $g \in S(\varphi)$. The group $S(\varphi)$ also acts on (partial) assignments of $\mathcal{V}$ as follows: for $g \in S(\varphi)$, $\alpha \in Ass(\mathcal{V})$, $g.\alpha = \{g.\ell \mid \ell \in \alpha\}$.

The next section presents how to compute the set of *generators* of a given formula.

## 3.1 Symmetry detection in SAT

For the detection of symmetries in SAT, we fist introduce the graph automorphism notion. Given a colored graph $G = (V, E, \gamma)$, with vertex set $V \in [1, n]$, edge set E and $\gamma$ a function that

apply a mapping : $V \to C$ where C is a set of *colors*. An automorphism of G is a permutation from its vertices $g : V \to V$ such that:

- $\forall (u, v) \in E \implies (g(u), g(v)) \in E$
- $\forall v \in V, \gamma(v) = \gamma(g(v))$

The graph automorphism problem is to find if a given graph has a non trivial permutation group. The computational complexity of this algorithm is conjectured to be strictly between P and NP. Several tools exists to tackle this problem like `saucy3` [5], `bliss` [4], `nauty` [6], etc.

There exists different ways to encode a SAT problems, which leads to different symmetries in these problem. When a symmetry depends on the structure of the problem, we say *syntactic* symmetries. In contrast, symmetries were *semantic*, when it is not inherent to the encoding. To find symmetries in SAT problem, the formula is transformed into colored graph and an automorphism tool is applied onto. Specifically, given a formula $\varphi$ with $m$ clauses over $n$ variables, the graph is constructed as follows:

- *clause nodes*: represent each of the $m$ clauses by a node with color 0;
- *literals nodes*: represent each of the $l$ literals by a node with color 1;
- *clauses edges*: connect each clause node to the node of the literals that appear in clause;
- *boolean consistency edges*: connect each pair of literals that correspond to the same variables.

Hakan: Explication du graph + informations num nodes num edges. Probleme reel battleship

Hakan: extension of permutation to clauses assignments ...

An optimization of this graph is possible with the usage of binary clauses i.e. a clause with only two literals. The clause node can be omitted and we connect the two literals. As we cannot distinguish between the optimized edge and boolean consistency edges, we must check if the produced permutations are spurious. To do so, as we ensure the permutation commutes with the negation it suffice to check: $\forall x \in Supp(g), g.\neg x = \neg g.x$. Roughly speaking, we check if the image of the negation of $x$ is equals to the negation of the image of $x$, for each element $x$ in the support of the permutation. In the previous example, the number of nodes in the graph is reduced to Hakan: COUNT NODES.

Hakan: optimisation du graph
Hakan: creation d'un probleme fil conducteur and utilisation de celui dans chaque partie, du calcul des symétries jusqu'au SBP

```
p cnf 12 21
−1 −2 0
−1 −3 0
−2 −3 0
−4 −5 0
−4 −6 0
−5 −6 0
−7 −8 0
−7 −9 0
−8 −9 0
−10 −11 0
−10 −12 0
−11 −12 0
1 4 7 10 0
2 5 8 11 0
3 6 9 12 0
1 6 8 10 0
2 4 9 11 0
3 5 7 12 0
1 5 9 10 0
2 6 7 11 0
3 4 8 12 0
```

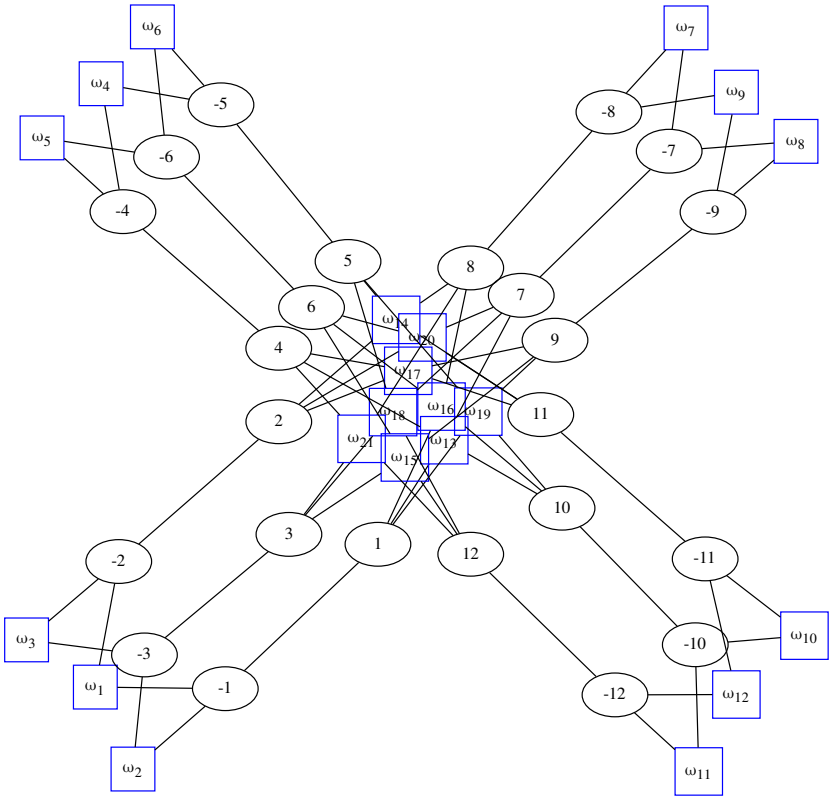


Figure 3.1: Example of constructed symmetry graph for a given CNF

When these graph is given to an automorphism tool like `bliss`, the following *generators* are obtained:
$g_0$: (2 3)(5 6)(8 9)(11 12)(-2 -3)(-5 -6)(-8 -9)(-11 -12)
$g_1$: (4 5 6)(7 9 8)(-4 -5 -6)(-7 -9 -8)
$g_2$: (4 7)(5 8)(6 9)(-4 -7)(-5 -8)(-6 -9)
$g_3$: (1 2)(5 6)(7 9)(10 11)(-1 -2)(-5 -6)(-7 -9)(-10 -11)
$g_4$: (1 10)(2 11)(3 12)(-1 -10)(-2 -11)(-3 -12)

## 3.2 Usage of symmetries

Since the set of permutations is obtained from the formula

We say that $g$ is a symmetry of $\varphi$ if $\alpha \models \varphi$ and $g.\alpha \models \alpha$

The optimal approach to solve a symmetric SAT problem would be to explore only one assignment per orbit (for instance each lex-leader). However, finding the lex-leader of an orbit is computationally hardHakan: CITATION. However, some practical algorithms exists to solve this problem.

*Symmetry breaking* aims at eliminating symmetry, either by *statically* posting symmetry

```
p cnf 12 21
−1 −2 0
−1 −3 0
−2 −3 0
−4 −5 0
−4 −6 0
−5 −6 0
−7 −8 0
−7 −9 0
−8 −9 0
−10 −11 0
−10 −12 0
−11 −12 0
1 4 7 10 0
2 5 8 11 0
3 6 9 12 0
1 6 8 10 0
2 4 9 11 0
3 5 7 12 0
1 5 9 10 0
2 6 7 11 0
3 4 8 12 0
```

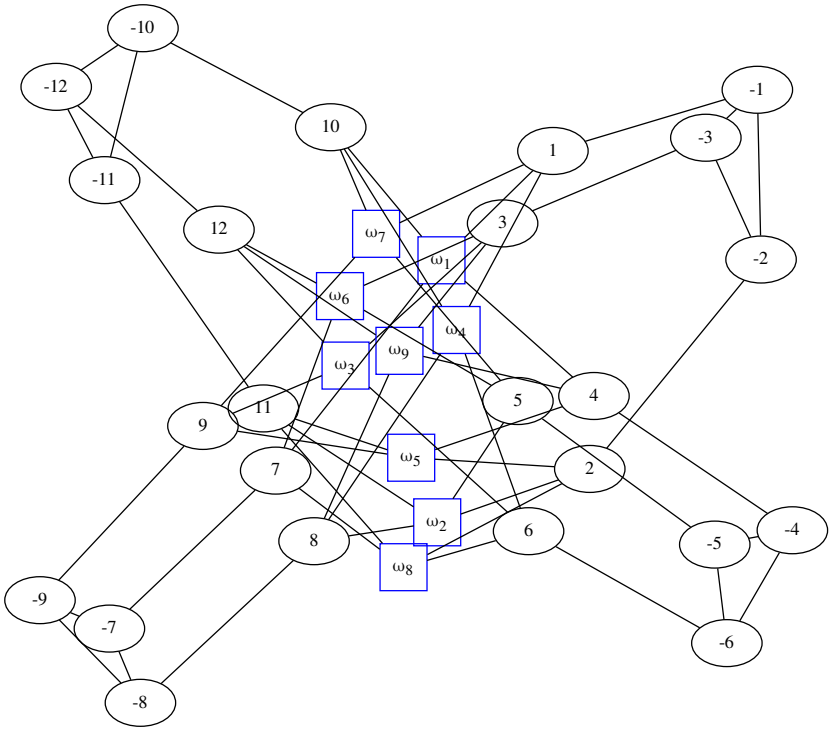


Figure 3.2: Example of constructed symmetry graph for a given CNF

breaking constraints that invalidate symmetric assignments, or by altering the search space *dynamically* to avoid symmetric search paths.

*static symmetry breaking* acts like a preprocessor which add *symmetry breaking predicates* (SBP) at the original formula and solve the augmented problem.

## Lexicographic leader

## Compact CNF encoding of Lexicographic leader

## Some special groups

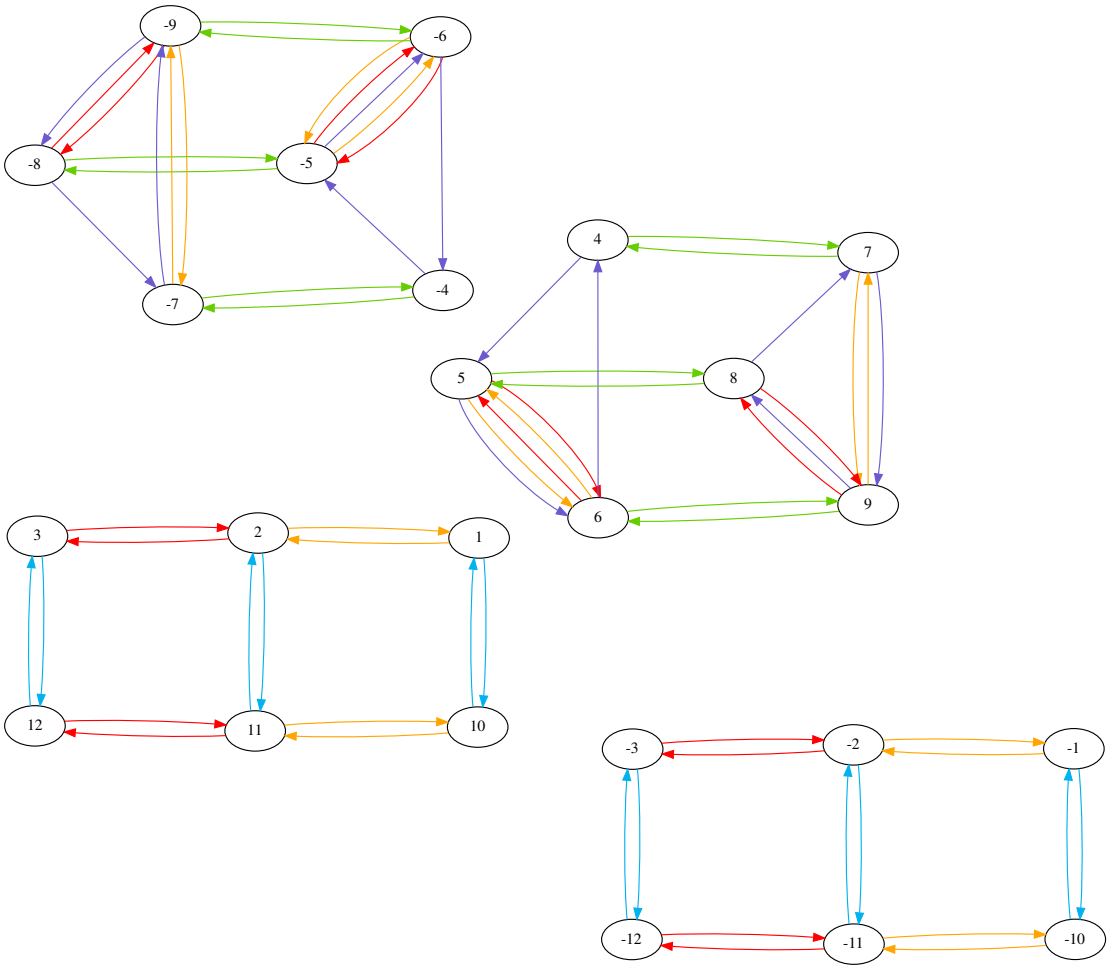


Figure 3.3: Orbits

CHAPTER 4

# CONCLUSION

This is conclusion.

# BIBLIOGRAPHY


[1] C. Barrett and C. Tinelli. Satisfiability modulo theories. In *Handbook of Model Checking*, pages 305–343. Springer, 2018.

[2] A. Biere, M. Heule, and H. van Maaren. *Handbook of satisfiability*, volume 185. IOS press, 2009.

[3] M. Davis, G. Logemann, and D. Loveland. A machine program for theorem-proving. *Commun. ACM*, 5(7):394–397, July 1962.

[4] T. Junttila and P. Kaski. Engineering an efficient canonical labeling tool for large and sparse graphs. In D. Applegate, G. S. Brodal, D. Panario, and R. Sedgewick, editors, *Proceedings of the Ninth Workshop on Algorithm Engineering and Experiments and the Fourth Workshop on Analytic Algorithms and Combinatorics*, pages 135–149. SIAM, 2007.

[5] H. Katebi, K. Sakallah, and I. Markov. Symmetry and satisfiability: An update. *Theory and Applications of Satisfiability Testing–SAT 2010*, pages 113–127, 2010.

[6] B. D. McKay. nauty user's guide (version 2.2). Technical report, Technical Report TR-CS-9002, Australian National University, 2003.

[7] F. Rossi, P. Van Beek, and T. Walsh. *Handbook of constraint programming*. Elsevier, 2006.